Opportunities and Entrepreneurship: Evidence on Advanced Labor Market Experience*

Matthew Pecenco[†], Carlos Schmidt-Padilla[‡], Hamilton Taveras[§]

September 17, 2024

Abstract

We study the role of high-value temporary managerial jobs in generating occupational transition and business creation. Exploiting randomized lotteries for these government contracts in the Dominican Republic, we show lottery winners transition away from formal employment and start new and growing firms in the medium-run. A selection model, estimated on repeated application choices, reveals significant heterogeneous effects driven by younger individuals and those with higher revealed preference for the position. Counterfactual policies that increase business creation come at the cost of reducing opportunities for individuals who otherwise would not become entrepreneurs.

*We are particularly indebted to Elisabeth Sadoulet, Reed Walker, and Edward Miguel for invaluable guidance and support on this project. We thank Susanna Berkouwer, Doris Chiang, Alain de Janvry, Kwabena Donkor, Benjamin Faber, Frederico Finan, Andrew Foster, Alexander Gelber, Marco Gonzalez-Navarro, Sean Higgins, Hilary Hoynes, Erin Kelley, Patrick Kline, Greg Lane, Ethan Ligon, John Loeser, Jeremy Magruder, Aprajit Mahajan, David McLaughlin, David Mao, Samuel Norris, Manaswini Rao, Jesse Shapiro, Christopher Walters, Shaoda Wang, Romain Wacziarg, and Jeffrey Weaver, as well as seminar participants at ABCDE, BSE Summer Forum, Brown, LACEA, NBER SI Productivity, Development, and Entrepreneurship, Northwestern Rookiefest, NEUDC, PacDev, RIDGE, UCLA Anderson, UCB (ARE, Econ, Haas), UPenn, U Chile, and the World Bank. This project would not have been possible without the support and assistance of Yokasta Guzmán Santos, Andres de la Rosa Batista, Marvin Cardoza, Sarah Sanchis, Juan Ariel Jímenez, Daniel Morales, and Marcio Báez. We thank Vilma Koury for alerting us to the *sorteo de obras.* Luis Almonte, Luis Mella Gómez, and Carmen Vargas provided superb research assistance. Funding for this project was graciously provided by CEGA, the J-PAL Governance Initiative, and the Weiss Program Family Fund. Pecenco acknowledges generous financial support from the Institute of Research on Labor and Employment through the Dissertation Fellows program. All views expressed are those of the authors and do not necessarily reflect the opinions of any of the funding or data-providing organizations.

[†]Department of Economics, Brown University. Contact: matthew_pecenco@brown.edu

[‡]Goldman School of Public Policy, University of California, Berkeley. Contact: cschmidtpadilla@gmail.com

[§]General Directorate of Internal Revenue, Ministry of Finance, Dominican Republic. Contact: Htaveras@dgii.gov.do

Firm creation plays a pivotal role in markets by introducing innovation, fostering competition, and driving job growth. However, in developing countries, entrepreneurial activity often lags across various dimensions, including formal firm creation and business growth, despite an abundant supply of self-employed individuals (La Porta and Shleifer, 2014). Many interventions designed to tackle this shortfall, by providing financial capital, business training, or assistance in hiring, have not substantively affected formal firm creation or growth.[1] Interventions may fall short because they fail to address constraints to business creation or do not target the individuals most likely to benefit.

In this paper, we provide new evidence on barriers to business creation along two key dimensions. First, we provide experimental evidence that temporary high-value managerial jobs—a different intervention from those commonly studied in the literature—significantly increase the creation of new, growing businesses. Recent literature has highlighted the importance of managerial job experience for entrepreneurship (Liang et al., 2018), yet identifying a source of exogenous variation has been elusive. Second, repeated, costly application choices for these positions and a model reveal important differential effects for applicants along dimensions not typically observable to researchers or policymakers, offering insights on how to better target these effects. This is significant because recent literature highlights the importance of targeting (Meager, 2022; Crépon et al., 2024) and the limitations of identifying potential entrepreneurs with observable characteristics (McKenzie and Sansone, 2019). Together, our results suggest managerial jobs can be a significant catalyst to business creation, but entrepreneurship programs aimed at increasing equity in entry rates may face a tradeoff with impacts.

We study these two questions—the effect of managerial jobs, and the importance of targeting—using a randomized lottery procurement scheme in the Dominican Republic (DR). Under this scheme, the government randomly assigns contracts to manage the construction of schools, hospitals, and other infrastructure projects to applicants from the population of licensed civil engineers and architects. From 2012 to 2015, the government allocated over 2,300 high-value contracts across a series of lottery events. Recipients primarily engage in managerial tasks, including hiring employees, purchasing materials, subcontracting, budgeting, and overseeing the execution of the work. Contract values are also assigned randomly within application groups and directly affect the amount of income received.

Our study addresses common limitations in experimental studies of labor market experiences in developing countries (McKenzie, 2017a). First, we study medium-run impacts

[1] Reviewed in Woodruff (2018).

by linking program randomization to high-quality administrative and self-collected survey data to measure a wide range of outcomes up to five years after the allocation of contracts. These include formal wage employment, income, firm creation, and future work with the government. We also collect information on firms owned by individuals in this sample (size, survival, revenue, etc.) to understand patterns of entrepreneurial quality. Second, we are uniquely able to study impacts across a wide swathe of the relevant population and over the career from ages 20-65 because the national scale of the program induced widespread entry into these lotteries. For example, over 80% of recently licensed engineers entered at least one lottery event.

We first use reduced-form evidence comparing lottery winners to nonwinners to document that the contracts lead to increases in subsequent measures of entrepreneurship in the medium-term. Lottery winners do not start firms at the time of contract receipt but shift into firm creation and ownership as the contracts finish. Five years later, winners are 7.8 percentage points (22%) more likely to have started a firm, predominantly in the construction sector. This effect is matched by a similar shift out of wage work over time, indicating the measured firm creation is not the result of the formalization of existing informal firms. Additionally, lottery winners see a substantial increase in net income, but the effect decays and incomes are similar across groups after five years.

The combined set of results suggests that the observed increase in business creation are more likely driven by managerial experience uptake. We directly show the firms are not created to access further public sector contracts or due to improved co-ownership networks (among winners). To differentiate between managerial experience and increases in accessible financial capital, we analyze heterogeneous effects along baseline individual covariates and from randomly assigned contract values. Younger individuals, rather than those with lower previous incomes, are more likely to start businesses, suggesting that less experienced individuals benefit more. Higher-value contracts, while increasing income, show small, insignificant, and negatively-signed estimates on firm creation and growth compared to lower-value contracts. Therefore, available capital does not appear to significantly impact entrepreneurship. Overall, these findings suggest that access to future capital or credit, though often associated with managerial jobs, may not be a key mechanism driving business creation in this context.

A central concern of interventions designed to generate new entrepreneurs is the quality of their newly created firms.[2] We find that firms created by lottery winners are more likely

[2]A growing literature has emphasized heterogeneity in entrepreneur quality and their resulting success. See,

to have hired employees and survived than firms created by nonwinners. Thus, the program may have beneficial impacts for long-term firm development.

To understand targeting, in the second part of the paper, we use a selection model to evaluate self-selection and generalize our treatment effects beyond the sample of lottery participants. We link panel data on individuals' application choices to uncover heterogeneity in preferences to participate along both observable (e.g., age, sex, past labor market history) and unobservable margins (e.g., motivation). Variation in lottery choices shifts individuals in and out of application over time, while the lottery randomization induces variation in treatment status across the preference distribution. We nest our structural preference estimates into a model of outcomes using a two-step control function approach, as in Walters (2018), to estimate aggregate treatment effects and identify which sources of heterogeneity drive our estimated effects.

We find significant selection effects. Individuals who are more advantaged along observable characteristics (e.g., males, high previous salary) are more likely to apply, but time-invariant observable traits account for only one-fifth of the variation in entry patterns, as compared to time-invariant unobservables. Contract impacts are largest for young individuals and those with high unobservable preference for contracts, consistent with Roy (1951) model-type selection. These selection effects are most pronounced when studying impacts on expanding firms, where the average impact of the intervention is *zero* on average in the population. At the same time, individuals with a high preference for these contracts are more likely to start businesses, even without receiving a contract.

While the government has multiple reasons for this procurement scheme, two stated goals of this policy are to democratize access to contracting and stimulate business growth (Aristy Escuder, 2016; Travieso Caraballo, 2016). With these goals in mind, we evaluate counterfactual policies that change event characteristics around the number of contracts, their value, and application costs. We find individuals are most responsive to the number of contracts (and hence potentially to their likelihood of receiving a contract) and the government faces a tradeoff. Increasing attractiveness of the events (by increasing contracts) draws in a population who would be less likely to start businesses, further democratizing access to firm creation, but at the cost of inducing less new firm creation, therefore stimulating less business growth. We provide quantitative results to understand this tradeoff.

Overall, these results show that managerial jobs can be an important path to en-

for example, Schoar (2010); Hurst and Pugsley (2011); Haltiwanger et al. (2013); Humphries (2016); Levine and Rubinstein (2017); Hombert et al. (2017); Gendron-Carrier (2018).

trepreneurship. However, numerous reasons suggest they may be under-provided in developing countries. Contract enforcement and monitoring issues of managers are cited as a reason by Indian firm owners to create these jobs only for their male family members (Bloom et al., 2013). More related to this work, if firms believe managerial experience will cause employees to create competing firms, they will under-provide these jobs and experiences. Each of these market failures would provide reasons behind the low skill accumulation from job experience in developing country labor markets (Lagakos et al., 2018).

This paper contributes to several areas of existing research. First, we contribute to the literature on the role of human capital, and, more specifically, managerial experience, to spur business creation and other measures of entrepreneurship. The existing literature has found mixed results of the importance of labor market experience (Evans and Leighton, 1989; Lazear, 2005; Silva, 2007; Humphries, 2016; Hincapié, 2017; Gendron-Carrier, 2018). This paper is most closely related to Liang et al. (2018), who—through a model and associated cross-country regression analysis—examine the role of managerial job experience in generating entrepreneurship. We see our contribution as establishing a causal relationship behind their mechanism.

Second, we contribute to research on entrepreneurship and labor markets in developing countries. Studies show mixed results—while intensive, individualized consulting benefits existing firms (Bloom et al., 2013; Bruhn et al., 2018), generalized training does not (reviewed in Quinn and Woodruff, 2019), and managerial internships benefit only those placed at high-quality firms (Abebe et al., 2019). We show business-creation effects of temporary experience in a managerial job. There have been a large number of evaluations of job experience and public sector employment programs in developing countries, but these programs are often for entry-level or low-skill employment (reviewed in McKenzie, 2017a).

Lastly, we contribute to the literature studying selection, and, in particular, self-selection in program take-up and returns for potential entrepreneurs in developing countries. For credit, Banerjee et al. (2019) shows experienced entrepreneurs have higher rates of returns, while Hussam et al. (2022) and Bryan et al. (2024) show peers and psychometric data can predict heterogeneous returns, respectively. Recent work shows self-selection in program application based on observable characteristics (Alatas et al., 2016), but self-selection in entrepreneurship has been historically more difficult to separate from other screening mechanisms used in program allocation (McKenzie, 2017b; Fafchamps and Quinn, 2017; Beaman et al., 2023). Our results showing selection in entry positively related to returns contrasts with the results in Abebe et al. (2021), likely explained by an absence of dynamic selection in this context. Our combination of lottery randomization with a selection model is most similar to Walters (2018) and van Dijk (2019).

Section 2 provides background on the Dominican procurement system. Section 3 details our data and sample, Section 4 describes the lottery-based results, Section 5 contains the selection model, and Section 6 estimates counterfactual policies. Section 7 concludes.

2 The *sorteo de obras*

In 2006, the Dominican government revamped its public procurement system by updating the protocols for its existing procurement schemes, creating new schemes, and instituting a new governing and supervisory body, the General Directorate of Public Procurement (*Dirección General de Compras y Contrataciones*, DGCP) (Travieso Caraballo, 2016). Since 2007, the DGCP has maintained the national registry of state suppliers (*Registro de Proveedores del Estado*, RPE) and managed records on all procurement processes in the country. Although each government agency is responsible for executing its own procurement processes, the DGCP establishes the protocols to follow for the different procurement schemes and publishes the processes in an online portal.

The same 2006 reform mandated the use of a new procurement scheme, the "lottery of works" (*sorteo de obras*), a true random allocation of government procurement contracts to applicants. This scheme applies to service projects within specific contractual size thresholds, which vary over time.[3,4] The steps to implement this procurement scheme are routine and shown in Figure A1. First, the procuring institution announces the opening of the procurement process, which includes posting it online. Second, all interested RPE members can sign up to participate in the process. Following the closing date to declare interest, the government institution running the process reviews all the participants and disqualifies those who do not meet the criteria. Third, all eligible participants are entered into the lottery of their choosing. Fourth, the selection of winners is a public affair. Entrants attend the lottery at a common location to enter their name into a randomization device, typically on a piece of paper placed into a transparent cylinder. A government official spins the cylinder

[3]Among other things, the new procurement scheme was meant to address the lack of democratization, transparency, and oversight of the country's bloated and corrupt procurement system (Artana et al., 2006). The program is increasingly advocated on the basis of providing greater opportunities to skilled workers who struggle to advance in their careers (Aristy Escuder, 2016; Travieso Caraballo, 2016).

[4]It is common for procurement agencies around the world to provide some equity regulations to incentivize new entrants. This is a more extreme example, but is not wholly unique. This paper shares the use of some form of randomness in procurement contract allocation with (Carrillo et al., 2023; Lee, 2021; Fadic, 2018), but is differentiated by studying different types of subjects, phenomena, and consequently *a priori* differences in mechanisms.

and pulls the winning entry (see Figure A2). Lottery winners are then required to take out an insurance policy within one week of the lottery date and begin the contractual period.

A single lottery event can offer anywhere from one to hundreds of contracts, which are often split into mutually exclusive "blocks." The blocks typically segment contracts by location, more rarely by occupation. Applicants may enter only one of these contract blocks per lottery event. Within the blocks, there are almost always multiple contracts offered. For each contract, first, second, and third place winners are chosen. If the first place winner cannot fulfill the contract, it is awarded to the second place winner, and similarly to the third if necessary. Within a lottery event-block group, all entrants have the same probability of winning one of the contracts and the first-place winners are chosen without replacement so that no entrant can receive multiple contract offers. While lottery losers are allowed to enter future events, winners are restricted from entering again until finishing their previous contract.

Eligibility requirements for the lottery events are generally consistent but may vary. Basic criteria include having relevant professional licensing (exequatur), registration as a state provider (RPE), being current on taxes, and not having an active government contract. Some lottery events impose either regional requirements (e.g., applicants must come from a particular region) or experience-based criteria (e.g., usually between 0-2 years of relevant experience), though many do not.

The costs of applying to these lottery events are nonpecuniary, primarily reflecting the opportunity cost of time. Applicants must prepare and submit paperwork to verify their eligibility and attend lottery events. Learning about the events and where to apply signifies a secondary potential cost. As there are multiple events, entrants who repeatedly enter are required to incur these application costs for each event.

Upon winning a lottery contract, the recipient assumes responsibility for constructing a pre-designed building, such as a school, as specified by the contracting agency (e.g., the Ministry of Education). The contracting institution provides the winner with building blueprints, a designated plot of land, and periodic supervision to ensure quality standards are met. As construction progresses and specific milestones are reached, the winner receives payments. The entire construction process is managed by the winner, who must handle budgeting, planning, procuring materials and labor, navigating bureaucratic procedures, managing workers, and making decisions to drive the project to completion.

In interviews conducted by a televised government program and in a focus group we hosted in 2018, winners described the contract as broadly beneficial, one even describing it as the “opportunity of a lifetime,” given limited prospects in the construction sector.[5] As shown in Table A1, when we surveyed winner and nonwinner lottery participants in 2019 about the potential benefits of winning a contract, 83% indicated that winning “helps the individual or firm learn construction skills.” This response suggests both a skills gap in the DR’s construction job market and an expectation for on-the-job learning after winning a contract. Additionally, many respondents believed that winning a contract would expand their social network for future work (76%), improve their chances of securing a public contract (64%), and make it easier to find private sector employment (60%), suggesting benefits from the bundle offered by the contract. We test each of these channels in Section 4.2.

Figure 1 shows several important features of the events. Panel A presents a histogram of entrants and winners across the lotteries. The ratio of entrants to winners is large: the contracts are oversubscribed by a factor of 26, indicating a high revealed preference valuation. More contracts were offered in the years 2012-2014 due to a high volume of government infrastructure spending corresponding to a school construction drive. Panel B documents the exact timing of lotteries and number of contracts offered in more detail. There are two salient details from this figure. First, there are a number of lottery events that happen over time. Second, the majority of contracts were from four of the lottery events.

The contract sizes vary greatly, from tens of thousands to millions USD, with the average project size around \$640,000 (Panel C). Within each lottery event-block, there is still considerable variation across the contracts (Panel D). This variation is random to individuals since they apply at the lottery event-block level. A one SD change in within event-block contract size, about \$250,000, maps directly into income SDs for the winner of \$25,000, since the winners’ pay is budgeted for 10% of the contract size.

3 Data

We collected most information on lottery events through “freedom of information” requests to the relevant government agencies responsible for the lottery procurement calls (e.g., Ministry of Education, Ministry of Public Works and Communications). From these agencies, we obtained records of all individuals who applied, qualified, and those who were randomly ranked in first, second, or third place in the lottery events. Additionally, from the Ministry of Education, we gathered data on the completion percentage for each awarded

[5] Testimonies from winning engineers can be viewed via this link and this one (last accessed on 21 August 2023). In one interview, a participant noted that, thanks to the job, “there are opportunities for everyone who previously could not participate because the conditions for entering the competition were somewhat prohibitive,” highlighting the positive value of the contract experience.

contract. These records form the basis for constructing the randomization and sample, as well as for use in our outcome analyses to study future lottery entry choices.[6]

To better understand the population of possible entrants and measure selection in entry to the lottery events, we supplement the above data with two primary sources. First, we collected data on all individuals who were registered as state providers with the DGCP at some point between January 2007 and December 2018. Second, we obtained data on all licensed architects and engineers who hold an *exequatur*, the official professional license granted by the Presidency's Legal Office upon completion of a tertiary degree. This dataset provides a comprehensive record of all individuals eligible to participate in the lottery events based on their professional qualifications. Together, these datasets define our samples of potential entrants and include information used to understand the participation decisions, such as licensing dates and other demographic information.

