NBER WORKING PAPER SERIES

WHAT PREDICTS THE GROWTH OF SMALL FIRMS? EVIDENCE FROM TANZANIAN COMMERCIAL LOAN DATA

Mia Ellis
Cynthia Kinnan
Margaret S. McMillan
Sarah Shaukat

Working Paper 31620
http://www.nber.org/papers/w31620

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2023

We are grateful to the National Microfinance Bank of Tanzania and its staff who made this research possible, especially Ineke Bussemaker, Richard Donatus, James Maiteron, Dionys Msavila, and Beatrice Mwambije. The authors also thank Josaphat Kweka (Talanta International, Economic and Social Research Foundation) for his efforts in-country, including facilitating the focus groups and survey work. We also acknowledge the excellent work of Ideas in Action, led by Samson Kiware, for their enumeration of the firm survey. We thank the International Growth Centre (IGC) and IFPRI-led CGIAR's research program Policies, Institutions, and Markets (PIM) for support. We thank seminar audiences at CSAE, Tufts, the IPA Entrepreneurship and Private Sector Development working group, the Gendered Effects of Globalization and Development conference, and especially Ama Baafra Abeberese for helpful comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

What Predicts the Growth of Small Firms? Evidence from Tanzanian Commercial Loan Data
Mia Ellis, Cynthia Kinnan, Margaret S. McMillan, and Sarah Shaukat
NBER Working Paper No. 31620
August 2023
JEL No. G14,J16,O16

ABSTRACT

Not all firms have equal capacity to absorb productive credit. Identifying those with higher potential may have large consequences for productivity. We collect detailed survey data on small- and medium-sized Tanzanian firms who borrow from a large commercial bank, which in turn raises funds via international capital markets. Using machine learning methods to identify predictors of loan growth, we document, first, that we achieve high rates of predictive power. Second, "soft" information (entrepreneurs' motivations for entrepreneurship and constraints faced) has predictive power over and above administrative data (sector, age, etc.). Third, there is a different and larger set of predictors for women than men, consistent with greater barriers to efficient capital allocation among female entrepreneurs.

Mia Ellis
IFPRI
1201 Eye Street
Washington, DC 20005
m.ellis@cgiar.org

Cynthia Kinnan
Department of Economics
Tufts University
177 College Ave #620
Medford, MA 02155
and NBER
cynthia.kinnan@tufts.edu

Margaret S. McMillan
Tufts University
Department of Economics
177 College Ave #669
Medford, MA 02155
and International Food Policy Research Institute
and also NBER
margaret.mcmillan@tufts.edu

Sarah Shaukat
Department of Economics and
Fletcher School
Tufts University
177 College Ave #652
Medford, MA 02155
sarah.shaukat@tufts.edu

An online appendix is available at http://www.nber.org/data-appendix/w31620

1 Introduction

It is well-established that not all firms are equally productive, nor equally able to absorb productive credit. Even within a particular location and sector, estimated rates of return can vary significantly as shown by, among others, De Mel et al. (2008); Hsieh & Klenow (2009); Banerjee & Duflo (2014); Beaman et al. (2020). Moreover, there is a growing recognition that this variation in returns is to some degree predictable, both by entrepreneurs themselves and by others in their communities (e.g. Banerjee et al. 2020, Hussam et al. 2022).

However, econometrically predicting which firms will grow and which will not remains challenging (McKenzie & Sansone, 2019). Banks themselves struggle to predict which firms should receive more credit, either due to placing incorrect weight on observables (Bryan et al., 2021) or due to within-firm misalignment of incentives (Rigol & Roth, 2021). The difficulty in predicting firms' borrowing outcomes was summed up by Banerjee et al. (2015): "The modest R-squareds in ... borrowing-likelihood models estimated in this [AEJ Applied Jan 2015] volume suggest .... it is difficult to predict microcredit use even with a rich set of baseline characteristics of potential borrowers."

The question of whether it is possible to identify inefficiencies in credit allocation, and to mitigate them, is of both policy and academic interest. Lenders, NGOs, governments and others who play a role in allocating capital would like to channel resources to those who are best able to absorb them. The aggregate consequences of improving the efficiency of capital allocation are potentially very large, as shown by Bau & Matray (2022).[1]

In this study, we aim to use the history of loan data (an outcome which is not subject to manipulation by firms)[2] to better understand why some small and medium sized firms grow and others do not. To do so, we worked with National Microfinance Bank (NMB), the second-largest commercial bank in Tanzania.

The experience of NMB illustrates the role of access to international capital in shaping credit access, particularly for female-owned firms. Prior to 2005, NMB was entirely owned by the government of Tanzania. NMB underwent privatization in 2005, and this was associated with a large influx of foreign capital from Rabobank in the Netherlands (Cull & Spreng, 2011). NMB's partnership with Rabobank not only increased its capital base, but also influenced the corporate culture at NMB by increasing the bank's awareness of the importance of lending to businesses who

[1]This is not, of course, to deny the importance of factors other than productivity/demand in the decision of how to allocate capital.

[2]For a discussion of "gaming" and manipulation, see Björkegren et al. (2020). Hussam et al. (2022) find that rankings of own and others' productivity are also subject to manipulation when there are stakes to doing so.

previously had limited access to capital, including female-owned businesses.[3]

Working with NMB and a local survey firm, we carried out a detailed survey of a sample of NMB's clients, i.e., firm owners. This data allows us to capture the constraints firms face, and owners' entrepreneurial motivations, including less tangible aspects which would be difficult to study using administrative data alone. Thus, our methodology provides a novel way to assess to what extent Rabobank and NMB's stated goal of reducing inequities in capital allocation has been successful.

Our aim is to determine the defining characteristics of high-growth small firms and how these characteristics differ from those of low/no growth small firms. Specifically, we focus on predicting growth in loan balances. This is of interest for several reasons. First, lenders are often interested in understanding which customers have the capacity to increase borrowing over time — indeed this was an interest of NMB. Additionally, given that most firms cannot achieve rapid growth in scale or profits solely relying on retained earnings, understanding which firms can increase borrowing is relevant to understand the potential for firm growth more broadly. Indeed, there is a long tradition of studying constraints to firm growth through the lens of capital markets, e.g. Evans & Jovanovic (1989); Banerjee & Newman (1993).[4]

Given the evidence of differential constraints faced by female entrepreneurs (e.g., Chiplunkar & Goldberg 2021; Hardy & Kagy 2018, 2020; Brock & De Haas 2022 and Morazzoni & Sy 2022), we separately examine male- and female run firms. Such information may be used by both public and private sector stakeholders to inform lending programs and public policies targeted at small firms and private sector development. While direct application of this information must be mindful of Lucas critique considerations (Lucas, 1976), this approach can indicate groups which face particularly high barriers to efficient capital allocation, which in turn may suggest interventions to "level the playing field."

We first document that typical performance metrics are correlated with loan growth outcomes — high-loan growth firms are more likely to pay taxes, employ more workers, and have higher monthly revenue and costs, as well as a higher value of collateral. They are operated by younger owners, on average. Next, given the fact that we have more potential predictors than we have observations, and to guard against the possibility of false discovery — finding spurious correlations in the data — we use a data-driven approach, using LASSO (least absolute shrinkage and selection operator) to optimally select predictive variables.

We demonstrate that characteristics "beyond the balance sheet" are predictive of loan growth. Some such characteristics are likely proxies for productivity, for instance

[3] See Section 3 for more background on NMB.

[4] We return to these issues in section 3.2.1.

reporting constraints related to lack of market access or lack of managerial skills. However, other "irrelevant" characteristics are also predictive: for instance, having experienced a traffic accident and experiencing issues with loan documentation being in English. We also validate that predicted loan growth (using the LASSO-selected characteristics) is predictive of firm profits, sales, stocks and employment.

When we disaggregate our results by gender, striking differences emerge. For male entrepreneurs, loan growth is largely predicted by characteristics such as record keeping and financing business growth with retained earnings — characteristics that closely relate to economic fundamentals. For the female sample, however, we find less predictive power of fundamentals, and instead find that challenges such as loan documentation being in English, the need to get signatures from a local government authority, and experiencing a theft affect female firm owners' loan growth. This finding that the characteristics predicting loan growth differ for men and woman, and that women have a larger set of predictors, reveal a degree of misallocation among this sample of commercial bank borrowers.

We are able to achieve high rates of predictive power in forecasting loan growth. While the pooled model yields an R^2 of 0.562, the gender-specific models achieve R^2s of 0.757 (men) and 0.869 (women). Thus, forcing the same prediction model to fit both male and female borrowers results in a worse fit, speaking to the different, and larger, set of predictors for women. In contrast, under the efficient benchmark, only fundamentals (current capital stock, savings rate, etc.) would matter, and the model would be equally predictive for women and men. These findings speak to the presence of greater barriers to efficient capital allocation for women, and to the need to explore supply- and/or demand-side interventions to facilitate more equitable access to capital, particularly for female entrepreneurs.

Related Literature

Our work relates to and builds on several literatures. One is the literature on predicting the productivity of microentrepreneurs. Like a number of recent papers (Banerjee et al., 2019; McKenzie & Sansone, 2019; Hussam et al., 2022; Beaman et al., 2020; Banerjee et al., 2020; Bryan et al., 2021; Coad & Srhoj, 2019), we examine the extent to which entrepreneurial performance can be predicted using characteristics observable to the econometrician. We build on this literature in several ways. First, we use a survey of firm owners which captures an unusually detailed degree of attributes. Our survey moreover includes a range of characteristics that, to our knowledge, have not previously been considered, including information about entrepreneurs' motivations for entrepreneurship and constraints faced in financing and growing their businesses. As discussed below, like several other recent papers (eg, Hussam et al. (2022) and Bryan et al. (2021)) we use machine learning (ML) methods to appropriately deal with our high-dimensional dataset in a way that avoids overfitting and researcher discretion.

In addition, our work relates to the body of literature relating dispersion in marginal product to misallocation, and studying correlates of such dispersion to shed light on the causes of misallocation (Hsieh & Klenow, 2009; Banerjee et al., 2019). We consider the selected predictors of firm growth potential, separately by the gender of the entrepreneur. This allows us to both generate potentially more accurate forecasts, and to use the results to shed light on differential degrees of market failure and misallocation affecting women vs. men, a topic also recently studied by Chiplunkar & Goldberg (2021) and Morazzoni & Sy (2022).

A paper particularly close to ours is Bryan et al. (2021), who use a combination of an RCT and novel psychometric and cognitive data to understand capital allocation and returns in Egypt. They find that those firms with the highest predicted treatment effects substantially increase profits, while profits drop for poor performers; and that loan officers' status quo capital allocation is notably inefficient. We, like Bryan et al. (2021), use novel data sources to investigate the efficiency of capital allocations. While they use a RCT to generate exogenous variation in loan supply, we take an alternate approach and focus on interpreting equilibrium capital allocations though the lens of a simple model. Our approach can thus complement the insights of both supply-side interventions (Bryan et al., 2021; Banerjee et al., 2015) as well as demand-side interventions (Bertrand et al., 2010; Brailovskaya et al., 2021).

Another paper closely related to, and complementary to, ours, is Coad & Srhoj (2019), who use ML methods to predict firms which will be high-growth in Croatian and Slovenian data. Like us, they employ ML methods to deal with high-dimensional data. Their large sample size gives them a high degree of statistical precision and their focus on high-income countries sheds light on predictors in contexts where markets may function more smoothly. In contrast, while our smaller cross-sectional dimension may give us less precision, we have a significantly larger number of potential explanatory variables, including respondents' experiences of financial difficulty, etc., which is important in allowing us to fully characterize predictors. Moreover, conducting this exercise using data from Tanzania, a country with limited financial market development, and distinguishing male- and female-run businesses allows us to shed light on the differential factors that determine men's and women's success, or lack thereof, in an emerging economy.

