Self-Employment, Credit Demand and Credit Constraints: Evidence from China

Jan Klingelhöfer and Rongrong Sun[1]

Abstract: This paper builds a three-stage sequential logit model to investigate bank credit demand and constraints faced by non-agricultural self-employed households in China. Self-employment contributes significantly to employment, poverty alleviation, entrepreneurship, innovation, and economic growth. However, credit constraints prevail with over half of self-employed households who need bank loans facing credit constraints, either discouraged from applying or rejected. Wealth, saving status, education, financial literacy, and risk tolerance significantly influence the likelihood of credit demand, application, and approval. Self-employed with the following features are more likely to demand loans: wealthy, without savings, rural, male, more financial literate, and less risk averse. However, costly application and perceived lower chances of approval result in a large proportion of discouragement from application. Once the self-employed have submitted application, the rejection rate is in general low (18%) and there is no evidence of gender discrimination in loan approvals. However, low income, less wealthy, middle-aged (aged 40s), and least educated households face the most significant credit constraints, hindering their financial mobility through self-employment. Targeted policy interventions, such as improving application procedures, financial literacy programs, loan guarantee programs, microfinance initiatives, or government subsidies, are crucial to address these disparities. Interestingly, rural residents are more likely to need and get bank loans compared to urban counterparts, even after controlling for other factors. Finally, our research finds that informal loans are often used alongside bank loans, implying that the actual credit constraints might be even severe than our study shows.

Keywords: self-employment, credit demand, credit application, credit constraints, sequential logit model
JEL-Classification: D13, G51

[1] Klingelhöfer: Nottingham Business School, Nottingham Trent University, jan.klingelhoefer@ntu.ac.uk; Sun: School of Business, University of Dundee, rsun001@dundee.ac.uk. Sun would like to acknowledge the Hong Kong Institute for Monetary Research (HKIMR) for their generous support and the opportunity to conduct this research as a Visiting Research Fellow. We are grateful to Florian Biermann, Sharada Davidson, Achim Hauck, Arye Hillman, one anonymous referee for their helpful discussions and comments. We also thank the seminar participants at various venues for their valuable insights. Any remaining errors are our own.

1. Introduction

Self-employment contributes significantly to employment, poverty alleviation, entrepreneurship, innovation, and economic growth (Balboni, Bandiera, Burgess, Ghatak, & Heil, 2021). In China, non-agricultural self-employment soared from near insignificance four decade ago to a remarkable 23.4% of total employment in 2019 (China's National Bureau of Statistics). This sector has emerged as a "key force underpinning China's resilient economic and job growth" (Xinhua News Agency, 2021)[2]. While research extensively documents the financing challenges faced by consumers and firms in general, this specific hurdle for self-employed individuals remains understudied and poorly quantified. This lack of attention, in part, stems from data limitations. Firm financing data is readily accessible, whereas the specific data on household financing of the self-employment activities is scarce.

To fill this research gap, this paper studies credit demand and credit constraints facing self-employed individuals in China, using two waves of the China Household Finance Survey (CHFS) from 2013 and 2015. The study focuses on non-agricultural self-employment, often referred to as "individual businesses". We begin with an overview of evolution of self-employment.[3] Then, drawing on the CHFS survey's rich detail, we unveil key characteristics of self-employment in China. This includes documenting possible reasons for market entry and exit, the link between self-employment and productivity (as indicated by income), and possible reasons for credit constraints.

To examine the factors impacting self-employed individuals' access to bank credit, we employ a three-stage sequential logit model. This model, often used in studies on education attainment due to its ability to capture nonlinear, ordered decisions (among others, see Mare, 1980), is ideally suited to analyze the sequential nature of credit acquisition. In this context, the three stages involve binary choices at each step: (1) demand (whether households need bank loans), (2) application (whether to apply given their need), and (3) granting (the bank's decision based on submitted applications). The advantage of this model lies in its simultaneous analysis of all three stages, accounting for both demand and supply sides. This helps to identify whether certain population groups face higher credit constraints due to their greater loan demand or because they are perceived as higher credit risks by banks. While this model has seen application in the US consumer loan market (Perraudin & Sorensen, 1992), its use in the context of self-employment credit in China represents a novel application.

The CHFS survey offers unique advantages for studying credit demand and constraints faced by individual businesses. Unlike other household loan data, the CHFS survey provides granular details on household financing for various activities, including non-agricultural self-employment (industrial and commercial) and others (agricultural, housing, vehicles, education, financial investment, credit cards, etc.). This granularity

[2] http://www.xinhuanet.com/english/2021-05/27/c_139972113.htm

[3] Due to data limitations in other countries, where self-employment datasets often encompass all types, including farmers, meaningful cross-country comparisons of non-agricultural self-employment are currently not possible.

allows for a focused examination of household bank loans specifically for non-agricultural self-employment activities. Such a granular examination of bank loans for specific activities in turn contributes to precise estimation given different features of different bank loan markets. Furthermore, the 2013 and 2015 CHFS survey data, among five publicly available waves, offer crucial information to track households' experiences throughout the loan process. We can then identify households that need, apply for, and receive/are rejected for self-employment loan applications. This enables a comprehensive analysis of all three critical decisions in the credit process: households' loan demand, their application decisions, and banks' credit granting decisions.

The 2013 and 2015 CHFS surveys unveil that around 14% of surveyed households run their own small businesses, mostly in service industries. Among these self-employed households, one-fifth need bank loans to fuel their operations.[4] However, credit constraints are prevalent: Over half (51%) of those who need loans face difficulties securing them. This includes households whose loan applications were rejected by banks and those who were discouraged from applying. In this paper, the discouraged borrowers are defined as individuals who need loans but choose not to apply when they are discouraged by the potential of rejection, concerns about future debt servicing, or a high application cost associated with complex application procedures or financial illiteracy (see, e.g., Chivakul & Chen, 2008; Jappelli, 1990). The high proportion of discouraged borrowers (around 78.5% of the credit-constrained) highlights that households face significant fixed costs when applying for loans.

Our statistical analysis identifies key economic and social-demographic factors that influence self-employed households' loan demand, application, and approval probabilities. Our key finding is that low income, less wealthy, middle-aged (aged 40s), and least educated self-employed households face the most significant credit constraints. This limited access to credit can significantly hinder their financial mobility through self-employment.

The more detailed findings are eightfold. First, wealth seems to be a more significant factor than income in our analysis. Wealthier households are more likely to seek bank loans and get approved by banks compared to less wealthy households. Second, both very low-income (i.e., negative income) and high-income households are more likely to seek loans, but higher income increases the chances of getting approved. Third, households use both formal bank loans and informal loans from friends, family, or informal financial institutions to finance their self-employed businesses. Households with access to informal loans are more likely to be discouraged from applying due to the application costs and perceived lower chances of approval. Fourth, households of different education do not show significant differences in loan demand, but banks tend to favor more educated borrowers, possibly because of their greater future earning potential and hence their greater loan repayment ability. Fifth, self-employed households in rural areas are more likely to demand

[4] This paper focuses on the bank credit market, leaving alternative financing channels (such as private lending, shadow banking, online lending, and crowdfunding) for future exploration.

and get loans compared to those in urban areas, even after controlling for other factors. Six, age does not monotonically influence household loan access probabilities. However, compared to the young and old groups, a large group of middle-aged households (in their 40s) are less likely to get loans despite their higher demand. Seven, males are more likely to apply for loans, but there's no significant difference in approval rates between males and females. Eight, households with higher financial literacy and risk tolerance are more likely to seek loans.

This study is related to three existing research strands. First, it is related to the well-established literature on imperfect capital markets, particularly the presence of quantity constraints on borrowing. While household credit constraints are well-documented, especially in consumer loans, understanding these limitations for self-employed individuals remains a significant gap. Built on a few exceptions that provide indirect evidence of credit constraints faced by the self-employed,[5] this study directly addresses this by examining credit demand and constraints specific to self-employed households in China.

Second, this study is related to the rich microcredit literature (e.g., Banerjee & Duflo 2004; De Mel, McKenzie, & Woodruff 2008; McKenzie & Woodruff 2006). This literature strand mainly focuses on microcredit programs' impact on poverty alleviation, often assuming widespread credit constraints among low-income individuals. By exploring the factors influencing credit demand and approval probabilities in the self-employment loan market, this study complements the microcredit literature, potentially shedding light on more effective targeted policy design.

Third, this paper is related to the self-employment literature. Existing research has documented considerable heterogeneities in self-employment rates across different population groups based on factors like economic conditions, demographics, and family backgrounds (cite the literature) (Zhang, Zhang, & Zhang, 2021). Understanding these differences can shed light on the nexus between self-employment and productivity growth. (Hout & Rosen, 2000) (Herreño & Ocampo, 2023). By documenting key characteristics of self-employed individuals in China, this study adds Chinse evidence to this ongoing research.

