Knockin' on the Bank's Door: Why Is Self-Employment Declining?

December 29, 2023

Abstract

The U.S. bank branch network has been contracting since the 2010s, limiting the access of self-employed borrowers to credit institutions. This paper analyzes this decline in the ability to obtain financing as a potential explanation for the decrease in U.S. self-employment. To evaluate the impact of the bank branch closings, I use a shift-share research design to assess self-employment exits using zip code variation in preexisting bank market shares. I disaggregate the self-employed into two categories: entrepreneurs whose businesses depend on business loans (incorporated self-employed) and other self-employed individuals (unincorporated self-employed). Using a Community Advantage Panel Survey, I find that the proximity of credit market institutions has heterogeneous effects on self-employment exits. Branch closings lead to the decline of incorporated business owners and do not affect unincorporated businesses. However, these effects are short-term (dissipating within a year since the shock) and very localized (within seven miles).

Keywords: Self-Employment, Access to Credit Market, Labor Mobility

JEL Codes: G21, G51, J62, L84, R12, R32

1 Introduction

Establishing new businesses primarily relies on the availability of credit market institutions (Petersen and Rajan (1994), Bates and Robb (2013), and Fracassi et al. (2013)). However, over the past two decades, the U.S. banking industry has undergone significant changes. Notably, the number of bank branches has declined, with most of these closures occurring in the recent decade (Fee and Tiersten-Nyman (2022)). This shift, largely driven by technological advancements and the evolution of the banking industry, has transformed lending practices, enabling lenders to remotely verify borrowers' credentials, assess risk, and monitor loan repayments. While this change has created a more competitive environment for lenders and more options for borrowers, there are several downsides.

The acceleration in the reduction of the bank branch network (Figure 1) has also increased the average distance between borrowers and lenders (Fee and Tiersten-Nyman (2022), Agarwal and Hauswald (2010), DeYoung et al. (2008); author's calculations in Figure 2), changing the nature of traditional local banking relationships.[1] As the distance increases, lenders might

[1]For instance, Berger et al. (2005) note the traditional role of smaller, community-based banks in local lending and their importance to local communities. Boot (2000) explores the concept of relationship banking, which emphasizes the importance of close, local interactions between banks and their clients.

struggle with assessing local economic conditions, borrower reliability, or collateral value. This reassessment might result in a higher perceived risk and potentially higher interest rates for borrowers.

Moreover, the lack of personal relationships between borrowers and lenders could lead to less flexibility during financial difficulties. Recent research indicates that the physical proximity of bank branches could promote entrepreneurship (Nguyen (2019)). This is attributed to the borrowing process for small-scale entrepreneurs, such as the self-employed, as it is still informationally intensive, with credit approvals primarily relying on soft information about the borrower (DeYoung et al. (2008)). In other words, a physical presence during a loan transaction continues to play an essential role in obtaining financing, and the reduced credit market access may negatively affect entrepreneurship.

In light of these changes, this paper examines the contraction of the U.S. bank branch network between 2004 and 2014 and its impact on self-employed borrowers. The empirical approach leverages a quasi-experimental setting, exploiting the spatial and temporal variation in branch closures across the United States. I am particularly interested in the potential supply shocks induced by branch closures, independent of changes in local demand for credit services. To isolate the supply shock, I use a statistical method to untangle changes in bank numbers, figuring out how much is due to banks' own decisions (supply) and how much is due to people's preferences (demand). Second, I construct a zip-code-year level measure of predicted supply shock, capturing the change in bank availability weighted by individual bank market shares. The large number of branch closures corresponds to a more severe supply shock. This measure incorporates the specific banks operating in each area and their relative share, providing a nuanced picture of the potential credit disruption self-employed borrowers face. By linking this supply shock measure to individual-level data on self-employment exits, I utilize a difference-in-differences framework with area-fixed effects to identify the impact of supply shock on the number of self-employed individuals relative to areas not experiencing closures. This isolates the net effect of reduced bank availability from local demand-driven changes.

My approach is grounded in the framework of Levine and Rubinstein (2017), in which individuals are split into two types based on their business's incorporation status—incorporated or unincorporated.[2] The authors show crucial differences between these two groups, suggesting

[2]The classification of businesses as entrepreneurial based on their legal structure is rooted in the theory proposed by Levine and Rubinstein. They argue that the key attributes of a corporation, notably limited liability and a separate legal identity, are especially beneficial for initiating sizeable projects, carrying high risk, and demanding outside financing. Levine and Rubinstein maintain that business experts are drawn to larger,

that incorporated businesses better represent entrepreneurship than the broader category of the self-employed demographic. While Levine and Rubinstein use volatility in home wealth to indicate credit condition fluctuations, I use a measure of local credit supply shocks. Notably, the findings highlight a fascinating heterogeneous impact: downturns in local banking markets cause those who run incorporated self-employed businesses to shut down, while unincorporated businesses remain largely unaffected.

The paper proceeds in two parts. In the first part, I discuss testable implications from an intuitive theoretical model that captures the role of local lending supply shocks in individual labor market decisions. The model is a three-sector Roy model based on Evans and Jovanovic (1989) and Levine and Rubinstein (2018) that shows the effect of branch accessibility on labor mobility among three labor states: incorporated self-employment, unincorporated self-employment, and paid employment. Incorporated self-employment comprises entrepreneurs, owners of firms that demand a higher amount of physical capital, and business loans. In contrast, unincorporated self-employment comprises non-entrepreneurs and owners of companies without business loans. In other words, the agent uses their own initial wealth to start a business. Agents enter self-employment based on their comparative advantage, which depends on their abilities and the number of available assets. These individuals face borrowing costs, including interest and non-interest costs (transportation costs, the time costs of arranging a loan, and any monetary initiation costs). The model assumes a negative local credit supply shock (more banks close their branches) associated with increased non-interest borrowing costs. The comparative static

riskier ventures; as a result, it's typical for more entrepreneurs to opt to incorporate their businesses. Their key findings indicate that individuals who start their businesses show higher learning abilities and self-esteem, earn significantly more per hour, and work longer than their salaried counterparts. Their study presents median annual and hourly earnings changes when individuals transition into or out of self-employment, both incorporated and unincorporated. For instance, a "smart and illicit" individual becoming an incorporated business owner sees a $7,000 increase in median annual residual earnings, which is 12% higher than the median residual earnings of their smart and unconventional salaried peers. These individuals see a substantial increase in earnings, both in absolute and relative terms, when they become incorporated business owners, more so than those with different cognitive and noncognitive traits. However, the situation for the unincorporated self-employed is quite distinct. This highlights the clear difference between entrepreneurship and other self-employment types and how various trait combinations are valued differently in various activities. Unlike incorporated business owners, "smart and illicit" individuals who choose unincorporated self-employment tend to experience a greater reduction in hourly earnings than those with different traits who also become unincorporated business owners. Being "smart and illicit" is positively linked to entrepreneurial success but negatively correlates with success in other self-employment ventures. Following this logic, I divide individual businesses into entrepreneurial and self-employed entities based on their either incorporated or unincorporated status.

shows that higher non-interest borrowing costs (decrease in loan accessibility or more severe supply shock) decrease the probability of starting an incorporated business and increase the probability of exiting incorporated self-employment.

In the second part of the paper, I estimate testable implications from the theoretical model. I focus on transitions out of self-employment by using an exogenous source of variation in local credit conditions. It tests a central prediction of the entrepreneurship model of Levine and Rubinstein (2018), diverging from Levine and Rubinstein's reliance on home wealth fluctuations. Additionally, my research differs from their work by delving into the effects of the consolidation of bank branches in the 2010s and its potential effect on the decline of self-employment in the US. Consequently, as access to funding becomes more limited, there could be a decline in the number of successful businesses, affecting local wealth and widening financial inequality.

The paper uses several sources of U.S. data between 2003 and 2014. The central database is the Community Advantage Panel Survey (CAPS) database,[3] and it allows me to identify all banks located near respondents' homes and to estimate geographical distances between individuals and all banks within the state. Information on bank locations comes from the Federal Deposit Insurance Corporation's (FDIC) Summary of Deposits (SOD) database.

I make three primary findings. First, bank consolidation negatively impacts the likelihood of loan approval and the stability of existing loan contracts. Specifically, a one standard deviation increase in anticipated credit availability correlates with a 0.5 percent increase in loan application rejections and a 0.3 percent surge in bankruptcy risk. Second, the impact of this credit supply shock varies among borrowers. Incorporated self-employed individuals face a significant risk of discontinuing their business pursuits. For example, a one standard deviation increase in projected credit supply corresponds to a 2.6 percent higher probability of exiting incorporated self-employment. Interestingly, the supply shock does not substantially impact the unincorporated self-employed. Last, these negative impacts are highly localized, dissipating within seven miles, and cause short-term (disappearing within two years) disturbances in entrepreneurial activity.

Taken together, the findings provide compelling evidence that major financial shifts, such as consolidation of bank branches, significantly influence entrepreneurial activity, complementing and expanding upon Levine and Rubinstein's research. These findings provide actionable insights for policymakers, emphasizing the significance of financial institutions in shaping en-

[3]This database comprises 11 years of a panel survey of approximately 5,000 low- and moderate-income homeowners and renters from 2003 to 2014. The UNC Center for Community Capital has collected the data with generous funding from the Ford Foundation.

trepreneurship.

The rest of the paper proceeds as follows. Section 2 provides background information and reviews the literature. Section 3 discusses testable implications from the theoretical model. Section 4 discusses the data used, and Section 5 introduces the empirical model. Section 6 discusses potential identification issues and solutions to them. Section 7 evaluates the alternative measure of the supply shocks. Section 8 explores the heterogeneous dynamic effects of supply shocks. Section 9 concludes.

2 Background and Literature

The U.S. self-employment sector, which spans numerous industries and professions, plays a crucial role in the economy. According to the U.S. Bureau of Labor Statistics (BLS), about 12 percent of the country's workforce is self-employed. These individuals contribute significantly to the economy, not just in terms of services and products but also through taxes and job creation when they hire additional workers.

There are two main types of self-employment: incorporated and unincorporated. The distinction hinges largely upon the legal and tax structures under which a person operates a business. To clarify, incorporated self-employment implies establishing a distinct, legal entity, such as a corporation or limited liability company, through which business is conducted. This grants the business a separate legal identity, meaning the owner's assets and liabilities are generally protected from business debts and lawsuits. Conversely, unincorporated self-employment lacks this separate legal entity, and the owner is personally liable for all business debts and liabilities. A detailed comparison of these two forms can be found in Appendix D.

The landscape of U.S. self-employment has seen notable shifts. While the total number of unincorporated self-employed individuals and total self-employment decreased between 2003 and 2015 (the sector has seen a decline of almost one million unincorporated self-employed), the number of incorporated self-employed individuals actually increased. However, the total rate of self-employed individuals as a percentage of the labor force has decreased by 1.1 percentage points over the same period (see Table 11, Figure 3).

The total share of self-employed workers in total employment decreased from 13.4 percent in 2003 to 12.3 percent in 2015.[4] The trend component of self-employed in the United States

[4]Author's calculations using data from the Current Population Survey (CPS). Between 2003 and 2015, the unincorporated self-employment rate fell from 9.09 to 7.8 percent. Over the 2003–2008 period, the incorporated self-employment rate rose from 4.3 to 4.8 percent. The rate then decreased to 4.6 percent in 2010 and remained

as a smoothed version of the self-employment numbers after removing cyclical fluctuations was estimated using the Hodrick-Prescott filter[5], has been decreasing steadily since 2003 (see Figure 4). The trend component allows us to see the underlying trend in the data, which is not obscured by short-term movements. The decline in the trend component of the self-employment rate is evident in both unincorporated and incorporated businesses. The average annual decline rate of the number of self-employed for both types has been 0.5%, after accounting for long-term trends' effects and the removal of business cycles.

But where did these individuals move? The transition matrix, Table 2, shows a lot of movement between different labor market statuses, even over a short period. Workers in paid employment are most likely to remain in paid employment (91.38%). However, a small percentage of workers in paid employment transition to incorporated self-employment (0.72%), unincorporated self-employment (1.08%), or non-employment (3.82%). More workers in incorporated self-employment stay in incorporated self-employment (56.26%) than transition to unincorporated self-employment (21.68%). Many also transition to paid employment (17.76%) or non-employment (4.3%). Workers in unincorporated self-employment are most likely to remain in unincorporated self-employment (55.12%). However, a significant percentage of workers in unincorporated self-employment also transition to paid employment (23.09%) or non-employment (5.19%). Most workers in non-employment stay in non-employment (70.9%), but a small number transition to paid employment (26.7%), incorporated self-employment (1.08%), or unincorporated self-employment (1.32%).

The existing literature provides limited reasoning for this decline. For instance, Hipple (2016) suggests that the drop in self-employment may be associated with a decline in agricultural employment. Schweitzer and Shane (2016) document another reason for this decline, finding that business cycles affect transitions in and out of self-employment,[6] while the recent fall in aggregate demand explains exits from entrepreneurship. In a departure from the literature, this paper posits that a potential explanation is the decline in the ability to obtain

at that level over the 2011–2015 period (see Figure 3).

[5]The Hodrick-Prescott (HP) filter is a popular statistical technique to de-trend statistics. The HP filter works by smoothing the data to remove short-term fluctuations while preserving the long-term trend. The algorithm of HP is the following. First, the data is transformed into a mean of zero and a standard deviation of one. The HP filter is applied to the transformed data. The transformed data is then inverted to obtain the de-trended statistics. The de-trended number of self-employed is a useful indicator of the underlying changes in the number of self-employed over time.

[6]Beckhusen (2014) also shows that transitions from self-employment to wage employment increased in the post-recession months.

financing.

A factor influencing the decision to become an entrepreneur is access to capital, combined with wealth and collateral constraints. Numerous studies have focused on the positive correlation between housing wealth and entrepreneurial activities (Fan and White (2003), Fairlie and Krashinsky (2012), Fort et al. (2013), Corradin and Popov (2015)). This body of literature shows that household-level credit constraints matter for the creation of new businesses (Evans and Jovanovic (1989), Holtz-Eakin et al. (1994), Gentry and Hubbard (2004), Cagetti and De Nardi (2006)). Other studies find that a bank loan is an essential source of financing for small businesses (Petersen and Rajan (1994), Bates and Robb (2013), Fracassi et al. (2013)) and that entrepreneurs often have to provide personal guarantees when they obtain financing (Berger and Udell (1998); Greenstone et al. (2014)).

Moreover, Herkenhoff et al. (2016) examine how access to consumer credit impacts employment prospects, earnings, and entrepreneurship. To isolate the causal effect of credit on labor market outcomes, the authors use bankruptcy flag removals[7] to separate a sizable discrete increase in credit access. They find that following the flag removal, there is (1) an increased flow rate into self-employment, (2) disproportionate borrowing by new self-employed entrants relative to other job transitioners, (3) an increased likelihood of starting an employer business, (4) an increase in startups entering capital-intensive and external finance-intensive industries, and (5) disproportionate borrowing by new employer businesses.

Levine and Rubinstein (2018) investigate puzzling gaps between theory and evidence regarding entrepreneurs' human capital, earnings, and liquidity constraints. They develop a theoretical three-sector Roy model to explore selection into entrepreneurship, self-employment, and salaried work based on human capital and liquidity constraints. They conclude that entrepreneurs—represented by the incorporated self-employed—are positively selected based on attributes like entrepreneurial talent, risk tolerance, salaried wages, and collateral. In contrast, the unincorporated self-employed represents the broader self-employment category and exhibits negative selection on the same attributes. Empirical findings from the study further distinguish between these two groups. The incorporated group (entrepreneurs) is positively selected based on talent and salaried wages and is influenced by collateral, while the unincorporated group

[7] A bankruptcy filing raises a "red flag" to potential lenders and affects future lending opportunities, preventing an individual from obtaining credit for a considerable period of time after filing. In addition, the associated interest rate will be exorbitant even when a debtor can obtain credit. Under the Fair Credit Reporting Act, a bankruptcy filing can remain on an individual's credit report for ten years, after which the bankruptcy can be removed.

(self-employed) shows opposite trends. The authors emphasize that these stark differences between entrepreneurship and self-employment must be considered when studying business cycles, noting that self-employment is countercyclical, while entrepreneurship is procyclical.

A growing strand of literature studies the importance of geographical proximity for bank lending, identifying two broad channels in the economics literature through which distance influences credit market transactions. First, studies on spatial rationing have established a correspondence between distance and credit rationing. A closer geographic distance provides banks privileged access to soft information that allows them to evaluate borrowers' creditworthiness, thereby permitting them to gain a cost advantage for monitoring over more remote competitors who may not enjoy the same degree of access to such information (Hauswald and Marquez, 2006). The information effect of distance has been shown empirically to facilitate ex-ante screening and ex-post monitoring of borrowers in bank lending, giving well-informed banks a competitive edge and market power (Petersen and Rajan (2002); Brevoort et al. (2010); Guiso et al. (2004); Sufi (2007); Qian and Strahan (2007)).

One benefit for borrowers who are located closer to their banks is that inefficient rationing might be reduced with privileged access to information. Using a unique data set of all loan applications by small firms to a large bank, Agarwal and Hauswald (2010) document that the closer a firm is to its branch office, the more likely the bank will offer credit. However, borrowing from closer banks corresponds to, for example, higher interest rates (Agarwal and Hauswald, 2010). The reason is that borrowers are informationally captured by lenders with privileged access to soft information to the extent that such information cannot be credibly communicated to outsiders (Dell'Ariccia and Marquez, 2004).

Second, a shorter lender-borrower distance potentially benefits both parties since it reduces transaction costs. For example, lenders can benefit from the reduced cost of obtaining soft information. Examples of such costs for a potential borrower include transportation costs incurred in applying for a loan and the time and effort spent personally interacting with loan officers or looking for a suitable loan. In classical models of location differentiation (Salop, 1979), borrowers incur distance-related transportation costs from visiting their banks, while banks price loans uniformly if they cannot observe their borrowers' locations or are prevented from charging different prices to different borrowers (Freixas and Rochet, 2008). However, if banks observe their borrowers' locations and offer interest rates based on that information, they may engage in spatial price discrimination (Lederer and Hurter Jr, 1986).

For example, Degryse and Ongena (2005) document that loan rates decrease with the distance between the firm and the lending bank and increase with the distance between the firm and competing banks, suggesting that transportation costs cause spatial price discrimination. Nguyen (2019) shows that bank branch closings during the 2000s led to a persistent decline in local small business lending. The author asserts that the effect is very localized and dissipates within six miles of the borrower's location.

This paper contributes to the literature on self-employment in two ways. First, it underscores the diversity of the self-employed. A growing body of research suggests that self-employed individuals should not be viewed as an aggregated group. Block and Sandner (2009) insist that using push-and-pull factors, which determine the selection into self-employment, allows the self-employed to be divided into two groups: "necessity" entrepreneurs who are self-employed due to the lack of other options and "opportunity" entrepreneurs who seek to bring new ideas to the market or avail themselves of other market advantages. Moreover, Levine and Rubinstein (2017) distinguishes between unincorporated and incorporated businesses, emphasizing that the latter, defined by specific cognitive and non-cognitive traits, serves as a more accurate representation of entrepreneurship. They note that these entrepreneurs tend to outearn typical salaried workers, whereas most self-employed individuals are incorporated with varied traits and earnings patterns. My results draw attention to the differential impact of local credit shocks on these categories of the self-employed. Incorporated self-employed individuals face a significantly higher risk of discontinuing their business operations due to supply shock. In contrast, those who are unincorporated appear to be less affected. This suggests that our understanding of entrepreneurship, as explored by Levine and Rubinstein (2017), must account for the heterogeneity within the self-employed population.

Secondly, while existing literature often explores binary choice models to understand the decision-making process of being self-employed, with a significant focus on individual wealth shocks as drivers for entrepreneurial ventures, this study adopts an innovative perspective. It examines the influence of local supply shocks, such as the closure of financial institutions, on the decision to exit self-employment. A one standard deviation rise in the supply shock (equivalent to 0.15 bank closures) leads to a 2.6 percent increased risk of leaving incorporated self-employment at the zip-code level. This effect is even greater in areas more acutely affected, reaching 3.9 percent. Notably, this magnitude is comparable to the well-studied wealth shock: Holtz-Eakin et al. (1994) discovered that inheritance of $100,000 increased the odds of moving from wage employment to self-employment by 3.3 percentage points[8]. This comparison demon-

[8]Following the pioneering work of Evans and Jovanovic (1989), many studies estimate binary choice models of being an entrepreneur, which includes some measure of individual wealth. Studies find that unexpected financial gains have a positive, significant, and sizeable influence on the chances of entering self-employment.

strates that the influence of external factors on exiting self-employment can be as substantial as the documented impact of individual wealth shocks on entering it, raising questions about how policies and support systems can be tailored to address the diverse drivers of career choices in this sector.

3 Testable Implications of the Theoretical Model

The theoretical framework examines the impact of supply shocks triggered by increased non-interest borrowing costs on labor market choices. The literature (Degryse and Ongena, 2005) distinguishes three factors underlying the role of bank branch closures: transportation costs (borrower's increased cost of obtaining credit), lender's monitoring costs, and the creation of asymmetric information. Proximity gives the bank insight into the local economic conditions affecting a potential borrower. Since the bank has soft information about the borrower, information asymmetry may either reduce the likelihood of lending or increase the loan price for the borrower (Carling and Lundberg, 2005). The model proxies the branch network contraction through the increased non-interest cost of borrowing due to the longer geographical distance between a borrower and a lender.

To emphasize the result of negative supply shocks through the bank closure and the increase in borrower-lender distance on labor market choices, I use an intuitive three-sector Roy model of labor market decisions based on the works of Evans and Jovanovic (1989) and Levine and Rubinstein (2018). The model captures the essential features of the relationship between distance and labor mobility among three labor states: two types of self-employment (incorporated, ISE, and unincorporated, USE) and paid employment (or salaried employment) (PE). The model is a discrete-time model. Individuals select among the three labor states in each period, where changes between states include known switching costs. Given the information available at the beginning of each period, individuals choose their labor states, borrowings, and consumption to maximize their discounted expected utility. They face borrowing costs, including interest and non-interest costs ψ (transportation costs, the time costs of arranging a loan, and any monetary initiation costs), which is a function of the borrower-lender proximity.

However, these effects tend to decrease as the value of the windfall increases. For instance, Blanchflower and Oswald found that in 1981, a British individual who inherited £5,000 was twice as likely to be self-employed compared to someone who did not receive any inheritance. Similarly, In Sweden, Lindh and Ohlsson (1996) determined that receiving lottery winnings increased the likelihood of self-employment by 54 percent, and an average-sized inheritance did so by 27 percent.

A shorter borrower-lender distance makes non-interest costs cheaper. A larger borrower-lender distance decreases credit market access and makes large-scale, incorporated self-employment less attractive since such business opportunities often require external financing. Unincorporated self-employment as a small-scale business sector is less dependent on external funding but still may experience contraction if the borrower-lender distance increases. All estimations are shown in the "Theoretical Model" section in Appendix A.

The theoretical model yields testable implications on the relevance of non-interest cost of borrowing on the probability of labor market choices. The key testable implication of the model is an increase in the non-interest cost of borrowing (higher borrower-lender distance or worse accessibility of banks) is attributed to the decreasing probability of incorporating a business and an increase in the likelihood of exiting self-employment.

Corollary 1. Higher non-interest costs of borrowing disincentive self-employed to incorporate businesses.

Higher non-interest borrowing costs could discourage self-employed individuals from incorporating. With incorporation, business debts become separate from personal finances, but borrowing becomes costlier.

Corollary 2. Higher non-interest costs of borrowing increases the probability of switching from incorporated self-employment.

Rising non-interest costs, like transportation or time spent acquiring loans, could push self-employed individuals towards quitting. These costs bite hard for established businesses, potentially forcing closure or career changes. Higher non-interest borrowing costs can make it more difficult for incorporated self-employed individuals to expand their businesses. This is because they will have to spend more revenue on servicing their debts. This can leave them less money to invest in new equipment, hire new employees, or market their products or services. As a result, they may be less competitive and more likely to fail.

4 Data

4.1 Main Database

This paper combines several sources of U.S. data. The primary source is the CAPS database collected by the UNC Center for Community Capital. The database comprises 11 years of panel survey data provided by approximately 5,000 low- and moderate-income homeowners and renters during the period 2003–2014. The homeowners recruited to participate in the CAPS received mortgages between 1999 and 2003 through the Community Advantage Program, and the participating renters were recruited to match these homeowners with respect to their geographic proximity and income ceilings.

The main benefit of using this database is the opportunity to identify each individual's full geographical location. The database has been used in several works (e.g., Quercia et al. (2011), Manturuk et al. (2017)),[9] and it collects a wide variety of information. The data include demographics and family formation, mobility and housing tenure choice, unemployment, wealth and asset holdings, social capital and civic engagement, and housing satisfaction.

The database contains a post-sample weighting used for the empirical analysis to minimize the impact of biases resulting from higher attrition across various demographic groups and selection to CAP. Appendix D4 provides more information about the survey design.

The CAPS database also contains self-reported information about individuals' primary work activity at the time of the survey. The survey has four categories of employment: workers in private businesses, government workers, self-employed workers, and workers in family businesses. The first two categories are combined as "paid employed workers." Due to absolutely different liabilities, the self-employed and workers in a family business are considered different labor types. Therefore, I consider only self-employed workers in the analysis and exclude workers in a family business from the data sample.

4.2 Bank Offices

All information on bank offices comes from the FDIC's SOD data for 2003–2014 as of June 30 of the corresponding year. The SOD database provides an annual enumeration of all branches belonging to FDIC-insured institutions. As of September 2019, the FDIC provided deposit insurance at 5,256 institutions. The agency insures deposits in member banks up to US$250,000

[9] The full list of publications is available at https://communitycapital.unc.edu/files/2017/10/Paper_22929_extendedabstract_1348_0.pdf.

per ownership category. The SOD data include full street addresses of bank headquarters and bank branches, the total amount of assets, and the latitude and longitude of the bank headquarters and each bank branch since 2008.[10]

The maps in Figure 5 show the location of all banks and branches from the SOD database (red dots) in 2003 and 2014. The database contains information about 87,279 bank locations in 2003 and 94,521 bank locations in 2014. Thus, the number of bank branches increased during that period, but the locations were more concentrated. To show the difference in concentration between 2003 and 2014, Figure 5 also illustrates the hot spot analysis (blue areas), showing the z-score for the Getis-Ord Gi* statistic for each location in a dataset. Darker areas indicate where bank locations with high values cluster spatially.

Bank addresses were transferred to geocodes[11] and were linked to the home locations of CAPS respondents so that each respondent was linked to all banks within the same state (the minimum distance; distance to nearest 5, 10, or 15 banks). Bank locations are used to identify changes in the number of bank branches bearing respondents' homes and changes in the borrower-lender distance.[12]

4.3 Geographical Unit

The main geographical unit used in the analysis is the zip code level, as it is the smallest geographical identifier available in the FDIC's SOD database. The U.S. has roughly 33,000 zip codes, occupying a land area of 2,808,990 square miles. The smallest zip code is 00906, which is only 0.0032 square miles. In contrast, the largest zip code is 99557, with a large area of 13,431 square miles. The average land area of a zip code is around 90 square miles, which is similar to the size of Knoxville, Tennessee. Much like land area, there are vast differences in population size. For example, the most populated zip code is 00725 in Puerto Rico (over 144,000 residents), whereas the smallest zip code is 59921, found in Lake McDonald, Montana,

[10]Due to some limitations, the data should be interpreted carefully. For example, the FDIC's SOD database only includes data from FDIC-insured institutions and may not fully capture the entire banking sector in some areas.

[11]Some banks' addresses were dropped as incomplete. On average, branches with incomplete information account for less than 1 percent of the total number of branches.

[12]I use the SAS geocode procedure for estimating distance, as SAS allows distances to be estimated without sending sensitive information to external servers. However, one downside of using SAS is that it estimates the Euclidean distance between two locations. Boscoe et al. (2012) conclude that for nonemergency travel, the added precision offered by the substitution of travel distance, travel time, or both for straight-line (Euclidean) distance is largely inconsequential.

with a population of just one resident.

4.4 Other Sources

Other sources of data include the U.S. Census Bureau Statistics (County Business Patterns), BLS Local Area Unemployment data (county-level unemployment rates and civilian labor force size), the FDIC's assets and liabilities report (total number of employees per branch; full-time equivalent), American Community Survey (ACS) county-level demographic information, and Zillow (the average house value).

4.5 Explanatory Variables

As explanatory variables, I use a set of time-invariant individual characteristics, such as gender (= 1 if female), race (white, black, and other), and Hispanic ethnicity. I also use time-varying variables, including age, age squared, years of schooling, spouse's years of schooling, marital status, household size, the number of children under the age of 14 currently living in the household, the region, calendar year effects, the log of the value of large and durable assets (houses, land, other real estate, and cars), and the difference in log-income for wage and salary workers and self-employed (non-incorporated) at the county level. The latter variable is a substitute for individual income, which is likely endogenous. Unfortunately, the CAPS database does not contain detailed wages or personal income information, only aggregated household income.

As a measure of local economic conditions, I use the following metrics at the zip code level: total population; the share of the population with professional, scientific, management, and administrative education; the unemployment rate; and the average house value. Additionally, I use the number of business establishments at the county level as a proxy for local entrepreneurial activity.

4.6 Summary Statistics

Table 3 shows summary statistics and differences between different labor statuses (paid-employed workers, incorporated and unincorporated self-employed, and non-employed). The table includes variables such as age, gender (indicated by "Female"), marital status, number of kids, number of household (HH) members, race, ethnicity, years of schooling, and several financial metrics like distance to the nearest bank, number of banks within 5 miles, number of bank workers per 1000 population, total assets in 2014, and Supply Shock (z-score). The table provides the means for each variable with standard deviations in parentheses. The table highlights differences between the groups based on employment status and possibly between incorporated and unincorporated self-employed individuals. Stars (*) shows p-values for the t-test of mean differences between incorporated and unincorporated self-employment, where *** indicates significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Incorporated self-employed workers are, on average, older than their unincorporated counterparts. This could be due to several factors, such as the need for more experience and education to start a business that justifies incorporation or the financial stability that comes with age allowing individuals to take on the risks of incorporation. Males and married individuals tend to be more likely to be incorporated self-employed. Historically, women have been underrepresented in entrepreneurship, potentially due to factors like access to capital, societal expectations, and work-life balance challenges. Similarly, marital status can influence the incorporation of the business. Spouses may provide support and financial stability, allowing individuals to pursue incorporated entrepreneurial ventures. Incorporated self-employed workers typically have more years of schooling than unincorporated workers. This suggests that higher levels of education may be associated with the skills and knowledge needed to run a business and navigate the complexities of incorporation successfully.

Incorporated self-employed workers have significantly higher total assets than unincorporated. This could be because incorporated businesses tend to be larger and more profitable, allowing them to accumulate more wealth over time. Also, incorporated self-employed workers have better access to financial services, as evidenced by the higher number of banks per 1,000 population in their areas. This access to capital can be crucial for business growth and investment, further widening the gap between incorporated and unincorporated workers.

5 Empirical Analysis

My previous section discussed the theoretical foundation of this study, which examined the complex relationship between the non-interest cost of borrowing and individual labor decisions. The non-interest cost of borrowing includes the time and effort borrowers expend to obtain a loan, which can be measured by the accessibility of financial institutions. In this section, I begin the empirical estimation phase of the research, where I strive to validate the propositions advanced by my theoretical framework empirically. To assess the impact of the accessibility of financial institutions, I use two key measures: (1) lending shocks, measured by changes in the number of bank branches, and (2) the geographical proximity of branches, measured by the distance between borrowers and lenders. I also address potential identification challenges and propose solutions. The results illuminate the impact of the accessibility of financial institutions on the likelihood of obtaining a loan, the propensity to exit self-employment, and the differential impact on self-employment by business type. The main empirical specification follows Greenstone et al. (2020) and uses the measure of lending shocks to isolate the effect of changes in the accessibility of banks for borrowers. As an alternative measure, I propose the borrower-lender distance.