In its focus on "beyond the balance sheet" data for credit targeting, our paper also relates to recent literature on psychometric credit scoring (Arráiz et al., 2017), which has been highlighted as a potential tool to increase women's access to finance (Alibhai et al., 2022). In addition, our work draws on a recent literature on optimal methods for post-selection inference on selected covariates (Lee et al., 2016; Meir & Drton, 2017; Banerjee et al., 2021).

The rest of our paper proceeds as follows. Section 2 presents a stylized model of optimal loan growth which we use to characterize (in)efficiency in access to capital. Section 3 describes our setting and section 4 describes our administrative and survey data. Section 5 presents the results and section 6 concludes.

2 Conceptual framework

To formalize our predictions of how capital should accumulate in the absence of inefficiencies, we begin with a Cobb-Douglas production function à la Mankiw et al. (1992). The output of firm i at time t will be:

$$Y_{i,t} = K_{i,t}^{\alpha}(A_{i,t}L_{i,t})^{1-\alpha} \tag{1}$$

where Y is output, K is capital, L is labor, A is a labor-augmenting productivity term, and $\alpha \in (0, 1)$ is is the capital share. Then, optimal growth of capital per worker, $k \equiv (K/L)$ is

$$\dot{k}_{i,t} = s_i k_{i,t}^{\alpha} - (n_i + g_i + \delta_i)k_{i,t} \tag{2}$$

where the firm's savings rate is s_i, the growth rate of its labor force is n_i, its productivity grows at rate g_i and its rate depreciation is δ_i. We assume that these characteristics may vary across firms but not across time.

This simple framework leads to the following implications:

Proposition 1 *Capital (normalized by effective labor) optimally grows at a rate governed by the firm's savings rate, the rate of growth of its labor force, its rate of productivity growth and its rate of depreciation.*

Proposition 2 *Optimal capital growth is independent of all additional factors, including predictors of productivity **levels**.*

These implications are testable with cross-sectional data on the growth of capital (in our case, the growth rate of loan balances), labor, (proxies of) productivity growth, savings, depreciation, and "irrelevant" (i.e., excluded) characteristics. By identifying which factors are statistically predictive of capital growth, the econometrician can test whether factors — other than productivity growth, savings, depreciation, and the stock of capital — are predictive of capital growth. If other factors are in fact predictive, this implies that Proposition 2 is rejected and hence that capital allocation is inefficient.

The intuition for this result is that, with common fundamentals, capital should flow to each firm until it attains the common economy-wide marginal product of capital (MPK). This is analogous to the convergence prediction in the economic growth literature (see, e.g., Quah 1996). In the presence of firm heterogeneity, capital flows should equalize MPK, adjusted for differences in fundamentals (Conditional convergence). If we find characteristics other than fundamentals predicting capital access, this indicates that equalization of (adjusted) MPK across firms fails, a result we will refer to as misallocation.[5]

3 Setting and Research Design

3.1 Setting: Global capital flows and the National Microfinance Bank

National Microfinance Bank (NMB) is the second largest commercial bank in Tanzania, and has a long-standing commitment to financial inclusion. In addition, NMB is well integrated into international capital markets: in September 30, 2005, 49 % of the shares of NMB were sold to Coöperatieve Centrale Raiffeisen-Boerenleenbank B.A. ('Rabobank Group') of the Netherlands. Rabobank played a key role in turning NMB into the successful bank it is today. Foreign capital injections helped to maintain bank solvency, but equally important was Rabobank's focus on training NMB's workforce. Moreover, while privatization can improve banks' efficiency, this can come at the cost of reduced lending, especially to the more risky small- and medium-sized enterprises. An early analysis of NMB under Rabobank showed that by 2007, NMB had expanded its ratio of loans to assets, installed 100 ATMs country-wide, and ended the year as the most profitable bank in Tanzania (Cull & Spreng, 2011). Today, NMB has 226 branches, over 9,000 agents, more than 700 ATMs across the country and serves over 4 million customers.

Foreign capital (particularly Rabobank) has played an important role in shaping the corporate culture at NMB. Historically, Rabobank had a commitment to people in the Netherlands with very limited access to capital markets. This commitment carried over

[5]We note that our empirical framework also admits an alternate interpretation: if the available measures of productivity growth; savings; and depreciation which we have identified *ex ante* are imperfect, the additional characteristics which are found to empirically predict capital growth can shed light on proxies for productivity growth, savings, and depreciation.

to its work in Tanzania as evidenced by its leadership of NMB. For example, from 2015 to 2019, the Managing Director of NMB was a Dutch woman, Ineke Bussemaker, with a strong commitment to financial inclusion and gender diversity. In 2021, Rabobank sold its share in NMB to Arise, B.V., a consortium of Rabo Partnerships, Norfund, NorFinance and FMO (a Dutch parastatal bank). Arise has a somewhat broader agenda than Rabobank; it is committed to investing in sustainable, locally owned African financial institutions with the goal of contributing to economic growth and poverty reduction in Sub-Saharan Africa.

3.2 Identifying High-loan-growth borrowers

Our approach begins with data from NMB with anonymized information on loans taken by firm owners from 2013 to 2019. These data, which cover loans from NMB's micro and small enterprise lending facility, cover 31,355 loans over the seven-year period, representing 10,002 unique firms. We limit our sample to active borrowers in and around Dar es Salaam.[6] For a client to be included in our sample, we require that he/she be an active borrower who has taken at least one loan from NMB in 2018 or 2019.[7] Moreover, we set the criteria that the client should have taken at least four loans from NMB between 2013 and 2019. Including only those clients who had borrowed multiple times allows us to establish trends in firm growth. This left us with a sample of 1,331 firm owners that had taken 5.7 loans on average from 2013 to 2019.

As the next step, we classify firm owners as belonging to the high growth or the low/no growth groups using data on the initial loan size and the growth rate of loan sizes. As shown in Figure B.1, we split the sample of 1,331 firm owners into terciles of initial loan size, and then split each tercile of initial loan size into quartiles of growth rates of loan sizes. A firm owner is assigned to the high growth group if he/she is in the highest quartile of growth in loan sizes for any tercile of initial loan size, and to the low/no growth group if he/she is in the lowest quartile of growth in loan sizes for any tercile of initial loan size. Therefore, the high growth group consists of 351 unique firm owners and the low/no growth group consists of 325 unique firm owners. Out of these, 212 clients were in branches (of NMB) that were too far outside Dar es Salaam to be surveyed. The remaining 464 firm owners – 235 (50.6%) high growth and 229 (49.4%) low/no growth – make up the potential sample for our study.

[6]We limited our sample to borrowers in and around Dar es Salaam to streamline data collection and allow a larger sample size within a fixed budget.

[7]We only include active borrowers in our analysis for several reasons. First, it would have been difficult to track and survey borrowers who had not used NMB's services in years. Secondly, this avoids the need to make assumptions about the status of inactive borrowers. For example, an inactive borrower from NMB could mean that he/she has switched to some other bank, that the firm is able to finance investment using internal funds, or that the firm is not growing and cannot productively absorb funds.

3.2.1 Loan growth vs. capital growth

Our empirical strategy focuses on loan growth. Loan growth — both its level and its correlation with firm outcomes — is of interest in and of itself, as a measure of financial inclusion and as a determinant of the performance of the banking sector. Given the fixed costs of customer acquisition and screening for banks and financial institutions, there is a natural interest in identifying customers who can productively absorb additional credit.

Moreover, as highlighted by the conceptual framework (Section 2), in a growing economy and in the absence of constraints, firms will optimally increase capital over time. While some firms expand solely using retained earnings, firms, particularly those with a high rate of optimal capital growth, may be unable to meet their capital needs entirely from retained earnings. The microfinance literature has found high returns to credit specifically for the subset of borrowers who use it to expand existing firms (e.g., Meager 2019; Banerjee et al. 2020).

External financing may also provide other benefits, such as a signal of quality and implicit or explicit insurance in the form of the ability to restructure debt in the face of negative shocks. It is also the case that some firms benefit from access to credit that is steady over time, however, that is not our focus in this paper.

In addition, loan growth is a key predictor of *capital* growth. At the country, industry and firm levels, use of external financing is positively correlated with economic growth (Beck et al., 2008, 2022). Indeed, in Section 5.5 we validate that, in our data, loan growth is predictive of firm profitability and scale.

From the perspective of external validity, our findings are relevant to the dimension of capital access which flows through external financing, namely bank lending.

3.3 Focus Groups

Like Arráiz et al. 2017, Alibhai et al. 2022, and Bryan et al. 2021, we want to understand the role of factors "beyond the balance sheet" in predicting loan growth. To facilitate a survey design sensitive to these issues, we conducted focus groups with a small number of active borrowers to learn details about their experiences. We interviewed groups of firm owners with impressive growth in loan size (the high-growth group) as well as firm owners with little to no growth in loan size (the low/no-growth groups).

In order to simplify logistics, we limited our focus groups to four branches of NMB. We identified branches with at least 15 borrowers in the high-growth group and at least 15 borrowers in the low/no-growth group. We then chose four branches[8] based

[8] These were Magomeni, Temeke, Mlimani City, and the Airport.

on the variation in initial loan size among our two groups so as to have borrowers from each initial loan size quartile in each meeting. To choose our 15 borrowers from each group at each branch, we randomly selected two borrowers from each initial loan size quartile to get the first eight participants. We then randomly selected the remaining seven participants from among those borrowers not yet selected.

Invitations were sent to a total of 120 borrowers. In total, 42 firm owners from the high-growth group, and 25 firm owners from the low/no-growth group, attended the meetings. The gender balance of the focus groups reflected the gender balance of the borrowers such that roughly one-third of the focus group participants were female. The two types of firm owners were interviewed separately, and all meetings were conducted in Swahili. These meetings were semi-structured interviews based on a set of 11 questions. The list of these questions is included in Appendix A. Firm owners were asked about their line of business and how it had changed over time, motivation for starting their business, perceptions about borrowing, history of and motivation for borrowing, experience with MSME organizations, employees, plans for growth, and major challenges. We analyzed the qualitative information collected during these meetings; the findings helped inform the design of a subsequent longer survey with a large set of owner characteristics. We combined administrative data from the bank with the more qualitative information from our survey into a dataset that allowed us to assess the relative importance of "off-balance sheet data" in predicting loan growth.

3.3.1 Focus Group Results

High-growth firms had a somewhat greater range of business activities and were also more likely to have diversified their income by operating multiple businesses at the same time. High-growth firms are less likely to be "reluctant entrepreneurs" (Karaivanov & Yindok, 2017; Banerjee et al., 2020), in that they are more likely to have chosen their activities for specific and positive reasons. The low/no growth firms, on the other hand, were more likely to have been pushed into their line of work by external factors. In terms of borrowing practices, low/no growth clients were somewhat more likely to take loans for personal purposes such as school fees or building a house. High-growth clients, on the other hand, were more likely to report having specific business-related purposes for their borrowing. Furthermore, low/no growth clients were more likely to take a loan to recover from some negative external shock—for example, to replace inventory after a theft.

In terms of the challenges facing the two types of establishments, the low/no growth clients were much more likely to be struggling because of negative external shocks such as dishonest workers, theft, or demolitions. Female focus group members also reported frustration around having to, (1) work with male loan officers, (2) go through the street/village leader which meant disclosing financial information to a non-bank employee and, (3) complete onerous paperwork, often in English. As we show below, these "off-balance sheet" characteristics of owners are important predictors of loan growth.

In contrast, the main challenges identified by high-growth clients were related to competition from the informal sector, low sales turnover due to worsening economic conditions, and burdensome taxes and government regulations. The high-growth firms are more likely to have plans for future growth, and these plans tend to be more ambitious than those of the low/no growth clients. Finally, high-growth entrepreneurs appear to be better educated and have stronger managerial skills.

The survey, which we describe in the next section, reflects the various concerns raised during the focus group meetings.