To provide our main outcomes, we match these samples to fiscal data from the Directorate General of Internal Revenue (DGII) using unique national (tax) ID numbers, along with distinct names. For these individuals, we track gross, net, and wage income, as well as employer and number of employees (if applicable). We discuss our construction of gross and net income measures in Section A1. Additionally, we gather demographic information, such as age and domicile, from both DGII and DGCP records.

From DGII, we further compile an annual dataset tracking firm holdings for each individual in our sample, including the start and end dates of each holding, the firm's establishment date, and ownership stakes, spanning the years 2010 to 2018. We then collect balance sheet information on these firms including annual firm income, profits, and size. These data comprise our main measures of entrepreneurship and firm quality.

We supplement these firm-holding records with data from the National Office of Industrial Property (ONAPI), linking intellectual property records—such as commercial trademarks and business signs—to individuals and their firm holdings. We show our main results are robust to these alternative measures of firm ownership and activity in Section A2.

We match individuals in our samples to every procurement process overseen by the DGCP since 2007. The DGCP provided detailed information, including the contract start date, contracting institution, contract ID, procurement scheme, contract amount, and the winning supplier. This data allows us to study future work with the government.

We link all administrative datasets in this project using one of two unique administrative identifiers: the registered state provider ID or the national individual/firm ID. For records

[6] Some minor lottery events, particularly in recent years, are missing from our data.

missing these IDs (27.5% of event participation records), we string match full names to the dataset of registered state providers and the national IDs, both of which have high name uniqueness. We are unable to match 0.9% of the sample to a unique identifier, resulting in no tax information for these cases. We detail the matching procedure in Section A3.

To complement the administrative data, we conducted a survey, stratifying the random sampling by the lottery event-block, in line with the program's design. We sampled 2,038 individual-by-event observations from 1,925 unique individuals who participated in lottery events between 2012 and 2015. Within each event-block, we ensured an equal number of lottery winners and nonwinners, with a minimum of 10 nonwinners per block. The survey, conducted between late April and early July 2019, took place over the phone or in person. Ultimately, we surveyed 716 unique individuals, corresponding to 765 individual-by-event observations. The survey response rate was 36.8%, and response rates were similar between treatment and control groups. The survey data affords us more detailed information on mechanisms, particularly regarding direct management experience and loan activity.

We restrict our main analysis sample to three main lottery events in the years 2012-2013 based on the condition that we observe completion records and the median contract finished within two years from the lottery date.[7] We chose this to minimize the likelihood of overlap between time directly working on the government contracts and the time to measure outcomes afterward. Our main analysis focuses on individual entrants rather than firm entrants because of the sample size differences, the different outcomes for these two agents, and the large mean level differences in outcomes. If we did not, the individual sample would dominate the results due to their higher representation (87% of the sample).

Table 1 summarizes the characteristics of the entrant population in the full sample in column (1) and in the analysis sample in column (4). Each observation corresponds to our level of analysis: the entrant of a lottery event-block. All individual entrants are included. We describe the entrant population for the analysis sample here, which accounts for 32.5% of the overall sample. On average, entrants have entered 0.8 lotteries previously. The majority of entrants are male, which reflects that men are more represented in the engineering and architecture sectors. The vast majority are taxpayers in the year prior to the lottery (82%) and have never won a previous contract with the government (98%). In the year prior to the events, 10% of individuals are owners of firms.

[7] Panel B in Figure 1 shows there were 5 large events in addition to a number of smaller events.

4 Lottery analysis and main results

4.1 Empirical strategy

To estimate the effect of receiving a temporary high-value managerial position from the government on future career and business outcomes, we exploit the random assignment induced by the *sorteo de obras*. Our main specifications take the following form:

$$y_{iebt} = \beta 1[\text{winner}]_{ieb} + \gamma_{eb} + \varepsilon_{iebt} \tag{1}$$

where y_{iebt} is the outcome of interest at time t for entrant i who applied to the set of contracts at event e in block b. The fixed effect γ_{eb} restricts comparisons within the set of individuals who applied in the same lottery event and block, and hence, had the same probability of receiving a contract. The indicator $(\text{winner})_{ieb}$ is a 1 if individual i won one of the lotteries (i.e., came in first) in block b at event e and 0 if not. Time or period t is defined to be relative to the year of lottery occurrence. Consequently, period 0 and 1 correspond to the year of the lottery event and the year afterwards, respectively.

The program's design, a series of lottery events, is analogous to the setting in multiple recent papers (Cellini et al., 2010; Gelber et al., 2015). We adopt their data structure and create an entrant-by-application level dataset. This corresponds to multiple stacked panel observations for repeat entrants, with the observations in reference to the relative period structure. Across all specifications, we cluster at the individual level to account for correlated outcomes that arise because of the inclusion of the same entrant across multiple lotteries. All main estimates are intention-to-treat or the reduced-form of a 2SLS model in which we instrument for contract take-up.[8] Figures showing the evolution of treatment effects over time are run as period-by-period regressions of Equation 1.

Because lottery winners are restricted from entering other events if they have an unfinished contract, winning may affect the likelihood of future lottery participation. As a result, Equation 1 measures both the effect of lottery winning on subsequent career outcomes and the dynamic effect on future lottery winning, although impacts on future winning are limited.[9,10]

[8] The first stage is close to one; we discuss this later in this section.

[9] Lottery winning generates similar impacts on ever winning as winning in this event because the control groups win probability is low (see Figure A3).

[10] The approach developed in Cellini et al. (2010) can estimate treatment effects that remove the channel of future lottery winning. Their estimator comes at the cost of imposing assumptions on treatment effect homogeneity. Since the effect of lottery winning on subsequent lottery winning is modest, estimates using

The lotteries generate randomized priority lists with up to three positions, but our research design only exploits the winning position. As discussed in de Chaisemartin and Behaghel (2018), our estimator, labeled the "Initial Offer" (IO) estimator, is consistent but may result in slight efficiency losses. The overall first-stage coefficient from a regression of contract take-up on lottery winning is 0.94, so efficiency losses are negligible.

We modify the regression above to look at intensive margin variation in contract size. Our main specification takes the following form:

$$y_{iebt} = \beta 1[\text{winner}]_{ieb} + \tau 1[\text{winner}]_{ieb} M_{ieb} + \gamma_{eb} + \varepsilon_{iebt} \quad (2)$$

where M_{ieb} is an intensive margin characteristic of the job, such as the awarded contract size. Our interest in contract size is to understand whether there are potential capital and scale effects on future outcomes. For example, a small contract may not allow individuals and firms to overcome financial capital constraints that would be ameliorated by larger contracts. As discussed in Section 2, there is significant variability within an event-block; a one standard deviation difference in contract value is about one-third of the average contract value. This and other intensive margin characteristics of contracts are orthogonal to unobserved factors of the participants conditional on the event-block fixed effects, γ_{eb}. This is because multiple contracts are allocated randomly within the same event-block and applicants are not permitted to apply to specific contracts within an event-block.

We validate the randomization of lottery winning in Table 1. In columns (3, 6) of this table, we regress entrant observable characteristics on lottery winning and event-block fixed effects. Lottery winners are not statistically different from nonwinners along any of the dimensions we test. A joint test of whether these observable characteristics are correlated to lottery winning fails to reject the null hypothesis of no correlation ($p = 0.42$ overall and $p = 0.66$ for individuals in the analysis sample). Analogously, contract size is uncorrelated with the observable characteristics of lottery winners ($p = 0.94$).

This randomization procedure is unique for providing true random variation in important intensive-margin characteristics of contracts as well. Even under our fortuitous conditions, the exclusion restriction required for our estimates of intensive-margin effects may still be violated if other contract characteristics vary with contract size (e.g., the government provides additional assistance to these candidates). We are unaware of any such treatments.

that approach are very similar to those reported in the paper.

4.2 Main results

4.2.1 Career outcomes

Figure 2 shows the main sectoral effects of winning a lottery contract, illustrating the timing of the treatment and evolution of outcomes. Panel A shows the effect of winning a lottery contract on the likelihood of having an open contract (i.e., a started but not yet completed contract). The likelihood of having an open contract is highest in periods 0-1, falling to 35% by period 3. The rate of contract completion is slower in the final 2 periods, with 23% of lottery winners still having an open contract 5 years later. Thus, some individuals have not yet finished by the end of the sample period. Delays are partly due to the government's challenges in implementing this large-scale program, including delays in land acquisition and payments.

Panel B shows the evolution of wage employment. Although some manage contracts while working another job, many leave formal wage employment. In period 1, lottery winners are 7.2 percentage points less likely than the control group to be working for a wage in the private sector in that year. This decline in formal wage employment continues and stays relatively steady. After 5 years, lottery winners are 12 percentage points less likely to have any wage employment, a 24% reduction. This persistent decline in wage employment does not mirror the trend in project completion rates. Even as many lottery winners finish their contracts, they do not return to wage employment.

In Panel C, we see that lottery winners are more likely to start and own firms.[11] To reduce the possibility of double-counting firm creation in the case of multiple firm owners, we conservatively scale firm outcomes by firm ownership shares and the share of months in the year when the individual is an owner.[12] Lottery winners have similar firm ownership as nonwinners in periods 0-1, when they are most actively working on the contracts. This shows firm creation is not mechanically correlated with receiving a government contract, a point we return to in Section 4.4. Instead, in year 2, when individuals begin to finish

[11] Panel B in Table A2 shows that firm ownership and firm creation are almost the same in this sample. We define firm creation as registering ownership shares of a firm within 6 months of the firm's registration of existence. Using this definition, firm ownership and firm creation lead to indistinguishable estimates of the main effects, 0.078 and 0.069, respectively.

[12] The number of firms owned, called FO, is calculated using the number of firms owned, N, the share of firm j the individual owns, S_j, and the share of months in the year that the individual was a firm owner of firm j, M_j. We then construct the number of firms owned accounting for the share of ownership and time of ownership as the sum over their product, $FO = \sum_{i=j}^{N} S_j \cdot M_j$. This procedure more accurately represents firm ownership in the case of non-unique owners and reduces concerns of double-counting if, for example, lottery winners are more likely to co-own firms.

contracts, firm ownership becomes significantly positive. By period 5, lottery winners have 0.053 more firms on average, with no evidence of mean reversion or fade-out.

In Panel D, we examine the evolution of impacts on owned firm size, scaled by firm ownership shares. This measure corresponds to aggregate increases in employment as a function of firm ownership, excluding the owner. In periods 0-1, we find that the estimated effects on owned firm size are negative and insignificant, indicative of lottery winners substituting away from entrepreneurship in the short-term while focusing on their open contracts. Impacts on firm size become positive and highly statistically significant in periods 4-5 when the majority of contracts have finished. Throughout the study period, the effects continue to grow and seem to lag the effect on owned firms, as one would expect from expanding businesses.

Table 2 provides point estimates for measures of firm creation and firm characteristics, aggregated to the individual level and unconditional on ownership, following McKenzie (2017b).[13] Columns (1-2) show lottery winners are 7.8 percentage points (22%) more likely to ever become firm owners and become owners of 0.15 more firms, a 31% increase. Column (3) restricts firm ownership to 2018 and scales by the mentioned shares, finding a point estimate of 0.06, indicating that newly created firms are about one-third owned by a lottery winner. Columns (4-6) report firm characteristic outcomes. We see that winners are responsible for businesses with higher employment and higher incomes, both increasing by over 40%. We see no differential effects on firm profits, but the relatively low control group means suggest possible tax avoidance.

We provide further detail on firm creation and firm size in Panel A of Table A2. Using five firm size definitions, ranging from no employment threshold to at least 10 employees measured in 2018, we find lottery winning increases ($p < 0.05$) any ownership of firms with at least 0, 1, 3 employees, and positive but statistically insignificant increases for firms with at least 5 and 10 employees. Proportional increases range from 13-26% across all definitions. However, due to the rarity of larger firms in the population (7% for 5 employees, 4% for 10 employees), we cannot reject the null of no effect.

The newly created firms vary widely and are not exclusive to the construction sector. Panel B in Table A2 shows that while lottery winners are much more likely to become owners of construction firms, they are also more likely to become owners of firms in commerce and other sectors. Thus, this contract experience potentially lets them expand into other sectors where their skills are useful as well.

[13]Individuals who do not own firms receive a 0 for firm characteristics.

Panels E and F in Figure 2 report the effects of lottery winning on individual income and profits. In Panel E, throughout all periods 0-5, results on total income are positive and statistically significant. In periods 0-2, total income is highest because individuals are being paid the full contract amount as income. These effects decay to about a tenth of the size by the final period. In Panel F, net income differences are large in the first periods (approximately 200-300%), but the difference in net income decays and incomes are similar between the groups in period 5. Net income is somewhat hard to measure because in early periods some contract recipients report no costs despite clearly having project-associated costs (for further details, see Section A1). Furthermore, individuals who start new firms may be better able to shield their firm profits or income.[14] In general, these results suggest important benefits for the incomes of individuals in the sample that fade-out over time, potentially suggesting nonpecuniary benefits to entrepreneurship.

We document additional impacts on managerial experience and financial access using survey evidence. Columns (7-8) in Table 2 show lottery winners have more direct managerial experience overseeing more temporary employees and more subcontractors, measured over their entire careers. These experience increases are further supported by reports that experience itself was one of the most important features of the contracts for recipients (see Section 2). Panel D columns (1-2) in Table A2 also show contract receipt does not lead to statistically significant increases in the number of loans or loan amounts to businesses, suggesting no substantial crowding out or crowding in of financial capital. That said, we have limited power and cannot reject medium-sized increases in number of loans (i.e., 35%).

4.3 Heterogeneity and main mechanisms

In the previous section, we observed a shift from wage employment to starting new and expanding firms among lottery winners. The contracts provided more income and managerial experience than winners would have otherwise acquired, consistent with their roles as temporary managers that bundle income and business-like experience.

As a guide to possible mechanisms, we surveyed the winners about the benefits and drawbacks of participating in the lotteries (Table A3). The most commonly reported benefit of winning (62%) is valuable work experience and skills, potentially related to direct management skills or indirect ones (e.g., budgeting). Moreover, 42% indicated increased confidence in their abilities, noting that working on the lottery projects demonstrated their

[14]For example, in the United States, Feldman and Slemrod (2007) find that the self-employed underreport income to tax authorities by a third.

capacity to undertake larger ones (we consider this part of overall experience effects).[15] On the financial side, 55% mentioned they now have more money to invest in their careers, suggesting potential income benefits from the contracts as well. The table suggests other mechanisms, such as improved connections or ability to get new contracts are unlikely to be important, and few individuals report negative effects of the contracts.

To more directly study these mechanisms, in this section we study effects across heterogeneous groups and use alternative tests to help reveal whether income or managerial experience are more responsible for driving these changes. While the evidence points more strongly to the importance of managerial experience, since we study a bundled good, it is difficult to fully rule out effects of capital, nor would we want to. In Section 4.4, we discuss and show evidence against three prominent alternative mechanisms.

4.3.1 Heterogeneous impacts by baseline covariates

We begin by investigating the effects of contracts based on the individual's age at the time of the lottery, as age is a proxy for experience. In Table 3, we bin the age of the individual at the time of the lottery events into three bins (20-34, 35-49, 50-64), which roughly correspond to sample terciles. All outcomes are measured in 2018, the last post-period year of data.

The effects of the contracts are highly heterogeneous across age bins. The youngest individuals (20-34) experience the largest decline in formal wage employment (19 percentage points), followed by middle-aged individuals (11 percentage points), and the oldest group (6.6 percentage points). We reject that the effects are constant across the age bins ($p = 0.02$). Analogously, we find the largest increases in firm ownership for individuals who are young. On average, contract receipt leads to 0.13 firms created for the youngest group, 0.04 for the middle group, and an insignificant 0.025 for the oldest. Young recipients also own firms with higher revenues and employee counts, although the effect on profits is positive but not statistically significant. The results for the two older age groups are smaller and insignificant. Furthermore, the heterogeneity is not driven by differences in contact take-up, as column (6) shows little difference in the first-stage across the bins.

To provide a point of comparison to recent work studying the age of entrepreneurs in the United States (Azoulay et al., 2018), we examine rates of becoming a majority-stake firm owner further disaggregating by age for both control and treatment groups in Figure 3.[16]

[15] A literature discusses the importance of self-confidence for successful managerial and entrepreneurial endeavors (e.g., Church, 1997; Robertson and Sadri, 1993; Garaika et al., 2019; Hayward et al., 2010).

[16] Azoulay et al. (2018) primarily defines business founders to be owners who also work at the firm. As our

Among nonwinners, the relationship between age and firm ownership follows an inverted-U shape, with individuals in their mid-to-late 30s most likely to start firms. In contrast, lottery winners aged 20-26 show the largest increase in business ownership, followed by those aged 27-40. Older winners do not show a statistically significant increase. These results suggest that the youngest individuals seem to be the most constrained from business creation by the benefits provided by receiving a contract.

Table A4 expands the set of possible other heterogeneous impacts, and finds that age is the most important. We report reduced-form and IV models of heterogeneous effects by baseline characteristics one-at-a-time and jointly for the covariates of age, sex, previous income, past formal employment, and whether the individual is from the largest city. The IV models instrument for contract take-up with being a first-place winner. The RF and IV models results are quantitatively and qualitatively almost identical, indicating that variation in contact take-up does not explain heterogeneity. All variables besides age are insignificant both separately and jointly, in contrast to the strong effects by age in both specifications. Females have insignificantly positive heterogeneous estimates in both specifications. The lack of a negative effect for females contrasts other influential studies that shows females have low returns to interventions such as credit (De Mel et al., 2008). Heterogeneous impacts by past income are economically small when viewed as a separate interaction—a 1% decrease in past income increases the likelihood of being a business owner by 0.003 percentage points—and positive in the joint estimation. This is inconsistent with the program inducing firm creation for a set of individuals with credit constraints that arise from low past incomes. Past formal employment is also insignificantly related to outcomes.