5.1 The Measure of Lending Shocks

In this subsection, I discuss the construction of the supply shock measure. Basically, the agent lives in a town with several banks. Suddenly, some of them close down, leaving fewer options for everyone. This would be a supply shock – a change in the number of banks available, independent of how much people want to use them. I'm interested in how such shocks affect people's lives, particularly their choices about work. To study this, I need a way to isolate the supply shock from other factors like changes in local demand for bank services. First, I use a statistical method to untangle changes in bank numbers, figuring out how much is due to banks' own decisions (supply) and how much is due to people's preferences (demand). Then, I create a bank availability measure for each area and year that captures the supply shock. This measure considers which banks operate there and how big they are.

Bank branch closures could adversely affect local areas where they operate through shocks in credit supply, disruptions in lending relationships, and increased non-interest borrowing costs, potentially affecting small business lending and individuals' decisions to run a small business. To identify the role of branch closure on self-employment, I developed a modified version of the shift-share approach using the Bartik instrument (Bartik and Bartik Timothy, 1991). Modifying the "classical" shift-share instrument[13] is necessary because bank branch closures may reflect both bank-specific supply shocks (e.g., the merge-induced consolidation of banks, as described in Nguyen (2019)) and a local decline in the demand for loans due to reduced entrepreneurial activities in the area.

To isolate the supply shock, I follow the strategy proposed by Greenstone et al. (2020) and separate the contribution of demand and supply shocks to changes in the number of banks.

[13] This instrument for the bank in area j in period t is created as the sum for each bank in area j of the inner products of bank shares in area j and local banking industry growth rates at the national level between $t-1$ and t.

The first step is to estimate the following equation:

$$\triangle\left(X_{kjt}\right)=d_{jt}+s_{kt}+\epsilon_{kjt}. \tag{1}$$

The outcome variable is the percentage change of the number of bank k network of branches in geographical area j[14] between periods t and $t-1$. d_j is an area fixed effect and measures the variation in the number of branches due to changes in local economic conditions, and captures the effects of local demand for credit from bank k. s_k identifies banks' fixed effects, which measure the supply response of branches, and it is separated from the variation in demand for loans in the area. s_k are the parameters of interest, estimated for each period between t and $t-1$ beginning in 2003. ϵ_{kj} is the unobserved random shock.

I next create a quasi-experimental and conditionally exogenous proxy for supply shocks for each zip code level and year. I use the estimated bank-specific supply shocks, $\widehat{s_k}$, to construct a zip-code-level measure of the predicted supply shock that captures bank availability. For each area j and year t, I take the weighted average of the estimated bank k fixed effects from Equation 1, weighting by bank k's market share at period t in area j, m_{kj}, which is measured by the ratio of the number of branches of bank k to the total number of all branches in area j.

$$p_{jt}=\sum_k m_{kjt}\widehat{s}_{kt} \tag{2}$$

The zip-code-predicted shocks, p_{jt}, are estimated for every pair of periods $(t-1)$ and t, creating a time lag between supply shock and individual labor choices. For the convenience of interpretation, I standardize the predicted shock with a mean of zero and a standard deviation of one.

Figure 7 presents the dynamic of the average supply shocks over the years, showing that the average supply shock increases over time, correlating with bank branches' consolidation. The predicted shock is scaled so that a higher value indicates a more severe supply shock (e.g., more branches closed), and a lower value indicates a less severe supply shock (e.g., fewer branches closed). This standardization measures the relative supply shocks by zip code area so that some zip codes will be above or below the average supply shock area. Figure 8 visualizes branch closure intensity for 2004 (a) and 2014(b). Blue dots reflect more severe supply shock (more branches closed), whereas orange dots show less intensive supply shock (fewer branches closed).

[14] The zip code level is the minimal geographical area available in the SOD database, which is why it is chosen as a base geographical unit.

5.2 Supply Shocks and Access to Credit

The empirical analysis begins by addressing the role of supply shocks in the probability of obtaining credit. Unfortunately, the CAPS data do not contain information about business-related lending, and in a perfect situation, the effect of the supply shocks should be estimated through the business loans. Using the available data, I highlight two main groups of variables characterizing borrowers' credit constraints: the probability of obtaining different credit products and the probability of changes in the current loan contracts. The first group of variables is a set of binary indicators capturing whether the respondent, in the last 12 months, applied for a new credit card, opened a new charge card, or refinanced a mortgage. The second group consists of a set of dummy variables noting if, within the same period, the respondent experienced a credit application rejection, was asked to pay off a remaining loan balance, filed for bankruptcy, or had their credit card limit reduced.

To evaluate the effect of the supply shocks on access to credit, I estimate a linear probability model for each variable characterizing access to credit. The model includes geographical cluster fixed effects to control for unobserved characteristics that characterize the initial access to the credit market.

$$C_{ijskt} = \alpha p_{jt} + \gamma X_{it} + \delta I_{it-1} + \beta Z_{jt} + \epsilon_{st} + \tau_k + \varepsilon_{ijskt} \tag{3}$$

where C_{ijskt} is a dummy variable that equals one if respondent i in zip code j and state s during period t has taken any of the following actions in the last 12 months: applied for a new credit card, opened a new charge card, refinanced a mortgage, had an application for new credit rejected, had their available limit for credit cards reduced, was asked to pay off the remaining balance for a loan, or filed bankruptcy. p_{jt} is a predicted lending shock estimated in Equation 2.

The model includes a full set of state-by-year fixed effects, ϵ_{st}, the error term, ε_{ijskt}, and zip code characteristics Z_{jt}. τ_k identifies geographical clusters to control for unobserved spatial characteristics (full details in the section "Identification"). The state-by-year fixed effects indicate that comparisons between the groups of zip codes are made within the state for each year. I_{it-1} includes the previous period's labor status (salaried employment, incorporated and unincorporated self-employment, unemployment). X_{it} is the set of observed covariates, including dummies for gender, marital status, three race dummies (white, black, other), Hispanic ethnicity, four education dummies (less than high school, high school, some college, college and above), credit score, debt-to-income ratio, age, and a quadratic for age. Z_{jt} includes local economic conditions.[15] All regressions are weighted by CAPS individual-level weights to extrapolate the results, and standard errors are robust.

Table 4 shows the results of estimating Equation 3. They suggest a more severe supply shock decreases the probability of obtaining a new credit card, charge card, and mortgage refinancing. The results suggest that a stronger supply shock increases several risks. These include a reduction in access to credit, rejection of new credit applications, a reduction in the available credit card, being asked to pay off the remaining balance, and filing for bankruptcy.

5.3 The Role of Lending Shocks in Self-employment

This subsection presents a descriptive overview of the relationship between lending shocks and self-employment trends. It includes a discussion of the correlation between the main variables of interest (transition from both types of self-employment, the minimum borrower-lender distance, supply shock, and the number of branch closures). The event study approach presents another descriptive evidence that shows the connection between branch closure and switching out of self-employment. The last method is the linear probability model, which evaluates the exit from self-employment due to the severity of supply shock.

The analysis starts with the correlation matrix (Table 5) between variables of interest, such as the indicator of switching from incorporated and unincorporated self-employment, the minimum borrower-lender distance, supply shock, and the number of branch closures. The correlation matrix suggests weak to moderate linear relationships between the four variables. Unsurprisingly, there is a strong and significant correlation between the supply shock measure and the number of branch closures. The analysis reveals a negative and statistically significant association between transitions from incorporated and unincorporated self-employment. This implies that these two self-employment forms may substitute for each other in individuals' career paths.

Findings suggest severe supply shock contributes to increased borrower-lender distance and the transition rate away from incorporated self-employment. At the same time, there is a decline in the flow out of unincorporated self-employment.

Figure 9 shows that branch closures have a heterogeneous effect on the switching from self-employment based on the business type. Event time is defined as a year when the closure of a bank branch occurred at the zip code level, and there was no closure of branches two years

[15] The number of business establishments; the unemployment rate; the average value of the house; and the share of the population with professional, scientific, management, and administrative education at the zip code level.

before.[16] The predicted probability of exiting the incorporated self-employment trend is higher after a branch's closure. The result for unincorporated self-employed individuals is the opposite: the branch closure reduces the transition from unincorporated self-employment to other types of employment. The event-study analysis has no casual validity but can be used as descriptive evidence of different responses of self-employed to bank branch closures.

Another way to address this heterogeneous effect is to investigate the role of supply shocks in exiting self-employment using the following linear probability model:

$$Y_{ijst} = \alpha p_{jt} + \gamma X_{it} + \delta I_{it-1} + \beta Z_{jt} + \epsilon_{st} + \tau_k + \varepsilon_{ijskt} \quad (4)$$

The dependent variable Y_{ijst} measures self-employed' exit (or entry) at the individual level. Y_{ijst} is a dummy variable, which equals one if individual i in zip code j, spatial cluster k, and state s was an self-employed in the year $(t-1)$ and becomes a non-SE in the year t, and zero otherwise (exists self-employment in the year t). p_{jt} is a predicted lending shock estimated in Equation 2. The model includes a full set of state-by-year fixed effects, ϵ_{st}, the error term, ε_{ijskt}, and zip code characteristics Z_{jt}. State-by-year fixed effects indicate that comparisons between the groups of zip codes are made within states each year. For ease of interpretation, p_{jt} are standardized to have a mean of zero and a standard deviation of one.

I_{it-1} includes the previous period's labor status (salaried employment, incorporated and unincorporated self-employment, unemployment). X_{it} is a set of observed covariates, including dummies for gender, marital status, three race dummies (white, black, other), Hispanic ethnicity, four education dummies (less than high school, high school, some college, college and above), age, and a quadratic for age. τ_k identifies geographical clusters to control for unobserved spatial characteristics. Z_{jt} includes local economic conditions.[17] All regressions are weighted by CAPS individual-level weights, and standard errors are bootstrapped.

The coefficients of interest are α, which capture the difference in transitions between labor states due to the severity of supply shocks. Figure 10 confirms a strong heterogeneous effect of supply shocks on switching from or to self-employment based on the type. The figure validates

[16]The graphs show the predicted probabilities of switching from or to self-employment $\widehat{Y_{ijt}}$, estimated from the following equation: $Y_{ijt} = \beta_0 + X_{ijt}\beta_1 + \beta_2 \sum_{s=2003}^{2014} \mathbb{1}(year = s) + \gamma \sum_{t=-5}^{6} \mathbb{1}(event = t) + \xi_{ijt}$. The probability model controls for each year before and after the event (the categorical timeline), a quartic polynomial of current age, and calendar year fixed effects. The plotted dots are the coefficients on the timeline variable plus the average predicted share evaluated at timeline = 0 (which is the omitted category). Only four points before and after the event are shown.

[17]The number of business establishments, the unemployment rate, the average house value, and the population's share with professional, scientific, management, and administrative education at the zip code level.

that the severity of the supply shock decreases the predicted probability of switching to (or staying in) self-employment for incorporated self-employed individuals, while the effect is the opposite for those who are unincorporated self-employed.

5.4 The Dynamic Effect of Recent Lending Shocks on Self-employment

In my subsequent analysis, I use the results of Goldsmith-Pinkham et al. (2020) and Baker et al. (2022) to estimate the causal effect of bank branch closings on self-employment exits. Goldsmith-Pinkham et al. (2020) show that shift-share models with geographic fixed effects can be considered difference-in-differences models. This insight allows for the estimation of the causal effect of bank branch closings on self-employment exits, mitigating concerns about potential correlations between other variables and both the closings and self-employment exits. Baker et al. (2022) suggest a method for estimating the timing of the effects of bank branch closings. This is important because the effects may not be immediate. For example, it may take some time for self-employed to realize they no longer have access to credit, and it may take even longer for them to decide to exit self-employment.

The validity of difference-in-differences models depends on the assumption that no pre-treatment trends exist in the outcome variables. This means that the outcome variables should be trending in the same direction in both the treatment and control groups before the treatment is applied. To control for this, I include state-by-year and cluster fixed effects trends in my models. State-by-year fixed effects control for any common trends that affect all zip codes in a state in a given year. Cluster fixed effects control for any common trends that affect all zip codes in a cluster, such as a metropolitan area.

To further test whether the identification assumption is satisfied, I estimate the following dynamic equation:

$$Y_{ijkst} = \sum_{\tau=-4}^{\tau=5} \alpha_\tau \theta_\tau p_{j\tau} + \gamma X_{it} + \delta I_{it-1} + \beta Z_j + \tau_k + \epsilon_{st} + \varepsilon_{ijskt}. \quad (5)$$

I extend Equation 4 by including a set of dummies θ_τ that equal one in the τth year before or after the most recent branch closure. α are the parameters of interest; they are the coefficients on the interactions of the τth year predicted lending shock with year indicators. They also measure the impact of the lending shocks on self-employment in the year of the shock and all subsequent years relative to self-employed in the years before the shock and other zip codes. Thus, this is a difference-in-differences-style estimator. If the identification assumption is valid, the estimated coefficients on α for $\tau < 0$ should not be statistically different from zero. This dynamic approach also allows me to see if the banking reforms have any lagged effects on self-employment.

Table 14 presents the results of estimating Equation 5. The table shows a significant role of a supply shock in the closure of incorporated businesses. A one standard deviation increase in the predicted credit supply corresponds to a 2.6 percent increase in the likelihood of exiting incorporated self-employment. However, this effect is short-term and disappears a year after the most recent branch closure.

Figure 11 provides an illustration of these results. The estimates confirm a statistically significant relationship between the predicted lending shock and exiting self-employed, particularly incorporated self-employed, in the year of a bank branch closure. Meanwhile, the role of the supply shock in switching out unincorporated self-employment becomes insignificant. The figure also shows the lack of a pre-treatment trend, which is an essential condition of the validity of estimations.

I also examine the difference in the effect of supply shocks on geographical areas based on the supply shock's severity. I divide zip codes into quartiles and compare the effect of the supply shock among the top (above the 75th percentile) and the bottom (below the 25th percentile) quartiles.

$$Y_{ijkst} = \sum_{\tau=-a}^{\tau=b} \alpha_\tau \theta_\tau p^q_{j\tau} + \gamma X_{it} + \delta I_{it-1} + \beta Z_j + \tau_k + \epsilon_{st} + \varepsilon_{ijskt}, \tag{6}$$

where $p^q_{j\tau}$ is an indicator where zip code j experiences the most severe supply shock (belongs to the top quartile) or the least severe supply shock (belongs to the bottom quartile).

Figure 12, panel (a) confirms the strong relationship between the severity of the supply shock (higher number of branch closures) and the probability of exiting self-employment. For incorporated businesses located in the zip code under the severe supply shock, with a one standard deviation increase in the supply shock, the risk of exiting increases up to 4 percent (panel (b)). The results confirm the independence of unincorporated businesses from supply shocks regardless of the shocks' size (panel (c)).

6 Identification

6.1 Validity of the Instrument

The shift-share instrument allows me to isolate the exogenous component of a local closure of bank branches, and it decomposes the closure into multiple parts, each with a different endogeneity. The most exogenous component is the national or global banking industry trend, which is not correlated with unobserved factors that affect the local credit market outcome. The most endogenous component is the local closure, correlated with these unobserved factors. The shift-share instrument neutralizes this endogeneity by de-localizing it over space and time, meaning the instrument is constructed using national or global industry trends rather than local trends. This ensures that the instrument is not correlated with the unobserved factors that affect the local credit market outcome.

The shift-share instrument is valid as long as the local credit market outcome cannot affect the share in previous periods or national averages beyond the zip code in the current period. These conditions can hold if the shares are chosen to represent deep characteristics of zip code location, such as the demographic composition or industrial mix.

The only validity issue that may arise from using this approach is that the skills of bank managers used to manage investment portfolios may also be used to choose branch locations, suggesting that worse-than-average portfolio performances could be correlated with more severe supply shocks. However, Greenstone et al. (2020) provides empirical evidence suggesting no correlation between worse-performing banks and branch closures. To further address this validity issue, my empirical analysis uses a proxy for managerial skills and the number of bank workers per 1,000 population and the share of the population with professional, scientific, management, and administrative education, both at the zip code level.

Goldsmith-Pinkham et al. (2020) discuss the methodology used in an exposure research design that uses industry shares as instruments to measure the differential exogenous exposure to a common shock. They argue that this empirical strategy is equivalent to difference-in-differences in settings where the researcher has a pre-period. The primary identification worry is that the industry shares may predict outcomes through channels other than those posited by the researcher. The authors suggest that one way to assess the instrument's validity is to test for parallel pre-trends, meaning one should see similar trends in the outcome variable before the common shock, both for the treatment and control groups. If there are pre-existing differences in the trends, then this suggests that the common shock is not the only factor driving the change in the outcome variable. I address the pre-trend issue in the empirical analysis (Equation 5); Table 9 shows the results of the test for pre-existing similarities between treatment and control groups.

6.2 Omitted Variable Bias

The shift-share (or Bartik-style) research design uses pre-existing bank market shares across zip codes to predict changes in local credit supply due to bank branch closings. The assumption is that this variation is uncorrelated with unobserved factors that affect self-employment exits in a given zip code. Any trends or shocks affecting exits should not drive the pre-existing bank market shares. To address these concerns and demonstrate that omitted factors are unlikely to be the primary force behind the findings, I use the proposed techniques proposed by Oster (2019) and Diegert et al. (2022).

Oster's method leverages observable factors to glean insights into the selection of unobservable ones. It estimates "breakdown points," which measure the maximum influence of unobservable factors needed to overturn the initial model's findings. This is done by estimating breakdown points, which measure the maximum selection of unobservables relative to observables required to overturn a specific finding from the initial model. These breakdown points, denoted as δ, are presented in Table 8. According to Oster (2019), the breakdown point greater than one ($\delta > 1$) indicates that unobservable factors are unlikely to explain the initial results.

However, Oster's method assumes that the control variables are exogenous, which can potentially be restrictive. To address this limitation, I implement the methodology proposed by Diegert et al. (2022). This approach relaxes the exogeneity assumption by introducing a parameter, $\overline{r_x}$, which allows us to test whether the supply shock coefficient differs from the baseline results. According to Diegert's method, a $\overline{r_x}$ value less than 1 implies that the selection on non-observables is weaker than the selection on observables, suggesting limited influence of unobservable factors on the explained data. Table 8 summarizes these findings.

6.3 Geographical Clusters

Another identification issue is related to possible unobserved spatial characteristics that may affect exposure to branch closures. Geographical locations with a similar initial share of branch closures may share unobservables that affect self-employment dynamics over time. For example, banks may apply similar optimization practices to their branches according to local economic conditions.

To mitigate this issue, I use propensity score matching to divide individuals into geographical clusters. Due to the relatively small number of observations in the CAPS database, the goal is to cluster individuals based on their initial spatial access to bank branches. Accounting for the low mobility of respondents in the CAPS database (only 1.15 percent of the sample changed their place of living within a state, i.e., county and zip code, and 0.92 percent of the sample changed their state), I divide individuals into groups using their initial proximity of respondents' addresses to a bank branch.

The proximity is identified based on the minimal distance between the respondent's place of living and the nearest branch, the number of banks per capita at the county level, the average number of bank employees per branch at the zip code level, and other local socioeconomic conditions at the zip code level (percentage of those unemployed, public school quality in the neighborhood, percentage of the population with a college degree, number of business establishments, and the average home value). Overall, there are 20 geographical clusters in the data, and the analysis includes fixed effects associated with these clusters.

7 Alternative Measure. The Role of Bank-Borrower Distance

Alternative measure of local credit market supply shock is the changes in the borrower-lender distance. This approach has some benefits, including theoretical justification of this relationship, and distance might be exogenous to other factors affecting entrepreneurship, like individual risk appetite or business acumen. However, the challenge in using the borrower-lender distance as a measure of the credit supply shock is the difficulty of proving that changes in the distance between borrowers and lenders (which are used to approximate changes in credit supply) affect individuals' employment decisions only through their impact on credit availability. Other factors, such as changes in local economic conditions or the types of industries present in a region, could also be correlated with changes in lending distances and could independently affect employment decisions.

Distance plays a dual role in lending. On the one hand, a bank closer to a borrower enjoys significant local market power due to the borrower's higher transportation costs from switching to a more distant lender (Petersen and Rajan (1994), Degryse and Ongena (2005)). On the other hand, a bank actively monitoring a borrower incurs transportation costs from visiting the borrower on-site, and being closer to borrowers can lower monitoring costs. These two channels imply different relationships between distance and the cost of borrowing: interest rate and transportation cost. When transportation costs are borne by the borrower (i.e., the borrower visits the bank to negotiate loan terms), banks enjoy local market power over close clients, and interest rates should fall as distance increases (local market power hypothesis). If transportation costs are borne by the bank in its monitoring process (i.e., the bank visits the borrower to inspect the business), then the effect of distance on interest rates depends on the pricing regime. Under imperfect competition, close banks enjoy an information-based monopoly that allows them to extract rents from borrowers (Agarwal and Hauswald (2010)). Reducing bank-borrower driving time will lead to higher bank rents due to reduced monitoring-based transportation costs, but not higher observed interest rates. Alternatively, under perfect competition, banks price loans according to the marginal costs of monitoring borrowers. In this case, interest rates should decrease as distance decreases since the cost of monitoring decreases (monitoring hypothesis). Which of these channels dominates is ultimately an empirical question. The relationship between distance and interest rates is multifaceted and difficult to study because other factors can also affect interest rates, such as local market power, borrower characteristics, and the overall state of the economy.

Directly testing these two channels is difficult due to the endogenous matching between borrowers and banks. Since these matching criteria are unobserved, they can bias the cross-sectional association between distance and the cost of borrowing. The existing literature finds cross-sectional evidence both consistent with the local market power hypothesis (Petersen and Rajan (2002); Degryse and Ongena (2005)) and the monitoring hypothesis (Knyazeva and Knyazeva (2012); Bellucci et al. (2013)). In the setting of Agarwal and Hauswald (2010), lower distances increase interest rates and are associated with larger loan sizes. Large banks, for example, lend to more distant and bank-dependent firms (Berger et al. (2005), Schweitzer and Barkley (2017)). The distance itself is a matching criterion, and bank-borrower distance depends on numerous bank and firm characteristics (Beck et al. (2019)). Banks also strategically choose the location of branches (Kim and Vale (2001)), introducing another selection link between distance and interest rates.

Endogenous relationships between the borrower-lender distance, interest rate, and local economic conditions are challenging. To overcome these issues, I will use the instrumental variable correlated with the independent variable of interest (distance) but not with the dependent variable (labor market status). One promising instrumental variable is the exposure of individuals to the merger-induced consolidation of large banks. This is because bank mergers can change the geographical proximity between banks and borrowers, but large banks' decision to merge is made at the national level and exogenous to local economic conditions.

As in Nguyen (2019) and Garmaise and Moskowitz (2006), I use the exposure of individuals to merger-induced consolidation of large banks as an instrumental variable for changes in the borrower-lender distance. I consider mergers of large banks under the condition that both banks held at least \$1 billion in pre-merger assets. The instrumental variable equals one if the individual lives in the zip code where two large banks had branches, and these banks went through the merge-induced consolidation at the national level. The area of the zip code is determined by population, and the average size is 82.25 square miles (9 by 9 miles square). The exposure of individuals to merger-induced consolidation (MIC) serves as an exogenous shock to borrower-lender distances. These consolidations are typically driven by strategic, financial, or regulatory considerations that are independent of individual borrower characteristics or behaviors, ensuring the exogeneity of the instrument. This makes it a valid instrument, addressing the endogeneity concerns stemming from unobserved heterogeneity and omitted variable bias in the relationship between borrower-lender distance and the dependent variable. The merge-induced consolidation decreases bank competition in the geographical area, which may be followed by a branch closure (Nguyen, 2019) and an increase in the borrower-lender distance.

The main idea in the empirical analysis is to exploit variations in the borrower-lender distance (Distance) in the propensity to become self-employed (Y_{ijst}). The basic assumption would be that in areas where the borrower-lender distance is shorter, more individuals might choose to become self-employed, whereas in areas where credit market institutions are less available (or become less available), more individuals might choose salaried employment. In the first stage (Equation 7), I estimate the relationship between the borrower-lender distance[18] and the binary indicator of exposure to merger-induced consolidation of large banks at the zip-code level, average commute time to work, number of bank workers per capita[19], other individual characteristics, state-by-year fixed effects, and geographical clusters fixed effects. The predicted value of the distance from the first stage is used in the second stage (Equation 8), where I estimate the linear probability model of switching out of self-employment.

$$\text{Distance}_{ijskt} = \beta_1 \text{MIC}_{jt} + \beta_2 B_{ijt} + \beta_3 X_{ikt} + \beta_4 I_{it-1} + \beta_5 Z_{jt} + \epsilon_{st} + \tau_k + \varepsilon_{ijskt} \tag{7}$$

$$Y_{ijskt} = \alpha_1 \widehat{\text{Distance}}_{ijkst} + \alpha_2 X_{ikt} + \alpha_3 I_{it-1} + \alpha_4 Z_{jt} + \epsilon_{st} + \tau_k + \varepsilon_{ijskt} \tag{8}$$

The dependent variable Y_{ijskt} measures entrepreneurs' exit (or entry) at the individual level. Y_{ijst} is a dummy variable, which equals one if individual i in zip code j, spatial cluster k, and state s was an entrepreneur in the year $(t-1)$ and becomes non-entrepreneur in the year t and zero otherwise (switch to entrepreneurship in the year t). The Distance_{it} is the geographical

[18] In the main specification, I use the minimum distance between borrower and lender, but for the robustness check, I also estimate it for the average distance between the borrower and five nearest banks, ten nearest banks, fifteen, twenty and twenty-five.

[19] The idea is larger number of bank workers is associated with the higher availability of loan officers.

distance between the respondent's home and the nearest bank.[20] The model includes a full set of state-by-year fixed effects, ϵ_{st}, the error term, ε_{ijskt}, and zip code characteristics Z_{jt}. The state-by-year fixed effects mean that comparisons between the groups of zip codes are made within the state for each year. For ease of interpretation, the p_{jt} are standardized to have a mean of zero and a standard deviation of one. The variable I_{it-1} includes the previous period's labor status (salaried employment, incorporated and unincorporated self-employment, unemployment). The variable X_{it} is the set of observed covariates, including dummies for gender, marital status, three race dummies (white, black, other), Hispanic ethnicity, four education dummies (less than high school, high school, some college, college and above), age and a quadratic for age. τ_k identifies geographical clusters to control for unobserved spatial characteristics. Z_{jt} includes local economic conditions.[21] All regressions are weighted by the CAPS individual-level weights, and standard errors are bootstrapped.

The key coefficient of interest is α_1, which is conditional on MIC being a valid instrument that will yield a causal estimate of the effect of a one-percentage-point increase in borrower-lender distance on the probability of switching out of self-employment. The model compares whether the labor decision between exposed and unexposed areas depends on the borrower-lender distance. Higher borrower-lender distance increases transportation or informational costs, making lending more expensive. Following this, a negative estimate of α_1 suggests that severe credit supply shock disincentives individuals to stay in entrepreneurship more for closer borrowers.

Table 12 show results of Equation 8. It's the second stage results of the procedure. In the first stage, I used the instrumental variable that equals one if the individual lives in the zip code where two large banks had branches, and these banks went through the merge-induced consolidation at the national level. This instrument is used for endogeneity correction of the minimum distance between the borrower and lender[22]. Results indicate that the increase in the borrower-lender distance by 1% increases the chances to leave self-employment by 5.1% and by 3.8% to leave the incorporated self-employment. Results indicate that the changes in the distance do not affect unincorporated businesses. If one considers other distance measures (see Table 10). the increasing number of lending options for the borrower leads to a decrease in the

[20]As a robustness check, I also used the average distance to the five nearest banks, ten nearest banks, fifteen nearest banks, and twenty nearest banks. Results are in Table 10

[21]The number of business establishments, the unemployment rate, the average value of the house, the share of the population with professional, scientific, management, and administrative education at the zip-code level.

[22]As a robustness check, I also used the average distance to the five nearest banks, ten nearest banks, fifteen nearest banks, and twenty nearest banks. Results are in Table 10

switching probability. For example, the increase in the average borrower-lender distance to the ten nearest banks by 1% increases the chances of leaving self-employment by 1.2% and by 0.8% to leave the incorporated self-employment.

Figure 13 shows the estimated predicted probability of switching out of self-employment depending on the change in the minimum distance between the borrower and the nearest lender (Panel a) and the percentage increase in the borrower-lender distance (Panel b). Results indicate that changes in the borrower-lender distance significantly affect the decision to switch out of self-employment within seven miles (or less than a 45% increase in the distance) for both types of self-employed. These results are similar to findings by Nguyen (2019).

8 The Heterogeneous Dynamic Effect of Supply Shocks on Self-employment

This section examines the heterogeneity observed in the results from estimating Equation 5, exploring these effects in different dimensions. I first investigate how the credit supply shock affects self-employment based on individuals' wealth. The literature (Fan and White (2003), Fairlie and Krashinsky (2012), Fort et al. (2013), Corradin and Popov (2015)) highlights the importance of wealth in opening and sustaining a business. I divide individual wealth into quartiles and compare the effect of supply shocks among the wealthiest (above the 75th percentile) and the least wealthy (below the 25th percentile) individuals. The wealthiest incorporated self-employed individuals show an increased risk of exiting in the year of the most recent branch closure (Figure 14, panel (b)). In contrast, the poorest unincorporated self-employed individuals show a decreased risk of exiting (panel (b)). These results confirm evidence from Levine and Rubinstein (2017) that the selection into incorporated self-employment is based on the amount of collateral (wealth) and access to capital.

I then examine gender differences in the responses to credit supply shocks, comparing the results of Equation 5 for men and women. In the year of a branch closure, male self-employed have an increased risk of exiting (Figure 15, panel (a)). For incorporated self-employment, the risk of exiting self-employment is increased for both genders, but it is higher for males (panel (b)). This evidence supports the previous literature (Cowling et al., 2020) showing a lower demand for bank loans among women and a smaller reaction to credit supply shocks. There are no gender differences in the response of unincorporated self-employed to credit supply shocks.

I next investigate the educational differences in responses to credit supply shocks, comparing the results of Equation 5 for high school graduates, college dropouts, and college graduates. Only college graduates respond to the supply shocks in the year of a branch closure. When the supply shock increases by one standard deviation, the risk of leaving their incorporated businesses increases up to 6 percent (Figure 16, panel (b)). There are no educational differences in the response of unincorporated self-employed to credit supply shocks.

Last, I check the differences in the supply shock impact on native- and foreign-born self-employed. Incorporated self-employed individuals born in the U.S. have an increased risk of leaving self-employment compared to those who are born outside the country (Figure 17, panel (b)). This evidence is supported by the literature (Bruder et al., 2011) investigating immigrants' accessibility to the credit market.

9 Conclusion

This paper explores the impact of the contraction in the U.S. bank branch network and the subsequent limited access to credit institutions on the decline in U.S. self-employment. Using a quasi-experimental research design, I estimate the causal impact of bank branch closures on exits from self-employment. The empirical challenge lies in separating bank-specific supply shocks from the local decline in loan demand due to reduced entrepreneurial activities in the area. I isolate the portion of these closures due to supply-side factors by creating a measure of bank availability that interacts with estimated bank-specific supply shocks with their market share at the smallest available geographical area (zip code area).

I find that the contraction of the bank branch network negatively affects the probability of obtaining a loan and has implications on existing loan contracts. A one standard deviation increase in the predicted supply of credit corresponds to a 1.2 percent increase in the probability of being denied during a new loan application and a 7.2 percent increase in the risk of filing for bankruptcy.

The supply shock has a heterogeneous effect on borrowers. Incorporated self-employed individuals who own growth-oriented businesses requiring physical capital and business loans face a significantly high risk of exit from self-employment. A one standard deviation increase in the predicted credit supply results in a 2.6 percent increase in the likelihood of exiting incorporated self-employment. This effect is higher (3.9 percent) in geographical areas experiencing more severe supply shocks. However, the adverse effects of the supply shocks are very localized, dissipating within seven miles, and have short-term consequences for self-employment (disappearing within two years). In contrast, unincorporated self-employed individuals who own less capital-intensive businesses do not experience significant changes due to the supply shocks.

I also compare the impact of the supply shocks on exiting self-employment based on different individual characteristics, such as wealth, gender, education, and immigration status. The results are largely consistent with the previous literature. Wealthier, more educated, and native-born incorporated self-employed individuals face an increased risk of exiting self-employment due to supply shocks. Incorporated self-employed individuals of both genders experience an increased likelihood of exiting, but the magnitude is higher for men.

References

R. M. Adams, K. P. Brevoort, and J. C. Driscoll. Is lending distance really changing? distance dynamics and loan composition in small business lending. Distance Dynamics and Loan Composition in Small Business Lending (September 23, 2021), 2021.