4 Data and Methodology

In addition to using administrative loans data to classify borrowers into two groups and conducting focus groups, we conducted a detailed survey covering a sample of high and low/no growth firm owners in Dar es Salaam from December 2019 to February 2020. Out of the 464 firm owners in our potential sample for this study, we successfully interviewed 258 firm owners for an overall coverage rate of 56%.[9] Therefore, the sample for this study includes 137 high-growth firm owners (53.1% of the surveyed sample) and 121 low/no-growth firm owners (46.9% of the surveyed sample). The coverage rate is not differential across the samples with high-growth and low/no-growth firm owners (p-value 0.238). As each firm owner could report answers for multiple businesses in our survey, we have detailed data on 306 small- and medium-sized firms across the 258 firm owners.

To motivate our prediction exercise, Section 4.1 first discusses descriptive differences between high and low growth firms. We focus on variables that are classified as owners' characteristics and behavioral issues, challenges faced by the owner, firm characteristics, and financial variables.

4.1 Descriptive Statistics

Firm Owners

Table 1a, summarizes the characteristics of the firm owners in our sample. The average age for a business owner in our sample is 44.8 years. The majority of the

[9]See Appendix B for details on the sampling process.

firm-owners are educated and male: 53.9% of the firm owners have at least secondary education, and 32.9% of the entrepreneurs are female. The average monthly household income for the business owners is 2.1 million Tanzanian shillings (TSh), approximately US$ 860. Almost 96% of all firm owners own a house, and 50% own a car.

On average, entrepreneurs own 1.2 firms. Roughly half (52.7%) had operated any business in the past, prior to the current business(es). Entrepreneurs have 17 years of experience running a business on average. Using data on previous occupations, we find that 39.1% of the owners report being previously employed by a formal business, 8.5% by an informal business, and 7.4% by a family enterprise. Roughly one in eight (12.8%) report being employees of the government in the past.

To understand their motivation to be entrepreneurs, we asked firm owners why they started their businesses. The most common motivators include wanting to run a personal business or try a business idea, recognizing a good opportunity to exploit, wanting means to supplement their income, and believing they could make more money by working for themselves. Some entrepreneurs reported starting a business as they could not find a job elsewhere, or were motivated as their family was in the same line of business.

We also collected information on the number of loans the firm owners took in the past five years. We find that they took 3.8 loans on average in the past five years, with an average loan amount of 16.2 million TSh (roughly US$ 7000). They used 78.6% of these loans as working capital, 14.3% for investment, and 6.7% for personal expenses. It's worth noting that 95.7% of all firm owners in our sample use collateral for loans. The average value of available collateral is 94.1 million TSh (approximately US$ 40,500).

Firms

As each firm owner could report information for multiple businesses, we have data on 306 firms across the 258 firm owners in our sample. Table 1b includes the summary statistics for these 306 firms, out of which 53.9% are in the high-growth group, while the remaining 46.1% belong to the low/no-growth group. Only 33% of the firms are female-owned.

Our sample comprises small- and medium-sized firms in Dar es Salaam. The average number of workers employed by these firms is 3.3, and hired employees range between 0 and 68. Most firms (96.1%) run full-time; firms have been operating for 11.2 years on average. The most common firm type is retail, followed by wholesale and non-retail services. Roughly half (44.1%) of firms report using the internet for their business activities. Most firms use mobile phones for business, pay taxes, and have a license.

The average monthly revenue for firms is 14.8 million TSh (approx. US$ 6400), and the average monthly costs are 10.3 million TSh (approx. US$ 4450), yielding average monthly profits of 4.5 million TSh (approx. US$ 1950).

Firm owners were asked the monetary values of their business assets, such as undeveloped land, residential and non-residential buildings, plant and office equipment, and information communication technology (ICT) equipment. Across all assets, the average value of residential buildings is the highest at 34.6 million TSh (approx. US$ 15,000), followed non-residential buildings and undeveloped land valued at 10.6 million TSh (approx. US$ 4,600) and 8.5 million TSh (approx. US$ 3,700), respectively. The average value of a firm's business stocks is 29.4 million TSh (approx. US$ 12,750).

The average value of business premises owned by a firm owner is 105.1 million TSh (approx. US$ 45,200). Most firms are located in commercial areas and close to other formal businesses. Only a small fraction operates from the owner's home. The average number of clients or customers per day is 36 and ranges from 0 to 500. The average start-up cost is 9.9 million TSh (approx. US$ 4,300).

4.2 Methodology

4.2.1 Comparing high- and low-growth firm owners and firms

As discussed in Section 4.1, 53.1% of the firm-owners belong to the high-growth group and 46.9% to the low/no-growth group. Similarly, when looking at the firm-level data in Section 4.1, 53.9% of the firms are in the high-growth group, while the remaining 46.1% belong to the low/no-growth group. We want to understand how these two types of firm owners and firms differ and thus conduct t-tests to check if any differences between the two groups are statistically significant. Results appear in Table 2a for firm owner characteristics and in Table 2b for firm characteristics.

In Table 2a and Table 2b, column (1) includes the mean value for the variable in the left-most column for the high-growth group, while column (2) include the mean for the low/no-growth group. Column (3), reports the difference in means and a tests for equality of means.

High-growth firm owners are, on average almost seven years younger than low-growth owners. They are 11.7 percentage points less likely to be previously employed by the government and have fewer years of experience. High-growth owners are more likely to be in business due to their parents or relatives being in business. They have taken larger average loan amounts in the past 5 years and have a larger collateral value than low-growth owners.

On the firm side, as shown in Table 2b, we find that high-growth firms have been operating for fewer years than low/no-growth firms. They are more likely to pay taxes and to use the internet for business activities. They have a larger number of employees. In addition, high-growth firms have higher monthly revenue and costs than low-growth firms. Their start-up cost, value of business stocks, and value of residential buildings are considerably larger than that of low/no-growth firms. High-growht firms are less likely to operate from home and more likely to locate in commercial areas.

The two groups differ across a number of dimensions, as indicated by the t-tests. These statistically significant differences are suggestive, but concerns of false discovery must be kept in mind. This motivates the methodology for our paper, where we use machine learning methods to delve into the factors that predict growth of firms. We discuss our methodology in the next subsection.

4.2.2 Least Absolute Shrinkage and Selection Operator (LASSO) for Prediction

Our goal is to determine the predictors of growth for small- and medium-sized firms in a developing country context. Our survey of 258 firm owners based in Dar es Salaam has detailed data on the attributes of these entrepreneurs as well as their borrowing behavior, attitude, aspirations, and challenges they face as entrepreneurs. We collected information on around 180 questions in our survey.[10] We also have firm-level data for the firms owned by these business owners, containing characteristics such as sector, number of workers hired by the firms, their business costs and revenue, assets and stocks' value, and other variables.

To determine if some areas have a differential prevalence of high-growth firm owners/firms, we include neighborhood dummies representing the ward and district of the firm owner's residence as potential predictors of growth.[11]

We use the Least Absolute Shrinkage and Selection Operator (LASSO) regularization technique to deal with high-dimensional data, avoid overfitting and produce useful out-of-sample forecasts Tibshirani (1996). To avoid reseacher discretion, we include all the variables collected through our survey as potential predictors of high growth and use LASSO to select the variables that best predict our outcome of interest. A complete list of variables appears in Appendix D.

We minimize the following function, which adds a penalty term to the standard

[10]This results in many more than 180 variables that could potentially predict loan growth. As part of our survey, we asked firm owners to answer questions related to multiple businesses under their ownership. In addition to answering individual-level questions, each firm owner thus reported characteristics of multiple firms. Moreover, categorical variables are transformed into multiple dichotomous variables. As a result, the total number of variables is close to 1800.

[11]Dar es Salaam is sub-divided into five districts and has a total of 73 wards.

OLS regression function:

$$argmin_{\beta} \sum_{i=1}^{N} (Y_i - \beta X_i)^2 + \lambda(\sum_{k=1}^{K} |\beta_k|) \tag{3}$$

where N refers to the number of observations and k to the number of potential predictors. λ is the tuning parameter and is selected via cross-validation to minimize out-of-sample prediction error. That is, we estimate the LASSO logit model to obtain the covariates that best predict high-growth firm owners and firms.

We perform our analysis at two levels: both at the firm-owner-level and firm-level.[12] One advantage of separately examining owner-level characteristics is that firm-level characteristics may themselves be endogenously chosen in light of other constraints. For instance, Abeberese (2017) shows that Indian firms' sector choices are distorted by energy prices.

The dependent variable in our model is a high-growth indicator, and we use the complete set of variables from our survey to determine which are selected as predictors of growth. The variables and the estimated coefficients that are most relevant are retained while the others are removed from the model, i.e., thei coefficients are set to zero.[13]

After fitting the LASSO model, we we obtain a list of firm owner and firm level attributes that predict firm growth. As LASSO does not provide the standard errors required to perform statistical inference, we follow a post-selection inference procedure in line with recent literature (Meir & Drton, 2017; Lee et al., 2016; Zhao et al., 2017; Banerjee et al., 2021). Specifically, once we have the list of LASSO-selected predictors of growth, we estimate an OLS model using these covariates:

$$y = \mathbf{X}\beta + \epsilon \tag{4}$$

where y is a binary indicator of whether the firm owner or firm belongs to the high growth group, $\mathbf{X}$ is a vector of the selected covariates using LASSO, β is a vector of estimated coefficients on the selected covariates, and ϵ is the error term. The regression is estimated using both the firm owner-level and firm-level data. We cluster the standard errors at the firm owner-level when estimating the equation using firm-

[12]The firm owner-level dataset contains characteristics of the surveyed firm owners, while the firm-level data includes information on not only the firms but also the characteristics of the respective owners. Therefore, the firm-level data encapsulates all the available variables in our survey data.

[13]One of the limitations of LASSO is that it randomly selects one of two or more highly collinear variables. To corroborate our findings, we conduct a robustness exercise using the elastic net technique, which does not eliminate highly collinear variables. Reassuringly, the same set of predictors are selected when using both techniques (LASSO and elastic net) and for both the firm-owner and firm level data.

level data.

4.2.3 Using a Gender Lens

We use a gender lens and examine the predictors of loan growth separately for female vs. male entrepreneurs' firms. As reported in Table 1a, we can see that 32.9% of the firm owners in our sample are female.[14] Of these women, 54% belong to the high growth group and 46% to the low/no-growth group. Almost the same ratio holds for the 173 male firm-owners in our sample.[15] The firm-level data, summarized in Table 1b, shows that 33% (101/306 firms) of firms are women-owned, of which 54.5% belong to the high growth group.

We split the firm-owner sample into two sub-samples comprising male and female entrepreneurs and repeat the LASSO regression analysis detailed in section 4.2.2. We then identify the owner and firm characteristics that predict growth, separately by gender.

5 Results

In this section, we present several sets of results. As noted above, it is relevant to focus on firm-owner level characteristics due to the possibility that firm-level characteristics, such as sector or firm age, may themselves reflect the presence of constraints and as such could be through of as "bad controls." We therefore begin by examining the predictors of loan growth at the firm-owner level: for the sample as a whole and separately by gender.

We next broaden our consideration to include firm-level characteristics alongside firm-owner level characteristics; again we present these results for the sample as a whole and separately by gender.

For these analyses, we group variables into three broad categories: owner's characteristics and behavioral variables, challenges faced by the owner, and financial variables.

5.1 What should we expect?

The conceptual framework in Section 2 yields a strong prediction about what, under first-best capital allocation, should predict loan growth: only predictors/proxies of the firm's savings rate, the rate of growth of its labor force, its rate of productivity

[14]The gender split of borrowers in our sample mirrors the gender split in the NMB loans database for 2018 and 2019, where 35% of borrowers are female.