The main contributions of this paper are twofold. First, it offers a comprehensive portrait of the self-employment credit market in China by looking into both demand and supply sides. This reveals the key characteristics (in dimensions of income, wealth, financial literacy, risk tolerance, and social-demographic features) of the self-employed households who are more likely to need loans or face credit constraints, filling the gap in existing research. These findings shed light on the criteria used by banks in assessing credit applications and reveal the strong dependence of preferences on financial and demographic characteristics. Second, it presents a novel application of sequential logistic regression, modelling self-employed business loan decisions in three stages: demand, application, and credit grant. We also demonstrate that a simple

[5] Among them, Herkenhoff, Phillips, and Cohen-Cole (2021) find that self-employment increases monotonically with credit limit and credit scores, for example, after the removal of the bankruptcy flag. Using the Chinese data, Wang (2012) has evidenced an increase in self-employment following the housing reform in 1994, which alleviated credit constraints by enabling households to monetarize the value of their real estate assets.

logistic regression without disentangling the factor influences on the demand side from their impact on the supply side leads to confounding results.

Understanding these credit access disparities has significant policy implications. The most vulnerable group facing high credit constraints is characterized by low income, low wealth, being middle-aged, and least educated. Their chances of obtaining loans are significantly lower. For example, borrowers with middle school or college education are 10.8 and 15.6 percentage points more likely to obtain self-employment loans compared to those with no or primary education, while holding all other variables at the median values. Similarly, high-income, high-wealth, and young borrowers (in their 20s) are (11-13 percentage points compared to their counterparts valued at the median values) more likely to secure loans. Expanding access to credit for this vulnerable group (74% financially illiterate) is crucial not only for promoting economic growth through increased self-employment rates, but also for fostering social mobility and empowering marginalized communities. Targeted policy interventions, such as financial literacy programs, loan guarantee programs, microfinance initiatives, or government subsidies, are essential to address these disparities. Further research could explore the long-term impact of these interventions and optimize program designs for maximum effectiveness.

The paper is organized as follows. Section 2 provides an overview of China's individual businesses, exploring their historical evolution and contemporary dynamics. Section 3 introduces the CHFS surveys and discusses how credit demand, application, and granting are identified. Sections 4 and 5 set up econometric models and present the estimation results. Section 6 concludes.

2. Self-employment in China

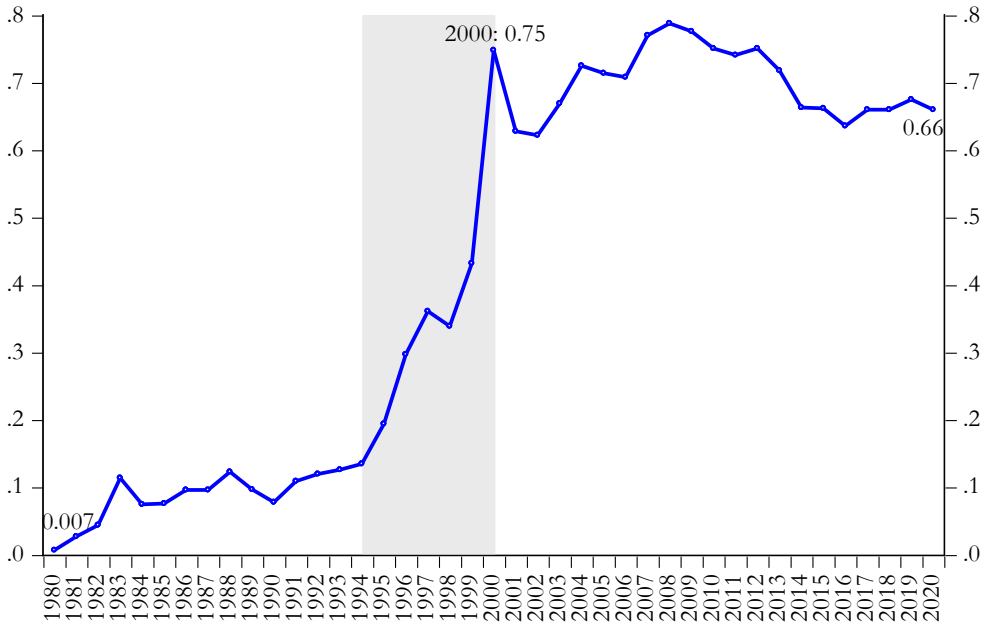
2.1 A historical perspective

The rise of self-employment played a crucial role in China's economic transformation. From the mid-1950s to 1978, private small businesses and trades were forbidden. In 1978, this policy was abandoned with the adoption of reform and opening-up policies. These reforms, which encouraged market liberalization and economic decentralization, created fertile ground for self-employment to flourish. In cities, the seeds of self-employment first took root among returning intellectuals and unemployed, who found opportunities in retail, services and other sectors. In rural areas, the household contract responsibility system boosted agricultural productivity, freeing up surplus labor. This led to the emergence of numerous small vendors in villages and towns. Recognizing the potential of this dynamic force, China resumed the registration management of individual industrial & commercial businesses in September 1979. Back then, these "geti gongshanghu", as they are called in Chinese (literally translating to "individually owned industrial and business entity"), numbered less than 10,000, constituting a mere 1% of the market entities.

In the following decades, their numbers grew rapidly, as showed in Figure 1, which tracks the evolution of self-employment as a proportion of the total market entities in China from 1980 to 2020. The initial five

years witnessed a sharp rise of this share from near zero to around 10%. However, the subsequent decade presented a period of stagnation, with the share hovering around 10%. This can be attributed to the continued dominance of state-owned enterprises (SOEs) during that period. Social safety nets like benefits and insurances were primarily employer-provided, making self-employment a less attractive option despite its potential for improved livelihoods. Additionally, societal stigma persisted, with self-employment often perceived as an undesirable career choice.

Figure 1: Share of the registered self-employed in the total market entities, 1980-2020

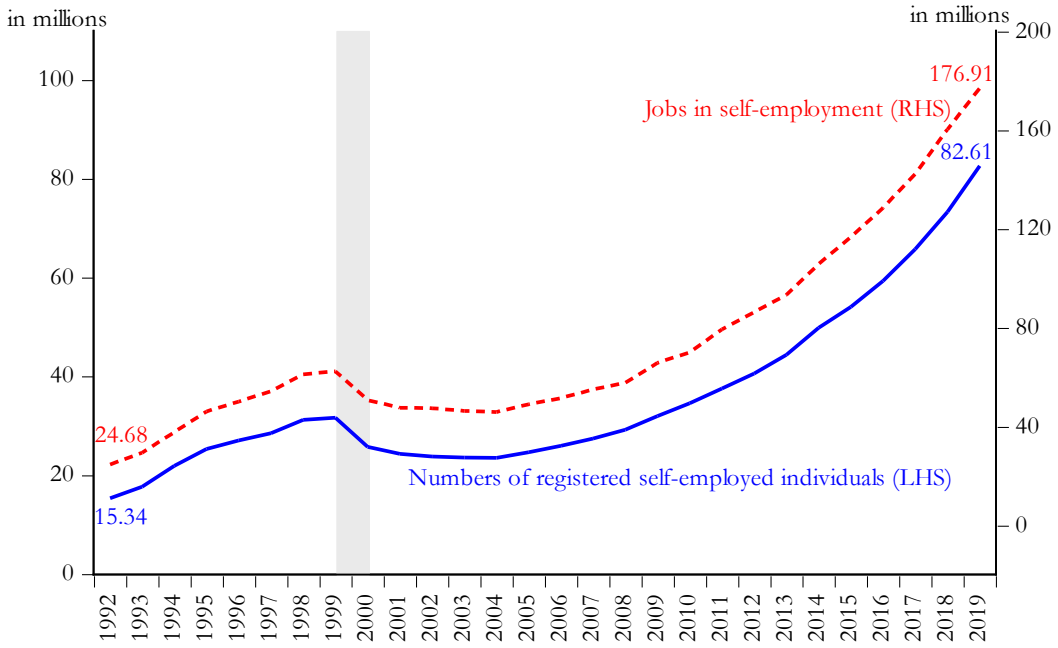


Note: The figure represents the share of registered self-employed businesses as a percentage of all market entities. The shaded area highlights the period of rapid self-employment growth from 1994 to 2000.
Source: Qichacha 1980-2020.08.

Deng Xiaoping's speeches during his 1992 southern tour marked a turning point for China's private economy. His words, emphasizing the importance of market forces and entrepreneurship, shifted policy perspectives[6] and ignited a surge of nationwide entrepreneurial spirit. Two waves of private sector pioneers in the mid-1990s also played a crucial role: Xiahai (party officials and intellectuals venturing into business) and Xiagang (laid-off SOE workers seeking new opportunities).[7] All these factors paved the way for a significant rise in self-employment in China. As Figure 1 illustrates, registered self-employed businesses soared from 13.5% to 74.8% of all market entities between 1994 and 2000. In recent years, this share has stabilized around two-thirds.