S. Agarwal and R. Hauswald. Distance and private information in lending. The Review of Financial Studies, 23(7):2757–2788, 2010.

A. C. Baker, D. F. Larcker, and C. C. Wang. How much should we trust staggered difference-in-differences estimates? Journal of Financial Economics, 144(2):370–395, 2022.

T. J. Bartik and J. Bartik Timothy. Who benefits from state and local economic development policies? 1991.

T. Bates and A. Robb. Greater access to capital is needed to unleash the local economic development potential of minority-owned businesses. Economic Development Quarterly, 27 (3):250–259, 2013.

T. Beck, S. Ongena, and İ. Şendeniz-Yüncü. Keep walking? geographical proximity, religion, and relationship banking. Journal of Corporate Finance, 55:49–68, 2019.

J. Beckhusen. Employment transitions among the self-employed during the great recession. Technical report, 2014.

A. Bellucci, A. Borisov, and A. Zazzaro. Do banks price discriminate spatially? evidence from small business lending in local credit markets. Journal of Banking & Finance, 37(11): 4183–4197, 2013.

A. N. Berger and G. F. Udell. The economics of small business finance: The roles of private equity and debt markets in the financial growth cycle. Journal of banking & finance, 22(6-8): 613–673, 1998.

A. N. Berger, N. H. Miller, M. A. Petersen, R. G. Rajan, and J. C. Stein. Does function follow organizational form? evidence from the lending practices of large and small banks. Journal of Financial economics, 76(2):237–269, 2005.

D. G. Blanchflower and A. Oswald. Self-employment and the enterprise culture, in r. jowell, s.witherspoon and l. brook (eds.), british social attitudes: The 1990 report, london, gower (1998).

J. Block and P. Sandner. Necessity and opportunity entrepreneurs and their duration in self-employment: evidence from german micro data. Journal of Industry, Competition and Trade, 9(2):117–137, 2009.

A. W. Boot. Relationship banking: What do we know? journal of financial intermediation. 2000.

C. Bopaiah. Availability of credit to family businesses. Small Business Economics, 11(1):75–86, 1998.

F. P. Boscoe, K. A. Henry, and M. S. Zdeb. A nationwide comparison of driving distance versus straight-line distance to hospitals. The Professional Geographer, 64(2):188–196, 2012.

W. J. Boyes, D. L. Hoffman, and S. A. Low. An econometric analysis of the bank credit scoring problem. Journal of Econometrics, 40(1):3–14, 1989.

A. Bracha and M. A. Burke. Who counts as employed?: informal work, employment status, and labor market slack. 2016.

A. Bracha, M. A. Burke, and A. Khachiyan. Changing patterns in informal work participation in the united states 2013-2015. Federal Reserve Bank of Boston Research Paper Series Current Policy Perspectives Paper, (15-10), 2015.

K. P. Brevoort, J. Wolken, and J. Holmes. Distance still matters: the information revolution in small business lending and the persistent role of location, 1993-2003. 2010.

J. Bruder, D. Neuberger, and S. Räthke-Döppner. Financial constraints of ethnic entrepreneurship: evidence from germany. International Journal of Entrepreneurial Behavior & Research, 2011.

M. Cagetti and M. De Nardi. Entrepreneurship, frictions, and wealth. Journal of political Economy, 114(5):835–870, 2006.

K. Carling and S. Lundberg. Asymmetric information and distance: an empirical assessment of geographical credit rationing. Journal of Economics and Business, 57(1):39–59, 2005.

S. Corradin and A. Popov. House prices, home equity borrowing, and entrepreneurship. The Review of Financial Studies, 28(8):2399–2428, 2015.

M. Cowling, S. Marlow, and W. Liu. Gender and bank lending after the global financial crisis: are women entrepreneurs safer bets? Small Business Economics, 55(4):853–880, 2020.

M. L. Cowling and M. Wooden. Does solo self-employment serve as a ‘stepping stone’to employership? Labour Economics, 68:101942, 2021.

A. Deaton. Instruments, randomization, and learning about development. Journal of economic literature, 48(2):424–55, 2010.

H. Degryse and S. Ongena. Distance, lending relationships, and competition. The Journal of Finance, 60(1):231–266, 2005.

G. Dell’Ariccia and R. Marquez. Information and bank credit allocation. Journal of Financial Economics, 72(1):185–214, 2004.

R. DeYoung, D. Glennon, and P. Nigro. Borrower–lender distance, credit scoring, and loan performance: Evidence from informational-opaque small business borrowers. Journal of Financial Intermediation, 17(1):113–143, 2008.

R. DeYoung, W. S. Frame, D. Glennon, and P. Nigro. The information revolution and small business lending: The missing evidence. Journal of Financial Services Research, 39:19–33, 2011.

P. Diegert, M. A. Masten, and A. Poirier. Assessing omitted variable bias when the controls are endogenous. arXiv preprint arXiv:2206.02303, 2022.

D. S. Evans and B. Jovanovic. An estimated model of entrepreneurial choice under liquidity constraints. Journal of political economy, 97(4):808–827, 1989.

R. W. Fairlie and H. A. Krashinsky. Liquidity constraints, household wealth, and entrepreneurship revisited. Review of Income and Wealth, 58(2):279–306, 2012.

W. Fan and M. J. White. Personal bankruptcy and the level of entrepreneurial activity. The Journal of Law and Economics, 46(2):543–567, 2003.

K. D. Fee and E. Tiersten-Nyman. Has bank consolidation changed people’s access to a full-service bank branch? Community Development Reports, (20220713), 2022.

T. C. Fort, J. Haltiwanger, R. S. Jarmin, and J. Miranda. How firms respond to business cycles: The role of firm age and firm size. IMF Economic Review, 61(3):520–559, 2013.

C. Fracassi, M. Garmaise, S. Kogan, and G. Natividad. How much does credit matter for entrepreneurial success in the united states? Available at SSRN, 2013.

X. Freixas and J.-C. Rochet. Microeconomics of banking. MIT press, 2008.

M. J. Garmaise and T. J. Moskowitz. Bank mergers and crime: The real and social effects of credit market competition. the Journal of Finance, 61(2):495–538, 2006.

W. M. Gentry and R. G. Hubbard. Entrepreneurship and household saving. Advances in economic analysis & policy, 4(1), 2004.

P. Goldsmith-Pinkham, I. Sorkin, and H. Swift. Bartik instruments: What, when, why, and how. American Economic Review, 110(8):2586–2624, 2020.

M. Greenstone, A. Mas, and H.-L. Nguyen. Do credit market shocks affect the real economy? quasi-experimental evidence from the great recession and 'normal'economic times. Technical report, National Bureau of Economic Research, 2014.

M. Greenstone, A. Mas, and H.-L. Nguyen. Do credit market shocks affect the real economy? quasi-experimental evidence from the great recession and" normal" economic times. American Economic Journal: Economic Policy, 12(1):200–225, 2020.

L. Guiso, P. Sapienza, and L. Zingales. Does local financial development matter? The Quarterly Journal of Economics, 119(3):929–969, 2004.

R. Hauswald and R. Marquez. Competition and strategic information acquisition in credit markets. The Review of Financial Studies, 19(3):967–1000, 2006.

J. Heckman and B. Singer. A method for minimizing the impact of distributional assumptions in econometric models for duration data. Econometrica: Journal of the Econometric Society, pages 271–320, 1984.

J. J. Heckman. The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process and some Monte Carlo evidence. University of Chicago Center for Mathematical studies in Business and Economics, 1987.

K. Herkenhoff, G. Phillips, and E. Cohen-Cole. The impact of consumer credit access on employment, earnings and entrepreneurship. Technical report, National Bureau of Economic Research, 2016.

C. Herpfer, A. Mjøs, and C. Schmidt. The causal impact of distance on bank lending. Management Science, 69(2):723–740, 2023.

S. F. Hipple. Self-employment in the united states. Monthly Labor Review, 133(9):17–32, 2016.

D. Holtz-Eakin, D. Joulfaian, and H. S. Rosen. Sticking it out: Entrepreneurial survival and liquidity constraints. Journal of Political economy, 102(1):53–75, 1994.

L. Jensen, A. R. Tickamyer, and T. Slack. Rural-urban variation in informal work activities in the united states. Journal of Rural Studies, 2019.

G. Jiménez, V. Salas, and J. Saurina. Organizational distance and use of collateral for business loans. Journal of Banking & Finance, 33(2):234–243, 2009.

S. Johnson. Are mortgage regulations affecting entrepreneurship?, 2016.

M. Kim and B. Vale. Non-price strategic behavior: the case of bank branches. International Journal of Industrial Organization, 19(10):1583–1602, 2001.

A. Knyazeva and D. Knyazeva. Does being your bank's neighbor matter? Journal of Banking & Finance, 36(4):1194–1209, 2012.

D. S. Lechmann and C. Wunder. The dynamics of solo self-employment: Persistence and transition to employership. Labour Economics, 49:95–105, 2017.

P. J. Lederer and A. P. Hurter Jr. Competition of firms: discriminatory pricing and location. Econometrica: Journal of the Econometric Society, pages 623–640, 1986.

R. Levine and Y. Rubinstein. Smart and illicit: who becomes an entrepreneur and do they earn more? The Quarterly Journal of Economics, 132(2):963–1018, 2017.

R. Levine and Y. Rubinstein. Selection into entrepreneurship and self-employment. Technical report, National Bureau of Economic Research, 2018.

T. Lindh and H. Ohlsson. Self-employment and windfall gains: evidence from the swedish lottery. The Economic Journal, 106(439):1515–1526, 1996.

A. Malkova and K. Sabirianova Peter. Labor informality and credit market accessibility. Available at SSRN 4404814.

K. R. Manturuk, M. R. Lindblad, and R. G. Quercia. A place called home: The social dimensions of homeownership. Oxford University Press, 2017.

C. Milani. Borrower–lender distance and loan default rates: Macro evidence from the italian local markets. Journal of Economics and Business, 71:1–21, 2014.

I. Murtazashvili and J. M. Wooldridge. A control function approach to estimating switching regression models with endogenous explanatory variables and endogenous switching. Journal of Econometrics, 190(2):252–266, 2016.

H.-L. Q. Nguyen. Are credit markets still local? evidence from bank branch closings. American Economic Journal: Applied Economics, 11(1):1–32, 2019.

D. S. Nightingale. Informal and nonstandard employment in the united states: implications for low-income working families. 2011.

A. Ono, Y. Saito, K. Sakai, and I. Uesugi. Does geographical proximity matter in small business lending? evidence from changes in main bank relationships. 2016.

E. Oster. Unobservable selection and coefficient stability: Theory and evidence. Journal of Business & Economic Statistics, 37(2):187–204, 2019.

L. E. Papke and J. M. Wooldridge. Panel data methods for fractional response variables with an application to test pass rates. Journal of Econometrics, 145(1-2):121–133, 2008.

P. M. Parente and J. S. Silva. A cautionary note on tests of overidentifying restrictions. Economics Letters, 115(2):314–317, 2012.

S. C. Parker. The economics of entrepreneurship. Cambridge University Press, 2018.

M. A. Petersen and R. G. Rajan. The benefits of lending relationships: Evidence from small business data. The journal of finance, 49(1):3–37, 1994.

M. A. Petersen and R. G. Rajan. Does distance still matter? the information revolution in small business lending. The journal of Finance, 57(6):2533–2570, 2002.

V. Prowse. Modeling employment dynamics with state dependence and unobserved heterogeneity. Journal of Business & Economic Statistics, 30(3):411–431, 2012.

J. Qian and P. E. Strahan. How laws and institutions shape financial contracts: The case of bank loans. The Journal of Finance, 62(6):2803–2834, 2007.

R. G. Quercia, A. Freeman, and J. Ratcliffe. Regaining the dream: How to renew the promise of homeownership for America's working families. Brookings Institution Press, 2011.

S. Rabe-Hesketh and A. Skrondal. Avoiding biased versions of wooldridge's simple solution to the initial conditions problem. Economics Letters, 120(2):346–349, 2013.

S. C. Salop. Monopolistic competition with outside goods. The Bell Journal of Economics, pages 141–156, 1979.

A. Sarker and B. Unel. The impact of bank expansion on self-employed business owners: Evidence from us states. Technical report, 2017.

M. Schweitzer and S. Shane. The ins and outs of entrepreneurship: Explaining the decline in self-employment. Federal Reserve Bank of Cleveland, 2015.

M. E. Schweitzer and B. Barkley. Is' fintech'good for small business borrowers? impacts on firm growth and customer satisfaction. 2017.

M. E. Schweitzer and S. Shane. The ins and outs of self-employment: An estimate of business cycle and trend effects. 2016.

R. S. Sobel and A. M. Dean. Has wal-mart buried mom and pop?: The impact of wal-mart on self-employment and small establishments in the united states. Economic Inquiry, 46(4): 676–695, 2008.

A. Sufi. Information asymmetry and financing arrangements: Evidence from syndicated loans. The Journal of Finance, 62(2):629–668, 2007.

B. Unel. Effects of us banking deregulation on unemployment dynamics. Journal of Macroeconomics, 65:103237, 2020.

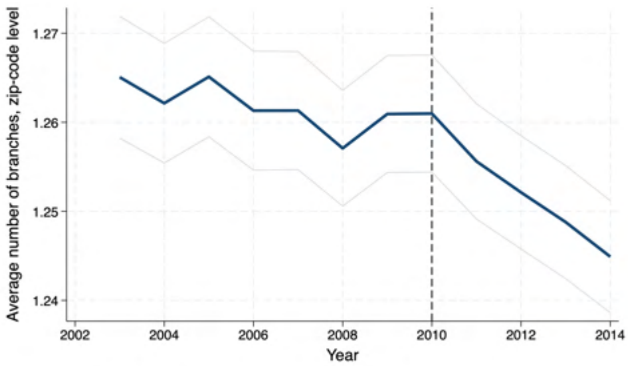
A. M. Wiersch and S. Shane. Why small business lending isn't what it used to be. Economic Commentary, (2013-10), 2013.

J. M. Wooldridge. Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. Journal of applied econometrics, 20(1): 39–54, 2005.

J. M. Wooldridge. Introductory econometrics: A modern approach. Nelson Education, 2016.

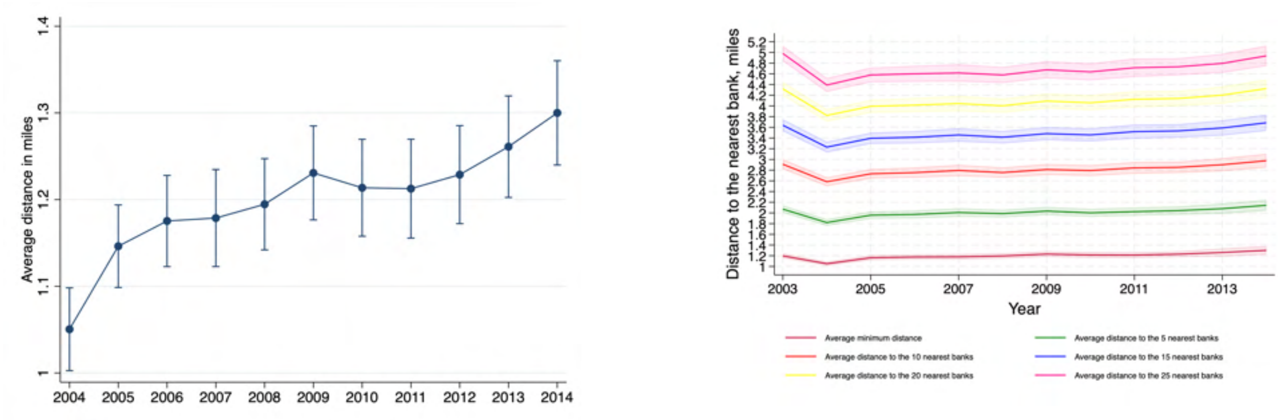
10 Figures

Figure 1: Average Number of Branches at the Zip-Code Level



Notes. This figure shows the dynamic of the average number of branches over the years; the interval ranges show 95 percent confidence intervals.

Figure 2: Average Distance to the Nearest Bank Branch



a) b)

Notes. The figure shows the average minimum distance between a borrower and a lender, estimated from the CAPS database and the FDIC's SOD database (miles). For each yearly observation on the graph (the dots), the interval ranges show 95 percent confidence intervals.

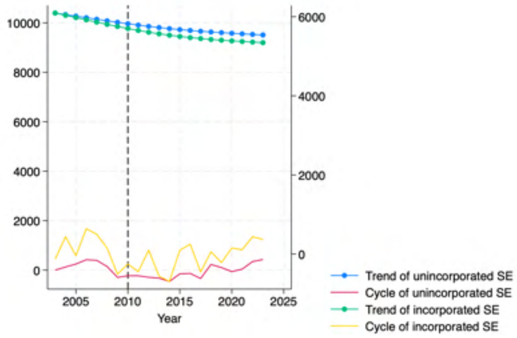
Figure 3: Dynamic of Self-Employment in the U.S.

a)

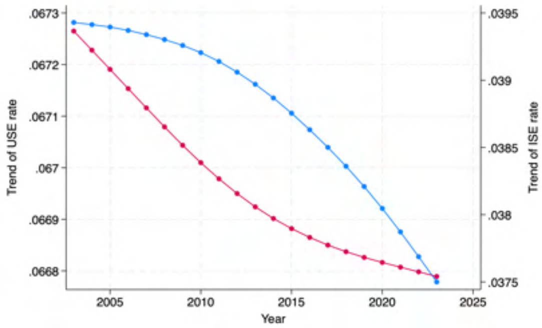
b)

Notes. Panel (a) shows the number of self-employed unincorporated (the left y-axis) and incorporated (the right y-axis) individuals from the Current Population Survey (thousands, seasonally adjusted). Panel (b) shows the self-employment rate as a proportion of the total employment made up of unincorporated (the left y-axis) and incorporated (the right y-axis) self-employed workers from the CPS (thousands, seasonally adjusted).

Figure 4: De-trended Dynamic of Self-Employment the U.S.



a)



b)

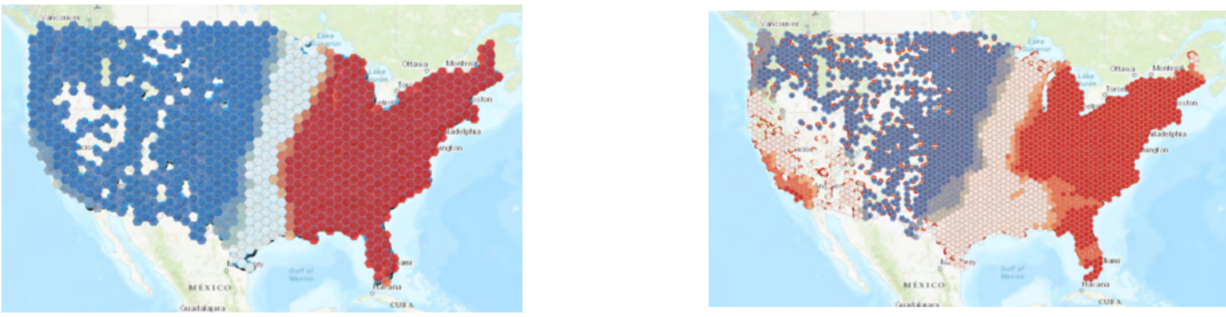
Notes. Both panels show the de-trended dynamics of self-employment using the Hodrick-Prescott filter. On the panel (a), the total number of self-employed individuals (y_t) (incorporated and unincorporated) was separated into a deterministic time trend (τ_t) and cyclical components (c_t), such that $y_t = \tau_t + c_t + \epsilon_t$, where ϵ_t is the error component. The Hodrick-Prescott filter is implemented to identify the time trend and the cyclical component. It's the optimization problem's solution that minimizes the cyclical component's variance subject to a smoothness penalty on the trend component. $\arg\min_{\tau_t} \sum_{t=1}^{T} \left((y_t - \tau_t)^2 + \lambda \sum_{t=2}^{T-1} [(\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1})]^2 \right)$, where λ is the smoothing parameter, controlling the trade-off between fitting the data closely and the smoothness of the trend component. For the annual data, λ is set to 6.25. Panel (a) shows separately the trend components and the cycle components for unincorporated (the left y-axis) and incorporated (the right y-axis) self-employed. Panel (b) shows only the trend component of the self-employment rate for unincorporated (blue dots, the left y-axis) and incorporated (red dots, the right y-axis) individuals.

Figure 5: Location of All Banks in the U.S.

a) 2003 b) 2014

Notes. The figures show the location of all banks and branches from the FDIC's SOD database (red dots). In 2003, there were 87,279 bank locations listed in the FDIC's SOD database; in 2014, there were 94,521 bank locations listed in it.

Figure 6: Location of All Banks in the U.S. and Hot Spot Analysis



a) 2003 b) 2014

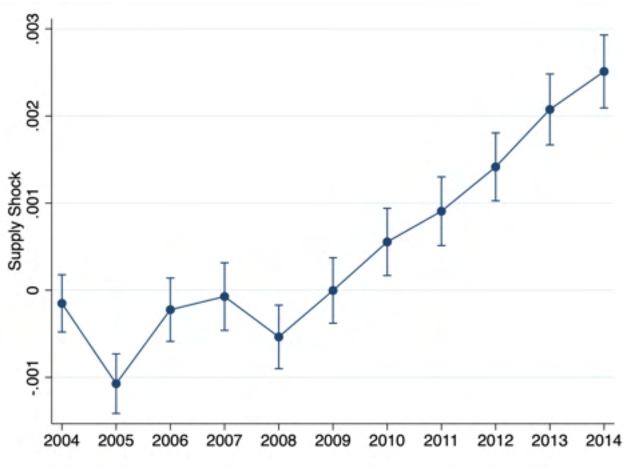
Notes. The figures show the location of all banks and branches from the FDIC's SOD database (red dots) and the hot spot analysis created in ArcGis (blue areas). In 2003, there were 87,279 bank locations listed in the FDIC's SOD database; in 2014, there were 94,521 bank locations listed in it. The hot spot analysis shows the z-score for the Getis-Ord Gi* statistic for each location in a dataset. Red areas tell where bank locations with high values cluster spatially; these areas are hot spots with 99 percent confidence, and white areas are not significant. Blue areas are cold spots with 99 percent confidence.

The z-score for the Getis-Ord Gi* statistic was calculated using ArcGIS tools. The formula for the z-score is

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j} x_j - \frac{\sum_{j=1}^{n} x_j}{n} \sum_{j=1}^{n} w_{i,j}}{\sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - \frac{\sum_{j=1}^{n} x_j}{n}} \sqrt{\frac{n \sum_{j=1}^{n} w_{i,j}^2 - \left(\sum_{j=1}^{n} w_{i,j}\right)^2}{n-1}}},$$

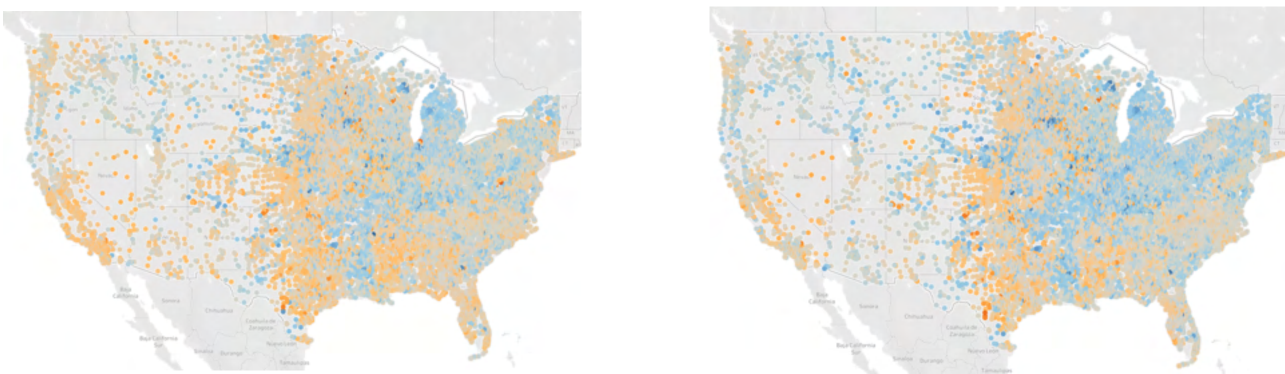
where x_j is the attribute value of feature j, $w_{i,j}$ is the spatial weight between features i and j, and n is equal to the total numbers of features. To be a statistically significant hot spot, a feature should have a high value and be surrounded by other features with high values as well.

Figure 7: Dynamic of Average Supply Shocks



Notes. This figure shows the dynamic of average supply shocks over the years; the interval ranges show 95 percent confidence intervals.

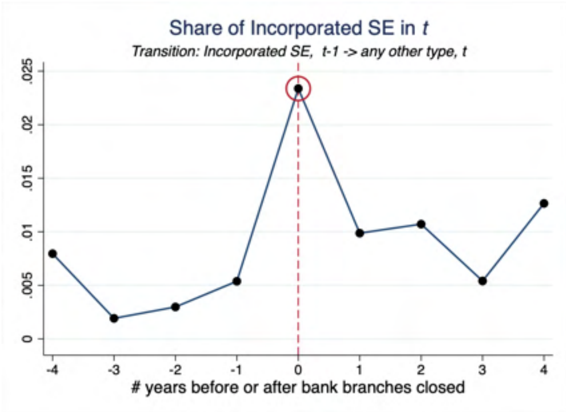
Figure 8: Branch Closure Intensity: Visualizing the Density of Closures Across Zip-Codes



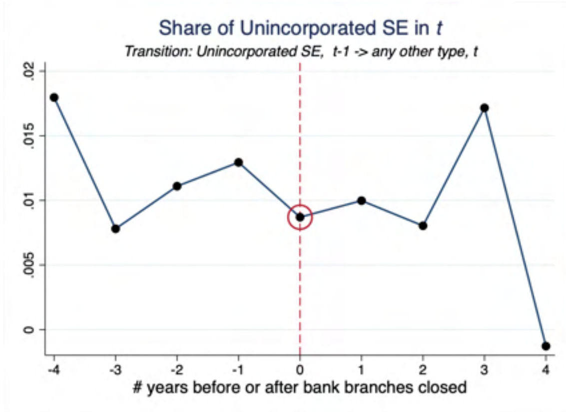
a) b)

Notes. The figures show the intensity of supply shocks in 2004 (a) and 2014 (b). Blue dots reflect more severe supply shock (more branches closed), whereas orange dots show less intensive supply shock (fewer branches closed).

Figure 9: Event-Study Analysis



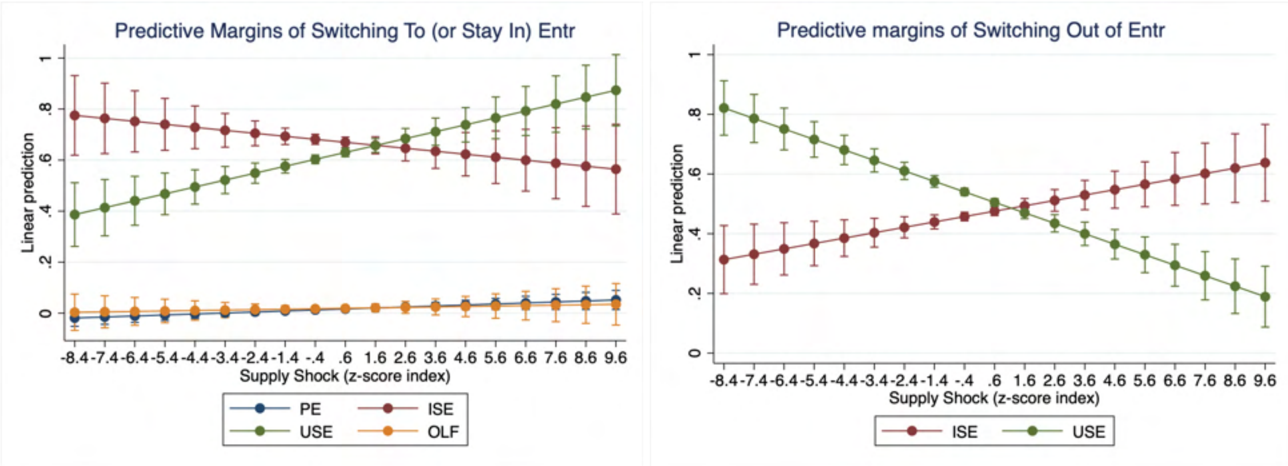
a)



b)

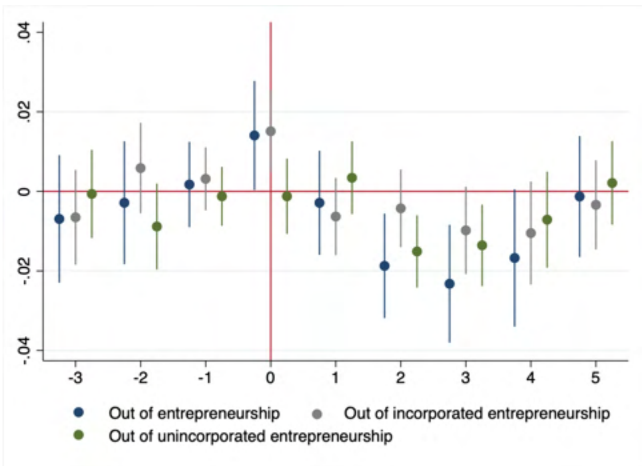
Notes. Each dot pertains to the predicted probability of the transition from/to self-employment based on an event-study analysis. Panel (a) depicts the predicted exit rate from incorporated self-employed to any other type of employment as a share of incorporated self-employed in t. Panel (b) plots the predicted exit rate from unincorporated self-employed to any other type of employment as a share of unincorporated self-employed in t. Event time is defined as the year when the closure of a bank branch occurred at the zip code level, and there was no closure of branches two years before. The probability model is linear and controls for each year before and after the event (the categorical timeline), a quadratic polynomial of current age, and calendar year fixed effects. The plotted dots are the coefficients on the timeline variable plus the average predicted share evaluated at timeline = 0 (which is the omitted category). Only four points before and after the event are shown.

Figure 10: Self-employment and Supply Shocks



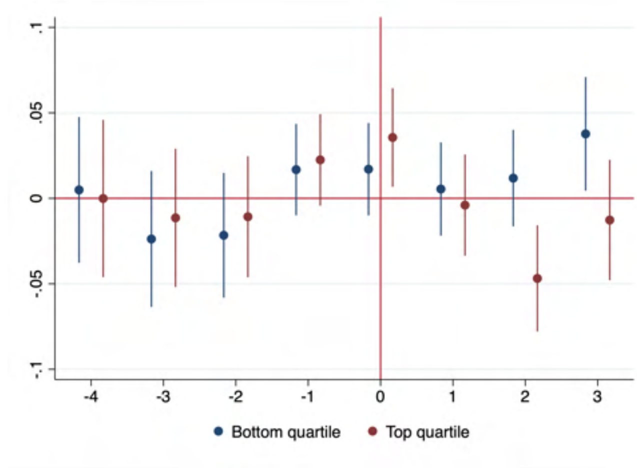
Notes. Panel (a) shows the predictive probability of switching to (or staying in) self-employment from Equation 4. The red line reports predictive probabilities for the incorporated self-employed, the green line for the unincorporated self-employed, the blue line for wage employees, and the orange line for individuals who are out of the labor force. Panel (b) shows the predictive probability of switching from self-employment from Equation 4. The red line reports predictive probabilities for the incorporated self-employed and the green line for the unincorporated self-employed.

Figure 11: Marginal Effect of Supply Shocks on the Predicted Probability of Exiting Entrepreneurship

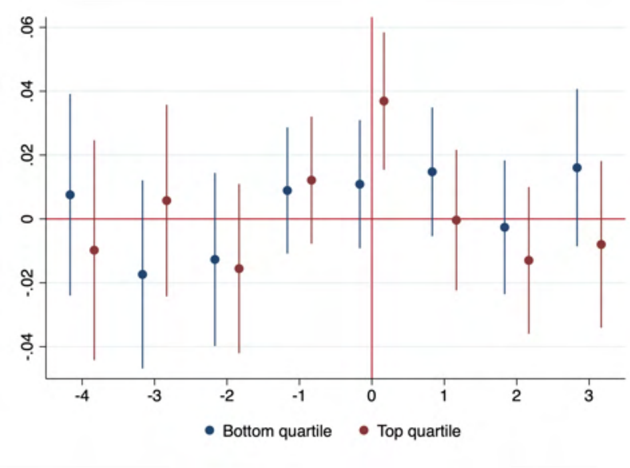


Notes. This figure shows the marginal effect of supply shocks on the predicted probability of exiting entrepreneurship, obtained from estimating Equation 5. Bars show 95 percent confidence intervals, and $\tau = 0$ is the year of the most recent closure of bank branches. All regressions are weighted by CAPS individual-level weights, and standard errors are bootstrapped.