[15]Female firm-owners are not differentially distributed across the high and low/no growth groups, as demonstrated by *t*-tests.

growth (*not* its productivity level) and its rate of depreciation. We will call this set of predictors $\mathbf{GP^{FB}}$ ("growth predictors, first best". These will be classified as financial variables under our rubric. Thus, under the maintained assumption that our data contains good proxies of the $\mathbf{GP^{FB}}$ variables, a testable prediction of first-best allocation is that no owner's characteristics, behavioral variables or challenges faced by the owner should be predictive of loan growth.

To the extent that this prediction is rejected, this tells us that there are frictions/ constraints which distort the growth of external capital (i.e. loans) away from the first-best benchmark. This may be the case for the sample as a whole or for a subsample (i.e., men or women). If one subsample is father from the first-best benchmark in the sense that financial variables are less predictive, this suggests that this group faces greater frictions to efficient growth of external capital.

It is also possible that our maintained assumption is rejected, i.e., that the financial variables in our data set do not contain a perfect set of empirical proxies for $\mathbf{GP^{FB}}$. It may also be the case that the model of firm growth laid out in Section 2 is not a perfect description of how the firms in our Tanzanian setting grow.[16] In these cases, variables beyond $\mathbf{GP^{FB}}$, in particular non-financial variables, may have additional predictive power. Call the set of variables that predict growth in this case $\mathbf{GP^{RW}}$ ("growth predictors, real world"). Even in this case, however, the set of variables which are predictive should not systematically differ for men and women: our estimated proxies of $\mathbf{GP^{RW}}$ should not systematically vary depending on whether we use the male or female samples to select it. If we do find that a different set of variables is predictive for women (and female-owned businesses) vs. men (and male-owned businesses), this suggests that the two groups do not face the same set of constraints in regards to efficient capital allocation. We will revisit these ideas, in light of our results, in Section 5.4.

5.2 Owner-level Predictors of Firm Growth

We begin by considering predictors of loan growth at the level of the firm owner. Table 3 reports the firm owner-level characteristics which are predictive of loan growth. Column 1 includes the post-LASSO estimated coefficients for the predictors of growth for firm owners in the pooled (male and female) sample, and columns 2 and 3 show the post-LASSO estimated coefficients for the predictors of growth for female and male entrepreneurs, respectively. Blank spaces in the columns indicate that the corresponding variable in the left-most column was not selected by LASSO as a predictor of growth for the respective sample.

[16] For instance, firms may not always act in a way that maximises (constrained) profits (Banerjee et al., 2023).

We group the selected variables into three broad categories: owner's characteristics and behavioral variables, challenges faced by the owner, and financial variables. The sign included at the end of each selected variable is the sign on the coefficient estimated using LASSO. Recall that LASSO selects covariates and estimates coefficients but does not provide the standard errors required for performing statistical inference. Therefore, we follow a post-selection inference procedure and estimate a multivariate linear regression of loan growth – our outcome of interest – on the selected variables (Lee et al., 2016; Meir & Drton, 2017; Banerjee et al., 2021).

We report the post-LASSO estimated coefficients with robust standard errors in column (1) of Table 3. We find that age is negatively associated with the probability of having a high-growth firm. Moreover, having plans to expand business is associated with a 4.9 percentage point increase in the likelihood of having a high-growth firm, on average and all else equal. Owners who maintain records of interest payments are 7.4 percentage points more likely to have a high-growth firm. However, those who do not wish to be part of any business association are 4.6 percentage points less likely to have growing businesses. Similarly, owners who are government employees are less likely to have a growing business.

Owners who face particular business challenges are less likely to have a high-growth business. For instance, owners who deal with insufficient working capital, lack of working space, and theft by a friend are less likely to grow. Lack of management skills when starting a business is also a challenge that hinders growth. However, dealing with insufficient market access when setting up a business correlates with higher success later in the business.

Lastly, we find that a 1 SD increase in the value of the collateral is associated with a 6.7 percentage point increase in the probability of being a high-growth firm owner, keeping other regressors constant.[17]

5.2.1 Owner-level Predictors of Growth by Gender

We next run the owner-level analysis separately on the sub-samples of men and women.

The predictors of having high-growth firms for female and male entrepreneurs are in the left-most columns of Table 3. Blank spaces in Table 3 mean that the corresponding variable in the left-most column was not selected by LASSO as a predictor of growth. The non-blank entries in columns (2) and (3) are the estimated coefficients and standard errors for the selected variables. The number of predictors selected for

[17] We transformed all the potential covariates following the cleaning procedure outlined in Appendix C before running LASSO regression analysis. All the variables were normalized to have a unit standard deviation, and thus the coefficients are interpreted accordingly.

women is higher than for men, and there is little overlap in the chosen variables. We return to these points below, after reviewing the results.

Female Firm Owners A one standard deviation increase in the female entrepreneur's age is associated with an 19.7 percentage point decrease in the likelihood of owning a high-growth firm, on average and all else equal. Women who record purchases of materials are more likely to have a growing business, while those currently employed by the government are less likely to grow. In addition, to determine if some areas have a differential prevalence of high-growth firm owners living there, we construct neighborhood dummies representing the ward and district of the firm owner's residence and include them as potential predictors of growth.[18] We find that female firm owners residing in the district Temeke are less likely to own high-growth firms.

Predictors of growth also include challenges faced by the female entrepreneurs. For example, women who believe getting signatures from the local government authority is a challenge with the borrowing process are less likely to own a high-growth firm. Similarly, women who believe dealing with loan documentation in English is challenging are 7.1 percentage points less likely to own a high-growth firm, on average and all else equal. In addition, women who face the problem of insufficient working capital are 11.1 percentage points less likely to have a high-growth firm, on average. Other challenges that predict the probability of having a high-growth firm include submitting audit records of a company's performance to obtain a loan, experiencing theft, and having low demand for products and insufficient capital when starting a business.

Male Firm Owners Looking at the results included in column (3) of Table 3. we can see that age is negatively associated with growth. In addition, a male entrepreneur who is optimistic and views his business as growing has a 12.5 percentage point higher likelihood of having a high-growth firm. Moreover, owners who maintain record of interest payments are 8.7 percentage points more likely to own a high-growth firm than those who do not, on average and all else equal.

In addition, we find that a 1 SD increase in the value of the collateral is associated with a 7.1 percentage points increase in the probability of having a growing firm, on average and all else equal. Average household income is also positively correlated with growth.

Looking at Table 3, it is clear that the list of predictors for owning high-growth firms is shorter and different for men than for women entrepreneurs. Interestingly, the covariates representing business problems or external challenges selected in the female-only sample are not chosen for male firm owners. Instead, most selected predictors

[18] Dar es Salaam is divided into five districts and has a total of 73 wards.

fall under the owner's characteristics and behavioral variables category. Moreover, financial variables are only selected as a predictor for men and not women.

5.3 Firm Level Predictors of Growth Status

We repeat the same analysis as discussed in section 5.2 but now using both firm characteristics and firm-owner characteristics. We now group the selected variables into four broad categories: owner's characteristics and behavioral variables, challenges faced by the owner, firm's characteristics and behavior, and financial variables. Analogously to Table 3, the left-most columns of Table 4 contain the list of predictors of firm growth selected for the full sample, and columns 2 and 3 show the coefficients for female- and male-run firms.[19]

The list of selected predictors still includes owner-level characteristics, challenges, and financial variables (discussed in Section 5.2) as predictors of firm growth in the presence of firm-level characteristics.[20] For instance, owner's age, recording interest payments, whether the owner has plans to expand the business by opening another branch or shop or if they believe the business can grow in the future are still selected as predictors of firm growth status. Similarly, not being a business association member and getting around using bicycles and motorcycle taxis predict firm growth status even in the presence of firm-level data. We do find evidence on a few additional predictors of growth status. These include maintaining records of worker payments and viewing the business as growing, both of which are positively correlated with growth. Furthermore, we find that firms with owners who hire employees with relevant experience and whose motivation to be in business is family involvement in the same field are more likely to grow.

Similarly, challenges such as insufficient working capital, lack of management skills, insufficient market access, and the requirement to submit audit records of the company's performance to get a loan continue to predict firm growth. Moreover, as shown in column (1) of Table 4c, we also find that the value of collateral is still selected as a predictor of firm growth. Some additional predictors include whether the owner has taken loans only from NMB and whether they have borrowed from the Small Industries Development Organization (SIDO). Taking loans only from NMB is positively correlated with growth, while borrowing from SIDO is negatively associated with growth.

Moving on to the next category – firm characteristics – several variables are selected as predictors of growth. For example, in column (1) of Table 4b, meeting suppliers at

[19] Note that the list of predictors spans three subtables (Table 4a–Table 4c).

[20] Fifty-seven percent of the predictors selected with firm owner-level data are still selected when we add firm-level characteristics as potential predictors of firm growth.

the place of business is associated with a 8.7 percentage point increase in the probability of belonging to the high-growth group. Similarly, if the firm type is wholesale, it has a 5.2 percentage point higher likelihood of growing, on average. Moreover, certain wards have a differential prevalence of high-growth firms, hinting at the possibility of spatial misallocation (for instance, the ward Mbezi Juu has a lower prevalence of high-growth firms).

5.3.1 Firm Level Predictors of Growth Status by Gender

We next present disaggregated results for male- vs. female-owned firms. We split the full sample of 306 firms into two sub-samples consisting of 101 female-owned firms and 205 firms owned by men. We then run LASSO regression analysis on these two sub-samples to identify predictors of firm growth. The results are included in column (2) for female-owned firms and in column (3) for male-owned firms in Table 4a–Table 4c. We summarize the results below.

Female-owned Firms As in the pooled case, some owner-level characteristics continue be selected as predictors of growth when we combine firm characteristics with firm owner-level information.[21] In particular, women's age continues to be negatively associated with firm growth: a 1 SD increase in owner's age correlates with a 15 percentage point decrease in the likelihood of a firm growing. Moreover, we find that women who start a business due to struggles with other business types are less likely to experience growth. On the other hand, firms whose owners plan to expand or were employed in family enterprises are more likely to experience growth. Likewise, female-owned firms with owners who carefully evaluate employee performance based on problems they create in business operations are more likely to succeed.

Looking at column (2) in Table 4b, we can see that several predictors fall under the 'challenges faced by the owner' category. Almost all of these variables (except one) are also selected to predict growth when just using firm owner-level data in section 5.2.1. Therefore, similar to what we find in section 5.2.1, female-owned firms that face the problem of insufficient working capital are 6.5 percentage points less likely to grow on average. Furthermore, firms whose owners believe dealing with loan documentation in English is complex have a 7.2 percentage points lower likelihood of growing. Other challenges that continue to differentially predict growth include theft, duration to get a loan, submitting audit records to get a loan, insufficient capital when starting a business, and getting a signature from the local government authority to obtain a loan.

[21] Sixty-nine percent of the predictors selected with firm owner-level data (discussed in section 5.2.1) are still selected when we add firm-level characteristics as potential predictors of firm growth.

However, we also find that facing low demand when starting the business hinders firm growth.

Moving on to firm characteristics (column (2) in Table 4b), we find that firm growth is predicted by whether the firm meets the suppliers at the place of business, whether the firm sells products outside Dar es Salaam, number of unskilled workers, and ownership of business premises – all of which are positively associated with growth. On the other hand, if the firm's location is due to lower rent or land prices, the firm is 12.6 percentage points less likely to be high-growth. Likewise, if the firm operates from home, the business is 6.5 percentage points less likely to grow.

Male-owned Firms Next, we look at the predictors of growth for male-owned firms (column (3) of Table 4a, Table 4b, and Table 4c). Comparing with the results in Section 5.2.1, we find that all of the predictors are still selected when we add firm characteristics to firm owner-level information. Similarly, we find that LASSO selects fewer and different predictors of growth for male-owned firms as compared to female-owned firms. Thus, unlike the case for women, challenges faced by men and their firms play a lesser role in determining growth.