Together, we find strong positive impacts for young individuals, while non-differential impacts for individuals based on sex, previous income, geography, or past formal employment. This points towards age-based factors such as experience or greater flexibility in occupational transition as potentially more likely to drive the impacts on business creation, which we further investigate below.

4.3.2 Heterogeneous impacts by event characteristics

To provide additional evidence on whether capital or experience is a driving factor in firm creation, we exploit heterogeneity in contract size. Since individuals do not apply for a single contract but rather a set of possible contracts, there is built-in heterogeneity in contract size that is randomly assigned to winners. A larger contract should provide more capital

data does not enable this, we increase certainty of the founder by restricting to majority-ownership stakes.

to a potential entrepreneur and thus may be more likely to relax any existing capital constraints. This is not a clean test, as managing a larger contract may provide more experience as well. That said, the difference in experience gained may not be large, as the projects are relatively homogeneous in terms of the steps necessary for completion. Each contract has a large fixed component of finding workers, materials, subcontractors, and budgeting, with a variable component that is the daily management and final execution.

Table 4 presents heterogeneous effects of contract size as measured in $200,000 increments.[17] We evaluate the effect of contract size on entrepreneurship outcomes including becoming an owner, firm income and firm size, on financial capital measured by net income, and on management experience measured by days spent in the contract, number of subcontractors managed, and number of temporary employees managed. In column (1), we show that the contracts lead to large increases in the likelihood of becoming a firm owner, but contract size has no effect on becoming an owner. The point estimate is small and opposite-signed than what we would expect if firm creation were driven by increases in capital. Columns (2-3) also show that contract size has little effect on aggregated firm income or firm size. In column (4), we show that contract size greatly increases net income received by the recipients.[18] These patterns show that greater potential financial capital does not translate into measurable entrepreneurial outcomes, but does not fully rule out that small amounts of capital may have beneficial impacts.

In columns (5-7), we show contract size is correlated with some proxies of managerial experience and not others. Column (5) shows that contract size is not correlated to average days spent in the contract. Column (6) shows contract size is also not correlated with number of subcontractors managed, although the main effect is not statistically significant due to the much smaller sample size of the survey. This provides some evidence in favor of the possible benefits of managerial experience in a contract, especially in terms of setting up a new project, for firm creation outcomes. Not all measures of managerial experience would be invariant to contract size. Indeed, column (7) shows the number of temporary employees is directly affected by the contract size because daily management tasks likely scale with contract size. Numerous other mechanisms are part of the bundle of benefits of a managerial job, such as positive updating about self, that we cannot directly test.

Figure A4 shows the effects of larger contract size on becoming a firm owner do not appear to be non-linear. We estimate the effects of contract size by quintiles of contract

[17] In 2018 exchange rates, this corresponds to about ten million Dominican pesos.

[18] The outcome variable is not winsorized in order to not impose a form on the relationship between net income and contract size.

size. The pattern is somewhat noisy and not indicative of a clear pattern, with some small contract sizes and some medium-sized contract sizes having large effects on firm creation. We cannot reject that the effects are homogeneous across the quintiles ($p = 0.59$).

4.4 Alternative mechanisms

4.4.1 Government-specific firm creation

One hypothesis is that individuals may start a firm to reduce personal business liability on their randomly-assigned government contract or to win future government contracts. We do not find evidence for either of these possible mechanisms.

Two pieces of evidence indicate that firms are not started to reduce personal liability during the lottery contractual period. First, at the beginning of the contractual period, the contract recipient is required to take out an insurance policy related to the work in the first week after the event. They are not allowed to have a firm sign the insurance policy on their behalf, so they are not allowed to reduce their liability in this manner. Figure 2 shows that lottery winners do not start firms at the time of the event, providing evidence that there is no incentive to start businesses at this time. Instead, lottery winners begin to start new firms two years after the lottery contracts were awarded, which corresponds to the time when recipients begin to finish their contracts.

In Panel C of Table A2, we examine the likelihood that the new firms of lottery winners are created to win more government contracts in the future. In column (1), we show that lottery winners start more firms. In columns (2-5), we analyze whether these firms have won more government contracts, aggregating firm outcomes to the owner level. We show that despite having many more firms, lottery winners are not more likely to own a firm that has ever won a randomly assigned government contract nor non-randomly assigned government contract, and have no difference in their cumulative value of government contracts. Additionally, the control group means show that very few of these firms have ever won any government contracts, indicating more broadly that this does not appear to be an important reason behind firm creation for this sample. Taken together, it does not appear that lottery winner firms are created to work further with the government along these major dimensions. Additionally, our results suggest one of the possible benefits of this program, stimulating a "competitive fringe" for the procurement process, does not occur.

4.4.2 Increasing business connections

Contracts in this context are allocated by randomly assigning applicants to a set of jobs. These jobs are independent but may allow entrants to interact, see each other's quality, and develop networks of firm co-ownership. Firm co-ownership is common, as is true of firm ownership around the world.[19] We show in two ways that it is unlikely that the business creation effects are driven by introducing a set of new potential firm co-owners.

The first piece of evidence against this channel is the comparison of any firm ownership compared to majority-share firm ownership. If co-ownership is an important factor of our results, we would expect individuals to take small non-majority ownership shares in firms. Panel B in Table A2 shows similar effects of lottery winning on ownership of any firm (7.8 percentage points) as when considering majority firm ownership (5.1 percentage points).

We also evaluate this potential mechanism directly by studying whether lottery winners in the same block are more likely to own firms jointly. We adopt a similar strategy as other papers that have looked at endogenous future group work outcomes as a function of schooling or place of residence (Zimmerman, 2019; Bayer et al., 2008).

We use a difference-in-differences approach, comparing whether lottery winners are more likely to co-own with other winners in the same event-block, while including co-variates of winning and event-block applicant groups to control for the impact of winning generally and rates of coownership between applicants without winning, respectively. The data is arranged at the entrant by entrant pair level. Individuals are considered linked if they are co-owners of the same firm, as derived from the firm ownership dataset. We include either full entrant block by entrant block fixed effects or separate fixed effects for the block of the application of each entrant to control for differences in both applicant group differences and application group interaction differences. We cluster standard errors at the entrant block by entrant block level to allow arbitrary correlation within groups.

Across three specifications, Table A5 shows that winners from the same lottery group have positive but insignificant differences in their firm co-ownership likelihood. The null effects suggest that firm co-ownership is not much more common among the lottery winner group than other potential firm co-owners and is not the driver of the firm creation results.

4.5 New firm quality

The welfare implications of new firm creation are dependent on firm quality. Firm creation may be entrepreneurial and lead to thriving businesses, or, at the other extreme, be

[19]The average firm ownership share in the data is 41%.

undesired and act as disguised unemployment. We assess these possibilities in this section.

Table 5 examines firm quality. We compare lottery winner firms to nonwinner firms at the firm level, making this a primarily descriptive rather than causal exercise. Differences arise from two potential sources. First, new, marginal firms are created as a function of contract experience. Second, firms that would have existed even without contract experience may be different. These forces may be in opposition in the case of negative selection but positive in the case of within-firm changes, or in the same direction if this experiment induces higher quality individuals to become entrepreneurs.

To make this firm comparison, we restrict our analysis to firms created by entrants in the same event-block and same year, to limit basic selection and time confounders. We follow the main measures of newly created firm quality used in Hombert et al. (2017), namely any firm hiring and firm survival in the first two years of the firm's existence. The rationale for these measures is that early firm actions, such as hiring, are predictive of firm hiring in the future. In the firms created by lottery nonwinners, 27% have hired within the first two years and 72% have survived.[20] Firms created by lottery winners are 4.8 percentage points or 18% more likely to have hired ($p < 0.10$) and are 5.1 percentage points or 7% more likely to have survived ($p < 0.05$) in their first two years. Revenues of firms created by lottery winners are insignificantly different from firms created by nonwinners. Together, the results suggest some important benefits to existing firm quality.

5 Evaluating selection into the program

In the previous section, we find that receiving a contract leads to large and persistent impacts on business creation and ownership. However, these effects are conditional on the sample of individuals who applied, raising important questions for policy and the literature on entrepreneurship-enhancing interventions. First, how important is selection in program application for the observed business creation effects? Second, if selection is important, how could the government increase entrepreneurship, and would this involve an equity-efficiency tradeoff?

To answer our two main questions, we analyze individuals' participation choices into lotteries to recover participation preferences in the population. We then document how treatment effects vary based on observed and unobserved heterogeneity in the decision to participate. Finally, combining both of these steps, we estimate the impact of counterfactual government program redesigns on individual entry and business creation patterns. We first

[20]The average firm size is 2 in the firm's second year.

motivate with descriptive evidence.

5.1 Descriptive evidence

To understand entry patterns along observable characteristics, we compare entrants to nonentrants drawing from exequatur records, which include all licensed civil engineers and architects. Figure A5 shows lottery entry rates are high across the population but recently licensed individuals are more likely to enter the lottery than individuals who have been licensed for more time. Specifically, 80% of those licensed between 2010-2012 applied at least once, as did about 50% of those licensed 20 years earlier. Overall, 57% of the eligible population licensed in the 20 years before the start of the lotteries participated at least once.

Since Figure 3 shows significant treatment effect heterogeneity by age, these patterns of observable selection are likely to be quantitatively important for the overall impacts of contract receipt. They also suggest *positive* selection along observable characteristics.

To further motivate the importance of self-selection and entry patterns for treatment effects, we investigate heterogeneous treatment effects by the number of times an individual previously entered a lottery event. Table A6 shows lottery-based reduced form regressions, as in Section 4.3.1. Individuals who enter more events are more likely to become firm owners after winning a lottery than those who enter fewer events. This is indicative of positive selection into the program in ways correlated with gains to firm creation (and along potentially unobservable characteristics). However, the reduced-form nature makes it difficult to fully evaluate this proposition. It is also unclear to the extent that the government can change the program attributes to induce differential selection and how changes to program allocation would affect equity in opportunity. The subsequent sections enable us to understand these forces more clearly.

5.2 Setup

We model lottery event application choices to characterize selection into this government program. Potential applicants i include individuals who have participated by entering any event and also non-participants from the set of licensed civil engineers and architects, who could have participated. These individuals may apply to the sequential events $e \in \{1,\ldots,E\}$. Within each event, applicants are required to make a mutually exclusive choice over the pre-grouped sets of contracts called "blocks", where B_e identifies the number of blocks in event e exclusive of the outside option. We denote blocks $b \in \{0,\ldots,B_e\}$, with $b = 0$ as the outside option of not participating. The number of blocks varies across events from one to thirty-two. We denote application choices as $A_i = (A_{i1},\ldots,A_{iE})$. For applicants to a certain block b in event e, the available contracts are randomly allocated with probability π_{eb}. A subsequent choice of contract acceptance is not modeled because almost all applicants accept, and hence the choice problem is exclusively over which block to choose. The outcome—specifically, whether an individual starts a firm or starts a firm with employees—is Y_i.[21]

5.3 Preferences

The utility of receiving a contract in event e and block b takes the form

$$U_{ieb} = X_i\beta + \delta W_{ieb} + \phi_{eb} + \theta_i + \nu_{ieb} \tag{3}$$

where X_i is a vector of time-invariant or predetermined observable characteristics of individual i, W_{ieb} are characteristics of the event-block that may also interact with the individual (e.g. distance from i to contracts in block eb), ϕ_{eb} is an event-block-specific utility accounting for observable features (e.g. number of contracts) and unobservable features (e.g. press coverage) of the events, and θ_i is time-invariant unobserved heterogeneity in individual i's preference for government contracts. The utility of not applying for a government contract in a specific event e is normalized to have mean 0

$$U_{ie0} = \nu_{ie0}. \tag{4}$$

As written, unobserved preferences, (θ_i, ν_{ieb}), are assumed to be additively separable and independent from covariates. Individual-specific heterogeneity in preference to have a contract, θ_i, is modeled to follow a normal distribution with mean μ_θ and variance σ_θ^2. Application-specific preference shocks, ν_{ieb}, are assumed to be iid draws from an extreme-value distribution across alternatives, events and people. This error structure provides the scale-normalization of the model. The inclusion of θ_i relaxes the assumption of uncorrelated preference shocks between contract lotteries within an event and across events as opposed to the outside option of not applying.

Individuals who win a lottery are not permitted to enter a lottery until they finish their past lottery contract, if any. In practice, this means few individuals are eligible to enter after they win a contract and empirically few do. Consequently, we do not include possible event choices after an individual is randomly chosen to win a contract.

[21] We restrict the sample to individuals licensed in the exequatur before the lotteries started in 2012 so that the sample has common application choices; this is 98.6% of the sample in the lottery analysis in Section 4.

5.4 Choices

Applicants make choices in each subsequent lottery event e based on their preferences. All applicants must engage in the costly process of learning about the events and possible contracts, preparing and submitting relevant documentation, and attending lotteries. We model this as applicants to the lotteries paying a fixed cost, C, for each application. Contracts are awarded with probability π_{eb}.[22] We assume conditional independence of choices after conditioning on the preference variables X_i, W_{ieb}, ϕ_{eb}, and θ_i, which rules out forward-looking behavior.[23] The choice facing potential applicants to event e is

$$A_{ie} = \arg\max_b \pi_{eb} U_{ieb} - C\,1[b \neq 0]. \tag{5}$$

The unobserved heterogeneity in preferences implies this is a form of mixed logit. Mixed logit is a workhorse model in the discrete choice literature because it can approximate any random utility model (McFadden and Train, 2000). To identify this equation, we observe variation in choice characteristics over time for individuals and variation across different events. Choices are assumed to be done period-by-period, so the repeated choice structure aids identification. Finally, we assume a parametric structure on the distribution of the unobservables, as is common. The formulation of the model with a single time-invariant random coefficient for the choices of applying to lotteries is analogous to a nested logit with one nest for all contract lottery choices.

5.5 Choice estimation

The likelihood of an individual making a series of application choices $a = \{a_1, a_2, \ldots, a_E\}$ based on observable characteristics is

$$L(A_i|X_i, W_i) = \int \prod_{e=1}^{E} \left[\frac{\exp(A_{iea_e})}{1 + \sum_{j=1}^{B_e} \exp(A_{iej})} \right] dF(\theta|X_i, W_i) \tag{6}$$

$$= \int q(a|X_i, W_i, \phi_{eb}, \theta) dF(\theta|X_i, W_i). \tag{7}$$

Since there is no closed-form solution for this integral, we estimate the model by Maximum Simulated Likelihood (MSL). The MSL estimator of $\Lambda = (\beta, \delta, \phi, \mu_\theta, \sigma_\theta^2)$ is

[22] The outside option is assumed to have probability 1.

[23] We view this assumption as relatively innocuous: forward-looking behavior in this context is unlikely because of the low probability of winning any one event, and because the existence or characteristics of future events are not known.

$$\Lambda^{MSL} = \arg\max_{\Lambda} \sum_i \ln \left(\frac{1}{R} \sum_{r=1}^{R} q(a|X_i, W_{ieb}, \phi_{eb}, \theta_i^r) \right) \tag{8}$$

where R is the number of simulations. We set R = 50; we observe little difference by increasing the number of repetitions.

The above procedure estimates coefficients on observable parameters and coefficients of the distribution of θ_i, namely μ_θ and σ_θ^2. To recover θ_i, we estimate posterior means of θ by simulation that are conditional on an individual's sequence of choices and the relevant choice characteristics[24]

$$\theta_i^* = E[\theta | A_i, X_i, W_{ieb}, \phi_{eb}]. \tag{9}$$

5.6 Outcomes

From the model of participation choice, we recover both observable and unobservable factors affecting entry. We include these parameters into an outcome equation characteristic of generalized Roy models. The estimating equation takes the form:

$$Y_i = \mu_0 + \mu_0^X X_i + \mu_0^\theta \theta_i^* + 1[\text{Winner}_i][(\mu_1 - \mu_0) + (\mu_1^X - \mu_0^X) X_i + (\mu_1^\theta - \mu_0^\theta)\theta_i^*] + \varepsilon_i \tag{10}$$

We consider two main outcomes, Y_i, whether an individual becomes a firm owner between 2012 and 2018, the full study period, or becomes a firm owner with at least one employee over this period. Parameter μ_0 measures the average effect without winning and $(\mu_1 - \mu_0)$ measures the population average differential effect of being a lottery winner since X_i and θ_i^* are demeaned. The relation between becoming a firm owner and observable characteristics are captured by μ_0^X, and as a function of unobservable characteristics, reflecting unobserved absolute advantage in firm ownership, is measured by μ_0^θ. Differential effects of lottery winning across observable characteristics are measured by $(\mu_1^X - \mu_0^X)$. Unobserved comparative advantage in lottery participation is measured by $(\mu_1^\theta - \mu_0^\theta)$. We compute standard errors using the methodology of Murphy and Topel (2002), which accounts for θ_i^* being a generated regressor for inference.

[24] While θ is independent of (W, X), they are not conditionally independent given A.

5.7 Identification

The outcome model in Equation 10 controls for selection using observable characteristics and a control function for unobservable factors. This type of selection model allows direct estimation of the population average treatment effect (Wooldridge, 2015) and heterogeneous effects of lottery winning across observable and unobservable characteristics. Either through reweighting or through simulation of counterfactuals, the parameters can be used to estimate effects for other subpopulations of interest, as we do in Section 6.2.

We discuss how the outcome equation is identified from three sources in this section and provide a more formal description in Section A4.1. First, the lottery randomization generates exogeneous variation in contract win for individuals with the same win probability. While this is broadly useful for estimating treatment effects for lottery entrants for the purposes of understanding heterogeneous effects, it also allows us to directly estimate the *untreated* mean outcome across the full population distribution of contract preferences from lottery losers. This lowers the burden for other sources of identification to identify aspects of population outcomes.

Second, we employ across-event information for identification. This nests two underlying assumptions. We first assume the menu of choices elicits information on individuals' valuation of contracts, which is implicit in the choice model formula in Equation 3. This requires individuals to have common valuations of characteristics of contracts, such as the number of contracts or contract size. We next assume that the effects of contracts are heterogeneous across events, but potential outcomes are unaffected by which event an individual wins. This allows us to use variation in choice information to study the impact of a common treatment.[25] Table 4 supports this homogeneity assumption by, somewhat uniquely, showing that business creation is unaffected by receiving a randomly higher- or lower-value contract. Together, these assumptions allow us to use variation in real-world choices similarly to the elicitation of the Becker et al. (1964) (BDM) mechanism, in which individuals reveal valuations for goods, and then receive the good from a second-step randomization.