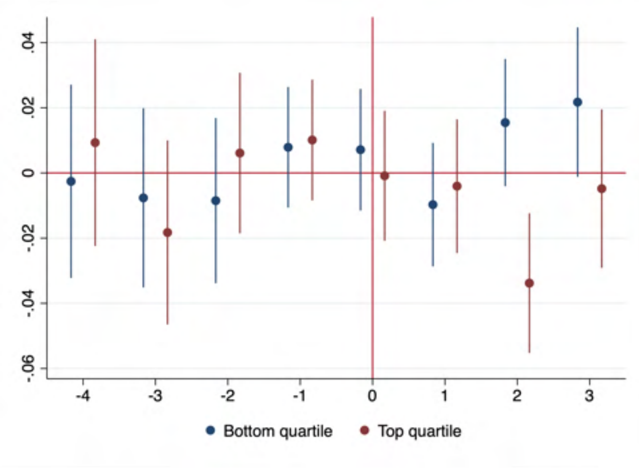
Figure 12: Marginal Effect of Supply Shock Quartiles on the Predicted Probability of Exiting Self-employment



a)



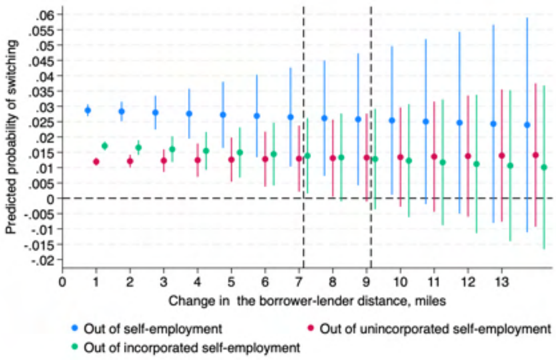
b)



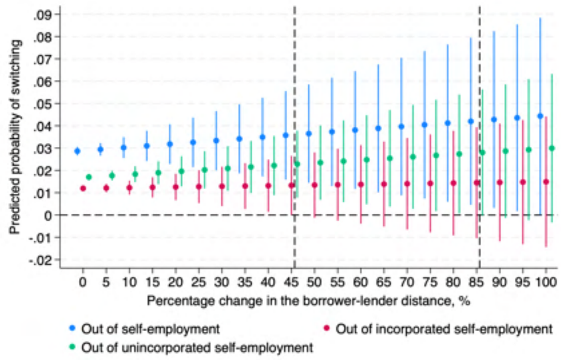
c)

Notes. This figure shows the marginal effect of supply shocks on the predicted probability of exiting self-employment, obtained from estimating Equation 6. Bars show 95 percent confidence intervals, and $\tau = 0$ is the year of the most recent closure of bank branches. Panel (a) shows the results of Equation 6 when the dependent variable is exiting self-employment (combining incorporated and unincorporated businesses). Panel (b) shows the results of Equation 6 when the dependent variable is exiting incorporated self-employment. Panel (c) shows the results of Equation 6 when the dependent variable is exiting unincorporated self-employment. All regressions are weighted by CAPS individual-level weights, and standard errors are bootstrapped.

Figure 13: The Marginal Effect of Branch Closure on the Predicted Probability of Switching



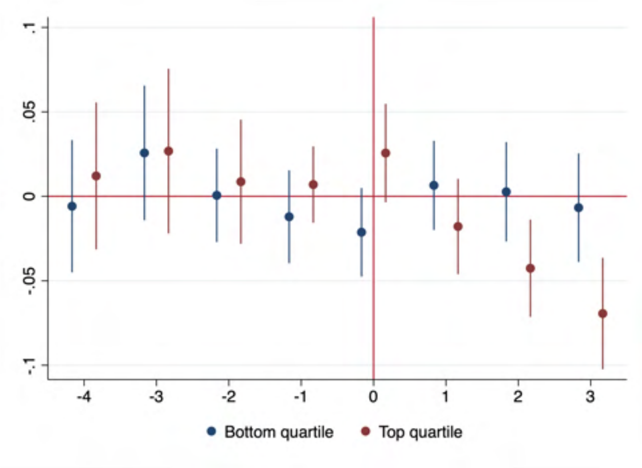
a)



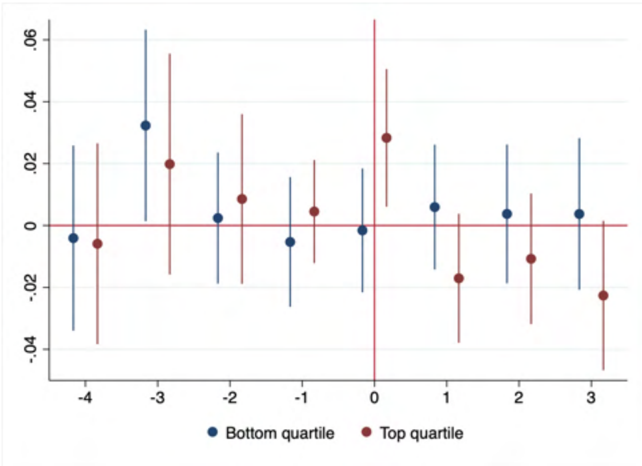
b)

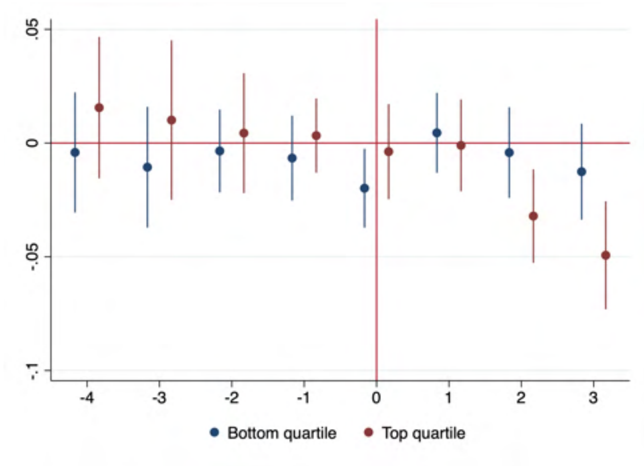
Notes. The figure shows the estimated predicted probability of switching out of self-employment depending on the minimum distance between the borrower and the nearest lender (Panel a) and the percentage increase in the borrower-lender distance (Panel b). For each estimated probability on the graph (the dot), the interval ranges show a 95% confidence interval.

Figure 14: Marginal Effect of Supply Shocks on the Predicted Probability of Exiting Self-employment, by Wealth



a)



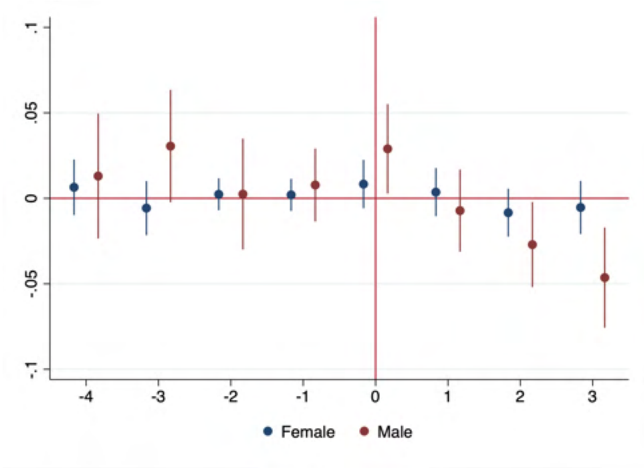


b)

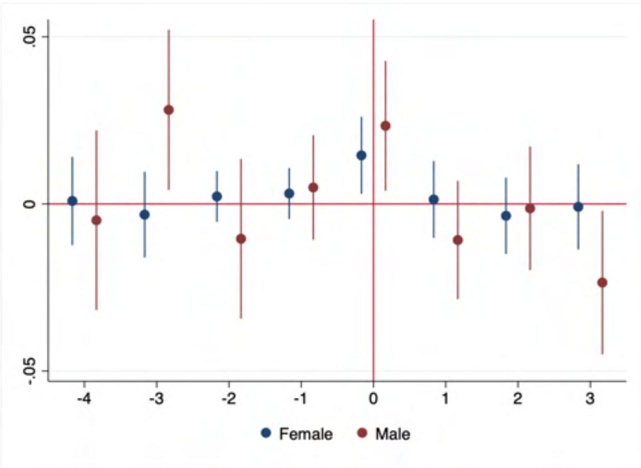
c)

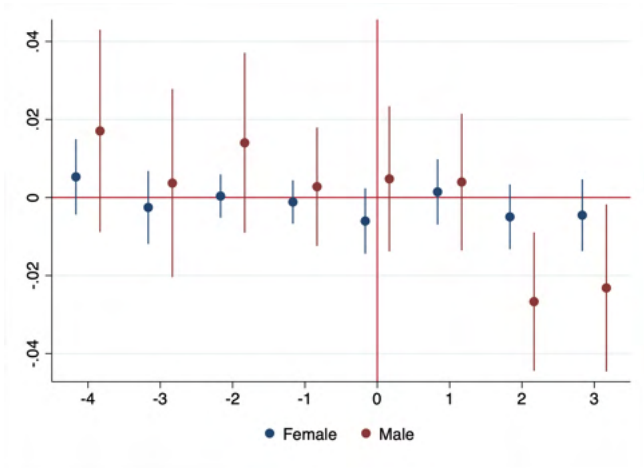
Notes. This figure shows the marginal effect of supply shocks on the predicted probability of exiting self-employment. Bars show 95 percent confidence intervals, and $\tau = 0$ is the year of the most recent closure of bank branches. Panel (a) shows the results of Equation 5 when the dependent variable is exiting self-employment (combining incorporated and unincorporated businesses). Panel (b) shows the results of the equation when the dependent variable is exiting incorporated self-employment. Panel (c) shows the results of the equation when the dependent variable is exiting unincorporated self-employment. All regressions are weighted by CAPS individual-level weights, and standard errors are bootstrapped.

Figure 15: Marginal Effect of Supply Shocks on the Predicted Probability of Exiting Self-employment, by Gender



a)

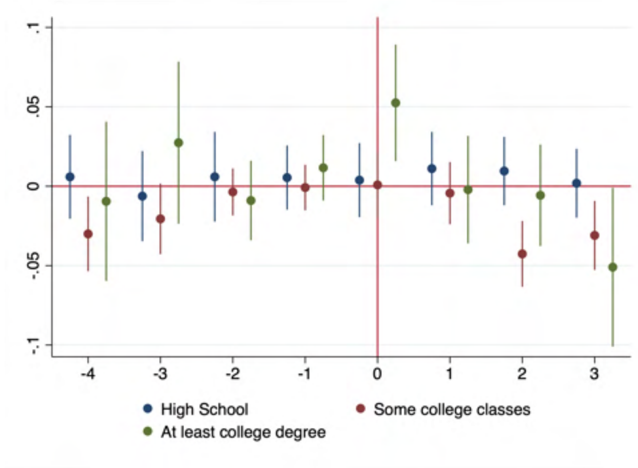




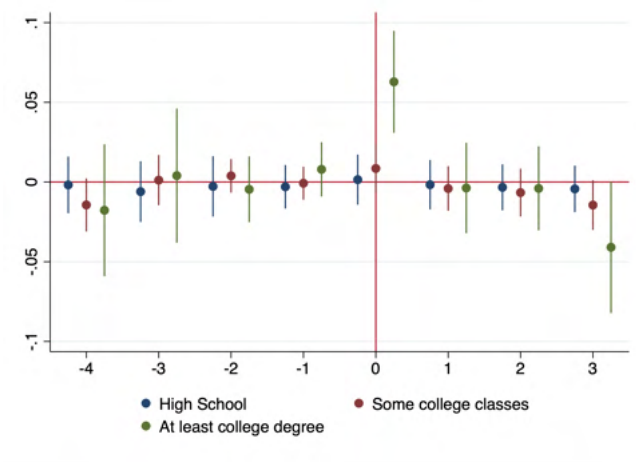
b) c)

Notes. This figure shows the marginal effect of supply shocks on the predicted probability of exiting self-employment. Bars show 95 percent confidence intervals, and $\tau = 0$ is the year of the most recent closure of bank branches. Panel (a) shows the results of Equation 5 when the dependent variable is exiting self-employment (combining incorporated and unincorporated businesses). Panel (b) shows the results of the equation when the dependent variable is exiting incorporated self-employment. Panel (c) shows the results of the equation when the dependent variable is exiting unincorporated self-employment. All regressions are weighted by CAPS individual-level weights, and standard errors are bootstrapped.

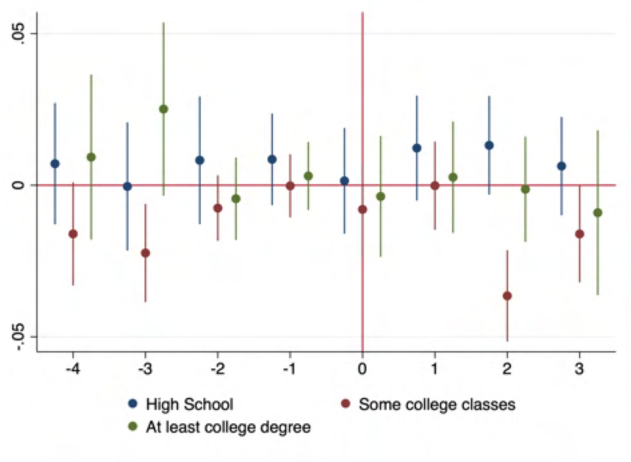
Figure 16: Marginal Effect of Supply Shocks on the Predicted Probability of Exiting Self-employment, by Education



a)



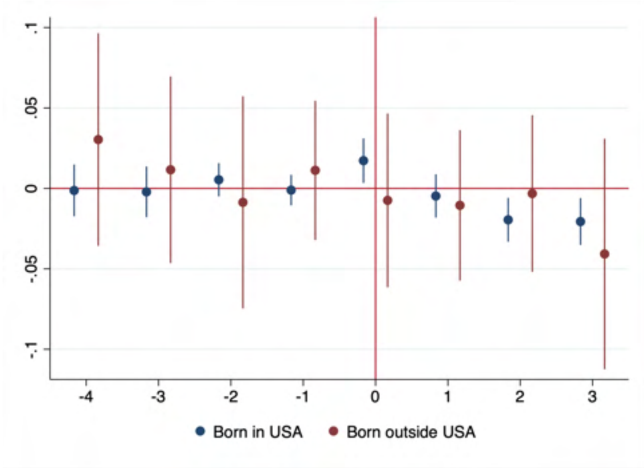
b)



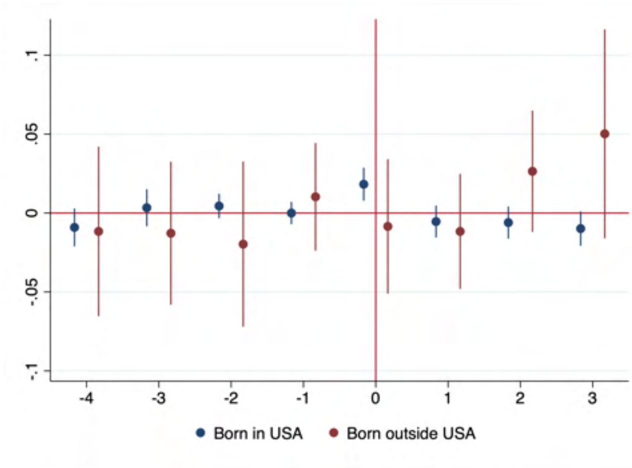
c)

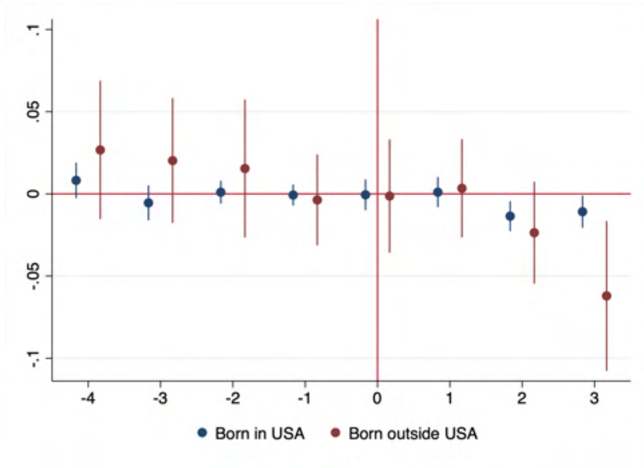
Notes. This figure shows the marginal effect of supply shocks on the predicted probability of exiting self-employment. Bars show 95 percent confidence intervals, and $\tau = 0$ is the year of the most recent closure of bank branches. Panel (a) shows the results of Equation 5 when the dependent variable is exiting self-employment (combining incorporated and unincorporated businesses). Panel (b) shows the results of the equation when the dependent variable is exiting incorporated self-employment. Panel (c) shows the results of the equation when the dependent variable is exiting unincorporated self-employment. All regressions are weighted by CAPS individual-level weights, and standard errors are bootstrapped.

Figure 17: Marginal Effect of Supply Shocks on the Predicted Probability of Exiting Self-employment, by Place of Birth



a)





b) c)

Notes. This figure shows the marginal effect of supply shocks on the predicted probability of exiting self-employment. Bars show 95 percent confidence intervals, and $\tau = 0$ is the year of the most recent closure of bank branches. Panel (a) shows the results of Equation 5 when the dependent variable is exiting self-employment (combining incorporated and unincorporated businesses). Panel (b) shows the results of the equation when the dependent variable is exiting incorporated self-employment. Panel (c) shows the results of the equation when the dependent variable is exiting unincorporated self-employment. All regressions are weighted by CAPS individual-level weights, and standard errors are bootstrapped.

11 Tables

Table 1: Summary Statistics for Trends in Self-Employment (Incorporated vs. Unincorporated)

	Year	Total Number, thousands	Rate (total number SE /LF)
Unincorporated	2003	10295	9.09%
	2015	9508	7.8%
Incorporated	2003	4956	4.3%
	2015	5486	4.5%
Total	2003	15252	13.4%
	2015	14994	12.3%

Notes. This table shows the total number of self-employed individuals and self-employment rate by year and type of self-employment. Data source: BLS, Current Population Survey.

Table 2: Transition Matrix Between Labor Markets, Based on Business Type (Incorporated vs. Unincorporated)

	Paid-employed at t+1	Incorp self-employed at t+1	Unincorp self-employed at t+1	Non-employed at t+1	Total at t+1
Paid-employed at t	94.38%	0.72%	1.08%	3.82%	84.7%
	19467	148	223	788	20626
Incorp self-employed at t	21.68%	56.26%	17.76%	4.3%	2.20%
	116	301	95	23	535
Unincorp self-employed at t	23.09%	16.59%	55.12%	5.19%	2.85%
	160	115	382	36	693
Non-employed at t	26.7%	1.08%	1.32%	70.9%	10.26%
	667	27	33	1771	2498
Total at t	83.81%	2.43%	3.01%	10.75%	100%
	20410	591	733	2618	24352

Notes. This table shows the average annual probabilities of transitioning from status j at time t to status k at period $t+1$. The definition of types of self-employment is based on the business type. The incorporated self-employed are owners of incorporated businesses, and the unincorporated self-employed are owners of unincorporated businesses.

Table 3: Summary Statistics

	Paid-E	Incorp SE	Unincorp SE	NE
Age**	39.51	40.02	41.11	42.68
	(10.28)	(9.66)	(9.82)	(11.61)
Female***	0.53	0.37	0.48	0.66
Married***	0.58	0.74	0.63	0.45
Number of kids***	1.03	1.34	1.05	1.08
	(1.18)	(1.34)	(1.25)	(1.25)
Number of HH members*	1.88	1.97	1.91	1.91
	(0.78)	(0.67)	(0.81)	(0.95)
Race: White***	0.64	0.76	0.68	0.50
Black	0.24	0.13	0.16	0.33
Other***	0.12	0.11	0.16	0.17
Ethnicity: Hispanic*	0.14	0.14	0.17	0.17
Years of schooling	14.73	15.01	14.71	14.27
	(3.52)	(3.56)	(3.52)	(4.08)
Distance to the nearest bank**, miles	1.22	1.47	1.29	1.1
	(1.56)	(1.78)	(1.62)	(1.39)
Number of banks within 5 miles***	36.13	40.43	32.21	31.72
	(40.04)	(59.30)	(42.27)	(42.24)
Number of bank workers per 1000 pop.	0.29	0.32	0.30	0.26
	(0.14)	(0.26)	(0.12)	(0.22)
Total assets***, 2014 $	15 801.6	43 507.91	25 598.05	9 019.31
	(45 536.39)	(118 841.8)	(73 625.28)	(35 357.12)
Number of branch closures***	0.035	0.036	0.029	0.023
	(0.193)	(0.241)	(0.169)	(0.159)
Supply Shock(z-score) ***	0.003	0.026	0.052	0.009
	(0.987)	(0.963)	(1.034)	(0.972)

Notes. This table shows summary statistics. Stars (*) shows p-values for the t-test of mean differences between incorporated and unincorporated self-employment, where *** denotes significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level. The number of banks per 1,000 population is calculated at the state level. The table indicates that the distance is in miles. The total assets are the sum of non-housing and housing assets adjusted for inflation.

Table 4: Supply Shocks and Access to Credit

Credit Products				
	Credit Card	Charge Card	Refinance Mortgage	
	(1)	(2)	(3)	
Supply Shock	-0.022***	-0.023***	-0.001***	
	(0.003)	(0.003)	(0.00002)	
Observations	33,160	33,201	16,804	
Reduction in Access to Credit				
	Application for new credit has been denied	Reduction in the Available Limit for Credit Cards	Being Asked to Pay Off Remaining Balance	Filing Bankruptcy
	(4)	(5)	(6)	(7)
Supply Shock	0.005***	0.009**	0.0008**	0.003***
	(0.00004)	(0.004)	(0.0003)	(0.0009)
Observations	8,568	8,385	8,571	29,559

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes. The table presents results from the linear probability model Equation 3. The dependent variables are dummy variables that equal one if the respondent applied for a new credit card, opened a new charge card, refinanced a mortgage, had an application for a new credit rejected, had the available limit for credit cards reduced, was asked to pay off the remaining loan balance, or filed bankruptcy in the last 12 months. The full estimates of Equation 3 are presented in Table 16.

Table 5: Correlation Matrix

	Out ISE	Out USE	B-L distance	Supply shock	Number of closures
Out ISE					
Out USE	-0.0131*** (0.0013)				
B-L distance	0.0031 (0.5632)	-0.0038 (0.4781)			
Supply shock	0.0039*** (0.0004)	-0.0032 (0.5672)	0.0031 (0.5785)		
Number of closures	0.0035 (0.4987)	-0.0031 (0.5478)	0.0056 (0.2889)	0.0189*** (0.0006)	

Notes. Each cell in the table contains a numerical value representing the correlation coefficient between the variables indicated by the respective row and column, with three asterisks (***) denoting statistical significance at the 1% level. Parentheses contain standard errors.

Table 6: Effect of Supply Shocks. Part I

	(1) Out of E	(2) Out of ISE	(3) Out of USE
Previous labor status			
PE	0.470***	0.466***	0.004
	(0.007)	(0.004)	(0.005)
ISE	0.524***	0.0004	0.523***
	(0.006)	(0.004)	(0.004)
USE	-0.004	-0.005**	0.0004
	(0.004)	(0.002)	(0.003)
Supply Shock:			
Supply Shock	0.017***	0.026***	-0.04
	(0.001)	(0.0008)	(0.123)
Other Explanatory Variables			
Number of bank workers	0.004	0.039*	-0.036
	(0.036)	(0.024)	(0.027)
Owner of house	-0.011***	-0.004**	-0.007***
	(0.003)	(0.002)	(0.002)
Number of banks	15.22*	2.641	12.85*
	(9.134)	(6.140)	(6.912)
Large durable assets	-0.002**	-0.002***	-0.0002
	(0.001)	(0.0008)	(0.0008)
Diff in wages	-0.003	-8.94e-05	-0.003
	(0.007)	(0.004)	(0.005)

Table 7: Effect of Supply Shocks. Part II

	(1) Out of E	(2) Out of ISE	(3) Out of USE
Age	-0.002	-0.0002	-0.001
	(0.001)	(0.001)	(0.001)
Age sq.	1.53e-05	3.18e-06	1.22e-05
	(1.09e-05)	(7.26e-06)	(8.22e-06)
Race (Black)	0.002	0.003	-0.001
	(0.004)	(0.003)	(0.003)
Race (Other)	0.003	0.003	-0.001
	(0.004)	(0.003)	(0.003)
Hispanic	0.0127***	0.006**	0.007**
	(0.004)	(0.003)	(0.003)
Female	0.003	0.003*	0.001
	(0.002)	(0.002)	(0.002)
Married	-0.002	0.003	-0.004
	(0.004)	(0.003)	(0.003)
1 child	-0.004	-0.002	-0.001
	(0.004)	(0.003)	(0.003)
Less than 4 kids	-0.004	-0.003	-0.001
	(0.003)	(0.002)	(0.002)
More than 4 kids	-0.0003	-0.004	0.004
	(0.011)	(0.007)	(0.008)
2 adults in HH	0.007**	0.001	0.006**
	(0.003)	(0.002)	(0.003)
Less than 4 adults in HH	0.006	-0.0001	0.006
	(0.005)	(0.004)	(0.004)
More than 4 adults in HH	0.011	-0.003	0.014*
	(0.011)	(0.007)	(0.008)
Local Economic Conditions			
Unemployment	-1.63e-07	-6.33e-08	-1.03e-07
	(1.37e-07)	(9.15e-08)	(1.04e-07)
% of college grads	0.018	-0.001	0.018
	(0.027)	(0.018)	(0.021)
lnGDP	0.024**	-0.0004	0.024***
	(0.01)	(0.007)	(0.008)
Number of business est	3.07e-07	9.68e-08	2.14e-07
	(2.81e-07)	(1.88e-07)	(2.13e-07)
Observations	36,984	36,635	36,526
R-squared	0.497	0.470	0.526

Table 8: Effect of Supply Shocks. Part III

	(1) Out of E	(2) Out of ISE	(3) Out of USE
Panel B: Sensitivity analysis (exogenous controls)			
Breakdown point	2.45	5.81	8.36
Panel C: Sensitivity analysis (endogenous controls)			
$\overline{r_x}$	0.983	0.983	0.884

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes. The table presents the results from estimating Equation 4. The dependent variable is a binary indicator of the individual switch from self-employment in year $(t-1)$ to non-SE in year t. Column 1 presents estimates for both types of self-employed, column 2 for the incorporated self-employed, and column 3 for the unincorporated self-employed. The following variables are included but not shown: spatial cluster fixed effects and state-by-year fixed effects. All regressions are weighted by CAPS individual-level weights, and standard errors are bootstrapped.

Table 9: Dynamic Effect of Supply Shocks

	Out of E	Out of ISE	Out of USE
Previous labor status:			
PE	0.464***	0.455***	0.012
	(0.013)	(0.009)	(0.009)
ISE	0.688***	0.003	0.687***
	(0.013)	(0.011)	(0.009)
USE	-0.003	-0.003	-0.0004
	(0.008)	(0.006)	(0.005)
Supply Shock:			
Supply Shock	0.142***	0.102***	-0.004
	(0.012)	(0.008)	(0.008)
Dynamic Effect:			
$\tau = -4$#Supply Shock	0.010	0.016	-0.006
	(0.013)	(0.009)	(0.009)
$\tau = -3$#Supply Shock	0.015	0.013	0.003
	(0.012)	(0.009)	(0.008)
$\tau = -2$#Supply Shock	0.017	0.012	0.005
	(0.012)	(0.009)	(0.008)
$\tau = -1$#Supply Shock	0	0	0
	(0)	(0)	(0)
$\tau = 0$#Supply Shock	0.029**	0.026***	0.003
	(0.013)	(0.009)	(0.009)
$\tau = 1$#Supply Shock	0.010	0.004	0.006
	(0.013)	(0.009)	(0.009)
$\tau = 2$#Supply Shock	-0.006	0.005	-0.011
	(0.013)	(0.009)	(0.009)
$\tau = 3$#Supply Shock	-0.011	-0.0007	-0.009
	(0.013)	(0.009)	(0.009)
$\tau = 4$#Supply Shock	-0.004	-0.001	-0.003
	(0.014)	(0.011)	(0.009)
$\tau = 5$#Supply Shock	0.014	0.007	0.007
	(0.013)	(0.009)	(0.009)
Observations	36,984	36,635	36,526
R-squared	0.623	0.502	0.713

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes. The table presents the results from estimating Equation 5. The dependent variable is a binary indicator of the individual switch from self-employment in year $(t-1)$ to non-SE in year t. Column 1 presents estimates for both types of self-employed, column 2 for the incorporated self-employed, and column 3 for the unincorporated self-employed. The following variables are included but not shown: dummies for the nth year before and after the recent closure of the bank at the zip code level, individual characteristics (dummies for gender, marital status, three race dummies (white, black, other), Hispanic ethnicity, four education dummies (less than high school, high school, some college, college and above), age (and a quadratic for age), spatial cluster fixed effects, state-by-year fixed effects, local economic conditions (the number of unincorporated self-employed, normalized by the prime age population), the number of business establishments, the unemployment rate, the average value of houses, and the share of the population with professional, scientific, management, and administrative education at the zip code level). All regressions are weighted by CAPS individual-level weights, and standard errors are bootstrapped.

Table 10: Role of Borrower-Lender Distance. Alternative Measures of The Distance

	Out of E	Out of ISE	Out of USE
$\text{Log}(\widehat{\text{Distance}})$	0.0517*	0.0381*	0.0140
	(0.0311)	(0.0210)	(0.0234)
$\text{Log}(\widehat{\text{Distance 5}})$	0.0149*	0.0100*	0.0049
	(0.0084)	(0.0057)	(0.0064)
$\text{Log}(\widehat{\text{Distance 10}})$	0.0124*	0.0083*	0.0042
	(0.0072)	(0.0048)	(0.0055)
$\text{Log}(\widehat{\text{Distance 15}})$	0.0119*	0.0079*	0.0039
	(0.0069)	(0.0046)	(0.0053)
$\text{Log}(\widehat{\text{Distance 20}})$	0.0111*	0.0081*	0.0041
	(0.0070)	(0.0047)	(0.0053)
Observations	36,984	36,635	36,526

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes. The table presents the results from the second stage of 2SLS procedure estimating Equation 8 for different distance measures, such as the minimum borrower-bank distance (Log(Distance)), the average distance to the five nearest banks (Log(Distance 5)), to the ten nearest banks (Log(Distance 10)), to the fifteen nearest banks (Log(Distance 15)), and to the twenty nearest banks (Log(Distance 20)). The dependent variable is a binary indicator of the individual switch from self-employment in year $(t-1)$ to non-SE in year t. Column 1 presents estimates for both types of self-employed, column 2 for the incorporated self-employed, and column 3 for the unincorporated self-employed. Distance measures are predicted values from the first stage (see Equation 7). The following variables are included but not shown: individual characteristics (dummies for gender, marital status, three race dummies (white, black, other), Hispanic ethnicity, four education dummies (less than high school, high school, some college, college and above), age (and a quadratic for age), spatial cluster fixed effects, state-by-year fixed effects, local economic conditions (the number of unincorporated self-employed, normalized by the prime age population), the number of business establishments, the unemployment rate, the average value of houses, and the share of the population with professional, scientific, management, and administrative education at the zip code level). All regressions are weighted by CAPS individual-level weights and standard errors are bootstrapped.

Table 11: Role of borrower-lender distance. First Stage.