[6] The evolving policy perspectives on China's private sector are reflected in the changing language used in the National Congress of the Communist Party of China. From a mere "supplement" in 1982 to an "important component" in 1997, the private sector finally was recognized as parallel to the state-owned sector through the "two unswervingly" principle in 2002.
[7] China is a society with ingrained societal hierarchies and traditional views about career prestige. The very character, "Xia" used in "Xiahai" and "Xiagang", denotes a descent from a higher position, subtly reflecting the historical context of societal hierarchies.

Figure 2: Numbers of the registered self-employed and jobs offered, 1992-2019



Note: A sharp drop in 2000, highlighted, is due to the fact that the inactive licenses were cancelled.
Source: China's National Bureau of Statistics.

This rapid expansion of self-employment is mirrored in employment figures, as shown in Figure 2. By 2019, over 82 million registered self-employed businesses employed nearly 177 million people, a stunning 22.4% of China's total employment. Self-employment has become the nation's primary job creator. Its importance to China's economy is reflected in the title of the latest revision of self-employment regulations in 2022 (preceded by the other two in 1987 and 2011) "Regulations on Promoting the Development of Individually Owned Industrial and Commercial Businesses". This sector, known for its small scale, flexible operations, and adaptability, has primarily fueled growth in the service industry, including "Wholesale & retail," "hospitalities," and "household services, repair & other services".

In addition to its contribution to job creation, the rise of China's self-employment sector has contributed to thriving entrepreneurship by fostering both physical and human capital accumulation (Wu 2017; Zhang 2007 Interview with *Trade-Off* published on *China Youth Daily*). For many first-generation entrepreneurs who started individual businesses in the 1980s and 1990s, their initial profits ("first pot of gold") became the seed capital for future enterprises. Renowned business leaders like Jack MA (Alibaba) and LEI Jun (Xiaomi) all share an early experience of self-employment.

2.2 Self-employment market entry and exit

Given the crucial role of self-employment in entrepreneurship, understanding the motivations and obstacles behind starting and ending such businesses is essential. The CHFS survey offers valuable insights into the motivations behind starting and ending self-employed businesses.

Market entry: According to the 2015 and 2017 CHFS surveys, a significant portion, around 14-16% of surveyed households, are engaged in self-employed businesses. The CHFS survey sheds light on the motivations behind starting such businesses. Panel (A) of Table 1 reveals the key reasons reported by respondents, with a consistent pattern across the 2015 and 2017 surveys (and similar patterns observed in all five rounds of the CHFS survey). A substantial portion (around 21%) report involuntary self-employment, likely due to limited job opportunities elsewhere. As explored in Zhu, Lee, Hong, and Shi (2023), this might be particularly relevant for internal migrants facing challenges in entering the formal job market. Involuntary self-employment can have negative consequences, potentially leading to lower income, job insecurity, and limited access to social benefits.

The majority (around 58%) choose self-employment for reasons associated with greater control and income potential: seeking work flexibility (29-33%) and the desire to earn more (25-28%). Additionally, 12-14% seek self-employment due to self-interest or the desire to be an entrepreneur. Another 2% interstage self-employment as a family business.

Table 1: Entry and exit motivations (based on the CHFS surveys)

(A) Market entry			(B) Market exit		
	2015	2017		2017	2019
1. More freedom and flexibility	32.7	29.2	1. Personal or family reasons	36.7	36.0
2. Possibility to earn more in business	25.5	28.1	2. Poor management	31.7	37.4
3. Not able to find other work opportunities	21.2	20.9	3. Funding difficulties	19.0	17.5
4. Ideal job/Entrepreneurial drive	13.8	11.8	4. Industry competition	17.1	15.2
5. Family business	1.5	2.2	5. Policy changed /Government reasons	8.3	10.8
			6. Short-term projects	3.4	2.7
			7. High tax	2.4	2.3
Number of observations	5,926	5,672	Number of observations	3,005	2,834

Notes: This table shows the percentage breakdown of the main reasons households started and ended self-employed businesses based on the CHFS surveys. Panel (A) focuses on market entry motivations, based on the survey question "Why did your household start a business?" (Question b2001a). Panel (B) explores market exit reasons, based on the survey question "Why did you end the business you previously run?" (Question b2000f), where respondents could select multiple reasons.
Source: The 2015, 2017, and 2019 CHFS surveys.

Market exit: Market exit is not rare. The CHFS surveys reveal that a substantial portion (35% and 42% in 2017 and 2019, respectively) of households with self-employment experiences were no longer engaged in such businesses by the survey year.[8] Given this high exit rate, it would be of particular policy implications to understand the reasons that drove these closures. Panel (B) of Table 1 sheds light on the top reasons for these closures. Personal or family reasons (e.g., childcare, eldercare, health issues) are the most common cause (around 36-37%). Poor management (32-37%) follows closely, highlighting the importance of business planning and leadership skills. Funding difficulties (18-19%) are another significant hurdle.

[8] This ratio is defined based on two key questions (b2000b and b2000e). Question b2000b asks whether the household participates in self-employment businesses, while question b2000e captures the number of self-employment projects a household has been involved in.

Competition also plays a role (15-17%), potentially due to factors like price wars or declining customer base. Finally, policy changes (8-11%) can disrupt operations and force closures.

Understanding these exit reasons can inform policy interventions to support the self-employment sector. For instance, policymakers could consider programs that help self-employed individuals manage personal or family challenges. Initiatives to improve business skills and financial literacy could also address the issue of poor management. Additionally, promoting regulatory stability or policies that ease financial burdens on self-employed businesses might help reduce closures.

An interesting future exploration would be to compare the most common reasons for market entry (e.g., desire for flexibility and income) with the reasons for exit. This comparison might reveal mismatches between expectations and reality for self-employed individuals, informing more targeted support programs.

2.3 Self-employment and productivity

This section presents a preliminary investigation into the relationship between self-employment and productivity using last year's annual income as a proxy. Figure 3 shows the joint distribution of household income and self-employment rate, using the full 2015 CHFS dataset. The self-employment rates are calculated for 25 equal-sized quantiles of the income distribution. The figure reveals a distinct hook-shaped pattern, with a long, rising tail towards high incomes.[9] The self-employment rate starts high at 20.6% at the low-income spectrum, likely representing "subsistence entrepreneurs" driven by necessity. This rate dips sharply and then gradually rises with income, reaching a peak of 31.5% among the top 4% of earners.

This pattern aligns with the concept of coexistence of both subsistence and productive entrepreneurs.[10] However, the substantial rise in self-employment among high income earners suggests a potentially favorable ratio of productive entrepreneurs. Assuming earnings directly reflect productivity, the increasing presence of self-employed individuals in high-income brackets could have a positive impact on productivity and economic growth in the longer term.[11]

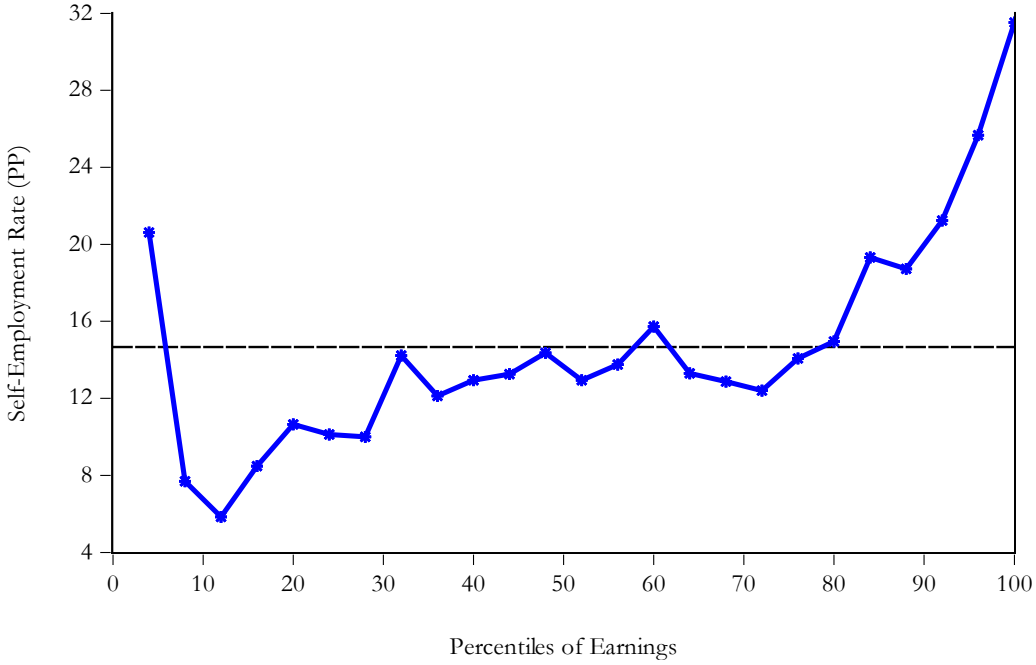
If we link income with business entry motivations (Panel (A) of Table 1), we find that self-employed individuals motivated by entrepreneurial spirit tend to have the highest average income, while those who are involuntarily self-employed have a lower average income.