As we find in section 5.2.1, we can see that age is negatively correlated with growth (it is selected by the LASSO model, although statistically insignificant in the post-selection inference). The number of years since the business was started is also negatively correlated with growth: a 1 SD increase in years of experience is correlated with a 7.6 percentage point decrease in high growth probability. Firms with owners who are not part of business associations or are willing to sell their business are less likely to grow, while firms whose owners have informal personal networks are more likely to succeed. Also, as in the aggregated sample, maintaining record of interest payments and worker payments is positively correlated with growth. Male-owned firms whose owners view their business as growing are in fact more likely to have high growth.

The challenges which differentially predict growth in male-owned firms include insufficient market access when starting a business, lack of management skills, access or costs of finance when starting a business, and submitting business registration to obtain a loan. Turning to firm characteristics for male-owned firms, meeting suppliers at the place of business is associated with a 4.2 percentage point higher likelihood of growth. Similarly, firms with more employees or that transport products or services outside Dar es Salaam are more likely to experience growth.

Moving on to financial variables in column (3) of Table 4c, we can see that several variables are selected as predictors of growth (while only a few are selected for the female-owned sample in column (2) of the same table). Some financial variables that positively predict growth are the value of collateral, whether the owner uses profits to expand or start a new business, and average household income. Average monthly profit in good months is positively correlated with firm growth.

5.4 Assessing barriers to loan growth for men vs. women

Both at the level of firm owners and the level of firms, we find evidence that non-financial variables, and specifically variables related to challenges faced by firm owners, predict loan growth. Thus, the prediction of the framework in Section 2 is rejected. This finding is of *prima facie* relevance in terms of informing what variables are predictive of access to external capital. However, as noted in Section 5.1, this finding could either reflect constraints to optimal capital/firm growth, or could reflect the fact that firms are optimizing, but are not behaving in a way that is well-described by the simple model in Section 2.

We can gain additional insight by noting that these non-financial variables are more predictive of loan growth for women than for men. This finding rejects both the "first best" model of capital allocation (Section 2 as well as a model in which firms are unconstrained but not not narrowly optimize profits. In this case, a larger set of variables may be predictive, but this set would be parallel across male and female entrepreneurs. It instead appears that women face greater barriers than do men in efficiently accessing external capital markets, echoing other recent findings (Morazzoni & Sy, 2022; Chiplunkar & Goldberg, 2021).

In assessing this conclusion, we would like to know *how* much more predictive role do non-financial variables have for women vs. men. In order to assess this difference, we examine how different our conclusions about predicting firm growth would be if we did not consider "business challenge" variables as part of the potential set of predictors. We carry out this exercise in Table E.1 (for owner-level variables) and Table E.2 (for firm- and owner-level variables).

Table E.1 reveals that when business challenge variables are omitted from the owner-level set of predictors, the set of selected predictors and the associated R^2 fall markedly for women: the only predictor selected is the owners age, and the R^2 for the female sample falls from 0.541 with the full set of potential predictors to 0.165 with the limited set.

Comparing the same exercise for the male sample reveals that this is not a mechanical result. When business challenge variables are omitted in this case, the set of other predictors chosen increases and the R^2 also increases, from 0.351 to 0.427.[22]

Repeating this exercise for the set of both firm- and owner-level variables (table E.2 reveals a similar conclusion, with the R^2 for women-owned firms falling from 0.869 (full

[22]This may indicate that for men, the sole business challenge variable selected, insufficient market access, is highly collinear with other variables.

set) to 0.779 (limited set) while for male-owned firms the R^2 increases from 0.757 to 0.782.

Does this conclusion result solely from the fact that men might not face business challenges to the same extent as women? While this would be interesting in and of itself, it would change the interpretation of the finding that these variables were less predictive for men. We assess this possibility in Table F.1, which shows that, in fact, men report business challenges at *higher* rates than women for all responses other than "challenges with the borrowing process."[23] Thus the greater predictive value of these variables is not due to these challenges being rare for men.

In summary, for women the inclusion of non-financial variables leads to a quantitatively and qualitatively meaningful improvement in the ability to forecast loan growth. For men, if anything, the opposite is true. The fact that women's access to external capital is predicted by a markedly different set of variables suggest that both hypotheses discussed in Section 5.1 are rejected and that women face a meaningfully greater set of constraints.

5.5 Loan Growth Predicts Firm Profitability and Scale

We next validate how well our loan growth measure predicts firm profitability and scale. To do so, we use data on monthly profits, monthly sales, estimated stock value, and the number of employees recorded in the loan history for each firm owner in the administrative data from NMB. For 66% of the sample, the data records these variables more than once.[24] However, despite multiple entries, the values stay constant over time in 95.3% of the entries, perhaps due to the data being carried forward without changes. To deal with any changes in the values over time, we calculate the average of the monthly profits, monthly sales, estimated stock value, and the number of employees, and then test whether loan growth predicts these variables.

Therefore, to validate our loan growth measure, we first estimate equation (1) and obtain $\hat{y}$, the predicted likelihood of the firm owner/firm belonging to the high-growth group (based on the selected predictors of growth). Then, we regress average monthly profits (and monthly sales, stock value, and the number of employees) on the standardized predicted likelihood of the firm owner/firm belonging to the high-growth group. So, for example, if loan growth predicts profitability, we would expect a positive relationship between loan growth and average monthly profits.

[23]The male-female difference is only significant for 2 of the 14 variables (column 3).

[24]Two firm owners have no information on monthly profits, monthly sales, estimated stock value, and the number of employees in the NMB administrative data. Moreover, 85 firm owners have information on these variables at only one point in time, 161 have two data entries, and 10 have three data entries for each variable.

Results for the validation exercise are shown in Table 5a when predicting profitability, in Table 5b when predicting sales, in Table 5c when predicting the stock value, and lastly in Table 5d when testing whether loan growth is positively associated with firm size. We use firm owner-level data for columns 1 – 3, and for columns 4 – 6, we use firm-level data. We also include the gender split here, and columns 2 and 5 include the estimates for female entrepreneurs and female-owned firms. In contrast, columns 3 and 6 include the estimates for male entrepreneurs and male-owned firms, respectively.

We find that loan growth predicts profitability, sales, and stock value. In column 1 in Table 5a, a 1 SD increase in the predicted likelihood of the firm owner belonging to the high-growth group is associated with a 1.23 million Tanzanian Shillings (TSh) increase in monthly profits on average. Similarly, in column 1 in Table 5b, we find that a 1 SD increase in the predicted likelihood of the firm owner belonging to the high-growth group correlates with a 6.58 million TSh increase in monthly sales. Loan growth also correlates positively with stock value.

6 Discussion

Our results show that the first-best benchmark scenario in which all firms can equally access capital, and thus can equalize their marginal product and achieve economy-wide allocative efficiency, is rejected among our sample of Tanzanian firms, particularly among women. This in turn speaks to the presence of misallocation – even among the set of firms who have successfully borrowed from the formal banking sector. From a policy perspective, this indicates that the extensive margin of financial inclusion (i.e, becoming "banked") is not the only relevant margin: there is also an *intensive* margin, namely which firms are able to access and productively absorb larger amounts of capital. This intensive margin may be particularly important in contexts in which increasing the extensive margin of bank access has limited impacts (Dupas et al., 2014, 2018).

Our approach to identifying misallocation in cross-sectional data may serve as a useful complement to existing approaches, which, depending on modelling assumptions, may give widely varying estimates of the quantitative extent of across-firm misallocation (Hsieh & Klenow, 2009; Carrillo et al., 2022). In addition, by focusing on *equilibrium* capital allocations, our approach complements the strategy of using randomized or natural variation to shock either the supply side (as in, e.g., Bryan et al. 2021) or demand side (as in, e.g., Brailovskaya et al. 2021) of capital access.

From a methodological perspective, our results offer several contributions. They highlight the value of including "beyond the balance sheet" measures relating to the barriers, challenges, motives and opportunities that have shaped the paths of entrepreneurs and firms. In tandem, we show how machine learning (ML) methods can be used to deal with the challenges of high-dimensional data and concerns of overfitting. This complements other recent studies which also make use of ML methods to study MSMEs in different contexts (e.g., Hussam et al. 2022; Coad & Srhoj 2019; Bryan et al. 2021).

Several open questions are raised by our findings. One is whether (and which) interventions can reduce misallocation by increasing women's access to capital. Candidate interventions on the demand side might include mentoring, business/financial training, or helping entrepreneurs access new product or geographical markets. On the supply side, interventions to encourage loan officers to make impartial lending decisions, or to make female and male entrepreneurs' credit histories equally visible to lenders, may help to address unequal access to credit.

An open question is whether (and how) the types of nontraditional predictors identified by studies such as ours (as well as the psychometric and cognitive data studied by Bryan et al. 2021) can be directly used in lending decisions without leading to issues of strategic responses, gaming, etc. Insights from the literature on incentive-compatible mechanism design (Hussam et al., 2022; Björkegren et al., 2020) may come into play. Another open question is whether, had we collected more and/or different types of survey data from the entrepreneurs, we might have identified even more predictors of loan growth. These avenues remain as promising areas of future research.

References

Abeberese, A. B. (2017). Electricity cost and firm performance: Evidence from India. *Review of Economics and Statistics*, *99*(5), 839–852.

Alibhai, S., Cassidy, R., Goldstein, M., & Papineni, S. (2022). Evening the credit score? impact of psychometric credit scoring on women-owned firms' financial access and performance in Ethiopia.

Arráiz, I., Bruhn, M., & Stucchi, R. (2017). Psychometrics as a tool to improve credit information. *The World Bank Economic Review*, *30*(Supplement_1), S67–S76.

Banerjee, A., Breza, E., Duflo, E., & Kinnan, C. (2020). Can microfinance unlock a poverty trap for some entrepreneurs? *National Bureau of Economic Research WP 26346*.

Banerjee, A., Breza, E., Townsend, R., & Vera-Cossio, D. (2019). *Access to credit and productivity: Evidence from thai villages* (Tech. Rep.). Working Paper.

Banerjee, A., Chandrasekhar, A. G., Dalpath, S., Duflo, E., Floretta, J., Jackson, M. O., ... Shrestha, M. (2021). Selecting the most effective nudge: Evidence from a large-scale experiment on immunization. *NBER WP 28726*.

Banerjee, A., & Duflo, E. (2014). Do firms want to borrow more? testing credit constraints using a directed lending program. *Review of Economic Studies*, *81*(2), 572–607.

Banerjee, A., Fischer, G., Karlan, D., Lowe, M., & Roth, B. N. (2023). *Do microenterprises maximize profits? a vegetable market experiment in india* (Tech. Rep.).

Banerjee, A., Karlan, D., & Zinman, J. (2015). Six randomized evaluations of microcredit: Introduction and further steps. *American Economic Journal: Applied Economics*, *7*(1), 1-21. Retrieved from `http://www.aeaweb.org/articles.php?doi=10.1257/app.20140287` doi: 10.1257/app.20140287

Banerjee, A., & Newman, A. (1993). Occupational choice and the process of development. *Journal of political economy*, *101*(2), 274–298.

Bau, N., & Matray, A. (2022). Misallocation and capital market integration: Evidence from India. *Econometrica*.

Beaman, L., Karlan, D., Thuysbaert, B., & Udry, C. (2020). *Selection into credit markets: Evidence from agriculture in Mali* (Tech. Rep.).

Beck, T., Demirgüç-Kunt, A., & Maksimovic, V. (2008). Financing patterns around the world: Are small firms different? *Journal of financial economics*, *89*(3), 467–487.

Beck, T., Döttling, R., Lambert, T., & van Dijk, M. (2022). Liquidity creation, investment, and growth. *Journal of Economic Growth*, 1–40.

Bertrand, M., Karlan, D., Mullainathan, S., Shafir, E., & Zinman, J. (2010). What's advertising content worth? evidence from a consumer credit marketing field experiment. *The quarterly journal of economics*, *125*(1), 263–306.