	Log(Distance)	log (Distance 5)	log (Distance 10)	log (Distance 15)	log (Distance 20)
= 1 if leaves in zip-code exposed to MIC	0.0149**	0.0174**	0.0136**	0.000652**	0.00764**
	(0.0041)	(0.0085)	(0.0084)	(0.0073)	(0.0063)
Mean travel time to work in minutes	0.00744***	0.0271***	0.0328***	0.0340***	0.0333***
	(0.000596)	(0.000863)	(0.000862)	(0.000843)	(0.000825)
Number of bank workers per capita, county level	-1.815***	-0.919***	-0.628***	-0.501***	-0.388***
	(0.0385)	(0.0547)	(0.0546)	(0.0534)	(0.0523)
Owner of house	0.0412***	0.111***	0.0924***	0.0836***	0.0731***
	(0.00784)	(0.0117)	(0.0117)	(0.0114)	(0.0112)
Number of banks per capita, state level	903.4***	513.4***	426.6***	379.7***	370.7***
	(21.89)	(33.91)	(33.86)	(33.12)	(32.39)
Large durable assets	-0.117***	-0.0553***	-0.0280***	-0.0189***	-0.0136***
	(0.00260)	(0.00376)	(0.00376)	(0.00367)	(0.00359)
Diff in wages	-1.031***	-0.538***	-0.329***	-0.223***	-0.163***
	(0.0135)	(0.0195)	(0.0194)	(0.0190)	(0.0186)
Age	-0.186***	-0.101***	-0.0703***	-0.0599***	-0.0540***
	(0.00214)	(0.00312)	(0.00312)	(0.00305)	(0.00298)
Age sq.	0.00161***	0.000877***	0.000626***	0.000546***	0.000500***
	(2.26e-05)	(3.30e-05)	(3.29e-05)	(3.22e-05)	(3.15e-05)
Race (Black)	-0.661***	-0.360***	-0.279***	-0.251***	-0.245***
	(0.00745)	(0.0108)	(0.0108)	(0.0106)	(0.0103)
Race (Other)	-0.192***	-0.171***	-0.145***	-0.130***	-0.130***
	(0.0105)	(0.0152)	(0.0152)	(0.0148)	(0.0145)
Hispanic	-0.107***	-0.0820***	-0.0689***	-0.0612***	-0.0498***
	(0.0108)	(0.0157)	(0.0157)	(0.0154)	(0.0150)
Female	0.0598***	0.00809	0.00241	0.000854	-0.00212
	(0.00598)	(0.00865)	(0.00864)	(0.00845)	(0.00826)
Married	-0.583***	-0.288***	-0.174***	-0.129***	-0.105***
	(0.00882)	(0.0128)	(0.0128)	(0.0125)	(0.0122)
1 child	0.654***	0.356***	0.263***	0.216***	0.188***
	(0.00781)	(0.0113)	(0.0113)	(0.0110)	(0.0108)
Less than 4 kids	-0.0581***	-0.0324***	-0.0106	0.00730	0.0188*
	(0.00745)	(0.0108)	(0.0108)	(0.0105)	(0.0103)
More than 4 kids	0.693***	0.376***	0.348***	0.328***	0.313***
	(0.0264)	(0.0377)	(0.0376)	(0.0368)	(0.0360)
2 adults in HH	-0.0382***	-0.00306	-0.00318	0.00656	0.00469
	(0.00852)	(0.0124)	(0.0124)	(0.0121)	(0.0118)
Less than 4 adults in HH	0.906***	0.505***	0.345***	0.282***	0.243***
	(0.0106)	(0.0154)	(0.0153)	(0.0150)	(0.0147)
More than 4 adults in HH	0.722***	0.418***	0.326***	0.314***	0.282***
	(0.0258)	(0.0387)	(0.0386)	(0.0378)	(0.0369)
Unemployment	-2.03e-05***	-1.23e-05***	-8.86e-06***	-7.34e-06***	-6.37e-06***
	(2.95e-07)	(4.25e-07)	(4.25e-07)	(4.15e-07)	(4.06e-07)
% of college grads	6.498***	2.767***	1.469***	0.941***	0.649***
	(0.0328)	(0.0476)	(0.0476)	(0.0465)	(0.0455)
lnGDP	-0.962***	-0.471***	-0.343***	-0.255***	-0.197***
	(0.0257)	(0.0384)	(0.0384)	(0.0375)	(0.0367)
Number of business est	4.96e-05***	2.56e-05***	1.57e-05***	1.14e-05***	8.68e-06***
	(5.27e-07)	(7.60e-07)	(7.59e-07)	(7.43e-07)	(7.26e-07)
Observations	17,037	16,370	16,370	16,370	16,370
R-squared	0.774	0.348	0.250	0.228	0.220

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes. The table presents the results from the first stage of 2SLS procedure estimating Equation 7. The following variables are included but not shown: spatial cluster fixed effects and state-by-year fixed effects. All regressions are weighted by CAPS individual-level weights, and standard errors are bootstrapped.

Table 12: Role of borrower-lender distance. Main Model. Part I

	Out of E	Out of ISE	Out of USE
$\text{Log}(\widehat{\text{Distance}})$	0.0517*	0.0381*	0.0140
	(0.0311)	(0.0210)	(0.0234)
Previous Labor Status			
PE	0.482	0.480***	0.00232
	(0.326)	(0.00434)	(0.00527)
ISE	0.529***	0.000865	0.528***
	(0.00571)	(0.00411)	(0.00429)
USE	-0.00498	-0.00509**	0.000113
	(0.00337)	(0.00226)	(0.00255)
Other Explanatory Variables			
Number of bank workers	0.345*	0.277**	0.0700
	(0.198)	(0.133)	(0.149)
Owner of house	-0.0108***	-0.00467**	-0.00619***
	(0.00314)	(0.00211)	(0.00237)
Number of banks	-41.91	-37.78*	-4.164
	(32.96)	(22.19)	(24.83)
Large durable assets	0.00496	0.00290	0.00206
	(0.00411)	(0.00277)	(0.00309)
Diff in wages	0.0576*	0.0417*	0.0164
	(0.0338)	(0.0228)	(0.0255)

Table 13: Role of borrower-lender distance. Main Model. Part II

	Out of E	Out of ISE	Out of USE
Other Explanatory Variables			
Age	0.00890	0.00714*	0.00182
	(0.00606)	(0.00408)	(0.00456)
Age sq.	-7.41e-05	-6.00e-05*	-1.45e-05
	(5.25e-05)	(3.53e-05)	(3.95e-05)
Race (Black)	0.0380*	0.0282**	0.0100
	(0.0211)	(0.0142)	(0.0159)
Race (Other)Other	0.0114	0.0112**	0.000242
	(0.00740)	(0.00498)	(0.00558)
Hispanic	0.0190***	0.00933***	0.00990**
	(0.00528)	(0.00357)	(0.00398)
Female	-0.00150	-3.17e-05	-0.00144
	(0.00310)	(0.00208)	(0.00234)
Married	0.0319	0.0264**	0.00581
	(0.0199)	(0.0134)	(0.0150)
1 child	-0.0406*	-0.0282*	-0.0128
	(0.0214)	(0.0144)	(0.0161)
Less than 4 kids	0.00116	0.000442	0.000712
	(0.00375)	(0.00251)	(0.00284)
More than 4 kids	-0.0417	-0.0325*	-0.00907
	(0.0258)	(0.0174)	(0.0193)
2 adults in HH	0.00736**	0.00184	0.00555**
	(0.00336)	(0.00225)	(0.00253)
Less than 4 adults in HH	-0.0482	-0.0369*	-0.0118
	(0.0298)	(0.0201)	(0.0225)
More than 4 adults in HH	-0.0352	-0.0353**	-0.000129
	(0.0259)	(0.0174)	(0.0196)
Local Economic Conditions			
Unemployment	9.91e-07	7.65e-07*	2.30e-07
	(6.85e-07)	(4.62e-07)	(5.16e-07)
% of college grads	-0.355*	-0.262*	-0.0957
	(0.213)	(0.144)	(0.160)
lnGDP	0.0606**	0.0265	0.0345*
	(0.0253)	(0.0170)	(0.0191)
Number of business est	-2.55e-06	-1.96e-06*	-6.05e-07
	(1.68e-06)	(1.14e-06)	(1.27e-06)
Observations	14,785	14,626	14,614
R-squared	0.486	0.463	0.519

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes. The table presents the results from the second stage of 2SLS procedure estimating Equation 8. The dependent variable is a binary indicator of the individual switch from self-employment in year $(t-1)$ to non-SE in year t. Column 1 presents estimates for both types of self-employed, column 2 for the incorporated self-employed, and column 3 for the unincorporated self-employed. Variable $\text{Log}(\widehat{\text{Distance}})$ is instrumented from the first stage (see Equation 7 and Table 11). The following variables are included: spatial cluster fixed effects and state-by-year fixed effects. All regressions are weighted by CAPS individual-level weights, and standard errors are bootstrapped.

A Theoretical Model

In this model, time is discrete. In contrast to Evans and Jovanovic (1989) and Levine and Rubinstein (2018), I abstract to a stylized two-period model to clarify predictions, although the extension to the infinite horizon is straightforward. The first category of self-employed includes owners of businesses that demand entrepreneurial skills, physical capital, and business loans. Examples of the incorporated self-employed are computer programmers, lawyers, doctors, and real estate managers. For the second category, in contrast with Levine and Rubinstein (2018), who argue that this category of self-employed individuals requires "none or little entrepreneurial ability, physical capital, and liquidity" and is driven mainly through non-pecuniary benefits of self-employment, I use a different definition[23]. I model the unincorporated self-employed as non-entrepreneurs and owners of businesses who operate without business loans. In other words, the agent either does not have enough collateral to access the credit market, or they can use their own initial wealth to start a business (as long as it is not capital intensive). Examples of the unincorporated self-employed are tutors, babysitters, and maintenance workers.

Agents live for two periods - the present (t) and the future ($t+1$)[24]. Each period, the agent can be in one of three labor states: incorporated self-employed ($j = ISE$), unincorporated self-employed ($j = USE$), or paid employed worker ($j = PE$). The set of all possible choices for the agent:

$$\begin{aligned} J =&\{(PE_t, PE_{t+1}); (PE_t, ISE_{t+1}); (PE_t, USE_{t+1}); (ISE_t, PE_{t+1}); (ISE_t, ISE_{t+1}); \\ &(ISE_t, USE_{t+1}); (USE_t, PE_{t+1}); (USE_t, ISE_{t+1}); (USE_t, USE_{t+1})\} \end{aligned}$$

Each period employment type-specific uncertainty (ϵ_t^j and ϵ_{t+1}^j) realizes. For simplicity in

[23]In the literature (Evans and Jovanovic (1989), Levine and Rubinstein (2018)), the self-employed are divided broadly into two categories: entrepreneurs and other self-employed. Entrepreneurs possess and manage capital- and skill-intensive businesses, while other self-employed individuals operate ventures that do not necessitate substantial liquidity or human capital. In this paper, I'm using different divisions of self-employed into groups: incorporated and unincorporated. These types are different in legal structure, taxation, and liability protection. Appendix C.1 provides details about the differences between these types of self-employed. Appendix C.2 discusses the financial dependence of these types of self-employed. First, an inverse relationship was shown between the proportion of self-employed individuals and the alteration in the number of bank branches in the local area. Second, Appendix C.2.2 provides details from the Survey of Small Business Financing. Results show that incorporating a business increases the likelihood of using a business credit line, credit card, and shareholders' loan and decreases the likelihood of using the owner's personal credit card. The findings indicate that incorporating a business enhances the probability of utilizing a loan to purchase equipment, a business vehicle, or a business-related mortgage.

[24]Period t is the beginning of the agent's life, and she did not participate in the labor market at period $t-1$.

describing the model, assume that the error terms are known at the beginning of the first period so that the optimization problem may be started in terms of individual making all decisions at the beginning of the first period, so at the beginning of the period t, the agent makes the static labor decision $j \in J$ for both periods. If, instead, the random variable for period $t+1$ was unknown until the beginning of that period, then the model could be extended with the decisions in period t based on the error term distribution in period $t+1$. Assuming that the error term for the second period is not realized until after the first period decisions are made complicates the notation but does not change the nature of the problem or solution.

Individuals select in which labor market sector to participate based on their productivity (ability), θ, and the number of available assets, a_t. The agent can take a business loan only if her amount of assets exceeds the value a_b that the bank determines (see Fig.18).

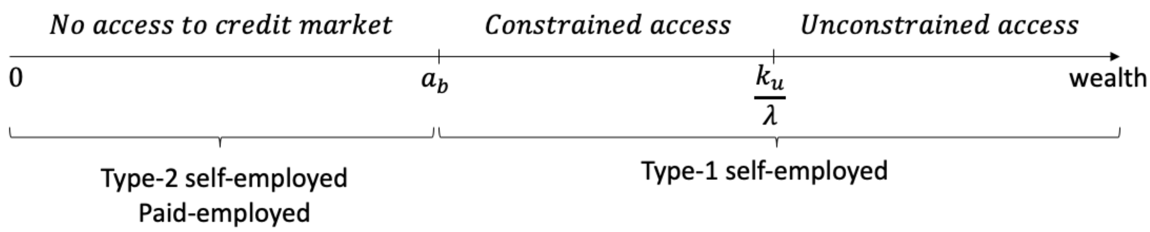


Figure 18: Wealth Distribution and The Access to Credit Market

Individuals with access to financial markets ($a_t > a_b$) may borrow to finance some capital acquisitions; others must finance their purchases from their beginning-of-period assets. At the end of the period, individuals (a) receive revenue from any business operations, paid-employment, and financial investments, (b) repay any loans and accrued interest, (c) pay switching cost in period $t+1$ if they changed labor market sectors from period t, (d) receive interest on savings, and (e) purchase the consumption goods. The agent is permitted to borrow at the period t against the next period's revenues at the period $t+1$, and any assets remaining at the end of the period t, a_{t+1}, are carried forward to the next period, $t+1$.

The individual selects the labor market sectors, her borrowing to purchase capital, and her consumption to maximize her expected discounted utility subject to the constraints below. The discounted utility for an individual who selects employment types j and who consumes c_t and

c_{t+1} :

$$\max_{c_t \geq 0,\ c_{t+1} \geq 0} V = U^j(c_t) + \beta U^j(c_{t+1}) + \epsilon_t^j + +\epsilon_{t+1}^j$$

where ϵ^j is the choice-specific error term.

The decision rule $\delta(j)$:

$$\delta(j) = \arg\max_j \left(U^j(c_t) + \beta U^j(c_{t+1}) + \epsilon_t^j + \epsilon_{t+1}^j\right)$$

Given this set-up, the budget constraints for a two-period version of the model can be most easily derived by first determining the constraint involving the assets at the end of period t, as dependent upon the decisions in that period, and then determining the constraint involving consumption in period $t+1$, as dependent upon the assets carried forward to the period and the decisions made during the period. For any labor market choice, the assets at the end of the period t consist of any revenue from business operations, paid employment, and interest on financial investments minus (a) consumption expenditures and (b) any repayment of business borrowings and the accrued interest, and the transaction costs of borrowing.

The budget constraint of self-employed worker (any type) at period t:

$$a_{t+1} + c_t = f(\theta, k_t) - r(k_t - a_t)\mathbb{1}((k_t - a_t) \geq 0) - (k_t - a_t)\mathbb{1}((k_t - a_t) < 0) - \psi_t \mathbb{1}((k_t - a_t) \geq 0)$$

where c_t is consumption in period t, $f(\theta, k_t)$- the production function, k_t - the amount of capital invested in the business at period t, a_{t+1} - savings in period t, r - the gross cost of capital (one plus the interest rate), θ - agent's productivity level. If $a_t \leq k_t$ the agent is a net-borrower, and $r(k_t - a_t)$ is the amount she repays at the end of the period; if $(a_t \geq k_t)$ she is a net-saver. The model also includes the non-interest cost of borrowing, ψ_t, that may contain not only the direct fees charged at closing, such as origination and application fees but also the indirect cost to borrowers, such as the value of time spent preparing application documents and traveling to the bank. The main goal of the theoretical model is to show how the decision rule, $\delta(j)$, may change if the non-interest cost of borrowing, ψ_t, increases.

The paid-employed worker is allowed to take a one-period non-business loan (e.g., consumer loan, mortgage) at a gross interest rate r_b (one plus the interest rate). The budget constraint of the paid-employed worker at period t:

$$a_{t+1} + c_t = \theta w_t + a_t + b_t(1 - r) - \psi_t \mathbb{1}(b_t \neq 0)$$

where c_t is consumption in period t, a_t - initial wealth at period t, a_{t+1} - savings in period t, w_t is the wage at period t, θ - agent's productivity level, b_t - a one-period non-business loan, r - the gross cost of loan (one plus the interest rate), ψ_t - non-interest cost of borrowing.

In the second period, the individual has no incentive to carry assets forward to the next period, so consumption equals the revenue from all sources minus (a) any repayment of business borrowing, accrued interest, and the transaction cost of borrowing, and (b) any switching cost as a result of a change from the type of labor market from the first period. If at period t, the agent was a paid worker, she pays the switching cost π_t^{PE} to any type of self-employed at period $t+1$. If she worked as unincorporated self-employed at period t, she pays the switching cost π_t^{ISE} to incorporated self-employed at period $t+1$.

The budget constraint of self-employed worker (any type) at period $t+1$:

$$c_{t+1} = f(\theta, k_{t+1}) - r(k_{t+1} - ra_{t+1})\mathbb{1}\left((k_{t+1} - ra_{t+1}) \geq 0\right) - (k_{t+1} - ra_{t+1})\mathbb{1}\left((k_{t+1} - ra_{t+1}) < 0\right) -$$
$$- \psi_{t+1}\mathbb{1}\left((k_{t+1} - ra_{t+1}) \geq 0\right) - \pi_{t+1}^{PE}\mathbb{1}\left(y_t = y^{PE}\right) - \pi_{t+1}^{ISE}\mathbb{1}\left((y_t = y^{USE}) \ \& \ (y_{t+1} = y^{ISE})\right)$$

where c_{t+1} - consumption in period $t+1$, $f(\theta, k_{t+1})$- the production function, k_{t+1} - the amount of capital invested in the business at period $t+1$, a_t - initial wealth at period t, a_{t+1} - savings in period t, r - the gross cost of capital (one plus the interest rate), ra_{t+1} - wealth in period $t+1$, θ - agent's productivity level. If $ra_{t+1} \leq k_{t+1}$, the agent is a net-borrower, and $r(k_{t+1} - ra_{t+1}))$ is the amount she repays at the end of the period; if $a_{t+1} \geq k_{t+1}$, she is a net-saver. If the agent decides to switch from paid employment to other labor states, she has to pay the cost of switching from paid employment, π_{t+1}^{PE} that may include the cost of the license, security deposit for office, cost of time invested in new skills. Being an incorporated self-employed requires higher entrepreneurial skills than being unincorporated self-employed; it means if the agent switches to incorporated self-employment, she has to pay the cost of switching, π_{t+1}^{ISE}, that includes the cost of time invested in new skills.

The budget constraint of the paid-employed worker at period $t+1$:

$$c_{t+1} = \theta w_{t+1} + ra_{t+1} + b_{t+1}(1-r) - \psi_{t+1}\mathbb{1}(b_{t+1} \neq 0)$$

where c_{t+1} - consumption in period $t+1$, a_{t+1} - savings in period t, r - the gross cost of capital (one plus the interest rate), ra_{t+1} - wealth in period $t+1$, w_{t+1} - the wage at period $t+1$, θ - agent's productivity level, b_{t+1} - a one-period non-business loan, ψ_{t+1}-non-interest cost of borrowing.

Given the additive nature of the business revenue and any loan costs in the constraints, the optimal capital decisions can be made separately from the consumption decisions. In either period, a self-employed individual with access to the capital market selects the amount of capital that maximizes her business profit or:

$$\text{Period } t: \quad \max_{0\le k_t \le \lambda a_t} [f(\theta, k_t) - r(k_t - a_t)\mathbb{1}((k_t - a_t) \ge 0) - (k_t - a_t)\mathbb{1}((k_t - a_t) < 0)]$$

$$\text{Period } t+1: \quad \max_{0\le k_{t+1} \le \lambda r a_{t+1}} [f(\theta, k_{t+1}) - r(k_{t+1} - ra_{t+1})\mathbb{1}((k_{t+1} - ra_{t+1}) \ge 0) -$$
$$- (k_{t+1} - ra_{t+1})\mathbb{1}(k_{t+1} - ra_{t+1}) < 0)]$$

Assuming that the production function is strictly concave in the capital and making the Inada assumption that the marginal product of capital is infinity at zero, the optimal level of capital is positive and finite and is denoted k^*. If a self-employed individual (unincorporated) does not have access to the capital markets and cannot finance the optimal capital with her own assets, then she uses all of her assets to purchase capital. The resulting revenue is greater than if she had used only part of the assets to purchase capital and had put the remainder in a financial instrument earning a gross return r. As in Evans and Jovanovic (1989), I assume that the agent can borrow an amount that is proportional to her wealth ($k_t \le \lambda a_t$ and $k_{t+1} \le \lambda r a_{t+1}$, where $\lambda \ge 1$). If the amount of wealth exceeds the amount required to finance ($\lambda a_t \ge k_t^*$ and $\lambda r a_{t+1} \ge k_{t+1}^*$), the remaining wealth is invested at the rate r, and it's "unconstrained case". If the optimal amount of capital is lower than the amount of wealth ($\lambda a_t \le k_t^*$ and $\lambda r a_{t+1} \le k_{t+1}^*$), the agent borrows the maximum available amount of loan, and it is "constrained case".

Agents sort into employment type j to maximize her utility:

$$V = U^j(c_t) + \beta U^j(c_{t+1}) + \epsilon_t^j + \epsilon_{t+1}^j$$

where ϵ^j is the choice-specific error term. All possible labor choices: $J = \{(PE_t, PE_{t+1});$ $(PE_t, ISE_{t+1}); (PE_t, USE_{t+1}); (ISE_t, PE_{t+1}); (ISE_t, ISE_{t+1}); (ISE_t, USE_{t+1}); (USE_t, PE_{t+1});$ $(USE_t, ISE_{t+1}); (USE_t, USE_{t+1})\}$

The decision rule $\delta(j)$:

$$\delta(j) = \arg\max_j \left(U^j(c_t) + \beta U^j(c_{t+1}) + \epsilon_t^j + \epsilon_{t+1}^j\right)$$

Assume that $U(c_t) = log(c_t)$ and $\epsilon^j \in N(0;1)$ for simplicity.

If the agent works as a paid-employed worker, her budget constraints:

$$\text{Period } t: \quad a_{t+1} + c_t = \theta w_t + a_t + b_t(1 - r_b) - \psi_t$$

$$\text{Period } t+1: \quad c_{t+1} = \theta w_{t+1} + ra_{t+1} + b_{t+1}(1 - r_b) - \psi_{t+1}$$

where c_t is consumption in period t, c_{t+1} - consumption in period $t+1$, a_t - initial wealth at period t, a_{t+1} - savings in period t, r - the gross cost of capital (one plus the interest rate),

ra_{t+1} - wealth in period $t+1$, w_t is the wage at period t, w_{t+1} - the wage at period $t+1$, θ - agent's productivity level, b_t - a one-period non-business loan at period t, b_{t+1} - a one-period non-business loan at period $t+1$, r_b - an gross interest rate of non-business loan, ψ_t - non-interest cost of borrowing at period t, ψ_{t+1} - non-interest cost of borrowing at period $t+1$.

If the agent works as an incorporated self-employed worker (only if $a_t \geq a_b, a_{t+1} \geq a_b$), she can borrow money from the bank. In this case, she maximizes her production function ($f(\theta, k_t) = \theta k_t^{\alpha}$, where $\alpha < 1$), making a decision about the amount of capital invested in the business, and she can use her wealth as collateral. In this case, the agent's budget constraints:

$$\text{Period } t: \quad a_{t+1} + c_t = \max_{0 \leq k_t \leq \lambda a_t} [\theta (k_t)^{\alpha} - r(k_t - a_t)\mathbb{1}((k_t - a_t) \geq 0) - (k_t - a_t)\mathbb{1}((k_t - a_t) < 0) -$$
$$- \psi_t \mathbb{1}((k_t - a_t) \geq 0)]$$
$$\text{Period } t+1: \quad c_{t+1} = \max_{0 \leq k_{t+1} \leq \lambda r a_{t+1}} [\theta (k_{t+1})^{\alpha} - r(k_{t+1} - ra_{t+1})\mathbb{1}((k_{t+1} - ra_{t+1}) \geq 0) - (k_{t+1} - ra_{t+1}) \cdot$$
$$\cdot \mathbb{1}((k_{t+1} - ra_{t+1}) < 0) - \psi_{t+1}\mathbb{1}((k_{t+1} - ra_{t+1}) \geq 0) - \pi_{t+1}^{PE}\mathbb{1}(y_t = y^{PE}) - \pi_{t+1}^{ISE}\mathbb{1}(y_t = y^{USE})]$$

where c_t is consumption in period t, c_{t+1} - consumption in period $t+1$, k_t - the amount of capital invested in the business at period t, k_{t+1} - the amount of capital invested in the business at period $t+1$, a_t - initial wealth at period t, a_{t+1} - savings in period t, r - the gross cost of capital (one plus the interest rate), ra_{t+1} - wealth in period $t+1$, θ - agent's productivity level. If $a_t \leq k_t$ or $ra_{t+1} \leq k_{t+1}$, the agent is a net-borrower, and $r(k_t - a_t)$ or $r(k_{t+1} - ra_{t+1}))$ is the amount she repays at the end of the period. If $(a_t \geq k_t)$ or $(a_{t+1} \geq k_{t+1})$, she is a net-saver, and $(k_t - ra_t)$ is the amount she has as wealth at the period $t+1$. If at period t the agent was a paid worker, she pays switching cost π_t^{PE} at period $t+1$. If she worked as unincorporated self-employed at period t, she pays switching cost π_t^{ISE} at period $t+1$. As in Evans and Jovanovic (1989), I assume that the agent can borrow an amount that is proportional to her wealth ($k_t \leq \lambda a_t$ and $k_{t+1} \leq \lambda r a_{t+1}$, where $\lambda \geq 1$). If the amount of wealth exceeds the amount required to finance ($\lambda a_t \geq k_t^*$ and $\lambda r a_{t+1} \geq k_{t+1}^*$), the remaining wealth is invested at the rate r, and it's "unconstrained case". If the optimal amount of capital is lower than the amount of wealth ($\lambda a_t \leq k_t^*$ and $\lambda r a_{t+1} \leq k_{t+1}^*$), the agent borrows the maximum available amount of loan, and it is "constrained case".

In constrained case $k_t^* = \lambda a_t$ and $k_{t+1}^* = \lambda r a_{t+1}$. In unconstrained case the amount of capital invested in the business is higher than the amount of capital in constrained case: $k_t^* = (\frac{r}{\theta})^{\frac{\alpha}{\alpha-1}}$ and $k_{t+1}^* = (\frac{r}{\theta})^{\frac{\alpha}{\alpha-1}}$. And if the agent uses her personal savings $k_t \leq a_t$ and it is lower than the amount of capital in the contained case, therefore the agent always prefers to borrow instead of paying out of pocket.

Budget constraints of unconstrained incorporated self-employed worker, where $\gamma = \left(\frac{r}{\theta}\right)^{\frac{1}{\alpha-1}}$:

$$\text{Period } t\text{:} \quad a_{t+1} + c_t = \theta\gamma^\alpha - r\left(\gamma - a_t\right) - \psi_t$$

$$\text{Period } t+1\text{:} \quad c_{t+1} = \theta\gamma^\alpha - r\left(\gamma - ra_{t+1}\right) - \psi_{t+1} - \pi^{PE}_{t+1}\mathbb{1}(y_t = y^{PE}) - \pi^{ISE}_{t+1}\mathbb{1}(y_t = y^{USE})$$

Budget constraints of constrained incorporated self-employed worker:

$$\text{Period } t\text{:} \quad a_{t+1} + c_t = \theta(\lambda a_t)^\alpha - ra_t(\lambda - 1) - \psi_t$$

$$\text{Period } t+1\text{:} \quad c_{t+1} = \theta(\lambda ra_{t+1})^\alpha - ra_{t+1}(\lambda - 1) - \psi_{t+1} - \pi^{PE}_{t+1}\mathbb{1}(y_t = y^{PE}) - \pi^{ISE}_{t+1}\mathbb{1}(y_t = y^{USE})$$

Further analysis will investigate the role of the non-interest costs of borrowing in labor market choices, and for the clarity of predictions, I will consider only the unconstrained case, although the extension to the constrained case is straightforward.

If the agent works as an unincorporated self-employed worker (if $a_t \leq a_b, a_{t+1} \leq a_b$), she cannot access the credit market. In this case, she maximizes her production function by making decisions about the amount of capital invested in the business, but she uses assets as capital. In this case, the agent's budget constraints:

$$\text{Period } t\text{:} \quad a_{t+1} + c_t = \max_{k_t \leq a_t}[\theta(k_t)^\alpha]$$

$$\text{Period } t+1\text{:} \quad c_{t+1} = \max_{k_{t+1} \leq ra_{t+1}}[\theta(k_t)^\alpha] - \pi^{PE}_{t+1}\mathbb{1}(y_t = y^{PE})]$$

Budget constraints of unincorporated self-employed worker:

$$\text{Period } t\text{:} \quad a_{t+1} + c_t = \theta(a_t)^\alpha$$

$$\text{Period } t+1\text{:} \quad c_{t+1} = \theta(ra_{t+1})^\alpha - \pi^{PE}_{t+1}\mathbb{1}(y_t = y^{PE})$$

A.0.1 Utility Functions for All Possible Labor Choices

- $\{ISE_t, ISE_{t+1}\}$

$$U^{(ISE_t,ISE_{t+1})} - \epsilon^{(ISE_t,ISE_{t+1})} = (1+\beta)log\left(\frac{(r^2+1)(\theta\gamma^\alpha - r\gamma) + r^3 a_t - \psi_{t+1} - r^2\psi_t}{r^2+r}\right) + \beta log(r)$$

- $\{ISE_t, PE_{t+1}\}$

$$U^{(ISE_t,PE_{t+1})} - \epsilon^{(ISE_t,PE_{t+1})} = (1+\beta)log\left(\frac{r\theta\gamma^\alpha - r^2(\gamma - a_t) - r\psi_t + \theta w_{t+1} + b_{t+1}(1-r_b) + \psi_{t+1}}{(r+1)}\right)$$

- $\{ISE_t, USE_{t+1}\}$

$$U^{(ISE_t,USE_{t+1})} - \epsilon^{(ISE_t,USE_{t+1})} = (\alpha+\beta)log\frac{(\theta\gamma^\alpha - r(\gamma - a_t) - \psi_t)}{r+\alpha} + (\alpha+\beta)log(r) + log(\alpha^\alpha\theta)$$

- $\{PE_t, PE_{t+1}\}$

$$U^{(PE_t,PE_{t+1})} - \epsilon^{(PE_t,PE_{t+1})} = (1+\beta)log\left(\frac{r\theta w_t + ra_t + rb_t(1-r_b) + \theta w_{t+1} + b_{t+1}(1-r_b)}{r+1}\right)$$

- $\{PE_t, ISE_{t+1}\}$

$$U^{(PE_t,ISE_{t+1})} - \epsilon^{(PE_t,ISE_{t+1})} = (1+\beta)log\left(\frac{r^2\theta w_t + r^2 a_t + r^2(b_t(1-r_b) - \psi_t) + \theta\gamma^\alpha - r\gamma - \psi_{t+1} - \pi^{PE}_{t+1}}{(r^2+r)}\right) + \beta log(r)$$

- $\{PE_t, USE_{t+1}\}$ (under assumption $\pi^{PE}_{t+1} = 0$[25])

$$U^{(PE_t,USE_{t+1})} - \epsilon^{(PE_t,SE_{21t+1})} = (1+\beta)log(\theta w_t + a_t + b_t(1-r_b)) + log\frac{r}{r+\alpha} + \beta log\frac{r^\alpha\theta\alpha}{r+\alpha}$$

- $\{USE_t, USE_{t+1}\}$

$$U^{(USE_t,USE_{t+1})} - \epsilon^{(USE_t,USE_{t+1})} = (1+\beta)log(a_t^\alpha) + log\frac{r\theta}{r+\alpha} + \beta log\frac{r^{\alpha+1}\theta^2}{r+\alpha}$$

- $\{USE_t, ISE_{t+1}\}$

$$U^{(USE_t,ISE_{t+1})} - \epsilon^{(USE_t,ISE_{t+1})} = (1+\beta)log\left(\frac{r\theta(a_t)^\alpha + \theta\gamma^\alpha - r\gamma - \pi^{ISE}_{t+1} - \psi_{t+1}}{(r^2+r)}\right) + \beta log(r)$$

- $\{USE_t, PE_{t+1}\}$

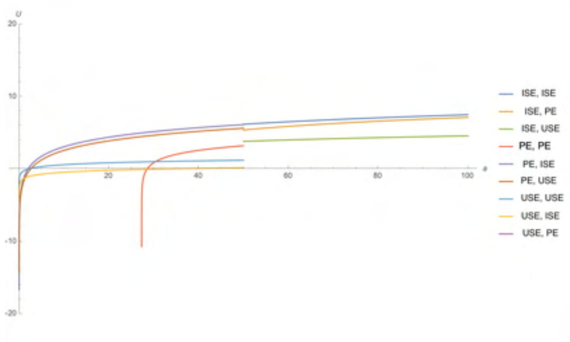
$$U^{(USE_t,PE_{t+1})} - \epsilon^{(USE_t,PE_{t+1})} = (1+\beta)log\left(\frac{r\theta(a_t)^\alpha + \theta w_{t+1} + b_{t+1}(1-r_b) - \psi_{t+1}}{(r+1)}\right)$$

[25] If $\pi^{PE}_{t+1} \neq 0$ it's impossible to find implicit functional form for $U^{(PE_t,USE_{t+1})}$.