[9] We obtain such joint distributions of similar pattern, using two subsamples for rural and urban residents.
It is worth noting that our analysis of the joint distribution of household income and non-agricultural self-employment rates is not comparable to the findings in Herreño & Ocampo (2023), where they include farmers within the self-employed category. Focusing on non-agricultural self-employment, our study sheds light on the dynamics of self-employment specifically within the service and commercial sectors, where individuals are more likely to be driven by entrepreneurial ambitions and profit potential. This focus explains why we observe a distinct hook-shaped pattern, unlike the U-shaped pattern (Figure 1) they observe due to the prevalence of subsistence farming among self-employed farmers in their study.

[10] Following established classification (among others, see Herreño & Ocampo, 2023; Zhang et al., 2021), subsistence and productive entrepreneurs are defined based on their earnings, which are assumed to reflect the value added by their businesses.

[11] Further nuanced research encompassing broader productivity and self-employment determinants is required to solidify this cautious optimism. This task is left for future studies.

Figure 3: Self-employment rate by income percentile (based on the 2015 CHFS survey)



Note: This figure shows the joint distribution of household income (last year's annual income) and self-employment, using the 2015 CHFS household dataset with altogether 5,471 households (from the total of 37,289) engaged in self-employment. The self-employment rates are calculated for 25 equal-sized quantiles of the income distribution. The horizontal dashed line marks the average self-employment rate of 14.7%.
Source: The 2015 CHFS survey.

2.4 Financing self-employment

This section addresses self-employment financing in two dimensions. First, we will show a recent trend of household self-employment leverage ratio and non-performing loan (NPL) ratio using the PBC's data (the People's Bank of China). Then, using the 2013 and 2015 CHFS survey data, we will show the self-reported reasons why households were discouraged from applying and rejected by banks.

2.4.1 Household leverage ratio and NPL ratio

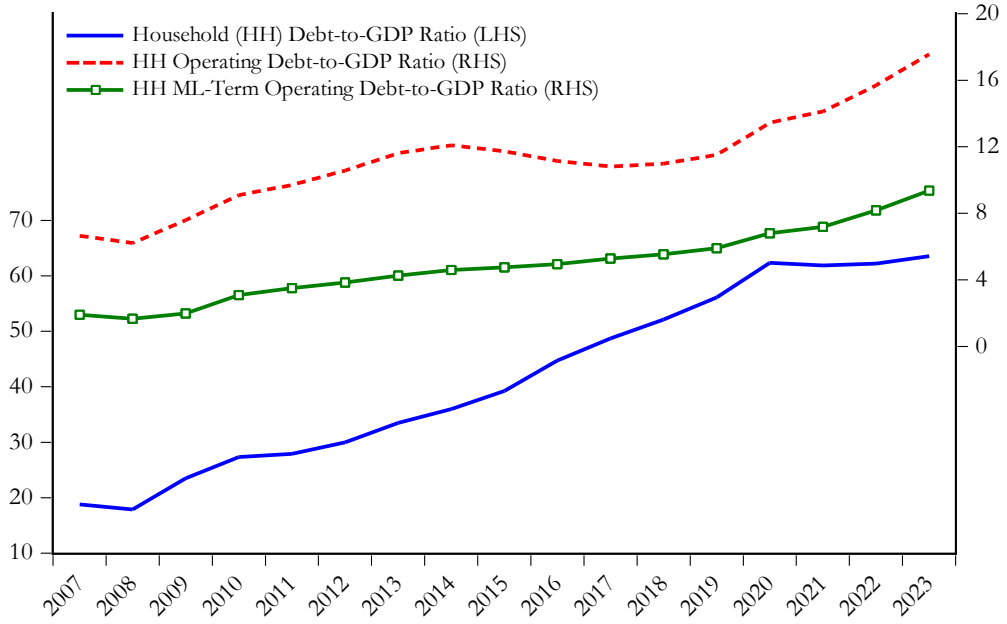
As China's household debt has soared in recent decades[12], a critical question arises: Are self-employment loans fueling a hidden risk bubble? To answer this, we delve into household loan categories, which are available from the PBC since 2007. As Figure 4 shows, the overall household leverage ratio has climbed from 19% in 2007 to 63.5% in 2023, driven primarily by consumption loans like mortgages. However, the remaining one-third, comprising operating loans for farming and self-employed businesses, has witnessed a similarly dramatic rise, nearly tripling from 6.6% to 17.6% in 2023.

[12] China's household leverage ratio has increased from around 7.5% in 1992 to 63.5% in 2023, according to China's Macro Leverage Ratio Database from the Center for National Balance Sheets (CNBS).This surge far outpaces the growth rates of other sectors, such as non-financial corporations, whose leverage has barely doubled, rising from around 90% to 169% over the same period.

While detailed data for self-employment loans is absent, analyzing medium- and long-term operating loans offers valuable insights as most farming loans are short-term. As Figure 4 reveals, the household self-employment leverage ratio, proxied by medium- and long-term operating loans,[13] has undergone a nearly fivefold increase, jumping from a mere 1.9% to 9.4% in 2023. This growth far outpaces the overall household leverage ratio, highlighting the rapid acceleration of self-employment borrowing.

Given the rapid ascent of self-employment borrowing, concerns emerge about potential risk. According to the PBC (China's Financial Stability Report), the overall household non-performing loan (NPL) ratio has always been at a low level. This ratio stood around 1.5-1.9% between 2013 and 2018. At the end of 2018, the balance of non-performing personal loans was 710 billion yuan, and the NPL ratio was 1.5%, which was 0.5 percentage points lower than the overall NPL ratio of bank loans.

Figure 4: Household leverage ratios, 2007-2023



Note: This figure shows the three household (HH) leverage ratios from 2007 to 2023: the overall household leverage ratio, the overall household operating leverage ratio, and the household medium- and long-term (in short as "ML-term") operating leverage ratio, which serves as a proxy for self-employment loans. All leverage ratios are defined as outstanding debt relative to annual nominal GDP.

Source: Authors' calculation, based on the PBC's various issues of "Summary of Sources And Uses of Credit Funds of Financial Institutions" (2007-2023).

Unfortunately, specific NPL data for household operating loans is absent. Notably, China's Financial Stability Report stopped updating the household NPL ratio in 2019, further hindering detailed analysis. However, a reasonable estimate based on low consumption NPLs[14] suggests that the NPL ratio for self-

[13] It is worth noting that this proxy might overestimate self-employment loans as household medium- and long-term operating loans might capture loans for other purposes like private firms.

[14] The household NPL ratios for three specific categories were published: personal housing loans (0.3% in 2018), personal credit card loans (1.6%), and personal auto loans (0.7%).

employed businesses could potentially range between 3% and 3.5% in 2018. While this may still be within a tolerable range, factors like economic downturns or industry vulnerabilities could lead to NPL increases in the future, necessitating further monitoring and analysis.

2.4.2 Barriers to loan acquisition

According to the 2013 and 2015 CHFS surveys, credit constraints are common for self-employed households. Over half (51%) of those who reported needing loans faced difficulties securing them. A substantial portion (78.5%) were discouraged from applying, while the remaining faced rejection by banks. Panel (A) of Table 2 highlights the reasons why self-employed borrowers with loan need chose not to apply, becoming "discouraged borrowers". Fear of rejection is the most prominent factor (36-48%). This is closely followed by application barriers (41-44%), where complex procedures and a lack of financial literacy ("do not know how to apply") create high application costs, discouraging potential borrowers. The 2015 CHFS survey identifies additional reasons that represent financial vulnerability: lack of collateral or guarantors (18%), concerns about interest rates (15%) and repayment (10%), and familiarity with loan officers (8%). All these factors turn out to be significant deterrents for a substantial number of borrowers.

Table 2: Discouragement and rejection reasons (based on the CHFS surveys)

(A) Discouragement from applying			(B) Self-reported reasons for rejection		
	2013	2015		2013	2015
1. Don't think the application will be approved.	47.6	36.2	1. No collateral	30.1	31.6
2. The application process is troublesome.	32.5	27.2	2. No guarantor	29.0	24.1
3. Don't know how to apply for a loan.	11.6	13.8	3. Not familiar with the loan officer	25.8	35.4
4. Don't have a guarantor or any collateral.		18.0	4. Low income and loan officer is concerned.	19.4	16.5
5. The interest rate is too high.		15.1	5. Policy reasons	7.5	6.3
6. Worried that I am unable to repay it.		9.6	6. Outstanding bank loan	3.2	0.0
7. Don't know the staff at the bank/credit union.		8.0	7. Bad credit history	0.0	3.8
Number of observations	372	312	Number of observations	86	79

Notes: This table shows the percentage breakdown of the main reasons for credit constraints for self-employed households, based on the 2013 and 2015 CHFS surveys. Panel (A) explores why households with loan needs (positive demand) chose not to apply for loans, based on responses to Question b3002a ("Why haven't you applied for a loan?"). Multiple choices were allowed. Panel (B) shows the percentage breakdown of the seven main (self-reported) reasons why loan applications were rejected, according to responses to Question b3003 ("Why was the loan rejected?").
Source: The 2013 and 2015 CHFS surveys.