Björkegren, D., Blumenstock, J. E., & Knight, S. (2020). Manipulation-proof machine learning. *arXiv preprint arXiv:2004.03865*.

Brailovskaya, V., Dupas, P., & Robinson, J. (2021). *Is digital credit filling a hole or digging a hole? Evidence from Malawi* (Tech. Rep.). National Bureau of Economic Research.

Brock, J. M., & De Haas, R. (2022). Discriminatory lending: Evidence from bankers in the lab.

Bryan, G., Karlan, D., & Osman, A. (2021). Big loans to small businesses: Predicting winners and losers in an entrepreneurial lending experiment.

Carrillo, P., Donaldson, D., Pomeranz, D., & Singhal, M. (2022). Allocative efficiency in firm production: A nonparametric test using procurement lotteries. *Working Paper*.

Chiplunkar, G., & Goldberg, P. K. (2021). *Aggregate implications of barriers to female entrepreneurship* (Tech. Rep.). National Bureau of Economic Research.

Coad, A., & Srhoj, S. (2019). Catching gazelles with a lasso: Big data techniques for the prediction of high-growth firms. *Small Business Economics*, 1–25.

Cull, R., & Spreng, C. P. (2011). Pursuing efficiency while maintaining outreach: Bank privatization in Tanzania. *Journal of Development Economics*, *94*(2), 254–261.

De Mel, S., McKenzie, D., & Woodruff, C. (2008). Returns to capital in microenterprises: evidence from a field experiment. *The quarterly journal of Economics*, *123*(4), 1329–1372.

Dupas, P., Green, S., Keats, A., & Robinson, J. (2014). Challenges in banking the rural poor: Evidence from Kenya's western province. In *African successes, volume iii: Modernization and development* (pp. 63–101). University of Chicago Press.

Dupas, P., Karlan, D., Robinson, J., & Ubfal, D. (2018). Banking the unbanked? evidence from three countries. *American Economic Journal: Applied Economics*, *10*(2), 257–97.

Evans, D. S., & Jovanovic, B. (1989). An estimated model of entrepreneurial choice under liquidity constraints. *The Journal of Political Economy*, *97*(4), 808–827.

Hardy, M., & Kagy, G. (2018). Mind the (profit) gap: why are female enterprise owners earning less than men? *AEA Papers and Proceedings*, *108*, 252–55.

Hardy, M., & Kagy, G. (2020). It's getting crowded in here: experimental evidence of demand constraints in the gender profit gap. *The Economic Journal*, *130*(631), 2272–2290.

Hsieh, C.-T., & Klenow, P. J. (2009). Misallocation and manufacturing tfp in china and india. *The Quarterly journal of economics*, *124*(4), 1403–1448.

Hussam, R., Rigol, N., & Roth, B. N. (2022). Targeting high ability entrepreneurs using community information: Mechanism design in the field. *American Economic Review*, *112*(3), 861–98.

Karaivanov, A., & Yindok, T. (2017). *Involuntary entrepreneurship: Evidence from Thai urban data.* (SFU Working Paper)

Lee, J. D., Sun, D. L., Sun, Y., & Taylor, J. E. (2016). Exact post-selection inference, with application to the lasso. *The Annals of Statistics*, *44*(3), 907–927.

Lucas, R. E. (1976). *The phillips curve and labor markets, v. 1 of carnegie–rochester conference series on public policy.* Amsterdam: North-Holland.

Mankiw, N. G., Romer, D., & Weil, D. N. (1992). A contribution to the empirics of economic growth. *The quarterly journal of economics*, *107*(2), 407–437.

McKenzie, D., & Sansone, D. (2019). Predicting entrepreneurial success is hard: Evidence from a business plan competition in nigeria. *Journal of Development Economics*, *141*, 102369.

Meager, R. (2019). Understanding the impact of microcredit expansions: A bayesian hierarchical analysis of 7 randomised experiments. *American Economic Journal: Applied Economics*, *11*.

Meir, A., & Drton, M. (2017). Tractable post-selection maximum likelihood inference for the lasso. *arXiv preprint arXiv:1705.09417*.

Morazzoni, M., & Sy, A. (2022). Female entrepreneurship, financial frictions and capital misallocation in the us. *Journal of Monetary Economics*, *129*, 93–118.

Quah, D. T. (1996). Empirics for economic growth and convergence. *European economic review*, *40*(6), 1353–1375.

Rigol, N., & Roth, B. N. (2021). *Loan officers impede graduation from microfinance: Strategic disclosure in a large microfinance institution* (Tech. Rep.). National Bureau of Economic Research.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288.

Zhao, S., Witten, D., & Shojaie, A. (2017). In defense of the indefensible: A very naive approach to high-dimensional inference. *arXiv preprint arXiv:1705.05543*.

Table 1a: Descriptive Statistics for Firm Owners

	(1) Mean	(2) SD	(3) Min	(4) Max	(5) N
Demographics					
Age (in years)	44.810	10.019	4	99	258
Education (at least secondary)	0.539	0.499	0	1	258
Female	0.329	0.471	0	1	258
Average household monthly income (in million TZS)	2.077	5.030	0	60	252
Number of current businesses	1.167	0.449	0	3	258
Firm-owner belongs to the high-growth group	0.531	0.500	0	1	258
Business owner's other occupations					
Owner previously employed by a formal business	0.391	0.489	0	1	258
Owner previously employed by an informal business	0.085	0.280	0	1	258
Owner previously employed in a family enterprise	0.074	0.262	0	1	258
Owner previously employed by the government	0.128	0.335	0	1	258
Motivation to be an entrepreneur					
Lost a previous job	0.058	0.234	0	1	258
Could not find job elsewhere	0.136	0.343	0	1	258
To support himself/herself and family	0.202	0.402	0	1	258
To try out a business idea	0.209	0.408	0	1	258
Believe can make more money working for own self	0.244	0.430	0	1	258
Had nothing else to do/no other means of survival/no better option	0.109	0.312	0	1	258
Parents/relatives were in business	0.136	0.343	0	1	258
Saw business to be a good opportunity	0.221	0.416	0	1	258
Always wanted to have own business	0.298	0.458	0	1	258
Encouraged by friends and relatives	0.074	0.262	0	1	258
To supplement income	0.275	0.447	0	1	258
Flexible hours	0.062	0.242	0	1	258
Home and car ownership					
Owner owns a house	0.957	0.202	0	1	258
Owner owns a car	0.500	0.501	0	1	258
Past businesses					
Owner has operated any past businesses	0.527	0.500	0	1	258
Number of years since first business (in years)	17.037	8.803	3	63	136
Loans, their source, and use					
Number of loans in past 5 years	3.771	1.314	0	5	258
Average loan amount in past 5 years (in million TZS)	16.200	11.844	0	48	252
Percentage of loans used for investment	14.326	25.924	0	100	258
Percentage of loans used for working capital	78.620	30.395	0	100	258
Percentage of loans used for personal expenses	6.667	15.641	0	100	258
Collateral for loans					
Value of available collateral (in million TZS)	94.130	107.596	0	900	250
Owner uses collateral for loans	0.957	0.202	0	1	258

Note: Column (1) reports the overall sample mean; column (2) reports the standard deviation (SD); columns (3) and (4) report the minimum and maximum respectively. Number of observations in column (5) less than 258 are due to missing values in the data. Some firm owners refused to answer certain questions or did not know the answers.

Table 1b: Descriptive Statistics for Firms

	(1) Mean	(2) SD	(3) Min	(4) Max	(5) N
Firm characteristics					
Firm age (in years)	11.167	7.184	0	40	306
Business runs full time	0.961	0.194	0	1	306
Firm type: manufacturing	0.020	0.139	0	1	306
Firm type: construction	0.036	0.186	0	1	306
Firm type: retail	0.529	0.500	0	1	306
Firm type: services	0.157	0.364	0	1	306
Firm type: wholesale	0.199	0.400	0	1	306
Firm type: agriculture	0.023	0.150	0	1	306
Firm type: agricultural processing	0.013	0.114	0	1	306
Firm type: bar	0.039	0.194	0	1	306
Firm type: lodge	0.010	0.099	0	1	306
Firm has a license	0.918	0.274	0	1	306
Firm pays taxes	0.863	0.345	0	1	306
Owner uses a mobile phone for business	0.944	0.229	0	1	306
Owner uses internet for business	0.441	0.497	0	1	306
Business classified as high-growth	0.539	0.499	0	1	306
Business owned by a female	0.330	0.471	0	1	306
Employment					
Total number of workers employed by the business currently	3.268	4.924	0	68	306
Workers are provided training	0.239	0.427	0	1	306
Business costs, revenue, and profits (in million TZS)					
Firm's average monthly revenue	14.806	31.731	0	400	306
Firm's average monthly cost	10.312	25.684	0	350	306
Firm's average monthly profits	4.494	8.836	0	62	306
Business assets and stocks (in million TZS)					
Asset value: Undeveloped land	8.479	43.782	0	600	301
Asset value: Residential buildings	34.557	74.576	0	700	302
Asset value: Non-residential buildings	10.558	73.135	0	900	300
Asset value: Machinery, plant, and other equipment	4.259	23.693	0	300	301
Asset value: Electricity generation equipment	0.346	2.017	0	25	304
Asset value: Transport equipment	4.842	25.570	0	400	303
Asset value: Furniture, fixtures, & office equipment	2.298	12.783	0	200	302
Asset value: ICT equipment	1.190	14.580	0	250	304
Firm's value of business stocks	29.410	52.060	0	450	292
Customers					
Firm's average number of clients per day	36.498	51.172	0	500	289
Firm location					
Business operates out of home	0.141	0.348	0	1	306
Business operates in a commercial area	0.784	0.412	0	1	306
Business is located near other similar formal businesses	0.725	0.447	0	1	306
Business is located near other different formal businesses	0.735	0.442	0	1	306
Business is located near similar informal businesses	0.232	0.423	0	1	306
Business is located near different informal businesses	0.271	0.445	0	1	306
Value of business premises (in million TZS)	105.138	156.406	0	900	66
Firm history					
Firm Start-up cost (in million TZS)	9.923	27.466	0	300	301

Note: Column (1) reports the overall sample mean; column (2) reports the standard deviation (SD); columns (3) and (4) report the minimum and maximum respectively. Number of observations in column (5) less than 306 are due to missing values in the data. Some firm owners refused to answer certain questions or did not know the answers.

Table 2a: Differences between High- and Low-Growth Firm Owners

	(1) Mean for high-growth firm-owners	(2) Mean for low-growth firm-owners	(3) Difference in means
Demographics			
Age (in years)	41.518	48.537	-7.019***
Education (at least secondary)	0.547	0.529	0.019
Female	0.336	0.322	0.013
Average household monthly income (in million TZS)	2.494	1.603	0.891
Number of current businesses	1.182	1.149	0.034
Business owner's other occupations			
Owner has other occupation besides running the business(es)	0.147	0.192	-0.045
Owner previously employed by a formal business	0.387	0.397	-0.010
Owner previously employed by an informal business	0.102	0.066	0.036
Owner previously employed in a family enterprise	0.095	0.050	0.045
Owner previously employed by the government	0.073	0.190	-0.117***
Motivation to be an entrepreneur			
Lost a previous job	0.051	0.066	-0.015
Could not find job elsewhere	0.161	0.107	0.053
To support himself/herself and family	0.175	0.231	-0.056
To try out a business idea	0.212	0.207	0.005
Believe can make more money working for own self	0.270	0.215	0.055
Had nothing else to do/no other means of survival/no better option	0.109	0.107	0.002
Parents/relatives were in business	0.175	0.091	0.084**
Saw business to be a good opportunity	0.197	0.248	-0.051
Always wanted to have own business	0.328	0.264	0.064
Encouraged by friends and relatives	0.051	0.099	-0.048
To supplement income	0.285	0.264	0.020
Flexible hours	0.073	0.050	0.023
Home and car ownership			
Owner owns a house	0.964	0.950	0.013
Owner owns a car	0.540	0.455	0.086
Past businesses			
Owner has operated any past businesses	0.533	0.521	0.012
Number of years since first business (in years)	15.384	18.952	-3.569**
Loans, their source, and use			
Number of loans in past 5 years (i.e., since 2014)	3.657	3.901	-0.244
Average loan amount in past 5 years (in million TZS)	21.787	9.856	11.931***
Percentage of loans used for investment	14.788	13.802	0.987
Percentage of loans used for working capital	78.569	78.678	-0.108
Percentage of loans used for personal expenses	5.912	7.521	-1.608
Collateral for loans			
Value of available collateral (in million TZS)	108.848	77.665	31.183**
Owner uses collateral for loans	0.942	0.975	-0.034

$*p < .10, **p < .05, ***p < .01$. Note: Column (1) includes the mean values for the variables in the left-most column for the high-growth group, while column (2) includes the means for the low/no-growth group. In column (3), we test if the differences in means are statistically significant.