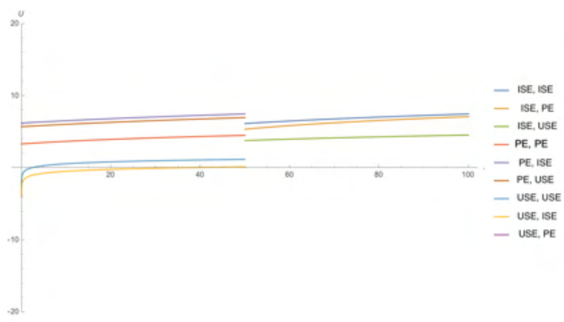
A.0.2 Relative Positions of Utility Functions

Let's consider an economy where $\beta = 0.9$, $\theta = 1.03$[26], $r = 1 + 0.025$[27], and assume that the production function θk^{α} has an decreasing returns to scale, $\alpha = 0.2$. In this economy $\gamma = 1.04$.

Fig. (a) shows the initial position of all utilities function ($a_b = 50$, $\gamma = 1.04$, $\alpha = 0.2$, $\theta = 1.03$, $r = 1 + 0.025 = 1.025$, $\beta = 0.9$, $r_b = 1.04$, $\psi_t = 0$, $\psi_{t+1} = 0$, $w_t = 0$, $w_{t+1} = 0$, $\pi^{PE}_{t+1} = 0.5$, $\pi^{SE1}_{t+1} = 0.5$). Under these assumptions about coefficients, the agent chooses $\{PE_t, SE_{t+1}\}$ if $a_t \leq a_b$ and $\{ISE_t, ISE_{t+1}\}$ if $a_t > a_b$.



a)



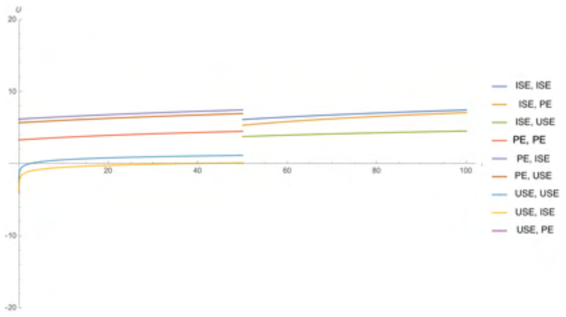
b)

Fig. (b) and (c) show positions of all utilities function if wages (for paid-employed worker) increase keeping all other parameters at the minimal level. Fig. (b) $w_t \uparrow$, Fig.(c) $w_{t+1} \uparrow$. ($a_b = 50$, $\gamma = 1.04$, $\alpha = 0.2$, $\theta = 1.03$, $r = 1 + 0.025 = 1.025$, $\beta = 0.9$, $r_b = 1.04$, $\psi_t = 0$, $\psi_{t+1} = 0$, $w_t \geq 0$ (Fig.(b)), $w_{t+1} \geq 0$ (Fig.(c)), $\pi^{PE}_{t+1} = 0.5$, $\pi^{SE1}_{t+1} = 0.5$). If wages increase in the first period, $w_t \uparrow$, the agent chooses $\{PE_t, ISE_{t+1}\}$ if $a_t \leq a_b$, and $\{ISE_t, ISE_{t+1}\}$ if $a_t > a_b$. Fig.(b) also shows such increase in $w_t \uparrow$ that the position of $U^{(PE_t, ISE_{t+1})}$ is above $U^{(ISE_t, ISE_{t+1})}$. It's possible that the agent prefers to be paid-employed worker in period t and incorporated self-employed in period $t+1$ regardless the amount of initial wealth, a_t. If wages increase in the second period, $w_{t+1} \uparrow$, the agent chooses $\{USE_t, PE_{t+1}\}$ if $a_t \leq a_b$, and $\{ISE_t, ISE_{t+1}\}$ if $a_t > a_b$. Fig.(c) shows the situation when the agent prefers to work as unincorporated self-employed at period t and paid-employed worker at period $t+1$ regardless the amount of initial wealth, a_t.

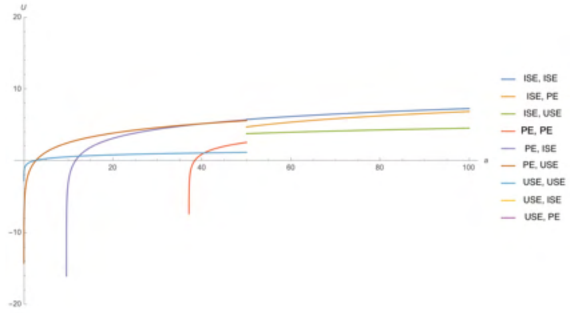
Fig. (d) and (e) show positions of all utilities function if the non-interest cost of borrowing increases keeping all other parameters at the minimal level. Fig. (d) $\psi_t \uparrow$, Fig.(e) $\psi_{t+1} \uparrow$. ($a_b = 50$, $\gamma = 1.04$, $\alpha = 0.2$, $\theta = 1.03$, $r = 1 + 0.025 = 1.025$, $\beta = 0.9$, $r_b = 1.04$, $\psi_t \uparrow$

[26] University of Groningen and University of California, Davis, Total Factor Productivity at Constant National Prices for United States, retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/RTFPNAUSA632NRUG, August 8, 2019.

[27] The Federal Reserve Interest Rate Decision on June, 2019

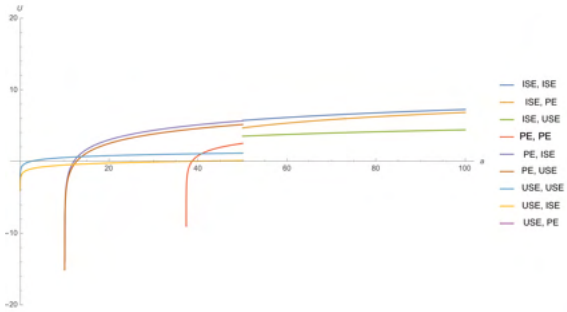


c)

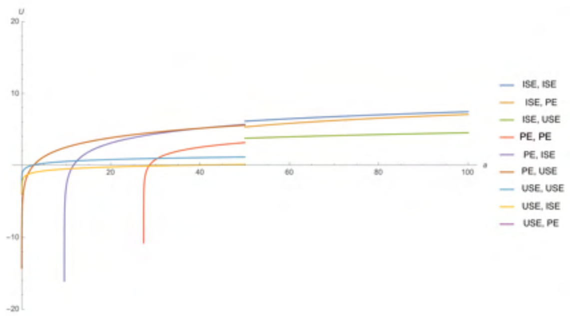


d)

(Fig.(d)), $\psi_{t+1} \uparrow$ (Fig.(e)), $w_t = 0$, $w_{t+1} = 0$, $\pi^{PE}_{t+1} = 0$, $\pi^{SE1}_{t+1} = 0$). If the non-interest cost of borrowing increases in the first period, $\psi_t \uparrow$, the agent chooses $\{USE_t, USE_{t+1}\}$ if $a_t \leq a_1 \leq a_b$, $\{PE_t, USE_{t+1}\}$ if $a_1 \leq a_t \leq a_2 \leq a_b$, $\{PE_t, ISE_{t+1}\}$ if $a_2 \leq a_b$ (where a_1 and a_2 are intersections of utility functions), and $\{ISE_t, ISE_{t+1}\}$ if $a_t > a_b$. If the non-interest cost of borrowing increases in the second period, $\psi_{t+1} \uparrow$, the agent chooses $\{USE_t, USE_{t+1}\}$ if $a_t \leq a_1 \leq a_b$, $\{PE_t, ISE_{t+1}\}$ if $a_1 \leq a_t \leq a_b$, and $\{ISE_t, ISE_{t+1}\}$ if $a_t > a_b$.



e)



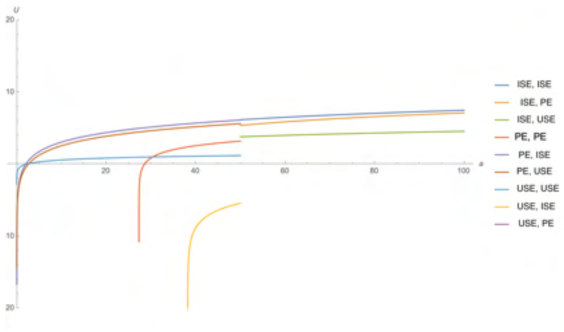
f)

Fig. (f) shows positions of all utilities function if the cost of switching from paid-employment, π^{PE}_{t+1} increases keeping all other parameters at the minimal level. ($a_b = 50$, $\gamma = 1.04$, $\alpha = 0.2$, $\theta = 1.03$, $r = 1 + 0.025 = 1.025$, $\beta = 0.9$, $r_b = 1.04$, $\psi_t = 0$, $\psi_{t+1} = 0$, $w_t = 0$, $w_{t+1} = 0$, $\pi^{PE}_{t+1} \uparrow$, $\pi^{SE1}_{t+1} = 0$). If the cost of switching from paid-employment increases, $\pi^{PE}_{t+1} \uparrow$, the agent chooses $\{USE_t, USE_{t+1}\}$ if $a_t \leq a_1 \leq a_b$, $\{PE_t, USE_{t+1}\}$ if $a_1 \leq a_t \leq a_b$, and $\{ISE_t, ISE_{t+1}\}$ if $a_t > a_b$.

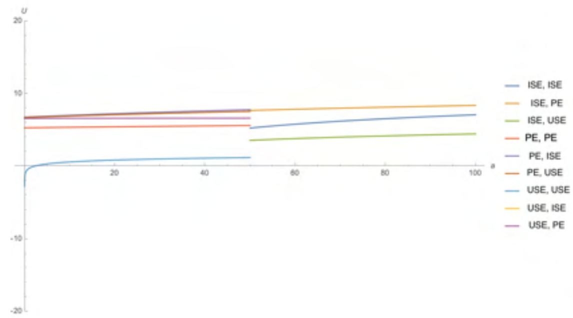
Fig. (g) shows positions of all utilities function if the cost of switching to incorporated self-employment, π^{SE1}_{t+1}, increases keeping all other parameters at the minimal level. ($a_b = 50$, $\gamma = 1.04$, $\alpha = 0.2$, $\theta = 1.03$, $r = 1+0.025 = 1.025$, $\beta = 0.9$, $r_b = 1.04$, $\psi_t = 0$, $\psi_{t+1} = 0$, $w_t = 0$, $w_{t+1} = 0$, $\pi^{PE}_{t+1} = 0$, $\pi^{SE1}_{t+1} \uparrow$). If the cost of switching to incorporated self-employment increases, $\pi^{SE1}_{t+1} \uparrow$, the agent chooses $\{USE_t, USE_{t+1}\}$ if $a_t \leq a_1 \leq a_b$, $\{PE_t, ISE_{t+1}\}$ if $a_1 \leq a_t \leq a_b$, and

$\{ISE_t, ISE_{t+1}\}$ if $a_t > a_b$.

Fig. (h) shows positions of all utilities function if all costs increase ($a_b = 50$, $\gamma = 1.04$, $\alpha = 0.2$, $\theta = 1.03$, $r = 1 + 0.025 = 1.025$, $\beta = 0.9$, $r_b = 1.04$, $\psi_t \uparrow$, $\psi_{t+1} \uparrow$, $w_t \uparrow$, $w_{t+1} \uparrow$, $\pi_{t+1}^{PE} \uparrow$, $\pi_{t+1}^{SE1} \uparrow$). In this case the agent chooses $\{PE_t, USE_{t+1}\}$ if $a_t \leq a_1 \leq a_b$, $\{PE_t, ISE_{t+1}\}$ if $a_1 \leq a_t \leq a_b$, and $\{ISE_t, PE_{t+1}\}$ if $a_t > a_b$.



g)



h)

A.0.3 The Role of Non-Interest Cost of Borrowing ψ_t in the Decision of Being Incorporated Self-Employed at Period t

The agent prefers to being as incorporated self-employed at period t and $t+1$ if she receives the highest utility function:

$$U^{(ISE_t,ISE_{t+1})} \geq \max\{U^{(PE_t,ISE_{t+1})}, U^{(USE_t,ISE_{t+1})}, U^{(PE_t,USE_{t+1})}, U^{(USE_t,USE_{t+1})}, U^{(PE_t,PE_{t+1})}, U^{(USE_t,USE_{t+1})}, U^{(ISE_t,PE_{t+1})}, U^{(ISE_t,USE_{t+1})}\}$$

To investigate the role of non-interest cost of borrowing ψ_t in the decision of being incorporated self-employed at period t, I fix a labor choice at the period $(t+1)$ as incorporated self-employed for simplicity to avoid an estimation of all nine possible labor choices. The extension of the model for other labor choices PE_{t+1} and USE_{t+1} is straightforward.

The agent prefers to works as incorporated self-employed at period t if she receives the highest utility function:

$$U^{(ISE_t,ISE_{t+1})} \geq \max\{U^{(PE_t,ISE_{t+1})}, U^{(USE_t,ISE_{t+1})}\}$$

In this subsection I show that the probability of being incorporated self-employed at period t declines if non-interest cost of borrowing at period t increases.

$$\frac{\partial Pr\left(U^{(ISE_t,ISE_{t+1})} \geq \max\{U^{(PE_t,ISE_{t+1})}, U^{(USE_t,ISE_{t+1})}\}\right)}{\partial \psi_t} \leq 0$$

The agent's task if she works as incorporated self-employed at period t and $t+1$:

$$\max_{c_t \geq 0, c_{t+1} \geq 0} \mathbb{E}U^{(ISE_t,ISE_{t+1})} = U(c_t) + \beta U(c_{t+1})$$

Agent's budget constraints ($a_t \geq a_b$ and $a_{t+1} \geq a_b$):

$$\text{Period } t\text{:} \quad a_{t+1} + c_t = \theta\gamma^{\alpha} - r(\gamma - a_t) - \psi_t \tag{9}$$

$$\text{Period } t+1\text{:} \quad c_{t+1} = \theta\gamma^{\alpha} - r(\gamma - ra_{t+1}) - \psi_{t+1} \tag{10}$$

where $\gamma = \left(\frac{r}{\lambda\theta}\right)^{\frac{1}{\alpha-1}}$, and $a_t \geq a_b$.

The first order condition for an internal solution of the agent's problem leads to the following Euler equation:

$$U^{'}(c_t) = \beta r^2 U^{'}(c_{t+1})$$

Assume that $\beta r = 1$ and $U(c_t) = log(c_t)$.

$$rc_t = c_{t+1} \tag{11}$$

Substitute the eq.11 in budget constraints eq.10 and eq.9:

$$r(\theta\gamma^{\alpha} - r(\gamma - a_t) - \psi_t - a_{t+1}) = \theta\gamma^{\alpha} - r(\gamma - ra_{t+1}) - \psi_{t+1} \Rightarrow$$
$$a_{t+1} = \frac{\theta\gamma^{\alpha}(r-1) + r\gamma + \psi_{t+1} - r^2(\gamma - a_t) - r\psi_t}{r^2 + r}$$

The agent's total utility function:

$$U^{(ISE_t, ISE_{t+1})} - \epsilon^{(ISE_t, ISE_{t+1})} = log(c_t) + \beta log(c_{t+1}) = log(c_t)(1+\beta) + \beta log(r) =$$
$$= log(\theta\gamma^{\alpha} - r(\gamma - a_t) - \psi_t - \frac{\theta\gamma^{\alpha}(r-1) + r\gamma + \psi_{t+1} - r^2(\gamma - a_t) - r\psi_t}{r^2 + r})(1+\beta) + \beta log(r) =$$
$$= log(A - \frac{\psi_{t+1}}{r^2 + r} - \frac{r^2\psi_t}{r^2 + r})(1+\beta) + \beta log(r) \Rightarrow$$
$$\frac{\partial U^{(ISE_t, ISE_{t+1})}}{\partial \psi_t} = \frac{(1+\beta)\frac{-r}{r+1}}{A - \frac{\psi_{t+1}}{r^2+r} - \frac{r^2\psi_t}{r^2+r}} \leq 0, \quad \frac{\partial U^{(ISE_t, ISE_{t+1})}}{\partial \psi_{t+1}} = \frac{(1+\beta)\frac{-1}{r^2+r}}{A - \frac{\psi_{t+1}}{r^2+r} - \frac{r^2\psi_t}{r^2+r}} \leq 0$$

where $A = \theta\gamma^{\alpha} - r(\gamma - a_t) - \frac{\theta\gamma^{\alpha}(r-1) + r\gamma - r^2(\gamma - a_t)}{r^2+r} = \frac{r^2(\theta\gamma^{\alpha} - r\gamma) + r^2 a_t(r-1) - (\theta\gamma^{\alpha} + r\gamma)}{r^2+r}$.

The increasing in the non-interest cost of borrowing in the period t or period $t+1$ decreases the total utility function.

The agent's task if she works as paid-employed at period t and as incorporated self-employed worker at period $t+1$:

$$\max_{c_t \geq 0, c_{t+1} \geq 0} \mathbb{E}U^{(PE_t, ISE_{t+1})} = U(c_t) + \beta U(c_{t+1})$$

Agent's budget constraints ($a_t \leq a_b$ and $a_{t+1} \geq a_b$):

$$\text{Period } t: \quad a_{t+1} + c_t = \theta w_t + a_t + b_t(1 - r_b) - \psi_t \tag{12}$$

$$\text{Period } t+1: \quad c_{t+1} = \theta\gamma^{\alpha} - r(\gamma - ra_{t+1}) - \psi_{t+1} - \pi^{PE}_{t+1} \tag{13}$$

The first order condition for an internal solution of the agent's problem leads to Euler equation:

$$U^{'}(c_t) = \beta r^2 U^{'}(c_{t+1})$$

Assume $\beta r = 1$, and $U(c_t) = log(c_t)$.

$$rc_t = c_{t+1} \tag{14}$$

Substitute the eq.14 in budget constraints eq.13 and eq.12 :

$$r(\theta w_t + b_t(1-r_b) - \psi_t + a_t - a_{t+1}) = \theta\gamma^\alpha - r\left(\gamma - ra_{t+1}\right) - \psi_{t+1} - \pi_{t+1}^{PE} \Rightarrow$$
$$a_{t+1} = \frac{r\theta w_t + rb_t(1-r_b) - r\psi_t + ra_t - \theta\gamma^\alpha + r\gamma + \psi_{t+1} + \pi_{t+1}^{PE}}{(r^2+r)}$$

The agent's total utility function:

$$U^{(PE_t,ISE_{t+1})} - \epsilon^{(PE_t,ISE_{t+1})} = log(c_t) + \beta log(c_{t+1}) = (1+\beta)log(c_t) + \beta log(r) = (1+\beta)\cdot$$
$$\cdot\, log\left(\theta w_t + b_t(1-r_b) - \psi_t + a_t - \frac{r\theta w_t + rb_t(1-r_b) - r\psi_t + ra_t - \theta\gamma^\alpha + r\gamma + \psi_{t+1} + \pi_{t+1}^{PE}}{(r^2+r)}\right)$$
$$+\beta log(r) =$$
$$= (1+\beta)log\left(\frac{r^2\theta w_t + r^2 a_t + r^2(b_t(1-r_b) - \psi_t) + \theta\gamma^\alpha - r\gamma - \psi_{t+1} - \pi_{t+1}^{PE}}{(r^2+r)}\right) + \beta log(r) =$$
$$= (1+\beta)log\left(G - \frac{r^2\psi_t}{(r^2+r)} - \frac{\psi_{t+1}}{(r^2+r)}\right) + \beta log(r) \Rightarrow$$
$$\frac{\partial U^{(PE_t,ISE_{t+1})}}{\partial\psi_{t+1}} = \frac{(1+\beta)\frac{-r}{r+1}}{G - \frac{\psi_{t+1}}{r^2+r} - \frac{r^2\psi_t}{r^2+r}} \leq 0; \quad \frac{\partial U^{(PE_t,ISE_{t+1})}}{\partial\psi_t} = \frac{(1+\beta)\frac{-1}{r^2+r}}{G - \frac{\psi_{t+1}}{r^2+r} - \frac{r^2\psi_t}{r^2+r}} \leq 0$$

where $G = \frac{r^2\theta w_t + r^2 a_t + r^2 b_t(1-r_b) + \theta\gamma^\alpha - r\gamma - \pi_{t+1}^{PE}}{(r^2+r)}$.

The agent's task if she worked as unincorporated self-employed at period t and as incorporated self-employed worker at period $t+1$:

$$\max_{c_t\geq 0, c_{t+1}\geq 0} U(c_t) + \beta U(c_{t+1})$$

Agent's budget constraints ($a_t \leq a_b$ and $a_{t+1} \geq a_b$):

$$\text{Period } t\text{:} \quad \text{s.t } a_{t+1} + c_t = \theta(a_t)^\alpha \tag{15}$$

$$\text{Period } t+1\text{:} \quad c_{t+1} = \theta\gamma^\alpha - r\left(\gamma - ra_{t+1}\right) - \psi_{t+1} - \pi_{t+1}^{ISE} \tag{16}$$

The first order condition for an internal solution of the agent's problem leads to Euler equation:

$$U^{'}(c_t) = \beta r^2 U^{'}(c_{t+1})$$

Under the assumption $\beta r = 1$, and $U(c_t) = log(c_t)$.

$$rc_t = c_{t+1} \tag{17}$$

Substitute the eq.17 in budget constraints eq.15 and eq.16:

$$r(\theta(a_t)^\alpha - a_{t+1}) = \theta\gamma^\alpha - r\left(\gamma - ra_{t+1}\right) - \psi_{t+1} - \pi_{t+1}^{ISE} \Rightarrow$$
$$a_{t+1} = \frac{r\theta(a_t)^\alpha - \theta\gamma^\alpha + r\gamma + \psi_{t+1} + \pi_{t+1}^{ISE}}{(r^2 + r)}$$

The agent's total utility function:

$$U^{(USE_t, ISE_{t+1})} - \epsilon^{(USE_t, ISE_{t+1})} = log(c_t) + \beta log(c_{t+1}) = (1 + \beta)log(c_t) + \beta log(r) =$$
$$= (1 + \beta)log\left(\theta(a_t)^\alpha - \frac{r\theta(a_t)^\alpha - \theta\gamma^\alpha + r\gamma + \psi_{t+1} + \pi_{t+1}^{ISE}}{(r^2 + r)}\right) + \beta log(r) =$$
$$= (1 + \beta)log\left(\frac{r\theta(a_t)^\alpha + \theta\gamma^\alpha - r\gamma - \psi_{t+1} - \pi_{t+1}^{ISE}}{(r^2 + r)}\right) + \beta log(r) =$$
$$= (1 + \beta)log\left(H - \frac{\psi_{t+1}}{(r^2 + r)}\right) + \beta log(r)$$

where $H = \frac{r\theta(a_t)^\alpha + \theta\gamma^\alpha - r\gamma - \pi_{t+1}^{ISE}}{(r^2+r)}$.

Assume that $\epsilon^j \in N(0;1)$, the probability of incorporated self-employment at period t:

$$Pr\left(U^{(ISE_t,ISE_{t+1})} \geq \max\{U^{(PE_t,ISE_{t+1})},U^{(USE_t,ISE_{t+1})}\}\right) =$$

$$= Pr\left(U^{(ISE_t,ISE_{t+1})} \geq U^{(PE_t,ISE_{t+1})}\right) \cdot Pr\left(U^{(ISE_t,ISE_{t+1})} \geq U^{(USE_t,ISE_{t+1})}\right) =$$

$$= Pr((1+\beta)log\left(A - \frac{\psi_{t+1}}{(r^2+r)} - \frac{r^2\psi_t}{(r^2+r)}\right) + \beta log(r) + \epsilon^{(ISE_t,ISE_{t+1})} \geq$$

$$\geq (1+\beta)log\left(G - \frac{r^2\psi_t}{(r^2+r)} - \frac{\psi_{t+1}}{(r^2+r)}\right) + \beta log(r) + \epsilon^{(PE_t,ISE_{t+1})}) \cdot$$

$$\cdot Pr((1+\beta)log\left(A - \frac{\psi_{t+1}}{(r^2+r)} - \frac{r^2\psi_t}{(r^2+r)}\right) + \beta log(r) + \epsilon^{(ISE_t,ISE_{t+1})} \geq$$

$$\geq (1+\beta)log\left(H - \frac{\psi_{t+1}}{(r^2+r)}\right) + \beta log(r) + \epsilon^{(PE_t,ISE_{t+1})}) =$$

$$= \Phi\left(\frac{1}{\sqrt{2}}\left(\underbrace{(1+\beta)log\left(A - \frac{\psi_{t+1}}{(r^2+r)} - \frac{r^2\psi_t}{(r^2+r)}\right) - (1+\beta)log\left(G - \frac{r^2\psi_t}{(r^2+r)} - \frac{\psi_{t+1}}{(r^2+r)}\right)}_{=J}\right)\right) \cdot$$

$$\cdot \Phi\left(\frac{1}{\sqrt{2}}\left(\underbrace{(1+\beta)log\left(A - \frac{\psi_{t+1}}{(r^2+r)} - \frac{r^2\psi_t}{(r^2+r)}\right) - (1+\beta)log\left(H - \frac{\psi_{t+1}}{(r^2+r)}\right)}_{=K}\right)\right)$$

$$\frac{\partial Pr\left(U^{(ISE_t,ISE_{t+1})} \geq \max\{U^{(PE_t,ISE_{t+1})},U^{(USE_t,ISE_{t+1})}\}\right)}{\partial\psi_t} = \Phi^{'}(J)\cdot(1+\beta)\frac{1}{\sqrt{2}}(\frac{\left(-\frac{r}{r+1}\right)}{A - \frac{\psi_{t+1}}{(r^2+r)} - \frac{r^2\psi_t}{(r^2+r)}} -$$

$$- \frac{-\frac{r}{r+1}}{G - \frac{r^2\psi_t}{(r^2+r)} - \frac{\psi_{t+1}}{(r^2+r)}}) \cdot \Phi(K) + \Phi^{'}(K)\cdot(1+\beta)\frac{1}{\sqrt{2}}\frac{\left(-\frac{r}{r+1}\right)}{A - \frac{\psi_{t+1}}{(r^2+r)} - \frac{r^2\psi_t}{(r^2+r)}} \cdot \Phi(J) =$$

$$= \underbrace{\Phi^{'}(J)\cdot(1+\beta)\frac{1}{\sqrt{2}}\left(-\frac{r}{r+1}\right)\frac{G-A}{(A - \frac{\psi_{t+1}}{(r^2+r)} - \frac{r^2\psi_t}{(r^2+r)})(G - \frac{r^2\psi_t}{(r^2+r)} - \frac{\psi_{t+1}}{(r^2+r)})}\Phi(K)}_{\leq ?\geq 0} +$$

$$+ \underbrace{\Phi^{'}(K)\cdot(1+\beta)\frac{1}{\sqrt{2}}\frac{\left(-\frac{r}{r+1}\right)}{A - \frac{\psi_{t+1}}{(r^2+r)} - \frac{r^2\psi_t}{(r^2+r)}} \cdot \Phi(J)}_{\leq 0}$$

Let's consider $(G-A)$ separately:

$$G - A = \frac{r^2\theta w_t + r^2 a_t + r^2 b_t(1-r_b) + \theta\gamma^\alpha - r\gamma - \pi^{PE}_{t+1}}{(r^2+r)} - \frac{r^2(\theta\gamma^\alpha - r\gamma) + r^2 a_t(r-1) - (\theta\gamma^\alpha + r\gamma)}{r^2+r} =$$

$$= \frac{r^2\theta w_t + a_t r^2(2-r) + r^2 b_t(1-r_b) + \theta\gamma^\alpha(2-r^2) + r^2\gamma - \pi^{PE}_{t+1}}{r^2+r}$$

It can be assumed that $r \leq 1.4 \leq \sqrt{2}$, because r is one plus the interest rate, and probably the interest rate will not exceed 40%. Also it can be assumed that $\pi^{PE}_{t+1} \leq r^2\theta w_t + a_t r^2(2-r) + r^2 b_t(1-r_b) + \theta$
So $G - A \geq 0 \Rightarrow G \geq A$.

If $G \geq A$, then $|\frac{\partial U^{(ISE_t,ISE_{t+1})}}{\partial\psi_{t+1}}| \geq |\frac{\partial U^{(PE_t,ISE_{t+1})}}{\partial\psi_{t+1}}|$, and

$$\frac{\partial Pr\left(U^{(ISE_t, ISE_{t+1})} \geq \max\{U^{(PE_t, ISE_{t+1})}, U^{(USE_t, ISE_{t+1})}\}\right)}{\partial \psi_t} \leq 0.$$

The probability of being incorporated self-employment at period t declines if non-interest cost of borrowing ψ_t increases.

A.0.4 The Role of Non-Interest Cost of Borrowing ψ_{t+1} in the Decision of Being Incorporated Self-Employed at Period $t+1$

To investigate the role of the non-interest cost of borrowing ψ_{t+1} in the decision of being incorporated self-employed at period $t+1$, I fix a labor choice at the period t as incorporated self-employed for simplicity to avoid estimation of all nine possible labor choices. The extension of the model for other labor choices PE_t and USE_t is straightforward.

The agent prefers to work as incorporated self-employed at period $t+1$ if she receives the highest utility function:

$$U^{(ISE_t, ISE_{t+1})} \geq \max\{U^{(ISE_t, PE_{t+1})}, U^{(ISE_t, USE_{t+1})}\}$$

In this subsection, I show that the probability of being incorporated self-employed at period $t+1$ declines if the non-interest cost of borrowing at period $t+1$ increases.

$$\frac{\partial Pr\left(U^{(ISE_t, ISE_{t+1})} \geq \max\{U^{(ISE_t, PE_{t+1})}, U^{(ISE_t, USE_{t+1})}\}\right)}{\partial \psi_{t+1}} \leq 0$$

The total utility of being incorporated self-employed at period t and $t+1$ is described in the previous subsection.