This phenomenon of discouraged borrowers can have a ripple effect. When a substantial number of potential borrowers don't apply due to these factors, it can limit the pool of loan applications that banks receive. This could restrict the availability of loanable funds in the economy, potentially impacting overall economic activity.

These findings highlight areas for potential policy interventions. Simplifying loan application processes and promoting financial literacy programs specifically designed for self-employed individuals could address some of the discouragement factors. Future research could delve deeper into the reasons behind the fear

of rejection and explore the effectiveness of existing programs aimed at improving loan access for self-employed borrowers.

Panel (B) of Table 2 focuses on the self-reported reasons for loan application rejections among self-employed borrowers. Financial vulnerability is a major hurdle. Insufficient collateral or guarantor (56-59%) and concerns about low income and repayment ability (17-19%) are the top reasons. These limitations can significantly restrict the financial mobility potential for this vulnerable group, limiting their ability to grow their businesses and build wealth. This creates a vicious cycle that perpetuates income inequality. Another challenge is the lack of familiarity with loan officers (28-35%), suggesting potential shortcomings in communication or outreach strategies by lending institutions. This highlights a similarity with the discouragement factors (Panel A): Both financial limitations and lack of familiarity appear as hurdles throughout the loan process. Finally, it is worth noting that policy factors (7%), bad credit history (4%), and other outstanding loans (3%) appear to be less prominent reasons for rejection compared to financial limitations.

This differs somewhat from the discouragement factors in Panel (A), where fear of rejection and lack of financial literacy were more prominent. This suggests that even when a borrower overcome their initial hesitation and apply, their financial profile might still lead to a rejection.

Further research could delve deeper into understanding how to address these challenges. Exploring alternative lending models designed specifically for self-employed borrowers or investigating ways for lenders to better assess the financial health of these businesses could be valuable areas for future studies. Additionally, research on strategies for loan officers to improve communication and outreach to self-employed borrowers could be beneficial.

Overall, these findings suggest that addressing fear of rejection, simplifying loan application processes, and improving financial literacy among self-employed borrowers could significantly improve their access to credit.

3. The CHFS survey data

The CHFS survey, launched in 2011 by the Survey and Research Center for China Household Finance (Gan et al., 2014), offers a valuable resource for nuanced analyses of Chinese household finance. This nationally representative, biannual cross-sectional survey features randomized participation and a steadily growing sample size across its five published rounds (2011-2019), ranging from 8,438 to 40,011 households.

To understand the factors impacting access to bank loans for self-employed households in China, we analyze data from the 2013 and 2015 rounds. These two rounds delve deeper into the experiences of self-employed borrowers, revealing not just loan holdings but also crucial information on whether households lack loans due to need, discouragement, or rejection.

Both surveys cover 29 provinces, with 32% of the households living in the rural area.[15] This paper focuses on self-employed bank credit, so we limit the sample to the self-employed. Combining the self-employed samples from both rounds, after limiting to working-age household heads and dropping the observations with excessively missing number of variables, yields a final sample of 7,461 observations.

Using information on households' outstanding business bank loans and the responses to Question 2 given under A.2 in the Appendix, we categorize households into four groups:

(1) those who do not need credit;

(2) those who need but are discouraged from applying;

(3a) those who apply and succeed;

(3b) those who apply and are rejected.

Households are regarded having positive demand if they belong to one of three groups, those with positive levels of bank loans (in 3a) and those with zero debt levels who reported rejection (in 3b) or discouragement (in 2). Households are discouraged from applying due to the prospect of possible rejection or household inability to afford debt services in the future, or from a high application cost due to troublesome application procedures or their financial illiteracy (see, e.g., Chivakul & Chen, 2008; Jappelli, 1990). The discouraged households, together with the rejected[16], are defined as credit constrained.

One caveat to this definition is that even if a positive debt holding is observed, we do not know whether the household may be partially constrained because there is no information about the lender's evaluation of the debt request. Similar to Crook (2001), Magri (2002) and Chivakul and Chen (2008), we treat positive debt holding as no credit constraint, bearing in mind that we may underestimate the probability of being credit constrained.

Another caveat is that the survey lacks information on the specific date of loan rejections, only mentioning if a household was "turned down". For simplicity, rejections are assumed to be recent. The fact that we consider current wealth and income although credit rejection may have occurred in earlier periods may bias the results in the direction of overestimating the probability of credit rationing if some households that had been credit constrained in the past were no longer constrained at the time of the interview. This may have some impact on the estimated effects of income and wealth, but most other variables are more slowly changing, suggesting that the timing issue is not critical for the estimation of the effect of those variables.

Table 3 reports the sample means of key variables across different groups. To minimize the impact of outliers, household income and wealth (two continuous independent variables) are winsorized at 1% tails. Around 21% of self-employed households reported positive loan demand for business purposes. However, around half (51%) face credit constraints, with the majority (78%) discouraged from applying (Column (3)).

[15] The original sample sizes are 28,141 and 37,289 households, respectively.

[16] Fig. A2 in the Appendix summaries the (self-assumed) reasons why an application was rejected.

This highlights the importance of identifying discouraged borrowers for an accurate assessment of credit constraints among self-employed households.

The two credit-constrained groups (discouraged and rejected households) share similar financial characteristics, including low average income and wealth, and low rates of savings and homeownership. This similarity provides the rationale for grouping them together and also suggests a high likelihood of loan rejection even if discouraged households formally applied. In contrast, indebted households (Column (5)) exhibit higher average income, wealth, and education attainment, potentially indicating factors preferred by banks when granting loans.

Table 3: Sample means of selected variables, different groups

Variables	Total Sample (1)	No Need (2)	Dicouraged (3)	Rejected (4)	Indebted (5)
Household variables					
Income (in ¥)	97,827	96,840	70,891	69,505	133,523
Wealth (in ¥)	1,300,312	1,241,944	979,964	1,055,460	2,065,857
Saving status (dummy, in %)	74.7	76.0	67.7	63.9	73.4
Homeownership (dummy, in %)	90.3	90.1	87.8	87.0	94.2
Access to informal self-employed business loans (dummy, in %)	18.1	11.5	50.4	57.4	34.5
Rural (dummy, in %)	20.3	19.3	23.5	19.5	25.0
Individual (household head) variables					
Age	43	43	42	44	42
Education level (in %)					
No or basic education	17.7	17.5	19.1	26.6	15.7
Secondary education	68.8	69.0	67.6	65.1	68.8
Tertiary education	13.6	13.5	13.3	8.3	15.5
Married (dummy, in %)	91.3	91.3	90.0	92.3	92.3
Male (dummy, in %)	78.3	77.6	80.9	82.8	81.1
No attention to econ & fin info (dummy, in %)	58.9	61.0	51.7	55.6	49.7
Risk averse (dummy, in %)	58.1	60.3	51.5	57.4	46.3
Number of cases	7,461	5,912	617	169	763
% of cases		79.2	8.3	2.3	10.2

Note: This table presents the sample means for key variables across different groups.
Source: The 2013 and 2015 CHFS surveys.

While demographic features like age, marital status, and gender appear similar across the groups, interesting differences emerge in financial attitudes. The loan-seeking group (Columns (3)-(5)), particularly the indebted group, exhibits a higher average risk tolerance and a greater interest in economic and financial news compared to those with no loan needs (Column (2)). A higher risk tolerance suggests a potential desire for higher returns on investment, potentially motivating them to seek loans. Additionally, a stronger interest in economic and financial news suggests potentially higher financial literacy, which can lead to a higher loan demand and potentially lower application costs.

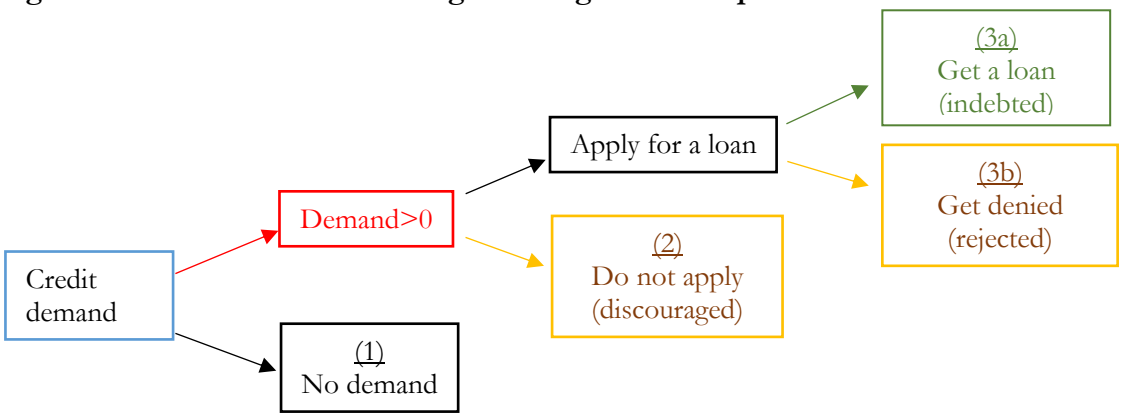
One interesting observation is about informal loans. More than half of credit-constrained households (52%) leverage informal loans, considerably exceeding the rate in the overall sample (18%) or among indebted

households (35%). This high reliance on informal loans suggests that credit-constrained households seek them as an alternative to formal bank loans when facing barriers to bank loans. Even among indebted households, 35% still use informal loans. While some may choose informal loans due to lower application or repayment costs (a topic for further research), others are likely forced to seek supplemental financing due to partial credit constraints. They may have received insufficient funds from banks, pushing them towards informal sources to bridge the gap. Therefore, the prevalence of informal loans as a complement to bank loans among indebted households serves as an indicator of unmet financial needs and the presence of partial credit constraints. This finding highlights the potential underestimation of the true extent of credit constraints faced by the self-employed, even after accounting for discouraged borrowers.