Table 2b: Differences between High- and Low-Growth Firms

	(1) Mean for high-growth firms	(2) Mean for low-growth firms	(3) Difference in means
Firm characteristics			
Firm age (in years)	10.127	12.383	-2.256***
Business runs full time	0.952	0.972	-0.020
Firm type: manufacturing	0.030	0.007	0.023
Firm type: construction	0.036	0.035	0.001
Firm type: retail	0.497	0.567	-0.070
Firm type: services	0.170	0.142	0.028
Firm type: wholesale	0.267	0.121	0.146***
Firm type: agriculture	0.036	0.007	0.029*
Firm type: agricultural processing	0.018	0.007	0.011
Firm type: bar	0.024	0.057	-0.032
Firm type: lodge	0.012	0.007	0.005
Firm has a license	0.939	0.894	0.046
Firm pays taxes	0.903	0.816	0.087**
Owner uses a mobile phone for business	0.958	0.929	0.028
Owner uses internet for business	0.509	0.362	0.147***
Employment			
Total number of workers employed by the business currently	4.115	2.277	1.839***
Workers are provided training	0.267	0.206	0.061
Business costs, revenue, and profits			
Firm's average monthly revenue (in million TZS)	19.111	11.029	8.082**
Firm's average monthly cost (in million TZS)	13.684	6.851	6.833**
Firm's average monthly profits (in million TZS)	5.040	4.020	1.019
Business assets and stocks			
Asset value: Undeveloped land (in million TZS)	5.554	11.843	-6.289
Asset value: Residential buildings (in million TZS)	43.986	23.647	20.340**
Asset value: Non-residential buildings (in million TZS)	6.597	15.084	-8.487
Asset value: Machinery, plant, and other equipment (in million TZS)	3.496	5.160	-1.665
Asset value: Electricity generation equipment (in million TZS)	0.256	0.450	-0.194
Asset value: Transport equipment (in million TZS)	6.313	3.151	3.162
Asset value: Furniture, fixtures, & office equipment (in million TZS)	1.828	2.841	-1.013
Asset value: ICT equipment (in million TZS)	0.290	2.229	-1.939
Firm's value of business stocks (in million TZS)	39.365	18.147	21.217***
Customers			
Firm's average number of clients per day	36.449	36.556	-0.108
Firm location			
Business operates out of home	0.097	0.191	-0.095**
Business operates in a commerical area	0.830	0.730	0.100**
Business is located near other similar formal businesses	0.709	0.745	-0.036
Business is located near other formal businesses in different activities	0.752	0.716	0.035
Business is located near informal businesses in the same activity	0.248	0.213	0.036
Business is located near informal businesses in other activities	0.261	0.284	-0.023
Value of business premises (in million TZS)	111.089	99.867	11.221
Firm history			
Firm Start-up cost (in million TZS)	13.662	5.566	8.095**

$*p < .10, **p < .05, ***p < .01$. Note: Column (1) includes the mean values for the variables in the left-most column for the high-growth group, while column (2) includes the means for the low/no-growth group. In column (3), we test if the differences in means are statistically significant.

Table 3: Predictors of High Growth Using Firm-Owner Level Characteristics

	(1) High Growth owners	(2) High Growth (female owners)	(3) High Growth (male owners)
Owner's Characteristics & Behavioral variables			
Owner's age (-)	-0.407***	-0.197***	-0.149
	(0.129)	(0.059)	(0.258)
Owner's age squared (-)	0.129		-0.080
	(0.097)		(0.208)
Owner views business as growing (+)	0.049		0.125***
	(0.033)		(0.033)
Owner has plans to expand business by opening another branch or shop (+)	0.049*		
	(0.027)		
Owner believes business can grow in future (+)	0.039		
	(0.034)		
Owner records interest payments (+)	0.074***		0.087**
	(0.026)		(0.033)
Owner records purchases of materials (+)	0.036	0.101*	
	(0.028)	(0.055)	
Owner does not want to be part of any business association (-)	-0.046*		
	(0.025)		
Owner not part of business associations as sees them as unbeneficial (+)	0.017		0.058***
	(0.014)		(0.012)
Owner owned a car before starting business (-)		-0.049	
		(0.035)	
Owner gets around using bicycles and motorcycle taxis (-)	-0.065**		-0.080***
	(0.026)		(0.028)
Owner currently employed by the government (-)	-0.049**	-0.065**	
	(0.019)	(0.030)	
Owner lives in district Temeke (-)		-0.093*	
		(0.048)	
Owner lives in ward Bunju (+)	0.074***		
	(0.027)		
Challenges Faced by the Owner			
Business problem faced by the owner: insufficient working capital (-)	-0.070**	-0.111**	
	(0.027)	(0.054)	
Submit audit records of company's performance to obtain a loan (+)	0.025	0.029	
	(0.031)	(0.027)	
Business problem faced by the owner: lack of working space (-)	-0.070***		
	(0.023)		
Owner's past businesses closed due to lack of managements skills (+)	0.046**		
	(0.023)		
Owner has experienced a traffic accident (+)	0.049***		
	(0.017)		
Get signature from the local government authority to obtain a loan (-)		-0.055	
		(0.042)	
Time to get a loan is a challenge with the borrowing process (+)		0.109**	
		(0.043)	
All documentation is in English is a challenge with the borrowing process (-)		-0.071*	
		(0.039)	
Owner has experienced theft by a friend (-)	-0.067**		
	(0.026)		
Owner has experienced theft by a stranger (+)		0.110*	
		(0.056)	
Challenge faced when started business: lack of mgmt skills (-)	-0.068***		
	(0.019)		
Owner faced low demand for products when first started business (-)		-0.066	
		(0.042)	
Challenge faced when started business: insufficient capital (-)		-0.051	
		(0.050)	
Challenge faced when started business: insufficient market access (+)	0.060**		0.070**
	(0.025)		(0.028)
Financial variables			
Value of collateral (TZS) (+)	0.067***		0.071*
	(0.025)		(0.040)
Owner uses loan to refurbish house (-)	-0.039		
	(0.026)		
Owner uses profits to expand business (+)			0.035
			(0.035)
Average household annual income (+)			0.056**
			(0.028)
N	258	85	173
R^2	0.424	0.541	0.351

Note: Robust standard errors are included in parentheses. $*p < .10, **p < .05, ***p < .01$. The left-most column contains the predictors selected using by the lasso. Column 1 includes the OLS estimates for the variables selected using lasso on the full sample of 258 firm owners. Column 2 and 3 include OLS estimates for the variables selected using lasso on the female and male only firm owner sample respectively. Missing entries in each column mean the corresponding variable in the left-most column was not selected for the respective sample. The sign included in () at the end of each selected variable in the left-most column is the sign on the coefficient estimated using lasso.

Table 4a: Predictors of High Growth Using Firm- and Owner-Level Characteristics

	(1) High Growth Firms	(2) High Growth (female-owned firms)	(3) High Growth (male-owned firms)
Owner's Characteristics & Behavioral variables			
Owner's age (-)	-0.155***	-0.150***	-0.165
	(0.024)	(0.036)	(0.113)
Owner's age - squared (-)			0.041
			(0.112)
Owner believes business can grow in the future (+)	0.084***		
	(0.024)		
Owner views business as growing (+)			0.082***
			(0.022)
Owner has plans to expand business by opening another branch or shop (+)	0.050**	0.062*	
	(0.022)	(0.037)	
Owner records interest payments (+)	0.091***		0.079***
	(0.029)		(0.024)
Owner records worker payments (+)	0.010		0.040
	(0.029)		(0.030)
Owner ever employed in a family enterprise (+)		0.068*	
		(0.037)	
Number of years since started first business (-)			-0.076***
			(0.026)
Owner does not want to be part of any business association (-)	-0.054**		-0.068***
	(0.023)		(0.024)
Owner gets around using bicycles and motorcycle taxis (-)	-0.051*		-0.042
	(0.026)		(0.025)
Owner gets around by walking (+)			0.044***
			(0.009)
Owner willing to sell company if made partner/owner (-)	-0.033		-0.080**
	(0.029)		(0.034)
Motivation to join network: learn from other entrepreneurs (-)	-0.069**		
	(0.032)		
Owner evaluates employee performance by the problems they create (+)		0.078***	
		(0.026)	
Owner hires employees with relevant experience (+)	0.040*		0.029
	(0.023)		(0.025)
Owner started multiple businesses as her other business were struggling (-)		-0.059***	
		(0.017)	
Owner chose business type as friends/relatives were in same line of business (-)			-0.050**
			(0.023)
Owner motivated to be an entrepreneur as parents/relatives were in business (+)	0.062**		0.072***
	(0.026)		(0.026)
Owner not part of business associations as sees them as unbeneficial (+)			0.037*
			(0.019)
Owner has informal personal networks for the business (+)			0.042*
			(0.022)
Discuss how to improve product quality/service at business association meetings (-)			-0.114***
			(0.025)
Discuss marketing strategies at business association meetings (-)			0.012
			(0.025)
Owner judge business environment by the prices she can charge (+)			0.035
			(0.027)
Owner reduces prices in face of competition (-)			-0.051**
			(0.020)
Owner lives in ward Bunju (+)	0.060**		0.060**
	(0.025)		(0.029)
Owner lives in ward Kibada (-)		0.008	
		(0.032)	
Owner lives in ward Kibamba (+)	0.045***	0.000	
	(0.017)	(0.021)	
Owner lives in ward Makongo (-)			-0.044***
			(0.014)
Owner lives in ward Charambe (-)		-0.040**	
		(0.016)	
Owner lives in district Kigamboni (-)		-0.097***	
		(0.033)	
Owner lives in district Temeke (-)		-0.122***	
		(0.028)	

Note: Standard errors clustered at the firm-owner level are included in parentheses. $*p < .10, **p < .05, ***p < .01$. The left-most column contains the predictors selected using by the lasso. Column 1 includes the OLS estimates for the variables selected using lasso on the full sample of 306 firms. Column 2 and 3 include OLS estimates for the variables selected using lasso on the female-owned and male-owned firms respectively. Missing entries in each column mean the corresponding variable in the left-most column was not selected for the respective sample. The sign included in () at the end of each selected variable in the left-most column is the sign on the coefficient estimated using lasso.

Table 4b: Predictors of High Growth Using Firm- and Owner-Level Characteristics (cont.)