Third, the outcome model is further identified through the exclusion of the distance covariates, W_{ieb}, from the outcome equation and are used in the choice equation to identify θ_i^* in Equation 9. These characteristics help trace out the support of the distribution of unobserved preferences. For example, a short distance to an attractive lottery event may

[25] Chassang et al. (2012) advocates a similar idea—induce individuals to make a series of choices based on probabilities of winning and varying costs, and then estimate treatment effects across the range of preference values. We do this using choice data over time in a real-world setting.

induce a low preference individual (i.e., θ_i^* is more negative) to enter and potentially win, while a long distance to an unattractive lottery event will only induce entry by individuals with high preference. We exclude this variation from our outcome model, assuming that the distance of a contract from an individual has no direct effect on whether one starts a business in the future. We view this as a reasonable assumption since relative distance has no direct economic relationship with business creation.

Together, these assumptions increase the support over which we estimate the treatment effects we recover from the choice model. Since our model estimates the average treatment in the population, to the extent to which we do not have full support in treatment across the population, we extrapolate using the form of our model (Kline and Walters, 2019). We make additional parametrizing restrictions to operationalize these ideas. Specifically, we control linearly and additively for observable factors X_i and unobservable selection θ_i^* in Equation 10, which is common in the selection literature.

A natural test of these assumptions (and additional functional form restrictions) is how well our model can reproduce analyses employing weaker assumptions, as in Section 4. We show in Section A4.2, we can closely reproduce the empirical reduced-form and LATE parameters and heterogeneity patterns in the population using this model.

5.8 Choice and selection model results

Table A7 provides the results from the estimation of participation choices. All coefficients are effects at the mean. Distance to the expected worksites strongly discourages entry with an estimated coefficient of -0.036 and standard error of 0.00028. All other individual characteristics are either time-invariant measures or are taken from the year 2011, the year before the beginning of the lotteries. Conditional on other characteristics, age increases entry although there are diminishing effects of age as individuals become older, with maximal entry predicted at age 54. Females are associated with weaker demand and individuals who are relatively more advantaged, with higher levels of employment and income, are associated with stronger demand. Previous business ownership is insignificantly related to participation. We also include event-block-specific dummies to account for supply-side program attributes (such as the number of contracts) and unobservable preference shocks across blocks. We use these estimated parameters in the counterfactual estimation in the following section.

Unobserved participation preference heterogeneity in the population, the standard deviation of θ, is large. The standard deviation of observable individual characteristics affecting participation is only 0.5, as compared to 2.8 for unobservable characteristics. This is even after including detailed observable characteristics previously discussed, which are major determinants of labor market choices. As a benchmark, a one standard deviation increase in unobserved preference for participation is equivalent to lotteries being 77 kilometers closer. On average, individuals prefer not to enter the lotteries, as indicated by the negative mean unobserved tastes, but the wide variation in underlying preferences in the population induces many to enter.

To see how an individual's observable and unobservable characteristics relate to firm creation with and without treatment, we examine Table 6. The four columns in this table correspond to two regressions, where columns (1-2) and columns (3-4) are each from one regression but split across columns. Columns (1-2) investigate any business creation and columns (3-4) investigate business creation with at least one employee. In column (1), we see the relationship between individual covariates and the likelihood of becoming a firm owner for nonwinners. Firm ownership is heterogeneous based on observable characteristics. Young individuals are more likely to start firms on average. Females are less likely. Individuals who were employed in 2011 and who had higher incomes are more likely to have started a firm. Past firm ownership is also correlated to new firm ownership. Along unobservable characteristics captured by θ_i^*, individuals with higher preferences for participation also have a higher likelihood of becoming a firm owner if they are nonwinners. That is, individuals appear to select into the program based on absolute advantage in business creation.

In column (2) of Table 6, we investigate the differential effect of being a lottery winner on becoming a firm owner. In the population, on average being a lottery winner insignificantly increases the likelihood of becoming a firm owner by 6 percentage points. The effect is highly heterogeneous by age. While females were much less likely to start firms without this program, females are not relatively less likely to start firms as a function of contractual experience. There is also no differential effect as a function of past income, providing another indication that income may not be the most important channel of new firm creation. Instead, individuals with higher unobserved preferences for participation are more likely to become firm owners after lottery winning, a characteristic of selecting into the program based on unobserved comparative advantage (Roy, 1951). Combining both results on unobservables, we see that those who participate in the lotteries are more likely to start a firm, even absent winning, and also more likely to start a firm as a result of winning.

Columns (3-4) investigate impacts on growing businesses and show fairly similar results with a few important exceptions. Rates of larger business creation for nonwinners are lower, at 15%, and much more heavily correlated to past business creation. Previous owners are more than 50% more likely to start a new expanding firm if they had previously started a business. Individuals with high θ_i^* are more likely to start firms even without winning a contract. The impacts of contract winning are exclusively driven by selection: the average effect of the contract in the population is 0.005, an economically and statistically insignificant effect. Instead, we find that age and unobservable selection drive the heterogeneous returns and we measure the magnitude of the effect of unobservable selection to be the same magnitude as the effect of any firm creation. This means that unobservable selection primarily drives expanding business creation rather than any business creation.

Overall, individuals select into the program consistent with future gains to firm creation. We also find that the age-based heterogeneous results are not the result of differential selection. The current program design of allocating contracts across a series of lottery events does already achieve selection by requiring individuals to incur participation costs across events. Given the positive selection based on gains to firm creation, the model suggests that potentially reducing the number of contracts allocated in the same event or increasing application costs would further select more individuals who will use this contractual experience to start new firms. We investigate these conjectures in the following section.

6 Counterfactuals

6.1 Supply-side parameters and counterfactuals

Governments worldwide use procurement schemes that involve at least some random allocation of contracts to applicants (e.g., Carrillo et al., 2023; Fadic, 2018; Lee, 2021).[26] When employing such schemes, governments make decisions that can influence who applies, such as choosing between holding one large event or two smaller ones. To evaluate these choices, two main questions are relevant to the government and individuals in this sector. Do alternative implementation choices affect business creation, and to what extent? If alternative policies are more efficient at increasing business creation, do they come at the cost of reducing equity in opportunity for participation? These questions may translate to other settings, such as the likely impacts of increasing or decreasing access to other entrepreneurship-enhancing programs.

In this section, we provide quantitative answers to these questions by first estimating the partial equilibrium impact of alternative government implementation choices on applicant utility. We then use these parameters to estimate the impact of various counterfactual

[26]The scale of the randomization in this context is unique.

government policies on entry and business creation in equilibrium.

From the estimation of Equation 5, we recover an alternative-specific constant for each event which takes the form, $\hat{\Phi}_{eb} = \phi_{eb} - C/\pi_{eb}$. This equation decomposes the estimated parameter into the event-block specific utility shifter, ϕ_{eb}, minus the entry cost rescaled by the win probability, C/π_{eb}. We model event-block utility as a linear function of observable and unobservable features, $\phi_{eb} = X'_{eb}\beta + \varepsilon_{eb}$. X_{eb} is a vector of characteristics of the blocks including the average value of the contracts, the number of the contracts, and the share of contracts relative to the province population size. We include both attributes of contract amounts (number and share) as potential applicants may value either dimension. Unobservable demand components, ε_{eb}, include unmodeled characteristics of the lotteries, such as press coverage, and characteristics of local demand.[27]

Because local demand may vary over time and across regions in ways that correlate with lottery attributes, we include both province and year fixed effects in our estimating equation. Therefore, our estimating equation is:

$$\hat{\Phi}_{eb} = X'_{eb}\beta - C/\pi_{eb} + \mu_t + \delta_p + \varepsilon_{eb}. \tag{11}$$

Section A4.4 discusses how this model is identified under a standard linear panel fixed effects exogeneity assumption. A key concern is that lottery characteristics are allocated in a way correlated with local labor market conditions; this appendix section shows that our regressors X_{eb} are uncorrelated to several variables likely correlated with labor demand. Therefore, this concern appears to be unwarranted in this context.

With our estimates $\hat{\beta}$ and $\hat{C}$ of the impact of government-controlled supply characteristics on entry, we conduct counterfactuals on the impacts of entry and business creation. We conduct counterfactuals that change the number of contracts, the average value of the contracts, and the application cost. These parameters can be thought of as either directly changing the lottery events or analogously setting up nearby events that cause individuals to enter events multiple times and repeatedly pay the cost for each smaller or larger event.

In this model, individuals respond both directly to counterfactual changes in program attributes and indirectly through the entry decisions of other potential applicants. The entry decisions of others are reflected in choices through the probability of contract winning, π_{eb}. Therefore, we must solve for an equilibrium entry that incorporates both of these forces. We do so numerically and do not find evidence of multiple equilibria across any of the

[27]To the extent our modeled lottery characteristics are correlated with unmodeled characteristics of the lotteries, we capture the partial correlation of the bundle of attributes of these contracts.

simulations.[28]

To simulate the counterfactuals, we sample a population 100 times larger than the observed population. We draw covariates with replacement from the distribution of observed covariates and take random draws from the estimated distribution of the unobservable. We allocate choices, wins, and outcomes based on the counterfactual lottery characteristics and equilibrium probabilities, and use these features to construct the relevant counterfactual-specific estimates.

6.2 Counterfactual results

Table A8 shows the impact of event-block characteristics on entry utilities from estimating Equation 11. Contract size has a positive albeit decreasing impact on block utility, but neither the linear nor the quadratic terms are statistically significant. We find the lottery share in the population to be a strong indicator of utility while the number of contracts is small and statistically insignificant. Finally, we estimate costs as the coefficient on the win probabilities in two ways: either as the current block's probability (column 1) or the lagged value from the previous lottery in this province (column 2). Both columns are similar and show the costs are robustly small. We choose column (1) as our preferred specification.

We use these estimated effects to define our counterfactuals. We consider changing three event-block attributes: the number of lotteries, the average value of contracts, and the entry cost. For each, we multiplicatively rescale the attributes across a range of values between 0.2 and 5. For example, we create counterfactual events with either one fifth the number of contracts or five times as many. We hold the number of contracts actually won fixed, so this counterfactual actually represents holding fewer but larger events.

Figure 4 shows the results of the counterfactuals in five figures. In all figures, we show three lines corresponding to the counterfactual rescaling of each of the three attributes; the x-axis is the size of the rescaling. Panel A shows counterfactual entry shares in the y-axis. Scaling up the number of lotteries and the average value of the contracts increases entry while scaling them down decreases entry at relatively similar rates. Proportional increases to the number of lotteries induce greater entry and proportional decreases induce less non-entry than scaling the lottery amounts. This occurs because increasing the number of lotteries also mechanically increases the win rate in the block, leading to more compli-

[28] An equilibrium is defined as a set of winning probabilities that equalize lottery entry conditions. Our setup of agents making application choices to lottery events is nested within the model structure found in Walters (2018). This paper shows the existence of equilibrium probabilities of more complicated dynamics in the context of charter school lotteries (in which schools must account for students receiving and accepting offers from competing programs, a feature we do not have to contend with) (see their Appendix D).

cated equilibrium changes. Decreasing the amenity values of the contracts has less of an effect on entry than increasing amenity values given the zero bound of these attributes and evidence that other determinants of entry are important. Finally, rescaling costs does not substantially affect entry shares since we estimate application costs to be small.

Panels B and C show counterfactual average treatment effects for winners (i.e., $Y_i(1) - Y_i(0)$) and outcomes for winners if they had not won (i.e., $Y_i(0)$). There are five key takeaways. First, broadly across all of the counterfactuals, reducing entry increases business creation and vice versa, consistent with positive selection across both observable and unobservable characteristics of entrants. Second, there are equity-efficiency tradeoffs: in addition to increasing business creation, reducing entry increases the share of entrants who would have started a new business even without winning a contract, an indication of general advantage. Third, our counterfactuals generate small to medium-sized differences in business creation: we observe increases in business creation by up to 3.5% for the policy maximizing business creation relative to the baseline business creation rate of 8.9 percent and observe policies with up to 11% reductions. By changing multiple parameters at once, the government would likely be able to induce greater selection patterns. These estimated impacts suggest that the government may pursue an emphasis on equity considerations given that the efficiency gains from selection do not appear to lead to transformative impacts. Fourth, we observe non-linear effects between entry and impacts across the counterfactuals. For example, changing the number of lotteries has a larger impact on changes to outcomes than changes to contract values, despite it not always inducing a larger impact on entry. Finally, we find almost no change in outcomes from policies that change the cost. This is interesting because selection models are often framed by discussions on unobserved entry costs or as ordeal mechanisms.

Panels D and E show treatment effects and the untreated mean for the outcome of starting a new business with at least one employee rather than the outcome of any business. This may be important because expanding businesses are likely to be most impactful for the market and for the government. Most importantly, we observe larger effects of selection from counterfactual program design for this outcome. The figures show increases in business creation by up to 9.7% for the business creation maximal policy relative to the baseline business creation rate of 5.4 percent and up to 56% reductions. For this outcome as well, we observe an equity-efficiency tradeoff although for some values of the counterfactuals we observe relatively small changes to the outcomes of winners when untreated.

Across many of these counterfactual policies, the government can increase business creation but often this comes at the cost of induced greater entry by individuals who would already be more likely to start businesses, rather than increasing equity in opportunity. Thus, the government must wiegh this tradeoff. In the current allocation procedure, the program attributes already induce a form of this selection, although future allocations could be redesigned to target other features. The government could also induce selection by targeting observables, such as age, which is common in other programs around the world.

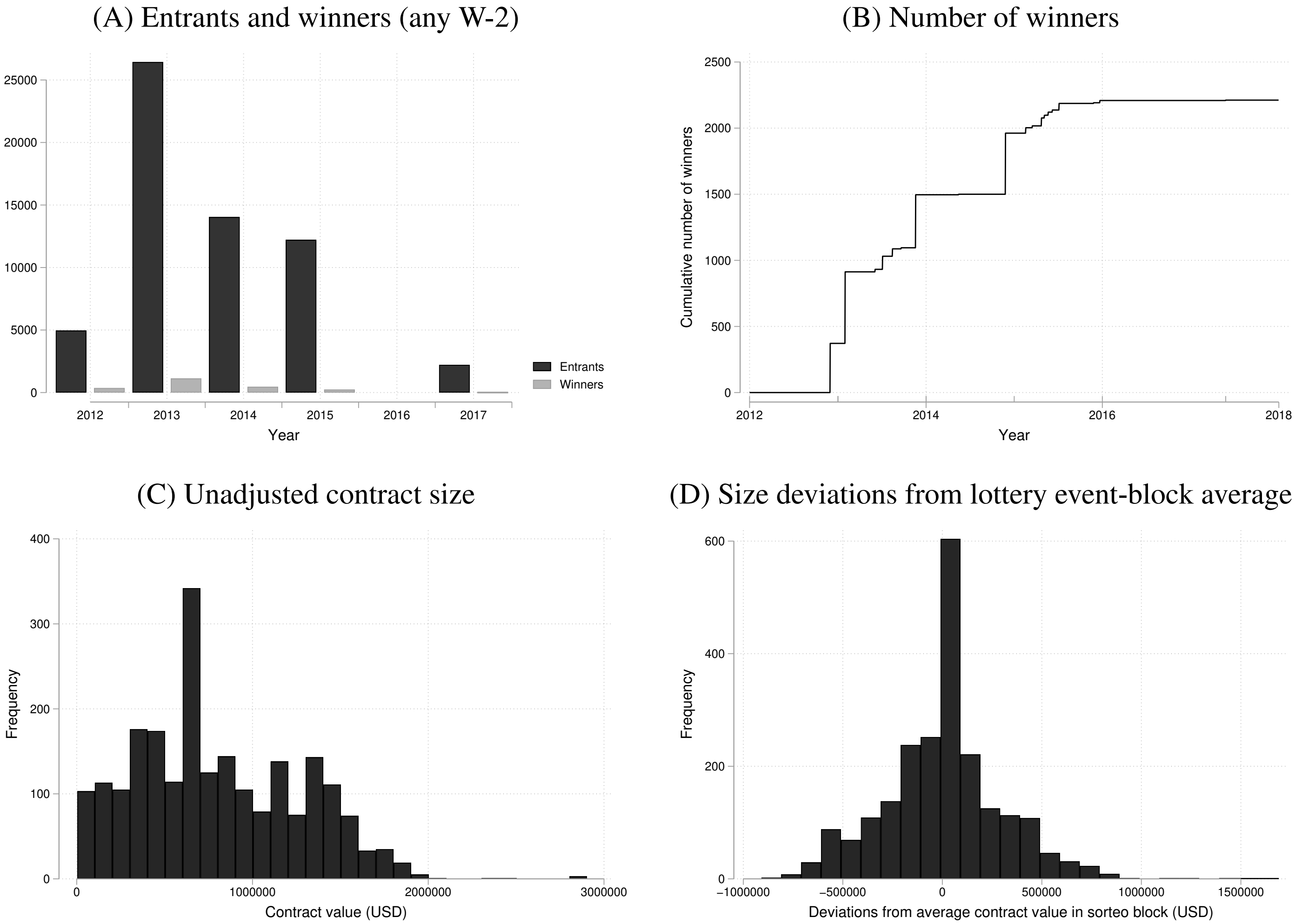
7 Conclusion

A substantial body of theoretical and empirical literature emphasizes the importance of managerial experience in catalyzing firm development in developing countries. In this paper, we provide new evidence on how opportunities to work in managerial roles affect firm creation. Through a fully randomized government program, we show that lottery winners given this opportunity shift out of formal wage employment and start new firms. We interpret these effects as most consistent with the bundle of benefits from managerial job experience, rather than increases in government contract access, increases in business networks, or increases in financial capital. The effects are driven primarily by the young, who are the most likely to benefit from new job experience. We then build a selection model to assess program targeting and find that individuals self-select into the program based on expected future gains from firm creation. Less attractive lottery features select individuals with higher marginal benefits for participation. However, this reduces opportunities for those unlikely to start firms, posing a tradeoff for the government.