4. Sequential logit model of multi-stage decision process

Fig. 5 reveals that observed debt levels arise from a multi-stage decision process involving both credit demand and supply. This process begins with household demand for loans: either present or absent. If demand exists, a household may choose not to apply when it is discouraged by application costs (such as time and efforts to prepare all the documents), the prospect of possible rejection, or concerns about future repayment. Then, after applying, the supply side (e.g., banks) decides whether to grant the loan. This intricate process highlights that there are many ways in which the observed debt holding may take a value of zero.

Fig. 5: Household credit obtaining: A 3-stage decision process



Notes: This figure illustrates a 3-stage decision process that determines a household's debt level. In Stage 1, households decide whether to seek credit, where households with no credit demand are identified as Group (1). In Stage 2, those who need a loan might encounter discouragement, leading to the "discouraged group" (Group (2)). Stage 3 involves the formal application process, with successful applicants becoming the "indebted group" (Group (3a)) and rejected applicants forming the "rejected group" (Group (3b)).
Source: Modified from Chivakul and Chen (2008: 6).

This household credit obtaining process arises from a sequence of decisions with two discrete choices (0 (failure) and 1 (success)) at each of three stages. This process can be well modeled using a sequence of discrete-choice methods like logit or probit, where logit models are often preferred for their computational

efficiency. By applying this approach to household credit decisions, we can unveil factors contributing to debt-free status, whether due to no need, self-discouragement, or loan denial.

Based on Mare (1980) and Perraudin and Sorensen (1992), we model the transition probability from stage $j-1$ to j in a logistic regression. Let $D_{ij} = 1$ if the ith household passes stage j and $D_{ij} = 0$ otherwise. The event $D_{j-1} = 0$ signifies that a household drops out at stage $j-1$; therefore, there is no meaningful event corresponding to the outcome $D_j = 1$ and $D_{j-1} = 0$. Let X_{ijk} denote the kth regressor determining transitions to stage j from stage $j-1$. The transition probability from stage $j-1$ to j is as follows:

$$\Pr\left(D_{ij} = 1 \middle| X_{ij1}, X_{ij2}, \dots X_{ijk}, D_{i,j-1} = 1\right) = \mathrm{P}_{j-1,j}\left(\beta_{j0} + \textstyle\sum_k \beta_{jk} X_{ijk}\right) \qquad (1)$$

where the β_{jk} are parameters to be estimated from the data. The function $\mathrm{P}_{j-1,j}(\cdot) = \frac{\exp(\cdot)}{1+\exp(\cdot)}$ is the standard logistic function. The equations are estimated by maximum likelihood procedures.

As pointed out by Mare (1980), the continuation probabilities are asymptotically independent of one another, which implies that Eq. (1) can be estimated separately for each transition on the appropriate subsamples that "passes" the previous stage's decision point. In our analysis of the loan securing process, the first stage uses the full sample to examine credit demand, revealing how various regressors influence credit demand. The second stage examines household decisions to apply for loans, focusing on the subset of self-employed households with positive demand. Finally, the third stage analyzes the supply side using self-employed population who have submitted applications, revealing factors influencing bank credit granting decisions.

5. Empirical estimations

5.1 Baseline sequential logit estimations

The baseline model is built on a sequential logistic regression, as specified in Eq. (1). In so doing, we examine how household characteristics influence decisions across three stages of credit acquisition, where each stage builds upon the success of the previous stage. The three stages represent key decision points:

(i) Credit demand: Households first assess their need for credit and decide whether to seek a loan.

(ii) Loan application: Households with a positive loan demand then decide whether to apply for a loan.

(iii) Loan approval: Banks assess loan applications and make the final decision to grant or reject credit.

The independent variables, X's, include household income (standardized), wealth (standardized), saving status, access to informal credit, homeownership, rural area status, household head age, gender, education level, marital status, no attention to economic & financial (econ & fin) information (used as a proxy for financial illiteracy), and risk attitudes. Other control variables include industry dummies (two main industries

(Wholesale & Retail Trades, Hotel & Catering Services) and others), year dummies (2013 or 2015) and regional dummies (east, middle, and west).[17]

In modeling the sequential logit model, we assume that some household features are not observable to banks. More precisely, it is assumed that banks do not have the information about the household's access to informal loans, financial illiteracy and risk attitudes. These three variables are included in the first two households' decisions whether to desire and apply for loans, but not in the banks' decision equation to whom to grant loans.

These independent variables influence credit demand and supply in various ways (see, e.g., Chivakul & Chen, 2008; Jappelli, 1990; Perraudin & Sorensen, 1992). Financial resources like income, wealth, savings, homeownership, and access to informal credit impact a household's ability and need to borrow, which banks also taken into considerations when making credit decisions. Demographic variables like age, gender, education, and marital status can influence both risk assessment and risk-taking behavior, affecting both credit demand and supply. Financial literacy influences borrowing decisions. Individuals with higher risk tolerance might be more likely to desire and apply for credit. The control variables account for the fix effects at the industrial, year and reginal level given that, for example, industry-specific capital requirements, the economic prospect and risk level in different years and regions may affect both credit demand and supply.

Panel (A) of Tables 4 presents the estimates of the sequential logistic regression based on Eq. (1), with the parameter estimates for households' decisions whether to demand bank loans in Column (A1), the estimates for households' decisions whether to apply for loans given they have positive demand in Column (A2), and the estimates for banks' decisions whether to approve or reject loan applications in Column (A3). Positive parameters mean that a household having the corresponding characteristic is more likely to have loan demand, more likely to apply for loans, and more likely to obtain loans upon application. In general, the parameter estimates seem convincing.

Loan demand (Column (A1)): The analysis of loan demand (Column (A1)) identifies key factors influencing households' decisions to seek bank loans. Households with higher wealth, no savings, access to informal loans, and higher risk tolerance are more likely to seek loans. This suggests that wealthier households are inclined to leverage their assets (including real estate) to access bank loans, while liquid assets such as savings are used to self-finance household businesses. Interestingly, access to informal loans doesn't necessarily prevent seeking formal loans. This highlights the potential complementary use of both financing approaches. Furthermore, limited financial knowledge (financial illiteracy) and lower risk tolerance are associated with a reduced likelihood of seeking loans.

[17] Ideally, to capture finer-grained regional nuances, using province-level dummies instead of broader regional dummies would be preferred. However, due to missing failure cases from Stage 2 to 3 (rejection by the bank) in some provinces, this option is not viable.

Loan Applications (Column (A2)): Only three factors significantly influence households' decisions to apply for loans (after initial demand) (Column (A2)): income, wealth, and access to informal loans. Conditional on positive demand, households with stronger financial resources (higher income and wealth) are more likely to submit formal applications, likely due to a higher perceived chance of loan approval. In contrast, access to informal loans has a negative correlation with submitting bank loan applications. This might be because informal financing options make some households more sensitive to high bank loan application costs or low bank loan approval prospect. Consequently, they might prioritize informal loans despite potentially needing formal financing as well.