The agent's task if she works as incorporated self-employed at period t and as paid-employed worker at period $t+1$:

$$\max_{c_t \geq 0, c_{t+1} \geq 0} U(c_t) + \beta U(c_{t+1})$$

Agent's budget constraints ($a_t \geq a_b$ and $a_{t+1} \leq a_b$):

$$\text{Period } t: \quad a_{t+1} + c_t = \theta \gamma^{\alpha} - r\left(\gamma - a_t\right) - \psi_t \tag{18}$$

$$\text{Period } t+1: \quad c_{t+1} = \theta w_{t+1} + r a_{t+1} + b_{t+1}(1 - r_b) - \psi_{t+1} \tag{19}$$

The first-order condition for an internal solution of the agent's problem leads to Euler equation:

$$U^{'}(c_t) = \beta r U^{'}(c_{t+1})$$

Under the assumption $\beta r = 1$, and $U(c_t) = log(c_t)$.

$$c_t = c_{t+1} \tag{20}$$

Substitute the eq.20 in budget constraints eq.18 and eq.19:

$$\theta\gamma^{\alpha} - r\left(\gamma - a_t\right) - \psi_t - a_{t+1} = \theta w_{t+1} + r a_{t+1} + b_{t+1}(1 - r_b) - \psi_{t+1} \Rightarrow$$
$$a_{t+1} = \frac{\theta\gamma^{\alpha} - r\left(\gamma - a_t\right) - \psi_t - \theta w_{t+1} - b_{t+1}(1 - r_b) + \psi_{t+1}}{(r+1)}$$

The agent's total utility function:

$$U^{(ISE_t,PE_{t+1})} - \epsilon^{(ISE_t,PE_{t+1})} = log(c_t) + \beta log(c_{t+1}) = (1+\beta)log(c_t) =$$
$$= (1+\beta)log\left(\theta\gamma^{\alpha} - r\left(\gamma - a_t\right) - \psi_t - \frac{\theta\gamma^{\alpha} - r(\gamma - a_t) - \psi_t - \theta w_{t+1} - b_{t+1}(1 - r_b) + \psi_{t+1}}{(r+1)}\right) =$$
$$= (1+\beta)log(L - \frac{\psi_{t+1}}{r+1} - \frac{r\psi_t}{r+1}) \Rightarrow \frac{\partial U^{(ISE_t,PE_{t+1})}}{\partial \psi_t} \leq 0; \ \frac{\partial U^{(ISE_t,PE_{t+1})}}{\partial \psi_{t+1}} \leq 0$$

where $L = \theta\gamma^{\alpha} - r\left(\gamma - a_t\right) - \frac{\theta\gamma^{\alpha} - r(\gamma - a_t) - \theta w_{t+1} - b_{t+1}(1-r_b)}{(r+1)} = \frac{r(\theta\gamma^{\alpha} - r\gamma) + r^2 a_t + \theta w_{t+1} + b_{t+1}(1-r_b)}{(r+1)} = \frac{r^2(\theta\gamma^{\alpha} - r\gamma) + r^3 a_t + r\theta w_{t+1} + r b_{t+1}(1-r_b)}{(r^2+r)} = A - \frac{r^3 a_t - (\theta\gamma^{\alpha} - r\gamma) - r\theta w_{t+1} - r b_{t+1}(1-r_b)}{(r^2+r)}$

The agent's task if she works as incorporated self-employed at period t and as unincorporated self-employed worker at period $t+1$:

$$\max_{c_t \geq 0, c_{t+1} \geq 0} U(c_t) + \beta U(c_{t+1})$$

Agent's budget constraints ($a_t \geq a_b$ and $a_{t+1} \leq a_b$):

$$\text{Period } t\text{:} \quad a_{t+1} + c_t = \theta\gamma^{\alpha} - r\left(\gamma - a_t\right) - \psi_t \tag{21}$$

$$\text{Period } t+1\text{:} \quad c_{t+1} = \theta(r a_{t+1})^{\alpha} \tag{22}$$

The first-order condition for an internal solution of the consumer's problem leads to Euler equation:

$$U^{'}(c_t) = \beta r U^{'}(c_{t+1})(r^{\alpha-1}\theta\alpha(\theta\gamma^{\alpha} - r(\gamma - a_t) - \psi_t - c_t)^{\alpha-1})$$

Under the assumption $\beta r = 1$, $M = [\theta\gamma^{\alpha} - r(\gamma - a_t) - \psi_t]$ and $U(c_t) = log(c_t)$.

$$r^{\alpha-1}\theta_2\alpha(M - c_t)^{\alpha-1}c_t = c_{t+1} \tag{23}$$

Substitute the eq.23 in budget constraints eq.21 and eq.22:

$$\theta(r a_{t+1})^{\alpha} = r^{\alpha-1}\theta\alpha(a_{t+1})^{\alpha-1}(M - a_{t+1}) \Rightarrow a_{t+1} = \frac{\alpha M}{\alpha + r}$$

The agent's total utility function:

$$U^{(ISE_t,USE_{t+1})} - \epsilon^{(ISE_t,USE_{t+1})} = log(c_t) + \beta log(c_{t+1}) = (1+\beta)log(c_t) + log(r^{\alpha-1}\theta\alpha(M-c_t)^{\alpha-1}) =$$

$$= \underbrace{(1+\beta)log\frac{rM}{r+\alpha} + (\alpha-1)log\frac{\alpha M}{r+\alpha} + log(r^{\alpha-1}\theta\alpha)}_{=N\geq 0} \Rightarrow \frac{\partial U^{(ISE_t,USE_{t+1})}}{\partial \psi_t} \leq 0$$

Assume that $\epsilon^j \in N(0;1)$, the probability of being incorporated self-employment at period $t+1$:

$$Pr\left(U^{(ISE_t,ISE_{t+1})} \geq \max\{U^{(ISE_t,PE_{t+1})}, U^{(ISE_t,USE_{t+1})}\}\right) =$$

$$Pr\left(U^{(ISE_t,ISE_{t+1})} \geq U^{(ISE_t,PE_{t+1})}\right) \cdot Pr\left(U^{(ISE_t,ISE_{t+1})} \geq U^{(ISE_t,USE_{t+1})}\right) =$$

$$= Pr((1+\beta)log\left(A - \frac{\psi_{t+1}}{(r^2+r)} - \frac{r^2\psi_t}{(r^2+r)}\right) + \beta log(r) + \epsilon^{(ISE_t,ISE_{t+1})} \geq$$

$$\geq (1+\beta)log\left(L - \frac{\psi_{t+1}}{r+1} - \frac{r\psi_t}{r+1}\right) + \epsilon^{(ISE_t,PE_{t+1})})\cdot$$

$$\cdot Pr((1+\beta)log\left(A - \frac{\psi_{t+1}}{(r^2+r)} - \frac{r^2\psi_t}{(r^2+r)}\right) + \beta log(r) + \epsilon^{(ISE_t,ISE_{t+1})} \geq$$

$$\geq N + \epsilon^{(ISE_t,USE_{t+1})}) =$$

$$= \Phi\left(\frac{1}{\sqrt{2}}\left(\underbrace{(1+\beta)log\left(A - \frac{\psi_{t+1}}{(r^2+r)} - \frac{r^2\psi_t}{(r^2+r)}\right) + \beta log(r) - (1+\beta)log\left(L - \frac{\psi_{t+1}}{r+1} - \frac{r\psi_t}{r+1}\right)}_{=O}\right)\right)\cdot$$

$$\cdot \Phi\left(\frac{1}{\sqrt{2}}\left(\underbrace{(1+\beta)log\left(A - \frac{\psi_{t+1}}{(r^2+r)} - \frac{r^2\psi_t}{(r^2+r)}\right) + \beta log(r) - N}_{=P}\right)\right)$$

$$\frac{\partial Pr\left(U^{(ISE_t,ISE_{t+1})} \geq \max\{U^{(ISE_t,PE_{t+1})}, U^{(ISE_t,USE_{t+1})}\}\right)}{\partial \psi_{t+1}} = \Phi^{'}(O)\cdot(1+\beta)\frac{1}{\sqrt{2}}(\frac{\left(-\frac{1}{r^2+r}\right)}{A - \frac{\psi_{t+1}}{(r^2+r)} - \frac{r^2\psi_t}{(r^2+r)}} -$$

$$- \frac{\left(-\frac{1}{r+1}\right)}{L - \frac{\psi_{t+1}}{r+1} - \frac{r\psi_t}{r+1}})\cdot\Phi(P) + \Phi^{'}(P)\cdot(1+\beta)\frac{1}{\sqrt{2}}\frac{\left(-\frac{1}{r^2+1}\right)}{A - \frac{\psi_{t+1}}{(r^2+r)} - \frac{r^2\psi_t}{(r^2+r)}}\cdot\Phi(O) =$$

$$= \Phi^{'}(O)\cdot(1+\beta)\frac{1}{\sqrt{2}}\left(\frac{-1}{r^2+r}\right)\left(\frac{(L - \frac{\psi_{t+1}}{r+1} - \frac{r\psi_t}{r+1}) - r(A - \frac{\psi_{t+1}}{(r^2+r)} - \frac{r^2\psi_t}{(r^2+r)})}{(A - \frac{\psi_{t+1}}{(r^2+r)} - \frac{r^2\psi_t}{(r^2+r)})(L - \frac{\psi_{t+1}}{r+1} - \frac{r\psi_t}{r+1})}\right)\cdot\Phi(P) +$$

$$+ \Phi^{'}(P)\cdot(1+\beta)\frac{1}{\sqrt{2}}\frac{\left(-\frac{1}{r^2+1}\right)}{A - \frac{\psi_{t+1}}{(r^2+r)} - \frac{r^2\psi_t}{(r^2+r)}}\cdot\Phi(O) =$$

$$= \underbrace{\Phi^{'}(O)\cdot(1+\beta)\frac{1}{\sqrt{2}}(\frac{L - rA}{(A - \frac{\psi_{t+1}}{(r^2+r)} - \frac{r^2\psi_t}{(r^2+r)})(L - \frac{\psi_{t+1}}{r+1} - \frac{r\psi_t}{r+1})})\cdot\Phi(P)}_{\leq ?\geq 0} +$$

$$+ \underbrace{\Phi^{'}(P)\cdot(1+\beta)\frac{1}{\sqrt{2}}\frac{\left(-\frac{1}{r^2+1}\right)}{A - \frac{\psi_{t+1}}{(r^2+r)} - \frac{r^2\psi_t}{(r^2+r)}}\cdot\Phi(O)}_{\leq 0}$$

Let's consider $(L - rA)$ separately:

$$L - rA = \frac{r^2(\theta\gamma^\alpha - r\gamma) + r^3 a_t + r\theta w_{t+1} + r b_{t+1}(1 - r_b)}{(r^2 + r)} -$$
$$- r\frac{r^2(\theta\gamma^\alpha - r\gamma) + r^2 a_t(r - 1) - (\theta\gamma^\alpha + r\gamma)}{r^2 + r} =$$
$$= \frac{r^2(\theta\gamma^\alpha - r\gamma)(r - 1) + r^4 a_t + r\theta w_{t+1} + r b_{t+1}(1 - r_b)}{(r^2 + r)} \geq 0$$

It means:

$$\frac{\partial Pr\left(U^{(ISE_t, ISE_{t+1})} \geq \max\{U^{(ISE_t, PE_{t+1})}, U^{(ISE_t, USE_{t+1})}\}\right)}{\partial \psi_{t+1}} \leq 0$$

The probability of being incorporated self-employment at period $t+1$ declines if non-interest cost of borrowing ψ_{t+1} increases.

A.0.5 The Role of Non-Interest Cost of Borrowing ψ_{t+1} in the Decision of Being Paid-Employed at Period $t+1$

In this subsection, I want to show that if the agent is an unincorporated self-employed worker (which means she does not have access to the credit market), she may prefer to switch to paid employment to be able to use credit products (e.g., mortgage) if the non-interest cost of borrowing declines. To show it I complicate the model by allowing employees to take a one-period loan in the bank. Paid-employed worker budget constraints:

$$\text{Period } t: \quad a_{t+1} + c_t = \theta w_t + a_t + b_t(1 - r_b) - \psi_t$$

$$\text{Period } t+1: \quad c_{t+1} = \theta w_{t+1} + r a_{t+1} + b_{t+1}(1 - r_b) - \psi_{t+1}$$

To investigate the role of non-interest cost of borrowing ψ_{t+1} in the decision of unincorporated self-employed at period t to switch to paid-employment at period $t+1$ - $USE_t \rightarrow PE_{t+1}$, I consider all possible labor choices that the agent has at period $t+1$ if she was unincorporated self-employed at period t -it's ISE_{t+1} and USE_{t+1}:

$$Pr(USE_t \rightarrow PE_{t+1}) = Pr(U^{(USE_t, PE_{t+1})} \geq \max\{U^{(USE_t, USE_{t+1})}, U^{(USE_t, ISE_{t+1})}\}$$

In this subsection I show that the probability of switching $USE_t \rightarrow PE_{t+1}$ declines if the non-interest cost of borrowing at period $t+1$ increases.

$$\frac{\partial Pr(USE_t \rightarrow PE_{t+1})}{\partial \psi_{t+1}} \leq 0$$

The agent's utility function if she works as an unincorporated self-employed worker at period t and a paid-employed worker at period $t+1$:

$$U^{(USE_t, PE_{t+1})} - \epsilon^{(USE_t, PE_{t+1})} = (1 + \beta)log\left(\frac{r\theta(a_t)^{\alpha} + \theta w_{t+1} + b_{t+1}(1 - r_b) - \psi_{t+1}}{(r+1)}\right) =$$

$$= (1 + \beta)log\left(Q - \frac{\psi_{t+1}}{(r+1)}\right)$$

where $Q = \frac{r\theta(a_t)^{\alpha} + \theta w_{t+1} + b_{t+1}(1 - r_b)}{(r+1)}$

The agent's utility function if she works as an unincorporated self-employed worker at period t and $t+1$:

$$U^{(USE_t, USE_{t+1})} - \epsilon^{(USE_t, USE_{t+1})} = (1 + \beta)log(a_t^{\alpha}) + log\frac{r\theta}{r + \alpha} + \beta log\frac{r^{\alpha+1}\theta^2}{r + \alpha} = R$$

where $R = (1 + \beta)log(a_t^{\alpha}) + log\frac{r\theta}{r+\alpha} + \beta log\frac{r^{\alpha+1}\theta^2}{r+\alpha}$

The agent's utility function if she works as an unincorporated self-employed worker at period t and incorporated self-employed worker at period $t+1$:

$$U^{(USE_t,ISE_{t+1})} - \epsilon^{(USE_t,ISE_{t+1})} = (1+\beta)log\left(\frac{r\theta(a_t)^{\alpha} + \theta\gamma^{\alpha} - r\gamma - \pi^{ISE}_{t+1} - \psi_{t+1}}{(r^2+r)}\right) + log(r) =$$

$$= (1+\beta)log\left(S - \frac{\psi_{t+1}}{(r^2+r)}\right) + log(r)$$

where $S = \frac{r\theta(a_t)^{\alpha}+\theta\gamma^{\alpha}-r\gamma-\pi^{ISE}_{t+1}}{(r^2+r)}$

Assume that $\epsilon^j \in N(0;1)$, the probability of being unincorporated self-employment at period t and paid-employed worker at period $t+1$:

$$Pr\left(U^{(USE_t,PE_{t+1})} \geq \max\{U^{(USE_t,USE_{t+1})}, U^{(USE_t,ISE_{t+1})}\}\right) =$$

$$Pr\left(U^{(USE_t,PE_{t+1})} \geq U^{(USE_t,USE_{t+1})}\right) \cdot Pr\left(U^{(USE_t,PE_{t+1})} \geq U^{(USE_t,ISE_{t+1})}\right) =$$

$$= Pr((1+\beta)log\left(Q - \frac{\psi_{t+1}}{(r+1)}\right) + \epsilon^{(USE_t,PE_{t+1})} \geq R + \epsilon^{(USE_t,USE_{t+1})}\cdot$$

$$\cdot Pr((1+\beta)log\left(Q - \frac{\psi_{t+1}}{(r+1)}\right) + \epsilon^{(USE_t,PE_{t+1})} \geq (1+\beta)log\left(S - \frac{\psi_{t+1}}{(r^2+r)}\right) + log(r) + \epsilon^{(USE_t,ISE_{t+1})}) =$$

$$= \underbrace{\Phi\left(\frac{1}{\sqrt{2}}\left((1+\beta)log\left(Q - \frac{\psi_{t+1}}{(r+1)}\right) - R\right)\right)}_{T\geq 0}\cdot$$

$$\cdot \underbrace{\Phi\left(\frac{1}{\sqrt{2}}\left((1+\beta)log\left(Q - \frac{\psi_{t+1}}{(r+1)}\right) - (1+\beta)log\left(S - \frac{\psi_{t+1}}{(r^2+r)}\right) + log(r)\right)\right)}_{U\geq 0}$$

$$\frac{\partial Pr\left(U^{(USE_t,PE_{t+1})} \geq \max\{U^{(USE_t,USE_{t+1})}, U^{(USE_t,ISE_{t+1})}\}\right)}{\partial \psi_{t+1}} = \underbrace{\Phi'(T)\cdot(1+\beta)\frac{1}{\sqrt{2}}\left(\frac{-\frac{1}{r+1}}{Q - \frac{\psi_{t+1}}{(r+1)}}\right)\cdot\Phi(U)}_{\leq 0} +$$

$$+ \underbrace{\Phi'(U)\cdot(1+\beta)\frac{1}{\sqrt{2}}\left(\frac{-\frac{1}{r+1}}{Q - \frac{\psi_{t+1}}{(r+1)}} - \frac{\left(-\frac{1}{r^2+r}\right)}{S - \frac{\psi_{t+1}}{(r^2+r)}}\right)\cdot\Phi(T)}_{\leq ?\geq 0}$$

Let's consider the last term separately:

$$\left(\frac{-\frac{1}{r+1}}{Q - \frac{\psi_{t+1}}{(r+1)}} - \frac{\left(-\frac{1}{r^2+r}\right)}{S - \frac{\psi_{t+1}}{(r^2+r)}}\right) = \frac{-\frac{1}{r+1}\cdot\left(S - \frac{\psi_{t+1}}{(r^2+r)}\right) + \frac{1}{r^2+r}\cdot\left(Q - \frac{\psi_{t+1}}{(r+1)}\right)}{\left(Q - \frac{\psi_{t+1}}{(r+1)}\right)\left(S - \frac{\psi_{t+1}}{(r^2+r)}\right)} =$$

$$= \frac{r\theta(a_t)^{\alpha} + \theta w_{t+1} + b_{t+1}(1-r_b) - \psi_{t+1} - r\theta(a_t)^{\alpha} - \theta\gamma^{\alpha} + r\gamma + \pi^{ISE}_{t+1} + \psi_{t+1}}{(r+1)(r^2+r)\left(Q - \frac{\psi_{t+1}}{(r+1)}\right)\left(S - \frac{\psi_{t+1}}{(r^2+r)}\right)} =$$

$$= \frac{\theta w_{t+1} + b_{t+1}(1-r_b) - \theta\gamma^{\alpha} + r\gamma + \pi^{ISE}_{t+1}}{(r+1)(r^2+r)\left(Q - \frac{\psi_{t+1}}{(r+1)}\right)\left(S - \frac{\psi_{t+1}}{(r^2+r)}\right)} = \frac{(c_{PE_{t+1}} - ra_{t+1} - \psi_{t+1}) - (c_{ISE_{t+1}} - ra_{t+1} - \psi_{t+1})}{(r+1)(r^2+r)\left(Q - \frac{\psi_{t+1}}{(r+1)}\right)\left(S - \frac{\psi_{t+1}}{(r^2+r)}\right)} =$$

$$\frac{c_{PE_{t+1}} - c_{ISE_{t+1}}}{(r+1)(r^2+r)\left(Q - \frac{\psi_{t+1}}{(r+1)}\right)\left(S - \frac{\psi_{t+1}}{(r^2+r)}\right)} = \frac{exp(U_{PE_{t+1}}) - exp(U_{ISE_{t+1}})}{(r+1)(r^2+r)\left(Q - \frac{\psi_{t+1}}{(r+1)}\right)\left(S - \frac{\psi_{t+1}}{(r^2+r)}\right)} \leq 0$$

The nominator is negative because the agent receives higher utility being incorporated self-employed that being paid-employed worker $U_{ISE_{t+1}} \geq U_{PE_{t+1}}$ if all else being equal.

The probability of switching form incorporated self-employment at period t to paid-employed at period $t+1$ increases if non-interest cost of borrowing ψ_{t+1} increases.

$$\frac{\partial Pr\left(U^{(USE_t, PE_{t+1})} \geq \max\{U^{(USE_t, USE_{t+1})}, U^{(USE_t, ISE_{t+1})}\}\right)}{\partial \psi_{t+1}} \geq 0$$

B Dynamic Choice Model

B.1 The Timeline of the Model

In the empirical model at the beginning of period t, the individual i living in area j can be in one of four employment states k, Y_{ikjt}: paid-employment ($k = 0$), incorporated self-employment ($k = 1$), unincorporated self-employment ($k = 2$) or non-employment ($k = 3$).

The individual i also has a set of time-varying individual characteristics X_{ijt} (e.g., age, years of schooling, marital status, the value of large, durable assets, local economic conditions,j, and etc.) and constant individual characteristics X_i (e.g., gender, race and Hispanic ethnicity), both sets denoted as X_{ijt} in the model. The agent observes the local credit supply shock occurred between period t and $t - 1$, p_{ijt}.

The accessibility of bank services may create incentives for the individual i to change the labor state. For example, if the use of bank services becomes cheaper for the self-employed, a person may make a decision to become self-employed and open a new business; or if self-employed individual i does not satisfy bank requirements (e.g., annual income requirements, collateral requirements, credit score, etc.), the person may make a decision to switch to paid employment to be able to use bank services. At the beginning of period t, one would observe a new employment state for individual i.

B.2 Dynamic Multinomial Logit Model of Employment

The purpose of this empirical analysis is to show how the local credit supply shocks, p_{ijt}, with explanatory variables, X_{ijt}, influence the transition between employment states $Y_{ikjt} \rightarrow Y_{ikjt+1}$, and I model this transition using the dynamic multinomial logit process with state dependence and unobserved heterogeneity. A similar approach to model the determinants of transitions between labor market states was implemented by Prowse (2012), Cowling and Wooden (2021), and Lechmann and Wunder (2017).

The probability of employment choice k can be expressed as follows:

$$Pr_{t+1}(k|X_{ijt}, p_{ijt}, \mu_{ik}) = \frac{exp(Y_{ikjt}\beta_{1k} + p_{ijt}\beta_{2k} + X_{ijt}\beta_{3k} + \mu_{ijk})}{\sum_{k' \in \{PE, SE_1, SE_2, NE\}} exp(Y_{ik't}\beta_{1k'} + p_{ijt}\beta_{2k'} + X_{ijt}\beta_{3k'} + \mu_{ijk'})} \quad (24)$$

The probability of employment choice k depends on the previous period job type, Y_{ikjt}, and the coefficient β_{1k} shows the mobility of individuals across employment states between t and $t + 1$. The error term, μ_{ik}, represents time-variant unobservables. As a proxy for non-interest borrowing costs, I use the measure of credit supply shock, p_{ijt}. A higher value of the CMA implies closer proximity to banking services. Since banks perform many functions, I assume that future unobserved shocks do not influence the availability of banking services and an individual's job choice in a given year. The credit supply shock, p_{ijt}, is assumed to be exogenous conditional on the time-constant unobserved effect (μ_{ijk}) and time-varying observed factors that may influence the opening of new bank offices in a given area, but I discuss later some potential solutions to possible violations of this assumption. β_{2k} is another coefficient of interest that shows the direct effect of the CMA on the type of employment. The agent's ability, θ, the value of available assets, a_{t+1}, and the wage of a paid employed worker, w_{t+1} are included in X_{ijt}, and I will discuss actual measures for these variables later.

The model can be re-written in a more conventional log-odds form by choosing the non-employed type as the base category:

$$ln\frac{Pr(Y_{ijt+1} = m,\ m \in \{PE, SE_1, SE_2\})}{Pr(Y_{it+1} = NE)} = Y_{ikjt}\beta_{1k} + p_{ijt}\beta_{2k} + X_{ijt}\beta_{3k} + \mu_{ijk} \tag{25}$$

In Equation 25, I allow unobserved heterogeneity μ_{ijk} to be correlated across employment states.

B.3 Identification

Several issues arise in estimating Equation 25. The first one is the possible correlation of the unobserved heterogeneity, μ_{ijk}, with explanatory variables. Following the literature of Murtazashvili and Wooldridge (2016), Papke and Wooldridge (2008), I add the Mundlak-Chamberlain device or the longitudinal average of time-varying explanatory variables, X_{ijt}, which models the permanent unobserved heterogeneity as a linear projection of the time average of time-varying characteristics. This approach accounts for the endogeneity of inputs with regard to unobserved factors like an individual fixed-effect model. Since the panel dataset used here is unbalanced, it is impossible to apply the original approach suggested by Wooldridge (2005) and include a complete history of lagged explanatory variables, $X_t^+ = (X_{ij2}, ..., X_{ijT})$ and $p_t^+ = (p_{ij2}, ..., p_{ijT})$. But I can apply the modified version of Wooldridge (2005) proposed by Rabe-Hesketh and Skrondal (2013), who suggests using the within-means of individual time-varying characteristics based on all periods excluding the first period observations, $\overline{X_i^+} = \frac{1}{T-1}\sum_{t=2}^{T} X_{ijt}$ and $\overline{p_i^+} = \frac{1}{T-1}\sum_{t=2}^{T} p_{ijt}$. $\overline{X_i^+}$ and $\overline{Z_i^+}$ are an analog to the Mundlak-Chamberlain device but without initial values. This approach accounts for the possible endogeneity of the CMA with regard to unobserved factors in a manner similar to the fixed-effect model.

The second issue is related to the initial conditions and occurs due to the correlation between μ_{ijk} and the initial observation Y_{ikj1}, and the endogenous initial conditions require a specification of the conditional distribution for Y_{ikj1}. I address this problem using the solution suggested by Wooldridge (2005) and Rabe-Hesketh and Skrondal (2013). This solution specifies the conditional distribution of μ_{ijk} via an auxiliary model that includes an initial dependent variable, Y_{ikj1}, the initial-period explanatory variables, X_{ij1} and p_{ij1}, and within-means of individual time-varying characteristics based on all periods excluding the first period observations, $\overline{X_i^+}$. The model specifies the following conditional density of the unobserved heterogeneity:

$$\mu_{ijk} = \pi_{1k} Y_{ijk1} + \overline{X_{ij}^+}\pi_{2k} + \pi_{3k} X_{ij1} + \overline{p_{ij}^+}\pi_{4k} + \pi_{5k} p_{ij1} + \eta_{ijk} = G_{ij}\pi_k + \eta_{ijk} \tag{26}$$

where $\overline{X_{ij}^+}$ is the within-means of individual time-varying characteristics based on all periods excluding the first period observations $\overline{X_{ij}^+} = \frac{1}{T-1}\sum_{t=2}^{T} X_{ijt}$; Y_{ikj1} is the initial-period dependent variable; X_{ij1} is the initial-period explanatory variables; $\overline{p_{ij}^+}$ is the within-means of the CMA index based on all periods excluding the first period observations $\overline{p_{ij}^+} = \frac{1}{T-1}\sum_{t=2}^{T} p_{it}$; p_{i1} is the initial-period CMA; η_{ik} is the error term where $Cov(\eta_{ijk}; \mu_{ikjt+1}) = 0$. The vector G_{ij} consists of the initial dependent variable, initial explanatory variables, and within-means of explanatory variables in subsequent periods. The approach suggested by Wooldridge (2005) and Rabe-Hesketh and Skrondal (2013) has several advantages. It does not require instruments because the initial conditions are not modeled separately (in contrast to Heckman (1987) approach). It can be applied to unbalanced panels, and the within-means terms $\overline{X_{ij}^+}$ and $\overline{p_{ij}^+}$ allow for the correlations between the explanatory variables and the unobserved heterogeneity, μ_{ijk}.

The substitution of Equation 26 into Equation 25 leads to the standard random-effects multinomial logit model:

$$ln\frac{Pr(Y_{ijt+1} = m,\ m \in \{PE, SE_1, SE_2)\})}{Pr(Y_{ijt+1} = NE)} = Y_{ikjt}\beta_{1k} + p_{ijt}\beta_{2k} + X_{ijt}\beta_{3k} + G_{ij}\pi_k + \eta_{ikj} \tag{27}$$

To estimate the model with the unobserved heterogeneity, η_{ikj}, I use a parametrical approach, assuming multivariate normality of the error term, $\eta_{ij} \sim N(0; \sigma_\eta^2)$.

B.4 Results

The results of the main model are reported in Table 14. The table presents the relative risk ratios from the dynamic multinomial logit model of employment with correlated random effects[28]. The

[28]Coefficients in a result of multinomial logit model don't have a direct interpretation as they are relative to the base category, and in this case using of the relative risk ratios (RRR) are more justified. The RRR of a coefficient indicates how the risk of the outcome falling in the comparison group compared to the risk of the outcome falling in the referent group changes with the variable in question. An RRR $>$ 1 indicates that the risk of the outcome falling in the comparison group relative to the risk of the outcome falling in the referent

dependent variable is the employment state at $t + 1$, with the non-employment state chosen as the base outcome and the paid employment state selected as the omitted lagged dependent category.

The previous labor status plays an essential role in the current job type. Predictably, for self-employed and non-employed individuals, the risk of working in the paid employment sector in period $t + 1$ relative to paid employed workers is small and negative (the odds are 0.191, 0.181, and 0.040, respectively). For both types of self-employed, the previous job in the self-employed sector increases the risk of being either type of self-employed. The relative risk ratio of being incorporated self-employed in period $t + 1$ relative to paid employed workers in period t is 50.034 for the incorporated self-employed and 13.791 for the unincorporated self-employed. The non-employed are less likely to switch to any self-employment. The relative risk ratio of being unincorporated self-employed in period $t + 1$ relative to paid employed workers in period t for the incorporated self-employed and 23.329 for the unincorporated self-employed.

The result of primary interest is the impact of the supply shock index. Specifically, a rise of one standard deviation in this index leads to a 1.037 increase in the likelihood of being paid employed. Conversely, this same increase in the supply shock index reduces the probability of becoming incorporated self-employed by 0.108. Additionally, a rise in the supply shock index heightens the chance of becoming unincorporated self-employed by 1.098.

The value of large, durable assets plays an essential role in being self-employed and increases the risk of being incorporated self-employed or unincorporated self-employed, but it does not significantly influence the risk of being a paid employed worker. Furthermore, the difference in log income for wage and salary workers and self-employed workers increases the risk of being paid employed, decreases the risk of being unincorporated self-employed, and does not influence the risk of being incorporated self-employed.

To show the average marginal effect (AME) of a one-unit improvement in the supply shock index on the size of sectors and the transition probabilities, I re-estimate the main model, including the interaction terms between the employment state, Y_{ikjt}, and the supply shock

group increases as the variable increases. In other words, the comparison outcome is more likely. An RRR < 1 indicates that the risk of the outcome falling in the comparison group relative to the risk of the outcome falling in the referent group decreases as the variable increases.

index, p_{ijt}:

$$ln\frac{Pr(Y_{ijt+1}=m,\ m\in\{PE,SE_1,SE_2\})}{Pr(Y_{ijt+1}=NE)}=Y_{ikjt}\beta_{1kj}+p_{ijt}\beta_{2kj}+Y_{ikjt}*p_{ijt}\beta_{3kj}+X_{ijt}\beta_{4kj}+$$
$$+F_{ij}\pi_{kj}+\eta_{ikj} \tag{28}$$

where $F_{ij}\pi_{kj}=\pi_{1kj}Y_{ikj1}+\overline{X_{ij}^{+}}\pi_{2kj}+\pi_{3kj}X_{ij1}+\overline{p_{ij}^{+}}\pi_{4kj}+\pi_{5kj}p_{ij1}$.

Table 15 shows the average marginal effect (AME) of the credit supply shock, p_{ijt}, on the probability of switching between employment states. The first row of Table 15 shows the share of sectors in the period $t+1$. An increase in p_{ijt} by one standard deviation decreases the likelihood of turning from paid employment to incorporated self-employment by 0.3 ppt, but increases the likelihood of unincorporated self-employment at period $t+1$ by 0.01 ppt. The effect of p_{ijt} on the likelihood of switching from incorporated self-employment to paid employment is substantial (2.9 ppt improvement). If the credit supply is worse by one standard deviation, incorporated self-employed are less likely to stay in the incorporated sector (-0.38 ppt) and less likely to switch to unincorporated self-employment (-1.1 ppt). It supports the evidence from the theoretical model (Corollary 2) that an increase in non-interest costs of borrowing (severe credit supply shock) is likely to increase the probability of switching from incorporated self-employment to paid employment. But for unincorporated self-employment, an increase in p_{ijt} by one standard deviation increases the probability of staying in the sector by 4.4 ppt, and switching to paid employment decreases by 3.5 ppt and to incorporated self-employment by 0.7ppt.