Our results suggest that incentivizing new experiences, especially in managerial roles, may lead to new firm creation, and also finds evidence in favor of the ability for costly admission procedures to target those with high potential for entrepreneurship. Each of these results are potentially important for developing country firm creation and labor markets. First, the results suggest a clear rationale why managerial job experience may be underprovided by firms—individuals may start competing firms. Public employment programs in developing countries often focus on low-skill jobs, which may alleviate temporary income shortages but contribute little to long-term skill development. Programs encouraging young people to gain advanced skills may be most beneficial. Finally, regarding targeting, recent initiatives like business plan competitions have yielded positive results, possibly because they induce self-selection among individuals with high marginal benefits. Greater evidence on mechanisms to induce self-selection in intervention targeting seems potentially fruitful.

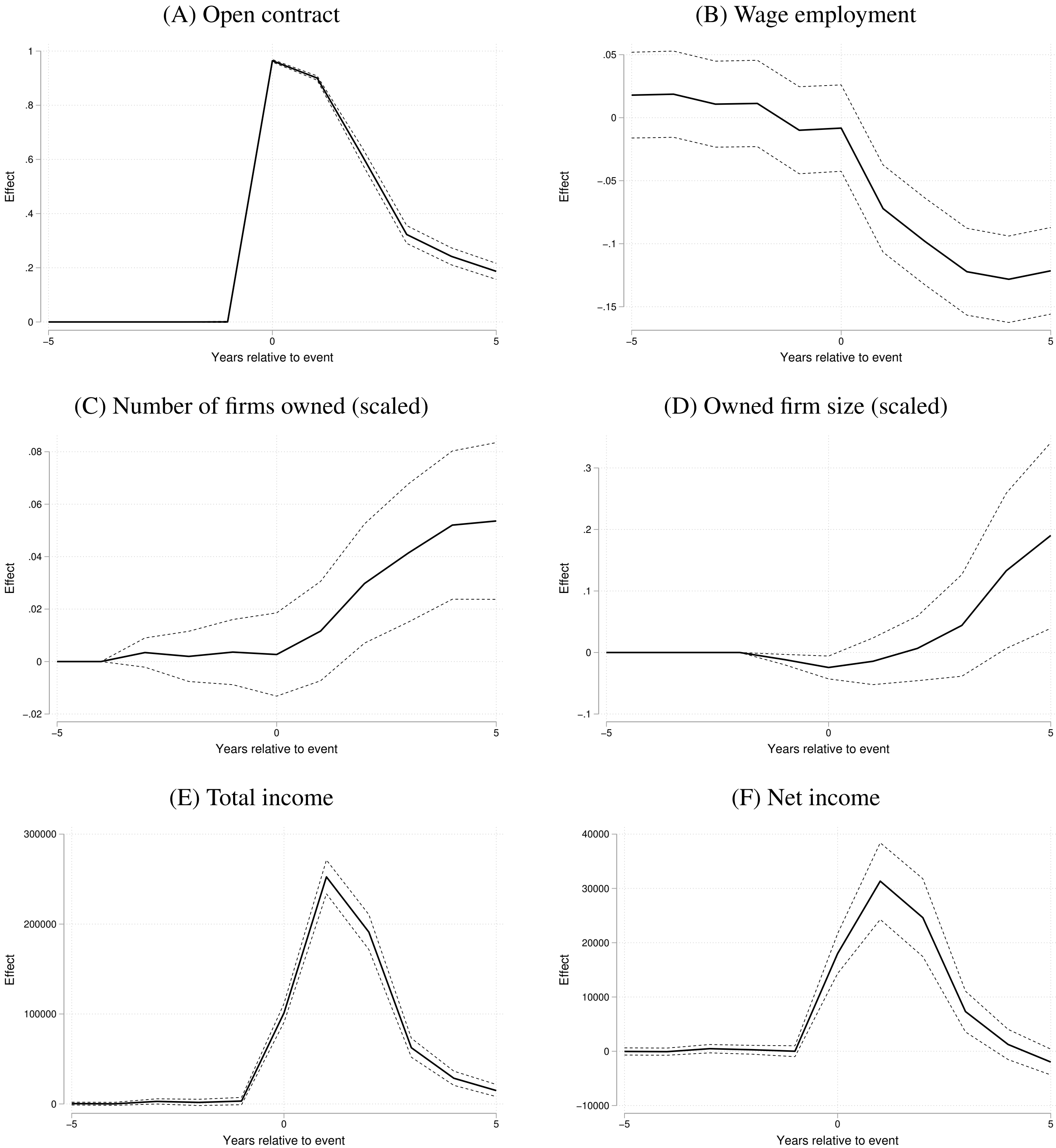
Figures

Figure 1: Lotteries over time and contract amounts



These figures present information on lotteries over time and variation across contracts. Panel (A) shows the number of entrants and winners by year of lottery event. Panel (B) shows the cumulative number of winners by date of lottery event. Panel (C) shows unadjusted contract sizes. Panel (D) shows deviations from lottery event-block average.

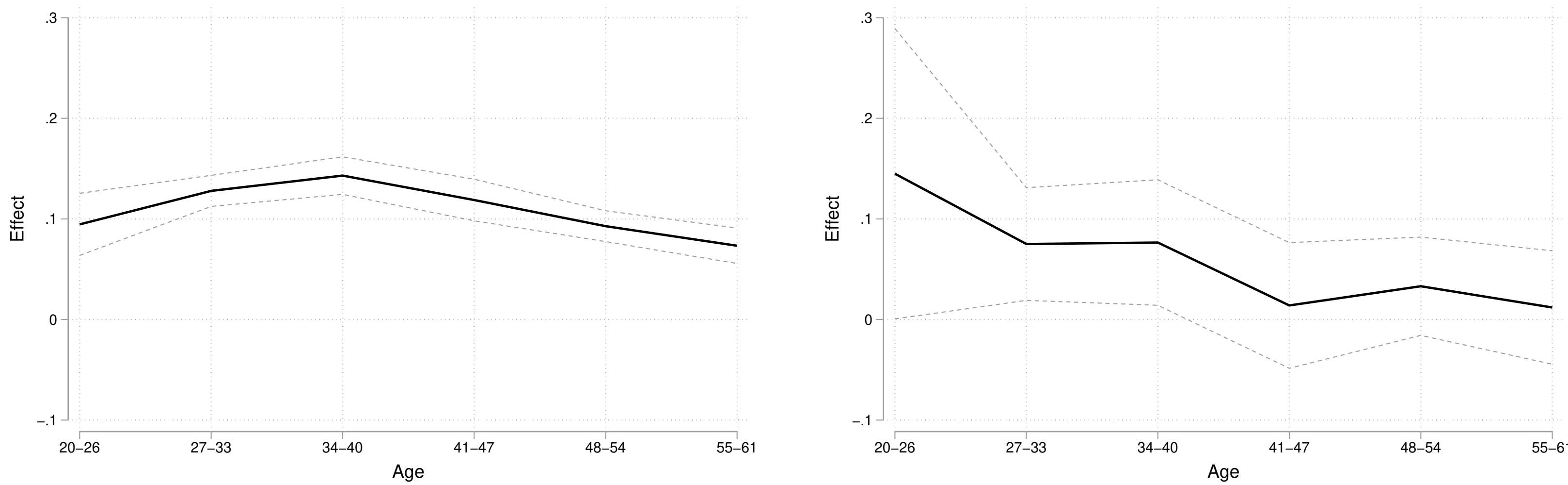
Figure 2: Sector participation for individuals



The figures display regressions of the outcome in the panel header on lottery winning. The sample includes individual entrants. Regressions are estimated separately for each year relative to the lottery event date and include event-block fixed effects. Dotted lines represent 95% confidence intervals for standard errors clustered at the individual level.

Figure 3: Age-based effects in majority-stake firm ownership

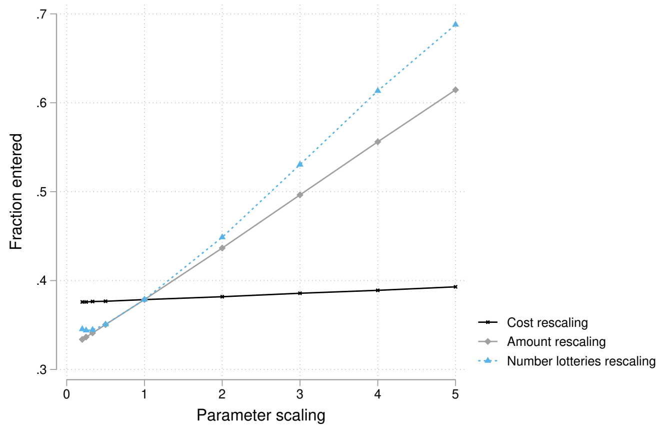
(A) Lottery nonwinner levels | (B) β



The figures display regressions of ever having owned a firm with a majority stake on the age that an individual entered a lottery event. Ages are binned in 7-year increments. The sample includes all individual entrants. Panel (a) report the results for the sample of lottery nonwinners. Panel (b) includes the full sample and we report the estimated treatment effect of being a lottery winner by age, controlling for both event-block and age fixed effects. Dotted lines represent 95% confidence intervals clustered at the individual level.

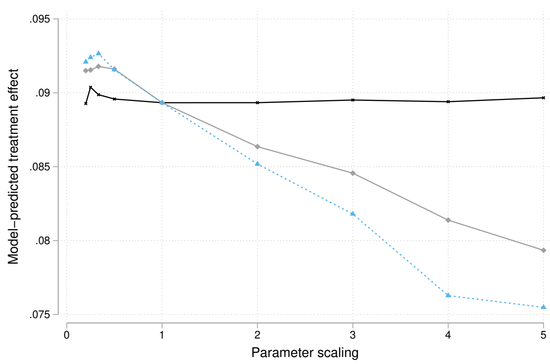
Figure 4: Counterfactual outcomes across supply parameters

(A) Fraction entered

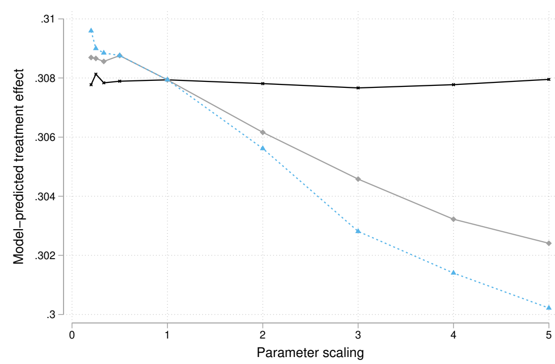


Outcome: Any firm

(B) $Y_1 - Y_0$, Winner

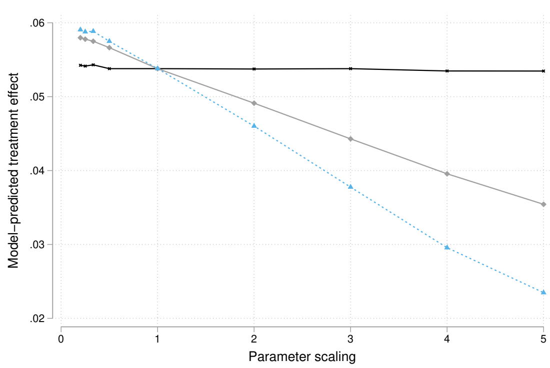


(C) Y_0, Winner

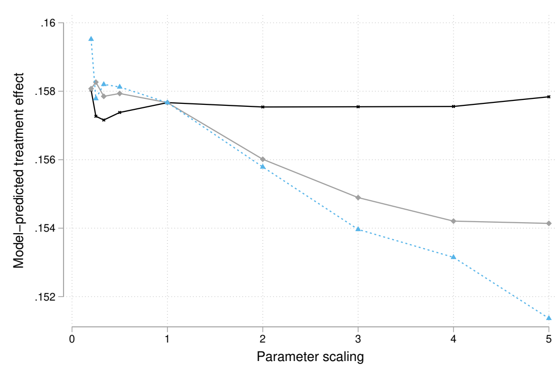


Outcome: Firm with at least one employee

(D) $Y_1 - Y_0$, Winner



(E) Y_0, Winner



These figures display counterfactual outcomes of changing lottery attributes. Lottery attributes (number of lotteries, entry cost, and contract amount) are changed one at a time and multiplicatively scaled by $\alpha \in \{0.2, 0.25, 0.33, 0.5, 1, 2, 3, 4, 5\}$. Panel A shows population entry rates. Panels B and C investigate any business creation between 2012-2018, while Panels D and E show any business creation with at least one employee. The simulated population is 100 times the size of the empirical population with observed covariates drawn at random with replacement and the unobservables drawn independently from the estimated distribution.

Tables

Table 1: Placebo tests for winning allocation

	All sorteos			Analysis sample		
	N	Var. Mean	Winner	N	Var. Mean	Winner
Number of sorteos, t=-1	59251	2	-.029	19284	.8	-.0038
			(.029)			(.015)
Firm	59251	.14	.0048	19284	.13	-.0077
			(.0075)			(.01)
Female	58849	.29	.0001	19062	.27	-.012
			(.0093)			(.013)
Filed taxes, t=-1	59251	.82	.0072	19284	.82	.011
			(.0081)			(.012)
Income, t=-1	59251	52,243	2,791	19284	52,269	-8,121
			(5,906)			(7,057)
Asinh Gross income, t=-1	59251	10	.0029	19284	10	.00055
			(.13)			(.19)
Ever won non-sorteo, t=-1	59251	.028	-.0031	19284	.016	-.0061*
			(.0031)			(.0037)
Ever firm owner, t=-1	59251	.13	.0088	19284	.097	.0073
			(.0073)			(.0099)
Joint p-value			.55			.48
Joint p-value, Individuals			.80			.47

Columns (1-3) correspond to all lottery events. Columns (4-6) correspond to the analysis sample. Columns (1, 4) report the sample size, columns (2, 5) report the sample mean, and columns (3, 6) reports the point estimate of an OLS regression of the entrant characteristic on lottery winning. Joint p-value comes from an F-test of joint significance of the variables in the rows on lottery winning. This value is presented for the full sample and for the individual-only sample. Controls include event-block fixed effects. Standard errors clustered at the individual/firm level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2: Individual firm ownership and aggregated firm outcomes

	Firm creation			Agg. firm outcomes			Experience	
	(1) Any firm Ever	(2) Num firms Ever	(3) Num firms 2018	(4) Employees 2018	(5) Income 2018	(6) Profits 2018	(7) Sub contractors	(8) Temporary employees
Winner	0.078***	0.15***	0.060***	0.17**	4187.4**	-55.7	7.60**	0.96*
	(0.017)	(0.032)	(0.014)	(0.076)	(1822.1)	(92.7)	(3.61)	(0.50)
Control Mean	.35	.49	.18	.39	7224	152	31	5.7
N	16855	16855	16855	16855	16855	16855	698	698

This table reports estimates of the effect of lottery winning on firm ownership, owned firm outcomes, and experience. Sample is restricted to individuals. Columns (1-2) report becoming a firm owner at some point after 2010. Columns (3-6) report aggregated firm results at the individual level for year 2018 and are scaled by participation and ownership period shares. Columns (7-8) reports survey-reported managerial experiences of the most subcontractors worked with and most temporary employees managed in their career. Controls include lottery event-block fixed effects, and in survey results also include age bins, and month of survey fixed effects. Standard errors clustered at the individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Employment, firm ownership, and contract take-up by age

	Employment	Firm ownership				Contract
	(1) Employed	(2) Num firms	(3) Income	(4) Profits	(5) Employees	(6) Take-up
Winner, 20-34	-0.19***	0.13***	8050.5**	130.0	0.46**	0.96***
	(0.032)	(0.028)	(3912.1)	(204.2)	(0.20)	(0.012)
Winner, 35-49	-0.11***	0.043*	3328.7	-112.1	0.12	0.94***
	(0.030)	(0.024)	(2861.8)	(155.1)	(0.11)	(0.014)
Winner, 50-64	-0.066**	0.025	3146.5	-174.3	0.020	0.93***
	(0.032)	(0.022)	(2949.1)	(150.6)	(0.084)	(0.015)
Joint p-value	0.02	0.02	0.55	0.47	0.12	0.25
Control Mean	.58	.18	7224	152	.39	.0019
N	15869	15869	15869	15869	15869	15869

This table shows estimates of the effect of lottery winning on employment, firm ownership, and contract take-up by age. Sample is restricted to individuals. All outcomes are from 2018. Column (1) registers formal employment. Columns (2-5) report aggregated firm results at the individual level for year 2018 and are scaled by participation and ownership period shares. Column (6) measures contract take-up overall. Joint p-value is from an F-test on joint equality of treatment variables. Controls include lottery event-block by age bin fixed effects. Standard errors clustered at the individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Effects of contract size on entrepreneurship, income, and management experience

	Entrepreneurship (2018)			Income (Cum.)	Experience		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Owner	Firm income	Firm size	Indiv. net	Days in contract	Sub contractors	Temporary employees
Winner	0.099***	7189.8**	0.38***	6740.0	1068.4***	0.80	-6.33
	(0.038)	(3414.3)	(0.14)	(23528.8)	(47.6)	(0.94)	(6.74)
Winner * Size ($200k)	-0.0065	-930.4	-0.067**	23681.5***	5.14	0.015	4.77**
	(0.010)	(830.3)	(0.030)	(7735.3)	(12.6)	(0.28)	(2.07)
Control Mean	0.35	7223.98	0.39	71076.22	0.00	5.70	31.44
N	16855	16855	16855	16735	16855	690	690

Sample is individuals. The treatment variables are winner and winner interacted with contract size. We assign contract size as 0 for control individuals and do not control for contract size. Controls include lottery event-block fixed effects. Standard errors clustered at the individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Created firm quality

	Hire		Survival		Revenue	
	(1)	(2)	(3)	(4)	(5)	(6)
Winner	0.048*	0.052*	0.051**	0.052**	-3580.3	-4473.0
	(0.028)	(0.028)	(0.022)	(0.022)	(4590.3)	(4566.3)
Firm type FE		X		X		X
Age tercile FE		X		X		X
Control Mean	.3	.3	.72	.72	44913	44913
N	4647	4591	4647	4591	4647	4591

This table compares firms created by lottery winners and those created by non-winners. The sample includes created firms after the date of the sorteo from participants in the analysis sample. Hire is a dummy equal to 1 if the firm has hired any employees within the first two years. Survival is a dummy equal to 1 if the firm continues to exist after 2 years as measured by filing tax records. Revenues come from the second year of firm existence. Controls include lottery event-block and year fixed effects. Standard errors two-way clustered at the firm and individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Selection model estimates of program effects

	Became owner: ≥ 0 employees		Became owner: ≥ 1 employees	
	(1) Non-winners	(2) Winner effect	(3) Non-winners	(4) Winner effect
Constant	.29***	.061*	.15***	.0051
	(.0051)	(.037)	(.0039)	(.026)
Age	-.0073***	-.022***	-.0022	-.011**
	(.0021)	(.0073)	(.0015)	(.0053)
Age Sq.	.000031	.00016**	4.6e-06	.000072
	(.000024)	(.000078)	(.000017)	(.000057)
Female	-.078***	.05	-.041***	-.0091
	(.0085)	(.038)	(.0062)	(.028)
Employed	.055***	-.053	.029***	-.035
	(.0077)	(.032)	(.0056)	(.023)
Income	1.4e-08***	-7.9e-10	4.5e-09***	5.5e-09
	(1.8e-09)	(8.7e-09)	(1.2e-09)	(6.3e-09)
Prev. owner	.16***	-.075	.54***	-.023
	(.014)	(.059)	(.01)	(.042)
θ^*	.009***	.023*	.0053***	.022**
	(.0017)	(.014)	(.0013)	(.0098)

This table reports selection-corrected estimates of the effect of winning a contract lottery on being a firm owner. The outcome for columns (1-2) is any firm ownership after 2011, while columns (3-4) report being a firm owner with at least one employee after 2011. Grouped columns (1-2, 3-4) come from a single regression with odd columns showing interaction terms between winner and covariates. All individual characteristics are taken from the year prior to the lottery events, 2011. Standard errors account for the estimation of θ. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

References

ABADIE, A. AND M. D. CATTANEO (2018): "Econometric methods for program evaluation," *Annual Review of Economics*, 10, 465–503.