Table 4: Credit demand and supply of the self-employed, sequential logit or logit

	(A) Seqlogit model						(B) Logit model	
	(A1) Seqlogit-1 Household's demand		(A2) Seqlogit-2 Household's application		(A3) Seqlogit-3 Bank's loan granting		Bank's loan granting	
Variables	β_{1k}	z-statistics	β_{2k}	z-statistics	β_{3k}	z-statistics	β_k	z-statistics
Household variables								
Income (standardized) (c.v.)	-0.012	(-0.36)	0.165**	(2.24)	0.195	(1.60)	0.065*	(1.75)
Wealth (standardized) (c.v.)	0.200***	(6.02)	0.398***	(4.92)	0.347***	(2.67)	0.303***	(8.64)
Saving status (d.)	-0.185**	(-2.54)	-0.061	(-0.49)	0.189	(0.97)	-0.177*	(-1.92)
Homeownership (d.)	-0.024	(-0.22)	0.264	(1.39)	0.648**	(2.18)	0.377**	(2.27)
Access to informal self-employed business loans (d.)	1.710***	(24.92)	-0.364***	(-3.25)				
Rural (d.)	0.252***	(3.20)	0.065	(0.48)	0.731***	(3.16)	0.400***	(4.13)
Individual (household head) variables								
Age (c.v.)	-0.002	(-0.63)	0.003	(0.49)	-0.011	(-1.06)	-0.012**	(-2.52)
Education (No or primary school (o.d.))								
Middle school (d.)	-0.047	(-0.55)	0.009	(0.06)	0.551**	(2.46)	0.131	(1.18)
College or more (d.)	-0.054	(-0.44)	-0.179	(-0.82)	0.953**	(2.56)	0.055	(0.35)
Married (d.)	-0.010	(-0.09)	0.192	(0.97)	-0.064	(-0.18)	0.153	(1.00)
Male (d.)	0.167**	(2.08)	-0.028	(-0.19)	-0.336	(-1.36)	0.158	(1.55)
No attention to econ & fin info (d.)	-0.306***	(-4.72)	0.110	(0.97)				
Risk averse (d.)	-0.312***	(-4.74)	-0.039	(-0.34)				
Year dummies	Yes		Yes		Yes		Yes	
Regional dummies	Yes		Yes		Yes		Yes	
Industry dummies	Yes		Yes		Yes		Yes	
Constant	Yes		Yes		Yes		Yes	
Obsrvations	7,461		1,549		932		7,461	
Pseudo R-Squared	0.123		0.064		0.079		0.047	

Notes: Panel (A) of this table presents the estimates of the sequential logit regression based on Eq. (1) for the self-employed business loans, while Panel (B) presents the estimates of the unconditional (using the full sample) logit regression for bank loan granting. Income and wealth are winsorized at 1% tails and standardized to have a mean of zero and unit variance. (c.v.) denotes a continuous variable; (d.) denotes a dummy variable (taking the value of 0 or 1); and (o.d.) denotes an omitted dummy (i.e., base category). Positive parameters imply that households with the indicated characteristic are more likely to have credit demand (Column (A1)), more likely to apply for loan given positive demand (Column (A2)), and more likely to get loans upon application (Column (A3)). Next to the estimated coefficients are z-statistics in parentheses. ***, **, and * indicate the significance at the levels of 1%, 5% and 10%, respectively.
Source: Authors' estimation using the 2013 and 2015 CHFS surveys.

Loan granting (Column (A3)): Column (A3) focuses on loan rationing, exploring the factors influencing banks' decisions to approve loans. The results show that wealth, homeownership, and higher education

(middle school and college) significantly increase the chance of securing a loan. These findings are consistent with research from other countries and theoretical predictions. Wealthier borrowers or homeowners are generally perceived as more creditworthy by banks, making them more attractive borrowers. Home can also serve as collateral, reducing the risk for banks and increasing the likelihood of loan approval. Similarly, banks favor borrowers with higher education (middle school or college), anticipating their greater future earning potential and loan repayment ability. The predicted probability suggests that, on average (holding other variables constant at the medium values), middle school graduates are 10.8 percentage points more likely to be approved compared to the least-educated group, and college graduates are 15.6 percentage points more likely.

General discussions: This 3-stage analysis controls for year, industry, and region. As expected, the estimates suggest significant year-, industry-, and region-specific heterogeneities in loan demand, application, and granting.

One caveat is that the Pseudo R-squared values are moderate (0.12, 0.06, 0.08), suggesting that other important factors not included in the model (due to unobservability), such as the size of the loan requested, potential future earnings, and repayment ability, could also play a significant role in loan access for self-employed households. Future research could explore how incorporating alternative data sources might improve the model's fitness to capture these unobserved factors.

Our 3-stage sequential logit model effectively identifies key financial and social-demographic characteristics that influence self-employed households' loan demand, application, and approval probabilities. These characteristics include income, wealth, saving status, access to informal loans, location, and household head demographics (gender, education, financial interests, and risk attitudes). The influence of these factors varies across the three stages (demand, application, approval). On the demand side, holding all else constant, self-employed households with more wealth, access to informal loans, male heads, higher financial interest, lower risk tolerance, lacking savings, and residing in rural areas are more likely to seek loans. Among those needing loans, households with higher income and wealth, but lacking informal loans, are more likely to apply formally despite the application costs. On the supply side, banks are more likely to approve loans for households with greater wealth, homeownership, rural residence, and more educated heads.

While our results confirm general theoretical expectations and findings from other countries, they also reveal several interesting nuances specific to the self-employed population in China. One key finding concerns the role of income and wealth in the loan process. Wealth significantly impacts all three stages: loan demand, application, and loan granting. In contrast, income only influences the decision to apply for a loan (after initial demand). Notably, income and wealth are positively correlated (correlation coefficient of 0.51). When we estimate the independent effects of income, excluding wealth, on loan demand, application, and granting (results not shown due to space limitations), we find positive and significant effects. However, these effects are explained away by wealth in our baseline model when wealth is included. This

suggests that wealth may capture underlying factors that influence both income and loan demand and acquisition, and that these factors are not fully accounted for by income alone.

Another interesting finding is that rural households are more likely to seek and secure bank loans for self-employed businesses compared to their urban counterparts, even after controlling for other demographics. This suggests an absence of systematic discrimination against rural residents, but the interpretation requires caution. This geographic differences in loan demand and acquisition could be due to factors unobservable in our dataset, e.g., specific loan types or loan sizes. Further investigation might be needed to understand the underlying reasons behind this finding.

The third interesting finding is that the analysis finds no significant association between the household head's age, marital status, and the likelihood of loan demand, application, or loan approval. Interestingly, males are more likely to initiate the loan seeking process compared to females. However, gender does not significantly impact loan approval decisions. This is somewhat inconsistent with prior research suggesting potential bias favoring females in loan approvals. However, a cautious interpretation is warranted due to the limited sample size for females (around 4%) and unmarried individuals (around 8.7%) as shown in Table 1. This limitation restricts the generalizability of the findings on gender and marital status. The potential influence of age on these three-stage decisions might be non-monotonic, which will be elaborated on in detail in the following sections.

Finally, the analysis cannot directly assess potential racial or ethnic discrimination in loan access due to missing ethnicity data in the 2015 survey. However, using the 2013 data with ethnicity information, we find no significant differences in loan demand, application, or approval between Han-Chinese and non-Han-Chinese groups. This suggests no ethnic bias in self-employment loan acquisition process in 2013. However, a cautious interpretation is necessary due to the overwhelming Han-Chinese majority (around 96%) in the sample. This limitation restricts the generalizability of the findings on ethnicity.

5.2 Comparing the sequential model with the simple logit model

The rich information in the CHFS surveys allows us to identify relevant sub-samples to trace loan demand and supply. This enables us to use the 3-stage logit model for more accurate estimation of loan access. For comparison, Panel (B) of Table 4 shows the results of a simple logit regression for the bank's loan rationing decision, estimated using the entire sample without considering loan demand and application. This simple logit regression produces significantly different results compared to Column (A3) in Panel (A). In Panel (B), the coefficients for education levels become insignificant, while other variables (income, wealth, saving status, homeownership, and age) show significance (though some at the 10% level). Notably, the coefficient for saving status even has the opposite sign.

These discrepancies arise because all these factors influence both households' demand for loans and banks' decisions to grant them. The simple logit model in Panel (B) fails to disentangle their impact on the demand

side from that on the supply side. It mixes the impact of these variables across different stages, presenting a distorted view of their true influence on banks' lending decisions.

This approach leads to confounding estimates of the true impact on the supply side. The significance level, sign, and magnitude of the coefficients can be inaccurate. For example, when a factor like saving status has opposing effects on demand and supply (negative on demand, positive on supply), the simple regression overlooks its positive (though insignificant) effect on loan granting because the stronger negative demand-side effect dominates. On the other hand, if a factor like income has the same positive effect on both loan application and granting, neglecting the demand-side effect could lead to an overestimation of its true impact on loan supply (appearing more significant than it actually is).

5.3 Extensions and group heterogeneities

The baseline model directly regresses on three continuous variables: income, wealth and age. This treatment assumes a linear influence of these variables on loan demand, application and loan granting. However, decisions might not be based solely on marginal changes in these variables. The extended model addresses this limitation by categorizing households into groups based on income, wealth, and age. This grouping strategy offers several advantages. First, it acknowledges that decisions are likely based on a household's position relative to others (in terms of income, wealth, and age), which is more plausible. Second, it allows for non-linear relationships between the variables and the outcomes. The baseline model's linear assumption might miss these potential nonlinearities. Third, the extended model facilitates the identification of heterogeneity. By exploring different groups, we can reveal variations in how income, wealth, and age affect loan access across different segments of the population.

To capture economic realities beyond simple intervals, we define income, wealth, and age groups based on economic criteria. We use the sample statistics, such as median and quartiles, to define six income groups (negative, low, medium-low, middle, medium-high, and high) and five wealth groups (low, medium-low, middle, medium-high, high). Four age groups (young, medium-young, middle-aged, old) reflect key life stages. While the exact cut-offs for each group involve some judgement, we aim to balance granularity and meaningful representation of economic disparities and life phases of the population.