	(1) High Growth Firms	(2) High Growth (female-owned firms)	(3) High Growth (male-owned firms)
Challenges faced by the Owner			
Challenge faced by the owner when started business: insufficient market access (+)	0.054**		0.055**
	(0.024)		(0.022)
Business problem faced by the owner: insufficient working capital (-)	-0.043*	-0.065*	
	(0.024)	(0.034)	
Owner experienced significant theft by a stranger (+)	0.063***	0.101***	
	(0.024)	(0.038)	
Submit audit records to company's performance to obtain a loan (+)	0.045*	0.002	
	(0.026)	(0.026)	
Get signature from the local government to obtain a loan (-)		-0.048	
		(0.038)	
Challenge faced by the owner when started business: lack of mgmt. skills (-)	-0.036*		-0.017
	(0.020)		(0.023)
Time to get a loan is a challenge with the borrowing process (+)		0.095***	
		(0.032)	
All documentation being in English is a challenge with the borrowing process (-)		-0.072**	
		(0.031)	
Challenge faced by the owner when started business: insufficient capital (-)		-0.041	
		(0.036)	
Challenge faced when started business: shortage of inputs (+)		0.098***	
		(0.028)	
Challenge faced when started business: lack of access to utilities (-)		-0.001	
		(0.016)	
Challenge faced by the owner when first started business: access or costs of finance (-)			-0.008
			(0.025)
Submit business registration to obtain a loan (-)			-0.093***
			(0.030)
Past businesses closed due to insufficient capital (-)			-0.025
			(0.026)
Owner has experienced demolitions from construction (+)		0.033	
		(0.026)	
Firm's Characteristics			
Owner meets suppliers at the place of business (+)	0.087***	0.028	0.042**
	(0.022)	(0.027)	(0.021)
Firm does not sell products/services outside of Dar es Salaam (-)	-0.031	0.009	
	(0.039)	(0.030)	
Total number of current employees (+)	0.026		0.025*
	(0.019)		(0.014)
Number of paid unskilled workers and laborers employed by the firm when started (+)	0.053**	0.053**	
	(0.022)	(0.026)	
Number of paid unskilled workers and laborers employed by the firm currently (+)	0.003	0.041	
	(0.023)	(0.051)	
Firm type: wholesale (+)	0.052**		0.047*
	(0.022)		(0.026)
Firm owner transports products outside of Dar es Salaam for selling (+)	0.038		0.073***
	(0.040)		(0.025)
Firm's customers are small traders in the area/town (+)	-0.002		0.032
	(0.024)		(0.025)
Firm's customers are large enterprises (+)	0.017		0.025
	(0.026)		(0.021)
Firm's customers are largers traders in the area/town (-)		-0.052**	
		(0.025)	
Owner travels outside Dar es Salaam to purchase firm inputs (+)		0.028	
		(0.026)	
Business premises owned by the government (-)	-0.012		0.003
	(0.021)		(0.023)
Business premises owned by a landlord (+)		0.019	
		(0.031)	
Firm's current location is due to cheap rent/land prices (-)	-0.043*	-0.126***	
	(0.023)	(0.033)	
Firm's current location is at home (-)		-0.065**	
		(0.025)	

Note: Standard errors clustered at the firm-owner level are included in parentheses. $*p < .10, **p < .05, ***p < .01$. The left-most column contains the predictors selected using by the lasso. Column 1 includes the OLS estimates for the variables selected using lasso on the full sample of 306 firms. Column 2 and 3 include OLS estimates for the variables selected using lasso on the female-owned and male-owned firms respectively. Missing entries in each column mean the corresponding variable in the left-most column was not selected for the respective sample. The sign included in () at the end of each selected variable in the left-most column is the sign on the coefficient estimated using lasso.

Table 4c: Predictors of High Growth Using Firm- and Owner-Level Characteristics (cont.)

	(1) High Growth Firms	(2) High Growth (female-owned firms)	(3) High Growth (male-owned firms)
Firm's profit increased due to lower input costs (-)			-0.088*** (0.015)
Firm's profit declined due to higher competition (-)		-0.052* (0.029)	
Firm's profit increased due to lower competition (-)			-0.032* (0.017)
Owner improved business premises in the past 3 years (+)		0.030 (0.033)	
Firm location: ward Kawe (-)		-0.060*** (0.020)	
Firm location: ward Kijichi (+)			0.033** (0.017)
Firm location: ward Mbezi Juu (-)	-0.060*** (0.017)		-0.042* (0.023)
Financial Variables			
Owner uses collateral for loans (-)			-0.031 (0.024)
Value of collateral (TZS) (+)	0.018 (0.023)		0.058* (0.031)
Loan taken from NMB in past five years (+)	0.031 (0.026)		0.036 (0.024)
Ever borrowed from SIDO (-)	-0.044** (0.020)		
Owner uses profits to invest in land (+)		0.037 (0.034)	
Owner uses profits to expand business (+)			0.013 (0.027)
Owner uses profits to start a new business (+)			0.042** (0.018)
Average household annual income in TZS (+)			0.052** (0.021)
Average monthly profit during good months (+)			0.074*** (0.017)
Owner has other income source from family contributions/remittances (-)			-0.069*** (0.019)
Monthly expenses for insurance (dummy for missing values) (+)	0.027 (0.017)		
Monthly repair expenses (dummy for missing values) (+)	0.070*** (0.018)		0.041*** (0.012)
Monthly expenses for fertilizer and other farm inputs (dummy for missing values) (+)	-0.015 (0.023)		
Monthly expenses for wages (dummy for missing values) (+)			0.052*** (0.013)
Average monthly profit (missing dummy) (-)		-0.032 (0.022)	
N	306	101	205
R^2	0.562	0.869	0.757

Note: Standard errors clustered at the firm-owner level are included in parentheses. $*p < .10, **p < .05, ***p < .01$. The left-most column contains the predictors selected using by the lasso. Column 1 includes the OLS estimates for the variables selected using lasso on the full sample of 306 firms. Column 2 and 3 include OLS estimates for the variables selected using lasso on the female-owned and male-owned firms respectively. Missing entries in each column mean the corresponding variable in the left-most column was not selected for the respective sample. The sign included in () at the end of each selected variable in the left-most column is the sign on the coefficient estimated using lasso.

Table 5a: Validating Loan Growth Measure Using Monthly Profits from NMB Administrative Data

	Firm-owner-level data			Firm-level data		
	(1) Profits (in 1000s TZS)	(2) Female firm owners: profits (in 1000s TZS)	(3) Male firm owners: profits (in 1000s TZS)	(4) Profits (in 1000s TZS)	(5) Female-owned firms: profits (in 1000s TZS)	(6) Male-owned firms: profits (in 1000s TZS)
Predicted high growth	1242.997***	586.864*	1074.761***	1036.360***	827.592***	1029.858***
	(358.652)	(304.645)	(216.893)	(145.612)	(234.596)	(190.084)
Constant	2075.068***	2656.840***	2190.685***	2388.608***	2486.730***	2791.103***
	(435.801)	(527.132)	(349.418)	(205.635)	(381.089)	(230.376)
N	256	84	172	304	100	204
R^2	0.110	0.047	0.087	0.106	0.117	0.091
Mean (low growth group)	2838.564	2618.103	2944.712	2783.376	2515.348	2914.539

Note: Robust standard errors are included in parentheses. $*p < .10, **p < .05, ***p < .01$; Columns 1 - 3 uses firm owner level data and columns 4 - 6 uses firm-level data. Columns 1 and 2 include the OLS estimates after regressing average monthly profits on the standardized predicted likelihood of whether a firm-owner/firm belongs to high-growth group. Columns 2 and 5 include the estimates for female entrepreneurs and female-owned firms, respectively, Columns 3 and 6 include the estimates for male entrepreneurs and male-owned firms, respectively.

Table 5b: Validating Loan Growth Measure Using Monthly Sales from NMB Administrative Data

	Firm-owner-level data			Firm-level data		
	(1) Sales (in 1000s TZS)	(2) Female firm owners: sales (in 1000s TZS)	(3) Male firm owners: sales (in 1000s TZS)	(4) Sales (in 1000s TZS)	(5) Female-owned firms: sales (in 1000s TZS)	(6) Male-owned firms: sales (in 1000s TZS)
Predicted high growth	6580.685***	2690.979	6447.336***	6377.960***	2614.798*	6838.649***
	(1475.537)	(1888.722)	(1096.124)	(1226.663)	(1559.878)	(1410.923)
Constant	12551.654***	15223.304***	12322.230***	12946.791***	14663.923***	15753.346***
	(1934.268)	(2693.969)	(2085.635)	(1437.285)	(2111.351)	(1644.482)
N	256	84	172	304	100	204
R^2	0.087	0.022	0.096	0.101	0.028	0.105
Mean (low growth group)	17048.36	13812.65	18606.3	16367.52	13245.5	17895.32

Note: Robust standard errors are included in parentheses. $*p < .10, **p < .05, ***p < .01$; Columns 1 - 3 uses firm owner level data and columns 4 - 6 uses firm-level data. Columns 1 and 2 include the OLS estimates after regressing average monthly sales on the standardized predicted likelihood of whether a firm-owner/firm belongs to high-growth group. Columns 2 and 5 include the estimates for female entrepreneurs and female-owned firms, respectively, Columns 3 and 6 include the estimates for male entrepreneurs and male-owned firms, respectively,

Table 5c: Validating Loan Growth Measure Using Stocks Value from NMB Administrative Data

	Firm-owner-level data			Firm-level data		
	(1) Stock value (in 1000s TZS)	(2) Female firm owners: stock value (in 1000s TZS)	(3) Male firm owners: stock value (in 1000s TZS)	(4) Stock value (in 1000s TZS)	(5) Female-owned firms: stock value (in 1000s TZS)	(6) Male-owned firms: stock value (in 1000s TZS)
Predicted high growth	12288.863***	-8043.486	6974.684***	12840.800***	-7685.988	14632.444***
	(3485.518)	(19045.023)	(2274.007)	(3685.989)	(18586.617)	(3480.282)
Constant	30209.893***	78370.321	26754.035***	27483.908**	70657.343	20413.520***
	(11481.770)	(62228.503)	(6922.310)	(12415.005)	(52431.917)	(3010.242)
N	256	84	172	304	100	204
R^2	0.003	0.001	0.013	0.005	0.001	0.068
Mean (low growth group)	47282.35	94664.62	24468.66	43187.73	83085.22	23663.42

Note: Robust standard errors are included in parentheses. $*p < .10, **p < .05, ***p < .01$; Columns 1 - 3 uses firm owner level data and columns 4 - 6 uses firm-level data. Columns 1 and 2 include the OLS estimates after regressing stocks value on the standardized predicted likelihood of whether a firm-owner/firm belongs to high-growth group. Columns 2 and 5 include the estimates for female entrepreneurs and female-owned firms, respectively, Columns 3 and 6 include the estimates for male entrepreneurs and male-owned firms, respectively.

Table 5d: Validating Loan Growth Measure Using Number of Employees from NMB Administrative Data

	Firm-owner-level data			Firm-level data		
	(1) Employees	(2) Female firm owners: employees	(3) Male firm owners: employees	(4) Employees	(5) Female-owned firms: employees	(6) Male-owned firms: employees
Predicted high growth	-0.015	0.084	0.096	0.463**	0.305	0.117
	(0.192)	(0.289)	(0.214)	(0.218)	(0.212)	(0.278)
Constant	2.455***	2.153***	2.343***	1.713***	1.905***	2.286***
	(0.412)	(0.444)	(0.539)	(0.257)	(0.332)	(0.465)
N	256	84	172	304	100	204
R^2	0.000	0.001	0.001	0.019	0.019	0.001
Mean (low growth group)	2.28	2	2.42	2.19	1.89	2.34

Note: Robust standard errors are included in parentheses. $*p < .10, **p < .05, ***p < .01$; Columns 1 - 3 uses firm owner level data and columns 4 - 6 uses firm-level data. Columns 1 and 2 include the OLS estimates after regressing average number of employees on the standardized predicted likelihood of whether a firm-owner/firm belongs to high-growth group. Columns 2 and 5 include the estimates for female entrepreneurs and female-owned firms, respectively, Columns 3 and 6 include the estimates for male entrepreneurs and male-owned firms, respectively.