ABEBE, G., A. S. CARIA, AND E. ORTIZ-OSPINA (2021): "The selection of talent: Experimental and structural evidence from ethiopia," *American Economic Review*, 111, 1757–1806.

ABEBE, G., M. FAFCHAMPS, M. KOELLE, AND S. QUINN (2019): "Learning Management Through Matching: A Field Experiment Using Mechanism Design," .

ALATAS, V., R. PURNAMASARI, M. WAI-POI, A. BANERJEE, B. A. OLKEN, AND R. HANNA (2016): "Self-targeting: Evidence from a field experiment in Indonesia," *Journal of Political Economy*, 124, 371–427.

ANGRIST, J. D. (1998): "Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants," *Econometrica*, 249–288.

ARISTY ESCUDER, J. (2016): "Análisis del impacto de la política de compras y contrataciones públicas en las MIPYMEs y mujeres en la República Dominicana," Technical Report TR-2016, Dirección General de Contrataciones Públicas, Santo Domingo, República Dominicana.

ARTANA, D., S. AUGUSTE, J. L. BOUR, F. NAVAJAS, M. PANADEIROS, AND R. M. GUZMÁN (2006): "El gasto público en República Dominicana," .

AZOULAY, P., B. F. JONES, D. KIM, AND J. MIRANDA (2018): "Age and High-Growth Entrepreneurship," .

BANERJEE, A., E. BREZA, E. DUFLO, AND C. KINNAN (2019): "Can microfinance unlock a poverty trap for some entrepreneurs?" Tech. rep., NBER.

BAYER, P., S. L. ROSS, AND G. TOPA (2008): "Place of work and place of residence: Informal hiring networks and labor market outcomes," *Journal of Political Economy*, 116, 1150–1196.

BEAMAN, L., D. KARLAN, B. THUYSBAERT, AND C. UDRY (2023): "Selection into credit markets: Evidence from agriculture in Mali," *Econometrica*, 91, 1595–1627.

BECKER, G. M., M. H. DEGROOT, AND J. MARSCHAK (1964): "Measuring utility by a single-response sequential method," *Behavioral science*, 9, 226–232.

BERRY, J., G. FISCHER, AND R. P. GUITERAS (2015): "Eliciting and utilizing willingness to pay: Evidence from field trials in Northern Ghana," .

BLOOM, N., B. EIFERT, A. MAHAJAN, D. MCKENZIE, AND J. ROBERTS (2013): "Does management matter? Evidence from India," *The Quarterly Journal of Economics*, 128, 1–51.

BRUHN, M., D. KARLAN, AND A. SCHOAR (2018): "The impact of consulting services on small and medium enterprises: Evidence from a randomized trial in mexico," *Journal of Political Economy*, 126, 635–687.

BRYAN, G., D. KARLAN, AND A. OSMAN (2024): "Big loans to small businesses: Predicting winners and losers in an entrepreneurial lending experiment," *American Economic Review*, 114, 2825–2860.

CARRILLO, P., D. DONALDSON, D. POMERANZ, AND M. SINGHAL (2023): "Misallocation in Firm Production: A Nonparametric Analysis Using Procurement Lotteries," Tech. rep., National Bureau of Economic Research.

CELLINI, S. R., F. FERREIRA, AND J. ROTHSTEIN (2010): "The value of school facility investments: Evidence from a dynamic regression discontinuity design," *The Quarterly Journal of Economics*, 125, 215–261.

CHASSANG, S., P. I. MIQUEL, E. SNOWBERG, ET AL. (2012): "Selective trials: A principal-agent approach to randomized controlled experiments," *American Economic Review*, 102, 1279–1309.

CHURCH, A. H. (1997): "Managerial self-awareness in high-performing individuals in organizations." *The Journal of applied psychology*, 82 2, 281–92.

CRÉPON, B., M. EL KOMI, AND A. OSMAN (2024): "Is it who you are or what you get? comparing the impacts of loans and grants for microenterprise development," *American Economic Journal: Applied Economics*, 16, 286–313.

DE CHAISEMARTIN, C. AND L. BEHAGHEL (2018): "Estimating the Effect of Treatments Allocated by Randomized Waiting Lists." *Available at SSRN 3175452*.

DE MEL, S., D. MCKENZIE, AND C. WOODRUFF (2008): "Returns to capital in microenterprises: evidence from a field experiment," *The quarterly journal of Economics*, 123, 1329–1372.

EVANS, D. AND L. S. LEIGHTON (1989): "Some Empirical Aspects of Entrepreneurship," *American Economic Review*, 79, 519–35.

FADIC, M. (2018): "Government Procurement and the Growth of Small Firms," .

FAFCHAMPS, M. AND S. QUINN (2017): "Aspire," *The Journal of Development Studies*, 53, 1615–1633.

FEIGENBAUM, J. J. (2016): "Automated census record linking: A machine learning approach," .

FELDMAN, N. E. AND J. SLEMROD (2007): "Estimating tax noncompliance with evidence from unaudited tax returns," *The Economic Journal*, 117, 327–352.

GARAIKA, G., H. M. MARGAHANA, AND S. T. NEGARA (2019): "Self efficacy, self personality and self confidence on entrepreneurial intention: study on young enterprises," *Journal of Entrepreneurship Education*, 22, 1–12.

GELBER, A., A. ISEN, AND J. B. KESSLER (2015): "The effects of youth employment: Evidence from New York city lotteries," *The Quarterly Journal of Economics*, 131, 423–460.

GENDRON-CARRIER, N. (2018): "Understanding the Careers of Young Entrepreneurs," .

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

Hayward, M. L. A., W. R. Forster, S. Sarasvathy, and B. Fredrickson (2010): "Beyond hubris: How highly confident entrepreneurs rebound to venture again," *Journal of Business Venturing*, 25, 569–578.

Heckman, J. (1990): "Varieties of selection bias," *The American Economic Review*, 80, 313–318.

Heckman, J. J. and E. Vytlacil (2005): "Structural equations, treatment effects, and econometric policy evaluation 1," *Econometrica*, 73, 669–738.

Hincapié, A. (2017): "Where are the Young Entrepreneurs? A Study of Entrepreneurship over the Life Cycle," Tech. rep., Mimeo, Washington University in St. Louis.

Hombert, J., A. Schoar, D. A. Sraer, and D. Thesmar (2017): "Can unemployment insurance spur entrepreneurial activity? Evidence from France," *HEC Paris Research Paper No. FIN-2013-1020*.

Humphries, J. E. (2016): "The Causes and Consequences of Self-Employment over the Life Cycle," Tech. rep., Technical Report, Mimeo, University of Chicago.

Hurst, E. and B. W. Pugsley (2011): "What do small businesses do?" Tech. rep., National Bureau of Economic Research.

Hussam, R., N. Rigol, and B. N. Roth (2022): "Targeting high ability entrepreneurs using community information: Mechanism design in the field," *American Economic Review*, 112, 861–898.

Jones, D. (2015): "The economics of exclusion restrictions in iv models," Tech. rep., National Bureau of Economic Research.

Kline, P. and C. R. Walters (2019): "On Heckits, LATE, and numerical equivalence," *Econometrica*, 87, 677–696.

La Porta, R. and A. Shleifer (2014): "Informality and development," *Journal of Economic Perspectives*, 28, 109–26.

Lagakos, D., B. Moll, T. Porzio, N. Qian, and T. Schoellman (2018): "Life cycle wage growth across countries," *Journal of Political Economy*, 126, 797–849.

Lazear, E. P. (2005): "Entrepreneurship," *Journal of Labor Economics*, 23, 649–680.

Lee, M. (2021): "Government purchases and firm growth," *Available at SSRN 3823255*.

Levine, R. and Y. Rubinstein (2017): "Smart and illicit: who becomes an entrepreneur and do they earn more?" *The Quarterly Journal of Economics*, 132, 963–1018.

Liang, J., H. Wang, and E. P. Lazear (2018): "Demographics and entrepreneurship," *Journal of Political Economy*, 126, S140–S196.

McFadden, D. and K. Train (2000): "Mixed MNL models for discrete response," *Journal of applied Econometrics*, 15, 447–470.

McKenzie, D. (2017a): "How effective are active labor market policies in developing countries? a critical review of recent evidence," *The World Bank Research Observer*, 32, 127–154.

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

MCKENZIE, D. AND SANSONE (2019): "Predicting Entrepreneurial Success is Hard," .

MEAGER, R. (2022): "Aggregating distributional treatment effects: A Bayesian hierarchical analysis of the microcredit literature," *American Economic Review*, 112, 1818–1847.

MURPHY, K. M. AND R. H. TOPEL (2002): "Estimation and inference in two-step econometric models," *Journal of Business & Economic Statistics*, 20, 88–97.

QUINN, S. AND C. WOODRUFF (2019): "Experiments and Entrepreneurship in Developing Countries," *Annual Review of Economics*, 11, null.

ROBERTSON, I. AND G. SADRI (1993): "Managerial Self-efficacy and Managerial Performance," *British Journal of Management*, 4, 37–45.

ROY, A. D. (1951): "Some thoughts on the distribution of earnings," *Oxford economic papers*, 3, 135–146.

SCHOAR, A. (2010): "The divide between subsistence and transformational entrepreneurship," *Innovation policy and the economy*, 10, 57–81.

SILVA, O. (2007): "The Jack-of-all-trades entrepreneur: innate talent or acquired skill?" *Economics letters*, 97, 118–123.

TRAIN, K. E. (2009): *Discrete choice methods with simulation*, Cambridge press.

TRAVIESO CARABALLO, S. E. (2016): "Estudio comparativo de los procedimientos de contratación en España, República Dominicana y Estados Unidos," Master's thesis, Universitat Politècnica de Catalunya.

VAN DIJK, W. (2019): "The socio-economic consequences of housing assistance," .

WALD, A. (1940): "The fitting of straight lines if both variables are subject to error," *The annals of mathematical statistics*, 11, 284–300.

WALTERS, C. R. (2018): "The demand for effective charter schools," *Journal of Political Economy*, 126, 2179–2223.

WINKLER, W. E. (1994): "Advanced methods for record linkage," .

WOODRUFF, C. (2018): "Addressing constraints to small and growing businesses," Tech. rep., Working Paper.

WOOLDRIDGE, J. M. (2015): "Control function methods in applied econometrics," *Journal of Human Resources*, 50, 420–445.

ZIMMERMAN, S. D. (2019): "Elite colleges and upward mobility to top jobs and top incomes," *American Economic Review*, 109, 1–47.

Online appendix

A1 Income and tax calculations

We calculate individual income and profits using data from multiple sources of DGII tax records. Individuals may report income using multiple forms—IR1 and IR18—and labor income may be reported on their behalf by their employer. Individuals are not required to file a tax form if their sole earnings are already reported by their employer. Therefore, we take the max of their reported income and income reported on their behalf.

For individual profits, we also use multiple sources to determine this value. First, we base profits on the tax form question of earned income minus costs. If individuals report a lower income than their employer-reported income, we add this difference to reflect their total net income across all earnings categories. Finally, we incorporate additional cost information. While we do not observe all of the costs that can be incorporated into the net cost amounts, we observe one relevant one in which all expenses should be incorporated into the net income caclulation: business input costs from Form 606. In the data, we observe multiple instances of lottery winners reporting the full value of their government contract as income and either reporting the same amount as their net income or not filling in their net income at all. This is despite the individual reporting positive amounts in their 606 form. We therefore create our preferred measure of net income which is the minimum of reported net income based on the above calculation or income minus 606-reported costs, as this constructed net income, from the DGII perspective, should be definitionally larger than the reported net income amount. In practice, we believe that the data still somewhat overstates the amount of net income received for a small share of lottery winners who report little cost information.

In Table A9, we show our main income regressions with our preferred net income variable and the reported net income variable unchanged with Form 606 information. Qualitatively, the results are very similar. We view our preferred profit variable as a better reflection of actual earnings based on our discussions with tax officials.

A2 Alternative firm measure

Our study relies heavily on administrative records from the DGII. A potential concern is that lottery winners might formalize their businesses or report activities more thoroughly than nonwinners, which could bias our results. One piece of evidence against informal firms formalizing is that individuals exit wage employment and move into firm creation at similar rates. To further address this issue, we utilize additional firm ownership data from ONAPI, which maintains records of firms that have trademarked their names. These records are independent of the Dominican tax system and reflect formalization unrelated to taxation.

Panel D in Table A2 confirms our main result: lottery winners are more likely to start firms. Lottery winners are more likely to have ever started a firm, own more firms, and hold a higher number of firm shares. In these specifications, firm shares are defined based on the portion of the year an individual is an owner, as ownership percentage information is unavailable. The point estimates and control group mean in this dataset are lower than in the main dataset because we can match only 68% of the sample and not all firms register with ONAPI.

A3 Matching between datasets

Linking between all administrative datasets in this project is completed by using the two unique administrative identifiers that apply to this population: the RPE, an identifier assigned to all registered state providers, and the national ID. The dataset on state providers includes both the RPE number and the national ID. We have access to one of these two identifiers across almost all of the outcome datasets. The main work in matching the other datasets, most notably the datasets on lottery event participation, is to assign identifiers when they are omitted. Since 72.5% of participation records come with the identifiers, the matching is only pertinent to the other 27.5%.

When administrative identifiers are missing from a dataset or an observation, we attempt to assign one of the two identifiers using the entrant's full name. Luckily, full names of individuals and firms in the DR are often unique. In the universe of registered taxpayers in the country, 90.5% of individuals have unique names.[1] The reason for the high name uniqueness in the Dominican Republic is that full names usually contain at least four parts, including first and middle, plus the father's and mother's last name. We are further usually matching into the set of 70,000 state providers from the list of over three million registered taxpayers in the DR, thereby mitigating concerns about name uniqueness.

To assign the identifiers using name, we usually start by trying to match to the registry of state providers. Our approach is automated matching with human assistance. For all matches, we standardize names by converting special characters to ASCII format and removing honorifics if included. We then assign Jaro-Winkler string distances between all

[1] Name overlap is more common in the United States (e.g., in Florida, only 83.9% of individuals have unique full names).

names within these datasets (Winkler, 1994). Jaro-Winkler scores are commonly used in string matching procedures similar to this one (e.g. Feigenbaum (2016)). The maximum value of a Jaro-Winkler score is 1. We use two criteria to define automatic matches. First, if two names have a Jaro-Winkler score of 0.99, and there is no other possible match with a Jaro-Winkler score within 0.03 units, we call this a match. Second, if the Jaro-Winkler score is above 0.95, and there is no other possible match within 0.07 units, we define this as a match. Each of these criteria is very restrictive; we use them to define matches when we have a high degree of confidence the match does not need to be reviewed. Finally, if there are no matches that fit these criteria, we output the five names with the highest Jaro-Winkler score to see if we can find a match. We then try to assign identifiers when obvious, or do a further set of matching using the full dataset of taxpayers, either manually or by a second application of the same method.

This parsimonious procedure uniquely assigns identifiers to almost all observations. Of the 59,861 observations in our event participation datasets, 43,428 come with both full name and either an RPE number or a national ID number. They did not require further matching. Overall, we were not able to match 526 observations to a national ID number, and hence, have no tax information for these observations.[2] All observations that are matched uniquely or matched to multiple identifiers are kept for the analysis sample. Since a relatively small sample is matched non-uniquely, we randomly choose one of the identifiers for our analysis sample. This generates a small amount of potential measurement error that we could correct by using probabilistic matching and associated regression methods, but our results would change little.

We modify the matching procedure for a few datasets. We use information from exequatur records to identify the universe of licensed engineers and architects in the DR. Since we are specifically interested in individuals who may not be state providers, we follow the procedure above but match directly to the DGII taxpayer records to recover the national ID. Of the 22,879 individuals who are licensed as civil engineers and architects, we match 92% to a *Cédula*. For records on *sorteo* completion status where we are only provided the name of the project, we hand-match given the small sample size.

Finally, we include a second dataset of firm creation from ONAPI. We follow the procedure above but with slight modifications to the match requirements since we did not hand match.[3] We match 67.9% of the original dataset to a record in the DGII taxpayer universe.

[2] They are not differential by treatment and control ($p = 0.92$), so we set them to 0.

[3] We did not want to do hand matching here because of the size of the dataset and its minor importance. We

97% of these matches are to unique taxpayers.

A4 Choice model and counterfactuals

A4.1 Model identification

This section further describes the identifying assumptions of the model in Section 5.7 using a simplified setting based on our context. We do so to highlight the assumptions used and identified parameters. Our primary goal is to identify selection effects and study counterfactual policy changes, although this methodology can identify general parameters such as the average treatment effect (ATE) in the population.