In addition to the AMEs (which are calculated across all individuals in the sample), I also estimate the marginal effects at the mean values of covariates and different values of p_{ijt}. These results are plotted in Figure 19. The figure shows the predicted probabilities of being incorporated or unincorporated self-employed at period $t+1$ at different values of the supply shock index. Results for unincorporated self-employed in the period $t+1$ (panel (a)) show that a more severe supply shock significantly increases the probability of staying in unincorporated self-employment. At the same time, incorporated self-employed tend to switch to unincorporated when a shock is small, but this probability decreases if the shock becomes more severe. Results for incorporated self-employed in the period $t+1$ (panel (b)) show that a more severe supply shock significantly decreases the probability of staying in incorporated self-employment. At the same time, unincorporated self-employed tend to switch to incorporated when a shock is small, but this probability also decreases if the shock becomes more severe.

Table 14: Dynamic Multinomial Logit Model of Employment Choices

VARIABLES	PE, $t+1$ (1)	Incorp SE, $t+1$ (2)	Unincorp SE, $t+1$ (3)
Incorp SE, t	0.191***	50.034***	12.107***
	(0.074)	(28.507)	(7.376)
Unincorp SE, t	0.181***	13.791***	23.329***
	(0.059)	(6.212)	(11.985)
NE, t	0.040***	0.390**	0.223***
	(0.007)	(0.148)	(0.087)
Supply Shock, t	1.037**	0.108**	1.098***
	(0.058)	(0.011)	(0.100)
Assets	1.095*	1.070**	0.987
	(0.062)	(0.119)	(0.121)
log (w_{PE})- log (w_{USE})	1.435***	1.751	0.756**
	(0.335)	(0.690)	(0.305)
Control variables, X_{ij} and X_{ijt}			
Age	1.250***	1.311***	1.310***
	(0.039)	(0.116)	(0.085)
Age sq.	0.997***	0.996***	0.997***
	(0.000)	(0.001)	(0.001)
Female	0.930	0.547***	0.719*
	(0.089)	(0.100)	(0.125)
Race:			
Black	0.749***	0.698	0.575***
	(0.074)	(0.154)	(0.121)
Other	0.670**	0.380***	0.832
	(0.107)	(0.122)	(0.256)
Hispanic	0.851	1.290	0.569*
	(0.146)	(0.444)	(0.171)
Married	1.013	0.952	1.283
	(0.131)	(0.282)	(0.328)

VARIABLES	PE, $t+1$ (1)	Incorp SE, $t+1$ (2)	Unincorp SE, $t+1$ (3)
Control variables, X_i and X_{it}			
Education:			
Some college	1.163*	1.427*	1.423*
	(0.104)	(0.303)	(0.257)
Bachelor's degree	0.795	1.550	0.662
	(0.189)	(0.769)	(0.384)
Grad degree	1.363**	2.344***	1.806**
	(0.186)	(0.640)	(0.456)
Number of kids:			
1 child	1.060	0.812	0.937
	(0.120)	(0.203)	(0.204)
Fewer than 4 kids	0.756**	0.835	0.676*
	(0.091)	(0.197)	(0.145)
More than 4 kids	0.566	0.874	0.849
	(0.252)	(0.571)	(0.438)
Number of HH members:			
2 adults in HH	1.283**	1.539	0.861
	(0.146)	(0.490)	(0.210)
Fewer than 4 adults in HH	1.359**	1.662	1.015
	(0.205)	(0.641)	(0.292)
More than 4 adults in HH	1.515	1.959	1.850
	(0.555)	(1.604)	(1.459)
Local Economic Conditions, L_{it}	Yes	Yes	Yes
Time-averaged variables $\overline{X_{ij}^{+}}$, $\overline{p_{ij}^{+}}$	Yes	Yes	Yes
Initial conditions, Y_{ij1}, X_{ij1}	Yes	Yes	Yes
Region FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes

VARIABLES	PE, $t+1$ (1)	Incorp SE, $t+1$ (2)	Unincorp SE, $t+1$ (3)
Unobserved effect, η_{ij}			
Var(η_{ij})	0.121	0.110	0.413
	(0.019)	(0.022)	(0.021)
Cov ($\eta_{i,PE}; \eta_{i,ISE}$)	0.211		
	(0.013)		
Cov ($\eta_{i,ISE}; \eta_{i,USE}$)	0.277		
	(0.013)		
Cov ($\eta_{i,PE}; \eta_{i,USE}$)	0.141		
	(0.016)		
Observations	11,922		

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

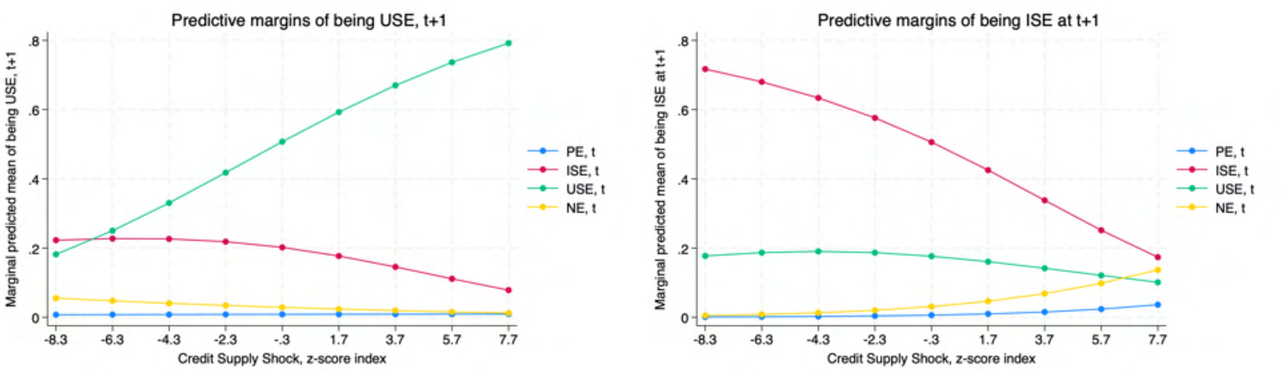
Notes. The table presents the relative risk ratios from the dynamic multinomial logit employment model with correlated random effects. The dependent variable is the employment status at time t+1, with the non-employed chosen as the base outcome. Columns present the full estimates of Equation 27 with correlated random effects, using the WRS solution to the endogeneity of the initial conditions. The following variables are included but not shown: year dummies, four regions, fixed effects for the first year of the stochastic process, the Mundlack device and the intercept, and regional economic characteristics. The omitted category of race is white, the number of kids is no kids, and the number of adults is one adult.

Table 15: Change in Transition Probabilities

	Paid-E, t+1	Incorp SE, t+1	Unincorp SE, t+1
Share of sector, t+1	83.81%	2.43%	3.01%
Paid-E, t	0.3pp	-0.2pp	0.01pp
	(0.003)	(0.0008)	(0.001)
Incorp SE, t	2.9pp	-0.38pp	-1.1pp
	(0.027)	(0.032)	(0.027)
Unincorp SE, t	-3.5pp	-0.7pp	4.4pp
	(0.027)	(0.023)	(0.029)
NE, t	-1.6pp	0.7pp	0.3pp
	(0.020)	(0.009)	(0.007)

Notes. This table shows the average marginal effect of an increase in the supply shock by one standard deviation in period t on the transition probability to labor states in period $t+1$ (paid employment, incorporated self-employment, unincorporated self-employment).

Figure 19: Dynamic of Self-Employment in the U.S.



a) b)

Notes. The figure shows the predicted probabilities of being in one of the two types of self-employment at different values of the credit supply index. The index is standardized with a mean of zero and a standard deviation of one. Predictions are obtained at sample means of covariates and based on the WRS model reported in Table 14.

C Additional Tables

Table 16: Supply Shocks and Access to Credit. Part I

VARIABLES	(1) Credit Card	(2) Charge Card	(3) Refinance Mortgage	(4) Application for new credit has been denied	(5) Reduction in the Available Limit for Credit Cards	(6) Being Asked to Pay Off Remaining Balance	(7) Filing Bankruptcy
Supply Shock	-0.022***	-0.023***	-0.001***	0.005***	0.009**	0.0008**	0.003***
	(0.003)	(0.003)	(0.00002)	(0.00004)	(0.004)	(0.0003)	(.0009)
PE	0.033	-0.025	0.010	0.036	0.065*	0.001	0.001
	(0.024)	(0.027)	(0.024)	(0.028)	(0.034)	(0.019)	(0.008)
ISE	0.017	-0.028	-0.002	-0.014	0.009	-0.0002	0.007
	(0.021)	(0.024)	(0.023)	(0.026)	(0.031)	(0.017)	(0.007)
USE	0.138***	-0.112***	-0.028*	0.023	0.018	0.023**	0.002
	(0.013)	(0.014)	(0.017)	(0.014)	(0.017)	(0.009)	(0.004)
Number of bank workers	-0.265**	0.289**	-0.009	-0.108	-0.262	-0.001	-0.043
	(0.128)	(0.143)	(0.138)	(0.155)	(0.183)	(0.103)	(0.041)
Owner of house	-0.167***	0.242	0.041	-0.140***	-0.033	-0.004	-0.031***
	(0.036)	(0.039)	(0.037)	(0.063)	(0.079)	(0.087)	(0.021)
Number of banks	-190.9***	-215.3***	-89.76*	-106.8***	-12.24	1.983	9.603
	(31.05)	(34.72)	(48.37)	(38.52)	(45.40)	(25.48)	(9.997)
Large durable assets	-0.097***	0.077***	0.015***	-0.015***	0.008	-0.002	-0.005***
	(0.004)	(0.004)	(0.004)	(0.005)	(0.006)	(0.003)	(0.001)
Diff in wages	-0.058**	0.061**	0.021	-0.032	-0.025	0.009	-0.019***
	(0.023)	(0.026)	(0.025)	(0.027)	(0.032)	(0.018)	(0.007)

Table 17: Supply Shocks and Access to Credit. Part II

VARIABLES	(1) Credit Card	(2) Charge Card	(3) Refinance Mortgage	(4) Application for new credit has been denied	(5) Reduction in the Available Limit for Credit Cards	(6) Being Asked to Pay Off Remaining Balance	(7) Filing Bankruptcy
Age	-0.001	0.006	0.004	-0.005	0.014**	0.008**	0.003**
	(0.004)	(0.004)	(0.005)	(0.005)	(0.006)	(0.003)	(0.001)
Age sq.	-0.000	-0.000	-0.000	0.000	-0.000***	-0.000**	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female	-0.020**	0.057***	-0.013	0.035***	0.008	0.001	0.005*
	(0.008)	(0.009)	(0.009)	(0.010)	(0.012)	(0.007)	(0.003)
Race (Black)	0.143***	-0.053***	-0.020	0.055***	0.004	0.048***	-0.000
	(0.014)	(0.015)	(0.015)	(0.016)	(0.019)	(0.011)	(0.004)
Race (Other)	0.034**	-0.026	-0.023	-0.007	0.030	0.013	-0.005
	(0.015)	(0.017)	(0.017)	(0.018)	(0.021)	(0.012)	(0.005)
Hispanic	0.042***	0.102***	-0.021	-0.038**	-0.042**	0.023*	-0.009*
	(0.015)	(0.017)	(0.017)	(0.018)	(0.021)	(0.012)	(0.005)
Married	-0.066***	0.097***	0.032*	-0.018	0.016	0.002	0.005
	(0.015)	(0.017)	(0.017)	(0.018)	(0.021)	(0.012)	(0.005)
1 child	0.049***	-0.032**	0.031*	0.009	0.015	-0.001	0.010**
	(0.014)	(0.016)	(0.016)	(0.017)	(0.020)	(0.011)	(0.005)
Less than 4 kids	0.066***	-0.031***	0.028**	0.002	0.000	0.006	0.000
	(0.010)	(0.012)	(0.012)	(0.013)	(0.015)	(0.008)	(0.003)
More than 4 kids	0.019	-0.129***	0.086**	0.010	-0.034	-0.011	0.010
	(0.039)	(0.043)	(0.040)	(0.047)	(0.056)	(0.031)	(0.012)
2 adults in HH	0.029**	-0.025*	-0.022	0.006	-0.014	0.006	0.004
	(0.012)	(0.013)	(0.014)	(0.014)	(0.017)	(0.010)	(0.004)
Less than 4 adults in HH	0.120***	-0.110***	-0.049**	0.061***	0.049*	0.000	0.012*
	(0.019)	(0.022)	(0.022)	(0.023)	(0.027)	(0.015)	(0.006)
More than 4 adults in HH	0.107***	-0.023	0.006	0.032	0.107**	0.024	0.030**
	(0.038)	(0.043)	(0.042)	(0.045)	(0.054)	(0.030)	(0.012)
Unemployment	-0.000	0.000***	0.000	-0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
% of college grads	-0.176*	-0.124	-0.062	0.157	0.093	-0.128	-0.003
	(0.099)	(0.111)	(0.116)	(0.120)	(0.142)	(0.080)	(0.032)
LnGDP	-0.167***	0.242***	0.040	-0.140***	0.033	0.004	0.031***
	(0.033)	(0.037)	(0.055)	(0.042)	(0.049)	(0.028)	(0.011)
Number of business est	0.000	-0.000***	-0.000	0.000	0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	33,160	33,201	16,804	8,568	8,385	8,571	29,559
R-squared	0.201	0.117	0.106	0.053	0.054	0.038	0.036

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

D Data Appendix

D.1 Types of Self-Employed

Table 18: Types of Self-Employment

	Incorporated	Unincorporated
Legal Entity	Operate their business as a distinct legal entity, such as a corporation (e.g., C-Corporation or S-Corporation) or a Limited Liability Company (LLC).	Usually operate as sole proprietors or in partnerships. They don't create a separate legal entity for the business.
Liability	Incorporation generally offers limited liability protection. This means that the individual's personal assets are typically shielded from the business's debts or legal liabilities.	There's no limited liability protection for sole proprietors. They are personally liable for all the business's debts and legal obligations. This can put their personal assets at risk.
Taxes	Corporations may be subject to double taxation. The corporation pays taxes on its earnings, and then shareholders also pay taxes on dividends [a].	There's no limited liability protection for sole proprietors. Sole proprietors report business income and expenses on Schedule C of their individual tax returns. Profits are subject to self-employment taxes (Social Security and Medicare). Partners in a partnership report their share of income and losses on their individual tax returns.
Regulation	Incorporated entities often face more rigorous regulatory requirements. They might need to have regular board meetings, maintain minutes, and meet other state-specific requirements.	Operating as an unincorporated self-employed individual typically involves fewer regulatory requirements and formalities than incorporated businesses. There's no need for board meetings or minutes, for instance.
Cost	Incorporating a business usually involves higher initial and ongoing costs, such as state registration fees, annual report fees, and costs associated with legal and accounting services.	Generally, starting and operating as an unincorporated self-employed individual is less costly regarding setup and maintenance.

[a] S-Corporations avoid this by passing income, losses, deductions, and credits through to shareholders who report this on their individual tax returns. Depending on elections and criteria, LLCs offer flexibility in tax treatment and can be taxed as sole proprietorships, partnerships, or corporations.

D.2 Access to Finance and Self-Employment

D.2.1 Changes in Number of Branches and Self-Employment.

In this subsection, I investigate the relationship between self-employment and the presence of bank branches, taking into account housing prices as a measure of individual wealth and local economic conditions. The analysis utilizes data from 2004 to 2014 from three sources: the change in the number of branches from the FDIC Summary of Deposits (SOD) for the years 2004 and 2014, the Federal Housing Finance Agency (FHFA) House Price Indexes (HPIs) for annual cumulative appreciation, and the County Business Patterns data for information on new establishments, self-employed, and small businesses (incorporated and unincorporated self-employed are presented as one variable). The dataset covers ten years, from 2004 to 2014, and includes various economic indicators at the county level.

Analysis reveals a reverse correlation between the rate of self-employment and the variation in the number of local bank branches. Generally, as the proportion of self-employed individuals in an area increases, the number of bank branches tends to decrease. However, this trend is influenced by housing prices. Precisely, in areas with lower housing prices, an increase in self-employment is often paralleled by a growth in bank branches. In contrast, in regions with medium to high housing prices, a rise in self-employment usually coincides with a decrease in bank branches.

Table 19 provides descriptive data showing changes in bank branches at the zip code level, correlating with the dynamics of self-employment. The table categorizes data into quartiles based on the proportion of self-employed individuals and housing price indices 2014. The table displays the change in bank branches from 2004 to 2014, with standard deviations included in parentheses.

The overall analysis reveals a prominent trend: as the number of self-employed people in an area increases, the number of bank branches usually decreases. But this doesn't happen everywhere in the same way. I observed that the correlation between the share of self-employed and the change in bank branches is somewhat contingent upon the housing values. More self-employed people in areas with lower housing costs correlate with more bank branches. The highest increase in bank branches is observed in counties with median housing prices and a low share of SE (1.811). This makes sense because banks want to be where their customers are. But in counties with higher housing costs, the opposite happens. More self-employed people can lead to fewer bank branches; the smallest increase is found in the county with a high share of SE and median housing prices (0.696). This might be because banks think opening new branches in these expensive areas is not worth it.

Overall, the correlation between the share of self-employed individuals and alterations in bank branches manifests differently under varying conditions of HPI, displaying an increase in branches under lower HPI and a decrease under higher HPI conditions.

Table 19: Changes in Bank Branches (2004-2014) by Self-Employment Rates and Housing Price Index (HPI) in 2014

	Low share SE	Moderate share SE	Median share SE	High share SE	Total
Low HPI	0.263 (0.701)	0.326 (0.801)	0.223 (0.615)	0.359 (0.781)	0.287 (0.718)
Moderate HPI	0.596 (2.302)	0.347 (1.001)	0.313 (1.261)	0.487 (1.164)	0.412 (1.494)
Median HPI	1.811 (17.883)	1.491 (11.686)	0.701 (5.174)	0.696 (1.893)	1.296 (12.451)
High HPI	1.098 (4.644)	0.836 (2.840)	0.866 (3.131)	0.793 (1.964)	0.923 (3.516)
Total	0.983 (10.023)	0.844 (6.963)	0.505 (3.647)	0.481 (2.083)	0.699 (6.414)

Notes.This table illustrates the changes in bank branches across quartiles of self-employed rate and the Housing Price Index (HPI) in 2014. Self-employed rate and HPI are broken into quartiles to visualize these variables' impact better. Numbers in parentheses represent standard deviations. The number of branches is derived from the FDIC Summary of Deposits for the relevant years; housing prices are from the Federal Housing Finance Agency House Price Indexes and County Business Patterns data for new establishments and small businesses. Covering a decade (2004-2014), the dataset encompasses various economic indicators at the county level.

D.2.2 The Financial Dependence of Different Types of Self-Employed.

Unfortunately, the main database used in the analysis (CAPS) does not have much information about sources of business finance. The financial dependence of different types of entrepreneurs can be analyzed using an additional source of data, the Survey of Small Business Finances (SSBF). The survey collects information on small businesses (fewer than 500 employees) in the United States. Owner characteristics, firm size, use of financial services, and the income and balance sheets of the firm are just some examples of the information collected. The survey contains six types of small businesses that can be divided into incorporated and unincorporated businesses[29]. Sole proprietorship and partnership are identified as unincorporated businesses. LLP, LLC, S-corporation, and C-corporation are incorporated businesses.

To identify the correlation between access to financial markets and types of business, I implement several empirical models. First, I estimate a linear probability model of using different credit products (business credit line, credit card, or personal credit card) based on the

[29]DEFINITION OF SOLE PROPRIETORSHIP: A sole proprietorship is an unincorporated business owned by a single person. The individual proprietor has the right to all the profits from the business and responsibility for all the business's liabilities.

DEFINITION OF PARTNERSHIP: A partnership is an unincorporated form of business in which two or more owners agree to split the business's profits/losses or capital in a particular fashion. There may be limited and/or general partners.

DEFINITION OF LIMITED LIABILITY PARTNERSHIP: A limited liability partnership (or LLP) is a business established as an entity under state law and owned by two or more owners where the partners' liability is limited to their investment in the partnership, and each of the partners has protection against personal liability for partnership liabilities except for liability due to partner's own negligence or of persons under partner's direction.

DEFINITION OF S-CORPORATION: An S-corporation is a corporation that does not pay tax on its income. To qualify for S-corporation status, the business (1) must be a U.S. corporation, (2) must have only one class of stock and no more than 75 shareholders, and (3) may not have certain organizations or non-U.S. citizens as shareholders.

DEFINITION OF C-CORPORATION: A C-corporation (or corporation) is a business that has been granted a state charter recognizing it as a separate legal entity having its own rights, privileges, and liabilities distinct from those of the individuals that form the business.

DEFINITION OF LIMITED LIABILITY COMPANY: A limited liability company (or LLC) is an entity established as a corporation under state law. The entity has the same limited liability characteristics as a corporation while often having the same tax characteristics as a partnership. LLC owners may actively participate in the organization's management and are protected against personal liability for the organization's debts/liabilities.

type of business:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 Z_{it} + \epsilon_{it}$$

where Y is the indicator of using a different credit product (business credit line, credit card, personal credit card, or shareholder's loan), and X is the indicator of having an incorporated business ($X = 0$ if the business is unincorporated, $X = 1$ if the business is incorporated). Z is the set of control variables, including a dummy for sampling strata (firm size*urban/rural*census division), a dummy for a year (2004 or 2005), number of owners, number of workers, number of sites, and total sales compared to the previous year ($). ϵ is the idiosyncratic error term.

Results show that incorporating a business increases the likelihood of using a business credit line, credit card, and shareholders' loan and decreases using the owner's personal credit card.

Table 20: Usage of Different Credit Products

	Business Credit Line	Business Credit Card	Personal Credit Card	Shareholder's Loan
Incorporated	0.642***	0.703***	-0.309***	0.246***
	(0.02)	(0.019)	(0.019)	(0.044)
Other Controls	Yes	Yes	Yes	Yes
Observations	21,200	21,200	21,200	21,200

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes. The table shows the results of the linear probability model where the dependent variable is the indicator of using a different credit product (business credit line, credit card, shareholders' loan, or personal credit card). Independent variables include the indicator of having an incorporated business and the set of control variables (a dummy for sampling strata (firm size*urban/rural*census division), a dummy for a year (2004 or 2005), number of owners, average owners experience, number of workers, number of sites, and total sales compared to the previous year ($)). All variables are weighted by survey sample weights.

Secondly, I delve into the purpose of the loan, examining the likelihood of borrowing for various objectives such as purchasing equipment, acquiring a business-use motor vehicle, or securing a mortgage for business needs. The model of analysis is the following:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 Z_{it} + \epsilon_{it}$$

where Y is the indicator of using a loan for different purposes (purchasing equipment, acquiring a business-use motor vehicle, or securing a mortgage for business needs), and X is the indicator of having an incorporated business ($X = 0$ if the business is unincorporated, $X = 1$ if the business is incorporated). Z is the set of control variables, including a dummy for sampling strata (firm size*urban/rural*census division), a dummy for a year (2004 or 2005), number of owners, number of workers, number of sites, and total sales compared to the previous year ($). ϵ is the idiosyncratic error term.

The findings indicate that incorporating a business enhances the probability of utilizing a loan to purchase equipment, a business vehicle, or a business-related mortgage.

Table 21: Purpose of Different Credit Products

	Equipment	Vehicle	Mortgage
Incorporated	0.483***	0.317***	0.172***
	(0.026)	(0.02)	(0.019)
Other Controls	Yes	Yes	Yes
Observations	21,200	21,200	21,200

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes. The table shows the linear probability model results where the dependent variable indicates using a different credit product (business credit line, credit card, or personal credit card). Independent variables include the indicator of having an incorporated business and the set of control variables (a dummy for sampling strata (firm size*urban/rural*census division), a dummy for a year (2004 or 2005), number of owners, number of workers, average owners experience, number of sites, and total sales compared to the previous year ($)). All variables are weighted by survey sample weights.

Finally, I looked at the dependence of the loan size on the type of business. The dependent variables are the average monthly expenses on the owner's personal or business credit card.

$$log(L_{it}) = \beta_0 + \beta_1 X_{it} + \beta_2 Z_{it} + \epsilon_{it}$$

where L is the loan size, and X is the indicator of having an incorporated business ($X = 0$ if the business is unincorporated, $X = 1$ if the business is incorporated). Z is the set of control variables, including a dummy for sampling strata (firm size*urban/rural*census division), a dummy for a year (2004 or 2005), number of owners, number of workers, number of sites, and total sales compared to the previous year ($). ϵ is the idiosyncratic error term.

Table 22: Size of the loan

	Expenses on Personal CC	Expenses on Business CC
Incorporated	0.751***	0.767***
	(0.034)	(0.033)
Other Controls	Yes	Yes
Observations	9,038	11,213

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes. The table shows the probit model results where the dependent variable is the indicator of using a different credit product (business credit line, credit card, or personal credit card). Independent variables include the indicator of having an incorporated business and the set of control variables (a dummy for sampling strata (firm size*urban/rural*census division), a dummy for a year (2004 or 2005), number of owners, number of workers, average owners experience, number of sites, and total sales compared to the previous year ($)). All variables are weighted by survey sample weights.

D.3 Measures of Distance in the Literature

Due to data sensitivity, the literature does not investigate the distance between individual borrowers and banks. To the best of my knowledge, only one paper investigated the distance between individual borrowers and banks: Malkova and Sabirianova Peter. The paper explores the expansion of branches and its effect on labor informality in Russia. It uses RMLS data, and the average distance is around 3.3 miles. Other papers that investigate the distance in the context of small firms show that the median distance between small firms and their supplier of credit is around 3-5 miles (Amel and Brevoort (2005) and Brevoort et al. (2010)). Similar to my estimations (around 1.2-1.6 miles), the average distance between small firms and branches was found in Ono et al. (2016) and Carling and Lundberg (2005), but in the context of Japan and Sweden.

Table 23: Borrower-Lender Distance in the Literature

Paper	Country	Ave Distance	Object of Research Interest
Corporate lending			
Petersen and Rajan (2002)	USA	115 miles	Interest rate
Degryse and Ongena (2005)	Belgium	1.4 miles	Interest rate
Agarwal and Hauswald (2010)	USA	10.25 minutes	Interest rate
Knyazeva and Knyazeva (2012)	USA	744 miles	Interest rate
Herpfer et al. (2023)	Norway	19.2 minutes	Interest rate
Milani (2014)	Italy	2.138 km	Loan Default
Jiménez et al. (2009)	Spain	353.84 km	Collateral size
Small Business lending			
Adams et al. (2021)	USA	4 miles	Change in distance
Bellucci et al. (2013)	Italy	3.2 miles	Interest rate
Brevoort et al. (2010)	USA	3.2 miles	Distance dynamics
DeYoung et al. (2008)	USA	4.99 miles	Loan Default
DeYoung et al. (2011)	USA	6.36 miles	Distance Dynamics
Ono et al. (2016)	Japan	1.2 miles	Firm-Bank Relationships
Carling and Lundberg (2005)	Sweden	1.6 miles	Credit Risk Management
Personal lending			
Malkova and Sabirianova Peter	Russia	5.4 km	Labor Informality

Notes. Table contains the average borrower-lender distance found in the previous literature for different types of borrowers (corporate clients, small businesses, and personal lending).

D.4 Data Description

The Community Advantage Program Study, launched in 1999, examines data collected since 1999 from homeowners who receive mortgages through the Community Advantage Program (CAP), a path-breaking mortgage initiative of Self-Help, the Ford Foundation, and Fannie Mae that has funded more than $4 billion of home loans to low- and moderate-income borrowers.

At the center of the CAP Study is data from the Community Advantage Panel Survey (CAPS), an 11-year panel survey of low- and moderate–income homeowners and renters fielded annually between 2003 and 2014. Initially, 3743 homeowners recruited to participate in CAPS received mortgages between 1999 and 2003 through the Community Advantage Program, and 1529 participating renters were recruited to match these homeowners concerning geographic proximity and income ceilings.

D.4.1 Survey Participation

The data sets reflect the dual-sample structure of the survey, with one series of data sets containing data for survey participants who were homeowners at the beginning of the survey period and received CAP mortgages (i.e., the owners' sample) and one series of data sets for survey participants who were renters at the beginning of the survey period (i.e. the renters' sample). Although some survey participants changed their residential tenure status during the survey, for data storage purposes, each case remains part of the sample to which it was assigned at baseline.

Table 24 shows statistics of the sample creation taken from the "CAPS data collection report" provided by the Center of Community Capital UNC. In 2003, 7223 homeowners were eligible for the CAPS as they received CAP mortgages. In 2004, these homeowners were matched by geographic proximity and income ceilings with 15934 renters. Table 24 shows different statuses of respondents. Cases for which no interview was completed included those that had moved out of the interviewing area or country. In this case, a new telephone number could not be found, and the respondent could not be located. Refusal reasons included disinterest in the study and an unwillingness to continue participating. Some cases were not interviewed because interviewers learned that the respondents were incarcerated, institutionalized, or incapable of completing interviews. Telephone interviewers could confirm the status of deceased respondents through spouses and relatives.

Table 24: Attrition Rate by Year

VARIABLES	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Homeowners												
Beginning panel	7223	3743	1670*	2537	3188	2795	2638	2592	2488	2325	2286	2145
Unable to contact	2111	553	195	353	653	166	275	353	397	197	300	262
Ineligible	344	358	13	5	170	28	17	23	8	21	19	13
Refusals	1025	218	151	61	286	228	117	128	65	52	43	39
Total eligible cases	4768	2832	1462	2179	2887	2780	2627	2585	2487	2320	2275	2140
Completed interviews	3743	2614	1283	2118	2079	2376	2229	2088	2018	2055	1924	1832
Attrition rate	79%	92%	88%	97%	72%	85%	85%	81%	81%	89%	85%	86%
Renters												
Beginning panel		15934	1529	1156	1214	1086	1061	1042	1015	958	944	901
Unable to contact		2113	250	108	250	57	105	135	142	45	97	81
Ineligible		8981	40	36	33	17	22	10	10	17	13	14
Refusals		3311	82	43	29	29	19	12	12	10	11	15
Total eligible cases			1521	1145	1188	1074	1046	1033	1009	951	939	894
Completed interviews		1529	1157	969	902	980	915	873	851	885	823	791
Attrition rate			76%	85%	76%	91%	87%	85%	84%	93%	88%	88%

Notes. This table shows statistics for participation in the survey. Most of the interviews are telephone surveys that happened between May and October in the given year. Except in 2005 (marked *), when homeowners participated in an in-home survey, we observed fewer eligible respondents.

D.4.2 Variables Description

Self-employed workers are a self-defined category by respondents. Self-employment is their primary job.

Paid-employed workers are a self-defined category and comprised of (a) employees in private business and (b) employees in government.

Type of business.

Age, female, year of survey. Self-explanatory.

Household income. It's a gross income, the sum of all HH members' income during the last year.

Race. A binary indicator for the following categories: White, Black, Hispanic, and other (Asian, Native American, etc).

Education. A categorical variable indicating the highest level of schooling completed by a respondent: [1] "Less than High School Diploma", [2] "High School Diploma", [3] "College graduate ", and [4] "Doctoral or professional degree. The first category is the base category.

Married. = 1 for legally married individuals (including those not living together) and 0 for other categories, including single, widowed, divorced, and living together without marriage.

Number of household members. Counts the number of household members who are presently living in the same household.

Number of kids per household. Counts the number of children under the age of 18 currently residing in the same household.

State population. Number of people living in the state in thousands. The population is taken from the 2010 Census.

Regions. Set of dummies for living in one of the four regions at the time of the interview. The regions are Midwest, Northeast, South, and West.

First wave fixed effects. Set of dummies for the starting year of the stochastic sequence or the year of entry to the estimation sample.

Distance to the nearest bank, miles. The distance to the nearest bank is the Euclidean distance between the location of the respondent's home and the bank's address. Source: The FDIC Summary of Deposits (SOD) data, various years.

Average distance to the ten nearest banks, miles. All distances between banks within the state and the location of the respondent's home were sorted, and the ten nearest banks were chosen. The sum of all distances is divided by 10.

Number of banks within 10 miles. Total number of banks within 10 miles from the location of the respondent.

Number of bank offices per 1,000 state population. The number of bank offices includes bank headquarters, credit organizations, branches, supplementary offices, and operational offices but excludes cash offices, cash desks, and mobile cash units. Source: The FDIC Summary of Deposits (SOD) data, 2010 Census.

Total assets. The amount of non-housing and housing assets adjusted for inflation at the moment of the interview.

Average house value. The average house value within the zip-code zone of the location of the respondent's home. Source: Zillow.