Assume potential applicants make choices to apply to each of two events $e \in \{1,2\}$. Individual-specific latent preferences to enter a lottery are defined $\theta_i \in [0,1]$, while the programmatic application benefits and costs are $c_e \in [0,1]$ and WLOG $c_1 > c_2$. Applicants are thus assumed to make application choices based on a standard latent index equation $A_{ie} = 1[\theta_i > c_e]$. We observe both binary entry choices $A_i = (A_{i1}, A_{i2})$.

Define potential outcomes $Y_i(d,z,e)$ to be the outcome for person i of treatment $D_i \in \{0,1\}$ for lottery winning instrument $Z_i \in \{0,1\}$ in event e. Potential treatment is defined $D_i(z)$, where we omit indexing by e. For event e, we can define the Wald (1940) instrumental variables estimand

$$IV(e) = \frac{E[Y|Z_i = 1, A_{ie} = 1] - E[Y|Z_i = 0, A_{ie} = 1]}{E[D|Z_i = 1, A_{ie} = 1] - E[D|Z_i = 0, A_{ie} = 1]}.$$

As in the actual experiment we analyze, assume lottery winning induces perfect takeup, $Pr(D_i(1) > D_i(0)) = 1$. Further assuming exogeneity/exclusion of the instrument Z_i, we can simplify the IV estimand for event e

$$IV(e) = E[Y_i(1,e) - Y_i(0,e)|A_{ie} = 1].$$

This parameter is the average treatment effect for applicants to event e. A lottery-based analysis that aggregates effects across events using a regression for example, estimates the

define a match if two names correspond with a Jaro-Winkler score of 0.985 and there is no other possible match with a Jaro-Winkler score within 0.015 units, or if the names correspond with Jaro-Winkler score of 0.95 and there is no other possible match within 0.04 units.

parameter

$$IV^1 = \sum_e w_e E[Y_i(1,e) - Y_i(0,e) | A_{ie} = 1]$$

where the weights sum to 1 and are estimator-specific (Abadie and Cattaneo, 2018; Angrist, 1998). This parameter corresponds to a weighted-average of average treatment effects for the population who enter the events. It does not correspond to the average treatment effect in the population.

Identification across events: We first assume that the events act like a menu of choices to elicit a "willingnes to pay" for contracts, akin to the Becker et al. (1964) (BDM) mechanism.[4] We then seek to use this information about relative preferences in the estimation of treatment effects. Heuristically, individuals reveal information about their valuation of the contracts, and the lottery permits us to study causal treatment effects for individuals with low and high valuations.

We must make two key assumptions. First, we assume a latent index equation for A_{ie} that allows us to map observable choices to well-defined subpopulations based on prefence measures, here θ_i. The form of this equation can be modified, but we require a common ordering of preferences across event characteristics. This type of latent index structure is common in the selection literature and is used in the local instrumental variables method of Heckman and Vytlacil (2005). Second, we assume that events e are excludable from potential outcomes, $Y(d,e) = Y(d)$.[5] In our empirical context, this assumption means that contract features like contract value or number of contracts allocated do not directly affect whether an individual starts a business. This assumption is consistent with Table 4, which shows no statistically significant heterogeneity in business creation across different randomly assigned contract values.

Using these assumptions, we can recover IV estimands based on observable choices,

$$IV\left(\sum_e A_{ie} = 2\right) = E[Y_i(1) - Y_i(0) | \theta_i > c_1] \tag{1}$$

$$IV\left(\sum_e A_{ie} = 1\right) = E[Y_i(1) - Y_i(0) | c_1 \geq \theta_i > c_2]. \tag{2}$$

[4] Chassang et al. (2012) discuss providing subjects a "menu of lotteries" due to its identifying power in assessing selection effects in the context of experimental design.

[5] In the BDM literature, this is similar to the assumption that prices paid for the treatment/good do not directly affect outcomes (Berry et al., 2015; Jones, 2015).

These two equations aid in identifying selection in two ways. First, comparing the IV estimands can identify how treatment effects vary with individual preference valuation. For k events or event-blocks, we can identify selection effects with polynomials of order up to $k-1$, although we use a linear specification in the main paper. Second, these equations show that an event with vanishing costs (i.e., $c_e = 0$) can induce entry in the full population, thus allowing us to estimate the ATE non-parametrically. This is similar to an "identification at infinity" assumption on entry costs (Heckman, 1990).

Identification with distance: We also make use of varying distances between potential applicants and lottery events as an instrument for participation. This description closely follows Walters (2018).

We make small modifications to the assumptions in this case. Assume individuals make choices $A_{ie} = 1[\theta_i + \beta b_{ie} > c_e]$ where b_{ie} is distance to event e. Define potential outcomes $Y_i(d,z,b)$ for treatment d, lottery win z, and distance b, and potential treatments $D_i(z,b)$. We can then write IV estimands

$$IV(b,e) = \frac{E[Y|Z_i = 1, B_{ie} = b, A_{ie} = 1] - E[Y|Z_i = 0, B_{ie} = b, A_{ie} = 1]}{E[D|Z_i = 1, B_{ie} = b, A_{ie} = 1] - E[D|Z_i = 0, B_{ie} = b, A_{ie} = 1]}$$

Suppose that there is a distance d^* such that

$$\lim_{b \to b^*} Pr[A_{ie} = 1 | B_{ie} = b] = 1$$

and $Y_i(d,z,b) = Y_i(d,z)$, that we can exclude distance from our potential outcomes, then we can rewrite

$$\lim_{b \to b^*} IV(b,e) = \frac{E[Y|Z_i = 1] - E[Y|Z_i = 0]}{E[D|Z_i = 1] - E[D|Z_i = 0]}.$$

Assuming further, prefect take-up and exploiting exogeneity/exclusion on contract instrument Z_i, we see

$$\lim_{b \to b^*} IV(b,e) = E[Y_i(1,e) - Y_i(0,e)].$$

This identifies the average treatment effect in event e, which if we assume that events are excludable from potential outcomes, we identify the average treatment effect. This strategy too requires an "identification at infinity" argument to identify the ATE.

Applying this to the empirical context: Our empirical application nicely fits this description. Individuals make choices over multiple events, which we restrict to 3 main events to maintain the same sample restriction, although 96 event-blocks also create variation in entry rates. The perfect first stage take-up in this context also limits the need for variation across events and by distance since unlike most applications, there is no selection at that stage.

There are some features of the empirical context that differ from the above description. Most notably, we do not observe the full set of choices A_i for all individuals i, since individuals who win an earlier lottery are not able to apply to all future events. This means we cannot directly calculate the quantities of Equation 1. We use the structure of our choice model to encode the choices we do observe for individuals and then we estimate selection effects across the unobservable control function from the choice model and observable characteristics as well.

A4.2 Choice and outcome model fit

In this section, we discuss choice and outcome model goodness-of-fit. Firstly, the structure of the choice model aids this. Equation 3 includes alternative-specific constants for each event-block combination. Consequently, the choice errors, v_{ieb}, are mean zero within each alternative (Train, 2009) and, thus, our choice model can match population event-block entry shares by construction.

We therefore evaluate the goodness-of-fit of our choice model to predict population heterogeneity. We predict entry using observable characteristics of individual covariates, distances, and win probabilities, integrating out the random coefficient of unobservable preference for entry (Train, 2009). Panel A of Figure A6 shows a scatterplot of the relationship of actual entry observed in the data by model-predicted entry rates at the event-level binned in ventiles, as in Walters (2018). We find the relationship closely tracks a plotted 45-degree line, indicating that the predictions match the empirical data well both in levels and heterogeneously. The slope of the relationship is 0.98 with an $R^2 = 0.997$ indicating that we are closely able to rationalize heterogeneity in choice preferences.

Panel B of Figure A6 plots a similar relationship for the model of outcomes from Equation 10. We plot the empirical likelihood of starting a business by the model-predicted likelihood of starting a business, split into ventiles separately for winners and nonwinners. As in our counterfactuals, these outcomes are generated by simulating the population by drawing with replacement from the distribution of observable characteristics and independently drawing from the distribution of unobservables, with a population 100 times the size of the empirical distribution. We simulate choices, lottery wins, and outcomes with parameters based on our empirical estimates. The relationship of the model prediction to empirical outcomes are similar for both groups—the results closely track the 45-degree line. For nonwinners, the slope is 1.004 with $R^2 = 0.995$, while for winners, the slope is 0.989 with $R^2 = 0.971$. The outcomes have more variance for winners than for nonwinners: the bottom of the distribution is similar for nonwinners, but the highest outcomes in the distribution of winners are much higher than for nonwinners. This indicates a positive average treatment effect and important heterogeneity, consistent with other results in this paper.

Together, our choice and outcome models fit the empirical data well. The following section provides additional validation that the model replicates untargeted empirical moments.

A4.3 Comparing reduced-form and model-predicted estimates

Table A10 compares our reduced-form analysis with model-predicted effects to provide additional validation of our model and also shows model-based results for commonly reported population parameters. Column (1) presents the lottery-based, reduced-form estimated treatment effects by including event-block fixed effects. Column (2) shows the model-based predictions, which we generate from a large population entrants, simulating choices and winners, and assigning outcomes based on simulated win status.

We compare our reduced-form results and model on two common parameters, a reduced-form and a LATE-like estimand, as an additional validation that the model can rationalize untargeted moments of the data. The first reduced-form parameter, called RF in the table, is a regression estimate of the main outcome of any business creation on lottery winning, controlling for event-block fixed effects.[6] The model-based estimate is very close to the true quantity, 0.084 compared with 0.082, and we clearly cannot reject that the lottery-based reduced-form estimate is different from the model-based estimate. We also show a version of a "LATE" parameter where we instrument for whether an individual ever won a contract. Across these outcomes, we also find the results are very similar for the empirical reduced-form and the model-based estimate, 0.085 compared to 0.087, respectively. Together, the results show the model accurately replicates the reduced-form findings.

We further use the model to extrapolate to other population-based estimands. To calculate the TOT, TNT, and ATE, we use our model-simulated outcomes of $Y_i(1) - Y_i(0)$ averaged over the simulated population that ever wins, never wins, and the full population, respectively. We find the TOT estimate, 0.088, to be very similar to the LATE estimate,

[6] The reduced-form parameter we estimate in this table (0.082) is similar to the estimate in column (1) in Table 2 (0.078). It differs slightly because the outcome here is any business creation between 2012 and 2018, while the version in Table 2 measures any business creation.

0.087. This is not surprising because almost everyone who wins a contract takes it and the likelihood of a control group individual winning a future contract is low. Thus, there is little selection driving differences between these parameters. It also indicates that OLS inverse-variance weighting across event-blocks has a limited effect compared with uniform weighting employed by the TOT calculation. The ATE and TOT parameters are similar to each other, but much smaller than the TOT, with the ATE being 0.061 and the TNT being 0.059. This is consistent with the positive selection we observed in Section 5.8, and shows that the program structure targets along the dimension of future firm creation. The close relationship between the average treatment effect and the treatment-on-the-not-treated parameters is reflective of the few contracts being awarded as compared to the size of the population. These parameters are not necessarily policy-relevant, so we calculate policy-relevant counterfactual effects based on shifting supply-side parameters in Section 6.

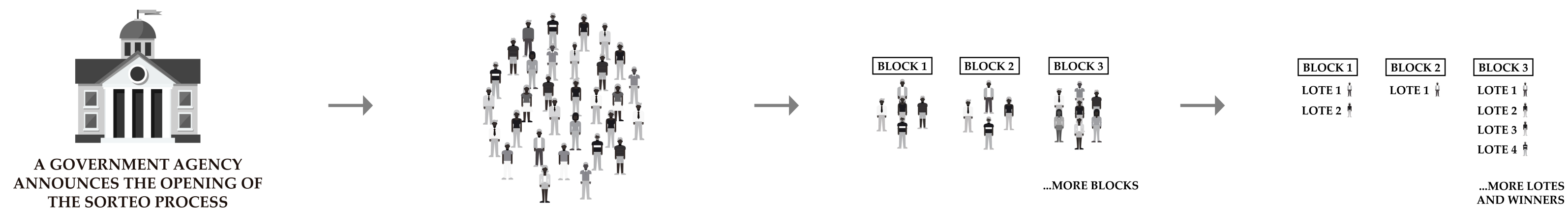
A4.4 Assessing panel-based analysis of supply-side parameters

In Section 6.1, we seek to understand how lottery block conditions affect utility for participants and ultimately entry. To do this, we use a panel fixed effect model that seeks to identify the effect of four characteristics of lotteries—lottery share in the population, number of contracts, average contract size, and win probability—on the entry utility in that event-block. We seek to identify causal effects assuming constant treatment effects in the population and a strict exogeneity asssumption. A key concern with this assumption is that the same event-block characteristics could be correlated with other local economic conditions that drive lottery entry, rather than the covariates themselves.

To help motivate the exogeneity assumption, columns (3-5) in Table A8 show regressions of the correlation of local economic conditions on the same lottery characteristics and time and province fixed effects, the same regression specification we use to estimate our effects on estimated utility. We measure local economic conditions of employment rates, hourly wages, and education rates from the Dominican employment survey, the ENFT.[7] We do not find that the lottery characteristics are correlated to population characteristics for any of the regressors, across all of the outcomes. This exogeneity seems plausible because the variation in characteristics of the lottery events over time aren't targeted towards time-varying changes in local conditions, but rather primarily towards deficiencies in future local schooling infrastructure around the country.

[7]We obtained yearly employment surveys (*Encuesta Nacional de Fuerza de Trabajo*, ENFT) for 2010-2016. Administered every April and relying on a two-stage cluster sample, the ENFT captures general demographic, education, employment, and socioeconomic characteristics of the country's households.

Figure A1: *Sorteo* chronology



A *sorteo* process begins by a government agency making a public call, which includes information on the blocks (if any), number of contracts, and minimum requirements to participate.

All interested participants (individuals and firms) present their documentation to the government agency making the call.

If there are multiple blocks in the lottery event, participants choose one to enter at the time of document submission. In the same event, a unique individual can only be listed in one block.

In public events and per block, the government agency draws the names of the winners for each contract. In a block, there can be multiple winners or just one, depending on the event.

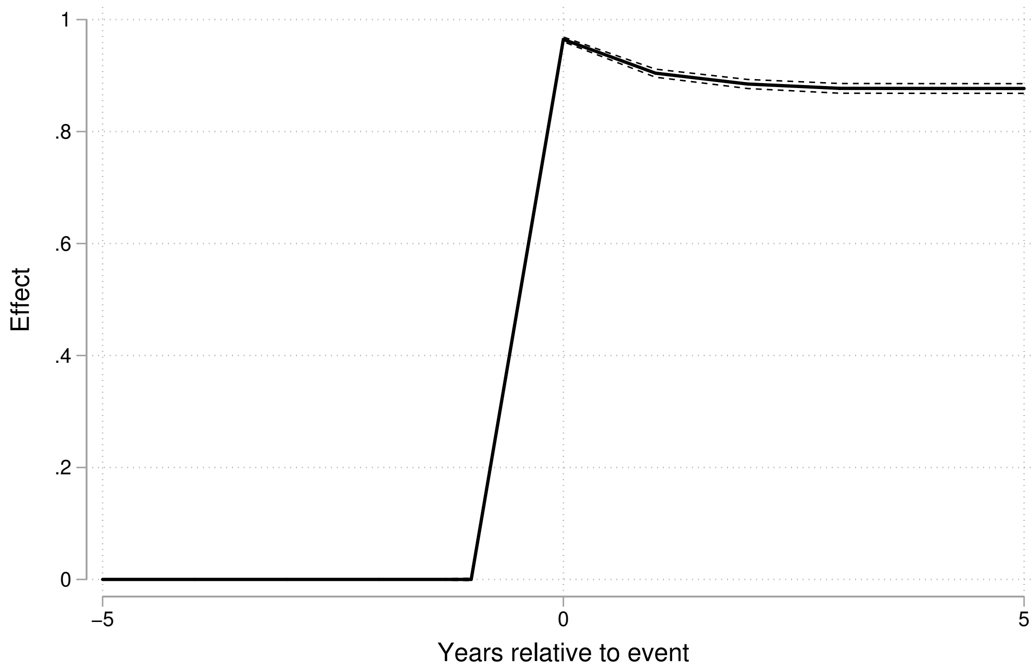
Figure A2: Lottery randomization

(a) Selection of lottery winners

(b) Winner celebrating

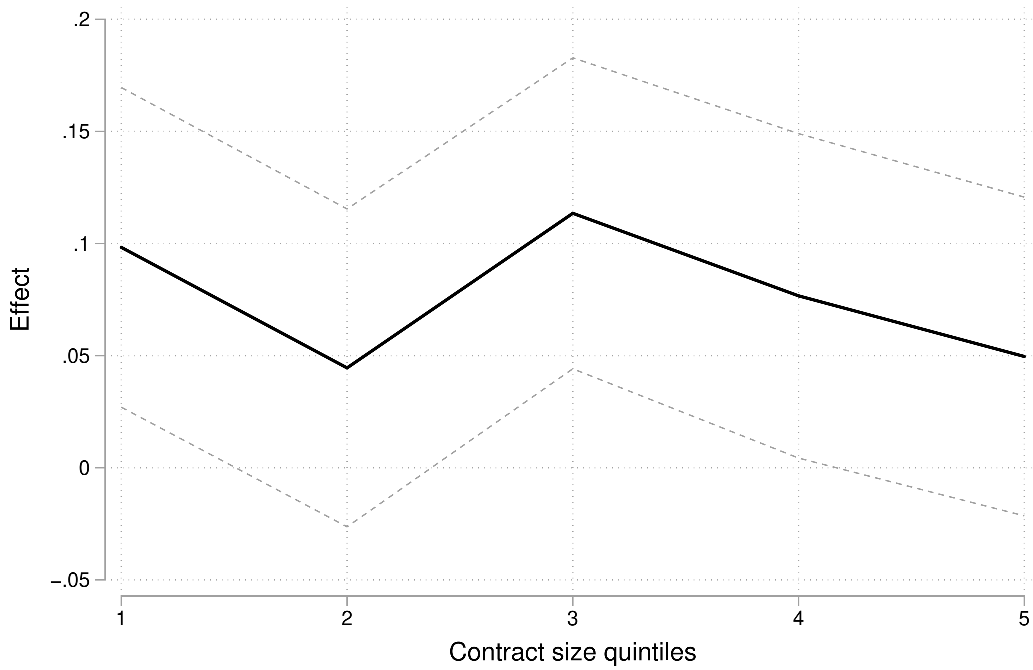
The images show the public lottery events. Panel (a) shows entrant cards in the transparent cylinder being spun/randomized. Panel (b) shows a winner whose card was pulled. **Source:** *Acento* and *Diario Libre*.

Figure A3: Ever winner by whether won in this year



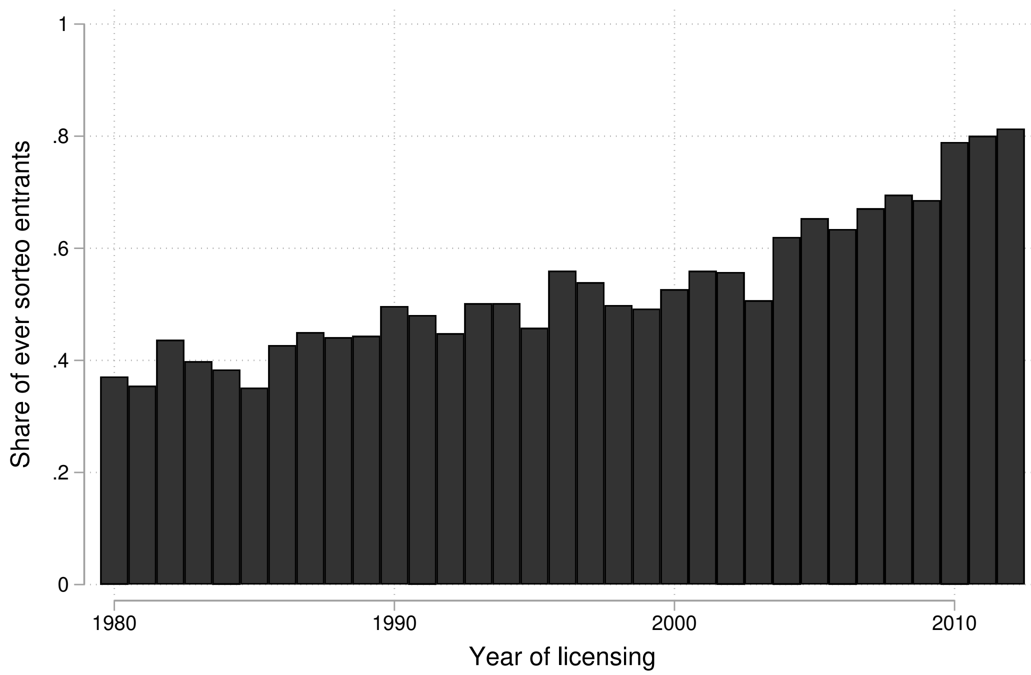
The figure displays regressions of whether an entrant has won a lottery up to year *t* on lottery winning. Regressions are estimated separately for each year relative to the lottery event date and include event-block fixed effects. Dotted lines represent 95% confidence intervals for standard errors clustered by individual.

Figure A4: Effect on firm ownership by contract size quintiles



The figure displays estimates from a regression of ever firm ownership on lottery winning interacted with quintiles of contract size. Regressions include lottery event-block fixed effects. Dotted lines represent 95% confidence intervals for standard errors clustered by individual.

Figure A5: Comparison of entrants and non-entrants

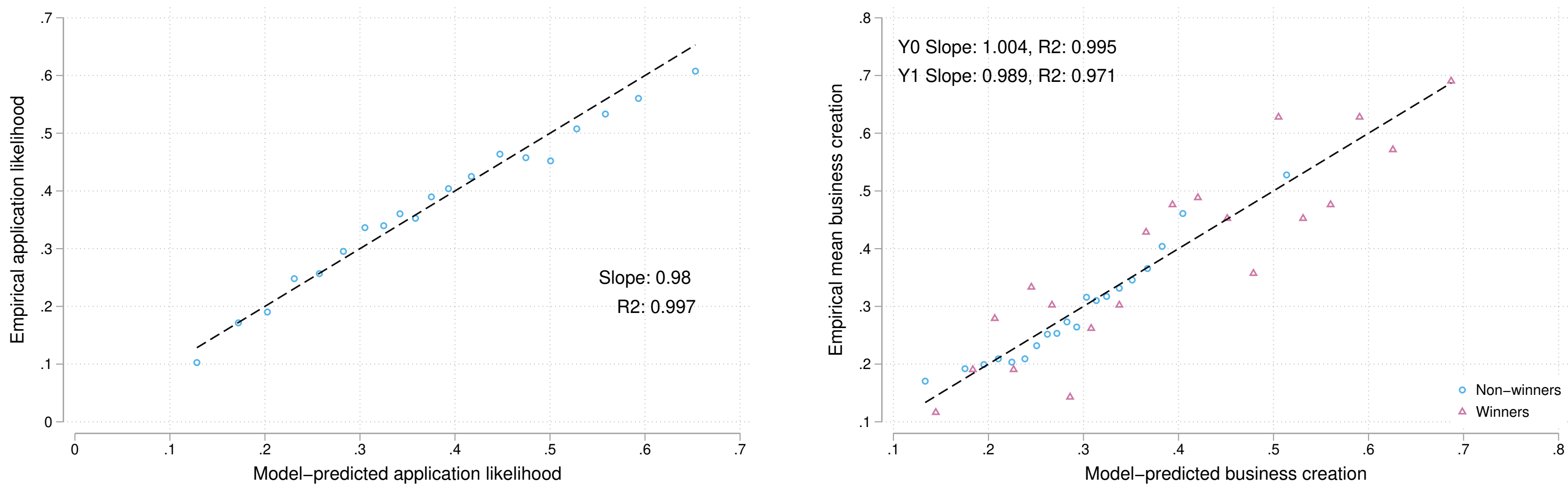


This figure shows the share of individuals ever entering a lottery by year of civil engineering or architectural licensing. The full sample is taken from Exequatur records, the registry of all licensed engineers and architects in the country.

Figure A6: Selection model evaluation

(A) Model vs empirical choices

(B) Model vs empirical outcomes



Panel A: Each point is a ventile of model-predicted and empirical application rates. The slope and R^2 reflect the ventiles. Panel B: Each point is a ventile of model-predicted and empirical application rate, separately measured for winners and nonwinners based on empirical data. The slope and R^2 for each group reflect the ventiles.

Table A1: Expected benefits from entering a *sorteo*

	No	Yes	N
Helps the individual or firm learn construction skills	.17	.83	681
Improves your social network to find future work	.24	.76	681
Makes it more likely that a future public contract will be awarded	.36	.64	681
Makes it easier to find a job in the private sector later	.40	.60	681

This table presents the reported anticipated benefits of entering a construction lottery. The data come from a survey of *sorteo* participants conducted in early 2019. Respondents were asked: "Hypothetically, for the average individual or firm that has not previously won a public works lottery, in your opinion, winning a public works lottery would..." Respondents were then prompted to select multiple from a provided list. The sample includes all surveyed winners and non-winners.

Table A2: Impacts on other firm outcomes

	(1)	(2)	(3)	(4)	(5)
	Panel A: Firm owner in 2018 by size				
	≥ 0	≥ 1	≥ 3	≥ 5	≥ 10
Winner	0.076***	0.032**	0.024**	0.009	0.008
	(0.017)	(0.013)	(0.011)	(0.010)	(0.008)
Control Mean	.34	.14	.093	.068	.041
N	16855	16855	16855	16855	16855
	Panel B: Firm ownernship by type				
	Starting owner	Majority owner	Constr-uction	Comm-erce	Other sector
Winner	0.069***	0.051***	0.064***	0.024**	0.028**
	(0.017)	(0.013)	(0.015)	(0.010)	(0.013)
Control Mean	.32	.11	.2	.058	.15
N	16855	16855	16855	16855	16855
	Panel C: Created firm government contracts				
	Num entered	Sorteo: Ever won	Non-sorteo: Ever won	Non-sorteo: Num won	Non-sorteo: Asinh(value)
Winner	0.140***	-0.000	0.006	0.004	0.001
	(0.040)	(0.003)	(0.006)	(0.013)	(0.015)
Control Mean	.2	.0055	.026	.055	.051
N	16855	16855	16855	16855	16855
	Panel D: Loans & Alternate firm registration (ONAPI)				
	Loans: Number	Loans: Amount	ONAPI: Any firm	ONAPI: Num firms	ONAPI: Num firm shares
Winner	0.097	0.816	0.033**	0.036*	0.034*
	(0.096)	(0.718)	(0.014)	(0.020)	(0.018)
Control Mean	.76	5.3	.16	.2	.18
N	698	698	16855	16855	16855

This table reports estimates of the effect of lottery winning on various firm outcomes. Panel A investigates binary outcomes of firm ownership in 2018 with employment $\geq X$ where X varies across the columns. Panel B investigates binary outcome based on the name: starting firms is defined as firm ownership within 6 months of firm registration; majority owner is defined as having a firm ownership share above 50%; other sector firms are neither construction nor commerce firms. Panel C investigates created firm interactions with the government. Column (1) reports all firm creation. Columns (2-5) report firm government contracts aggregated up to the individual level. *Sorteo* contracts refer to the randomized procurement scheme, while Non-*sorteo* contracts are non-randomized procurement schemes. Panel D investigates firm creation outcomes using an alternative source of firm ownership, ONAPI business records. Sample is restricted to individuals. Controls include lottery event-block fixed effects. Standard errors clustered at the individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: Perceived benefits and drawbacks from entering a *sorteo*

	No	Yes	N
Helped:			
It provided me valuable work experience and skills	.38	.62	292
I have more money to invest on my career or business	.45	.55	292
It showed me that I can do bigger projects	.58	.42	292
I have a better chance of getting contracts in the public sector	.78	.22	292
I have a better chance of getting contracts in the private sector	.78	.22	292
It helped my career in other aspects than the above	.97	.03	292
Hurt:			
I lost money because of it	.90	.10	292
It was a waste of time, I could have found a better job	.95	.05	292
It made me lose some connections	.95	.05	292
It helped my career in other aspects than the above	.97	.03	292

This table presents the perceived benefits and drawbacks of participating in a construction lottery. The data come from a survey of *sorteo* participants conducted in early 2019. Respondents were asked: "Did participating in the lotteries and/or obtaining a contract help or harm your professional or business development?" If they responded that it helped, they were prompted to select from the options in the first section of the table. If they responded that it hurt, they were prompted to choose from the second section. Respondents who indicated both help and harm were shown both sets of options. Those who answered "neither" were excluded from further prompts and are counted as "No" in the table. The sample is restricted to winners only.

Table A4: Heterogeneous effects of lottery winning on firm ownership

		RF		IV	
	(1)	(2)	(3)	(4)	(5)
	Control Mean	Separate Interactions	Joint Interactions	Separate Interactions	Joint Interactions
Age	42.84	-0.006***	-0.007***	-0.006***	-0.007***
		(0.001)	(0.001)	(0.001)	(0.002)
Female	0.28	0.022	0.001	0.022	0.001
		(0.040)	(0.040)	(0.042)	(0.043)
Ln income, t-1	10.62	-0.003	0.001	-0.003	0.001
		(0.003)	(0.004)	(0.003)	(0.004)
Formal employed, t-1	0.55	-0.040	-0.064	-0.038	-0.063
		(0.034)	(0.040)	(0.036)	(0.042)
From capital	0.45	-0.033	0.002	-0.035	0.001
		(0.035)	(0.036)	(0.037)	(0.038)
N			16418		16418

This table reports estimates of heterogeneous effects of lottery winning on being a firm owner. The sample is individuals. The dependent variable is whether the individual ever became a firm owner. Column (1) is the mean of the heterogeneity variable in the control group. Column (2) shows reduced form coefficients of the interaction terms $Winner * Variable$ estimated in separate regressions. Column (3) reports the same interaction terms estimated jointly. Columns (4-5) present heterogeneous IV coefficients on the variables $Takeup * Variable$ instrumented by the reduced-form regressors. Controls in all regressions include the full set of interacted variables and lottery event-block fixed effects. Standard errors clustered at the individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Firm co-ownership networks

	Firm co-ownership		
	(1)	(2)	(3)
Same block	.00014***		
	(.000033)		
Winners	.000061**	.000064**	
	(.000028)	(.000028)	
Same block * Winners	.0026	.0022	.0028
	(.0019)	(.0019)	(.0017)
Block and block FE	X		X
Block-block FE		X	
Sample Mean	.00	.00	.00
N	140564194	140564194	390286
Sample	All	All	Winners

This table analyzes firm co-ownership with other entrants in the same *sorteo*-block. Sample is from most completed sorteos. An observation is at the entrant by entrant link level. Same block means the links come from the same *sorteo*-block. Winners means that both linked entrants were winners. Fixed effects vary by regression. Block-block fixed effects are full interactions between all pairs of entrant *sorteo*-blocks. Block by block fixed effects are separate fixed effects for each entrant. Column (3) is restricted to only lottery winners. Standard errors are clustered at the block-block level and in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A6: Heterogeneous effects based on number of past lottery entries

	Ever firm owner	
	(1)	(2)
Winner * Num prev entries	.065*	.074**
	(.035)	(.035)
Sample Mean	.32	.32
Other interactions	No	Yes
N	12326	12287

This table reports estimates of the heterogeneous effect of lottery winning based on past number of lottery entries. The sample excludes the first lottery event. Column (1) includes the interaction between winner and number of past event entries, while column (2) includes additional covariates interacted with lottery winner. All regressions include covariates of number of past entries, age in 2011, employment in 2011, asinh income in 2011, male, and whether from the capital interacted with event-block fixed effects. Standard errors clustered by individual. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A7: Basic choice

	(1) Choice
Mean	
Distance (km)	-.036***
	(.00028)
Age	.34***
	(.016)
Age2	-.0032***
	(.00018)
Female	-.19***
	(.065)
Employed	.36***
	(.059)
Income	3.7e-08***
	(1.1e-08)
Prev owner	-.064
	(.11)
θ	-2.9***
	(.11)
SD	
θ	2.8***
	(.046)
Event-block FEs	Yes
LL	-53399
Chi2	8013
N	1425368

This table reports MSL estimates of the parameters of applicant preferences for lottery entry. The first panel shows effects at the mean. The parameter θ is the mean of a normally distributed random coefficient for entry, and the second panel shows the standard deviation of this parameter. The model additionally includes event-block-specific constants. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A8: Recover supply-side parameters and test exogeneity

	Event-block utility ($\hat{\Phi}_{eb}$)		Local labor conditions		
	(1)	(2)	(3) Secondary grad.	(4) Employed	(5) Hourly wage
Mean contract size	.21	.44	.017	-.0023	4.8
	(.28)	(.3)	(.012)	(.0084)	(3.7)
Mean contract size (sq)	-.047	-.078**	-.0018	.00045	-.38
	(.033)	(.038)	(.0013)	(.00094)	(.4)
Lottery share in pop.	11***	13***	-.0092	.12	21
	(2.2)	(1.8)	(.053)	(.11)	(28)
Number of contracts	.028***	.016**	.000041	.000015	-.0046
	(.0094)	(.0072)	(.000033)	(.000038)	(.0089)
$1/\pi$	.012*		-.000064	.00016	.076
	(.0064)		(.00022)	(.00028)	(.063)
$1/\pi$ (lagged)		.00089			
		(.00072)			
N	95	95	86454	46852	46852

This table shows estimates of supply-side and cost parameters in columns (1-2) and estimates of the relationship between lottery event-block characteristics and sociodemographic and economic characteristics of their respective province in the year prior in column (3-5). For columns (1-2), column (1) incorporates the current lotteries probability of winning a lottery while column (2) incorporates the lag of the probability from the previous event. The coefficient on $1/\pi$ is the cost variable c. In columns (3-5), the sample is respondents to the ENFT survey. The regressors are the share of contracts within the local population, number of contracts, average contract size and it's square, and the inverse probability of winning a contract, on average. All regressions include province fixed effects, with columns (1-2) also include event fixed effects while columns (3-5) also include year fixed effects. Standard errors are clustered at the province level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A9: Effects on individual's total and net income

	Total inc.	Net inc.	
	(1)	(2) Preferred	(3) Stated
Winner * Year 0	101148***	17992***	20927***
	(5324)	(1879)	(1849)
Winner * Year 1	252480***	31344***	42236***
	(9616)	(3617)	(3654)
Winner * Year 2	191057***	24613***	31071***
	(9877)	(3680)	(3668)
Winner * Year 3	62505***	7317***	8662***
	(5410)	(1928)	(1885)
Winner * Year 4	28595***	1254	2608*
	(4034)	(1425)	(1359)
Winner * Year 5	14891***	-1981	-925
	(3453)	(1216)	(1098)
Control Mean	32430	12089	12950
N	100194	100194	100194

This table analyzes the effect of lottery winning on individual total income and net income measures. Column (1) measures total income. Column (2) is the preferred measure of net income, calculated as the minimum of stated net income on tax filings and filed income minus reported local purchases for goods. Column (3) is reported net income. Sample is restricted to most completed sorteos. Each column is a separate regression. Controls include event-block by period fixed effects. Standard errors are clustered at the entrant by event-block level and in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A10: Comparing lottery-based and model-based analysis

	Lottery-based analysis	Model analysis
	(1)	(2)
RF	.082	.084
LATE	.085	.087
TOT		.088
ATE		.061
TNT		.059

This table compares estimates from the reduced-form, lottery analysis to the model-based analysis for multiple estimands. RF corresponds to the main reduced-form analysis of outcomes on lottery winning and event-block fixed effects. The LATE parameter is the same analysis incorporated into a 2SLS specification in which lottery winning instruments for ever winning. TOT is the treatment-on-the-treated estimate. ATE is the average treatment effect. TNT is the treatment-on-the-not-treated effect. The outcome is any business creation between 2012 and 2018. Standard errors are clustered at the individual level in the reduced-form analysis. The model estimates are derived from a simulated population which samples with replacements from the empirical distribution of observable covariates, and samples randomly from the distribution of unobservables. Choices, lottery winning, and outcomes are then simulated according to the estimated choice and selection estimates.