We re-run the sequential logit model using Eq. (1), replacing continuous income, wealth, and age variables with group dummies. The estimates for other variables remain comparable to those in the baseline (Table 4), while the model fit has slightly improved. The Pseudo R-squared values increased to 0.13, 0.08 and 0.09 for credit demand, application, and credit granting stages, respectively. This improvement suggests that the group-based approach better captures the underlying relationships in the data.

The estimated coefficients before the group dummies reveal interesting heterogeneities across income, wealth, and age groups. These group-specific effects are presented in detail in Tables 5-7.

5.3.1 Income group heterogeneity

Table 5 reports estimated income group differences in credit demand (Panel (A1)), application (Panel (A2)), and banks' credit granting (Panel (A3)). The reported parameters represent estimated differences between the row group and the column group (the reference group). *Z*-statistics are reported in parentheses below each estimate. Positive parameters indicate that households in the row group are more likely experience that event (loan demand, application, or approval) compared to the column group.

While statistical significance is crucial, it does not guarantee economic significance. Due to the inherent nonlinearity in logit models, estimated parameter values cannot be directly interpreted as probabilities. To circumvent this limitation and extract valuable insights, we can calculate the marginal probability change for specific outcomes by altering a given characteristic while keeping all other variables constant.

Table 5: Estimated income group heterogeneity for the self-employed

	Low income		Medium low income		Middle income		Medium high income		High income (>=300K) (484)	
	Parameter	Predited Pr.	Parameter	Predited Pr.	Parameter	Predited Pr.	Parameter	Predited Pr.	Parameter	Predited Pr.
(A1) Seqlogit-1: Household's demand										
Negative income (<0) (270)	0.433**	0.045	0.328*	0.035	0.645***	0.062	0.584***	0.058	0.230	0.026
	(1.98)		(1.77)		(4.40)		(3.75)		(1.24)	
Low income [0-1K) (239)			-0.105	-0.010	0.212	0.017	0.151	0.012	-0.202	-0.019
			(-0.51)		(1.22)		(0.82)		(-0.96)	
Medium low income [1K-10K) (445)					0.317**	0.027	0.256*	0.022	-0.097	-0.010
					(2.47)		(1.80)		(-0.55)	
Middle income [10K-100K) (4,225)							-0.061	-0.005	-0.415***	-0.036
							(-0.73)		(-3.10)	
Medium high income [100K-300K) (1,798)									-0.354***	-0.032
									(-2.68)	
(A2) Seqlogit-2: Household's application										
Negative income (<0) (270)	0.469	0.085	0.483	0.088	0.766***	0.149	0.264	0.046	-0.008	-0.001
	(1.31)		(1.62)		(3.27)		(1.02)		(-0.02)	
Low income [0-1K) (239)			0.015	0.003	0.297	0.064	-0.204	-0.040	-0.477	-0.086
			(0.04)		(1.01)		(-0.65)		(-1.27)	
Medium low income [1K-10K) (445)					0.283	0.061	-0.219	-0.043	-0.491	-0.089
					(1.31)		(-0.90)		(-1.55)	
Middle income [10K-100K) (4,225)							-0.502***	-0.104	-0.774***	-0.150
							(-3.30)		(-3.00)	
Medium high income [100K-300K) (1,798)									-0.272	-0.047
									(-1.04)	
(A3) Seqlogit-3: Bank's credit granting										
Negative income (<0) (270)	0.423	0.078	0.545	0.103	0.221	0.039	-0.176	-0.027	-0.587	-0.080
	(0.82)		(1.25)		(0.64)		(-0.46)		(-1.10)	
Low income [0-1K) (239)			0.122	0.026	-0.202	-0.039	-0.599	-0.105	-1.010*	-0.158
			(0.24)		(-0.47)		(-1.28)		(-1.67)	
Medium low income [1K-10K) (445)					-0.323	-0.065	-0.721*	-0.131	-1.132**	-0.184
					(-1.01)		(-1.93)		(-2.12)	
Middle income [10K-100K) (4,225)							-0.398	-0.066	-0.809*	-0.119
							(-1.56)		(-1.75)	
Medium high income [100K-300K) (1,798)									-0.411	-0.053
									(-0.88)	

Notes: This table presents the estimated income group differences based on the sequential logit model (Eq. (1)), using group dummies for income, wealth, and age. Households are categorized into six income groups based on self-reported annual income, with each group's income range and number of observations indicated in brackets following the group name. "Parameter" reports the estimated differences between the row group and the column group (the reference group). *Z*-statistics are in parentheses below each estimate. ***, **, and * indicate the significance at the levels of 1%, 5% and 10%, respectively. Positive parameters indicate that households in the row group are more likely experience that event (loan demand, application, or approval) compared to the column group. "Predicted Pr." reports the predicted probability differences between the two groups, holding all other variables at their medium values while varying income.

Table 5 includes "Predicted Pr." to show predicted probability differences between different income groups. Each difference reflects the change in probability when varying income groups (row group minus column group), while holding all other variables constant at their median values. The median values represent a middle-wealth ([500K-1M)) household with homeownership, savings, no informal self-employed business loans, residence in the eastern region, middle age ([40-49]), middle-school education, married, male, no financial interest, and risk aversion. By isolating income variations in this manner, we can effectively observe its relative effects on credit demand, application and granting.

Table 5 unveils a different pattern compared to Table 4's baseline model. While the baseline model identifies a single linear relationship between income and loan application at the second stage only, Table 5 finds a significant non-monotonic relationship through all stages: demand, application, and granting. This pattern is evident in the significant income group differences observed across all three panels, particularly pronounced between households at both income extremes (very low or very high) and the middle-income group.

At the lower end, negative-income households stand out for their strong credit need. Compared to all others, they exhibit a significantly higher desire and application rate for loans (Panels (A1) and (A2)). These discrepancies are particularly stark against the middle-income group. Predicted probabilities reveal a striking 6.2 and 14.9 percentage points difference in needing and applying for loans (conditional on positive demand), respectively, compared to their middle-income counterparts. This suggests a strong motivation among these households to leverage bank loans for financial improvement. Interestingly, banks appear more willing to lend to negative-income households compared to low- and middle-income groups, although the difference in approval rates is not statistically significant (Panel (A3)).

Similar trends emerge at the other end of the income spectrum. The high-income group displays higher demand and application rates compared to most other groups, particularly the middle-income group (Panels (A1) and (A2)). Predicted probabilities reveal a 3.6 and 15.0 percentage points advantage for the high-income group in needing and applying for loans, respectively. They also enjoy a higher, though sometimes statistically insignificant, probability of loan approval than all other groups (Panel (A3)).

5.3.2 Wealth group heterogeneity

Analogous to Table 5, Table 6 reports estimated differences between five wealth groups, with the column group as the reference group. Predicted probability differences are reported alongside estimated parameters. These differences reflect how varying household wealth impacts outcomes while holding all other variables constant at their medium values with income set to the middle-income range ([10K-100K)).

Table 6 reinforces the pattern observed in the baseline model (Table 4): a monotonic positive wealth-credit relationship. This means that wealthier households are more likely to experience each stage of the loan process (demand, application, approval) compared to less wealthy ones. This is evident in two key

observations. First, wealthier households consistently have higher rates of loan demand, application, and approval throughout all three stages (Panels (A1) - (A3)). The negative parameter values in the table (often statistically significant) indicate a higher likelihood for wealthier groups compared to the reference group. Second, the wealth gap amplifies the effect of wealth on loan access. The impact of wealth becomes stronger as the difference in wealth between groups increases, particularly noticeable in the application and approval stages. For example, the wealthiest group is 36.8 percentage points more likely to apply for loans (given they have a demand) compared to the poorest group, and 25.3 percentage points more likely to secure a loan once they apply.

Table 6: Estimated wealth group heterogeneity for the self-employed

	Medium low wealth		Middle wealth		Medium high wealth		High wealth (>=3M) (736)	
	Parameter	Predited Pr.	Parameter	Predited Pr.	Parameter	Predited Pr.	Parameter	Predited Pr.
(A1) Seqlogit-1: Household's demand								
Low wealth (<100K) (504)	-0.171	-0.010	-0.348**	-0.022	-0.496***	-0.034	-0.773***	-0.060
	(-1.17)		(-2.26)		(-3.11)		(-4.22)	
Medium low wealth [100K-500K) (2,428)			-0.176**	-0.012	-0.325***	-0.024	-0.602***	-0.050
			(-2.06)		(-3.51)		(-4.69)	
Middle wealth [500K-1M) (1,930)					-0.149*	-0.012	-0.425***	-0.037
					(-1.67)		(-3.45)	
Medium high wealth [1M-3M) (1,863)							-0.277**	-0.026
							(-2.40)	
(A2) Seqlogit-2: Household's application								
Low wealth (<100K) (504)	-0.590**	-0.146	-1.000***	-0.244	-1.100***	-0.266	-1.613***	-0.368
	(-2.30)		(-3.68)		(-3.90)		(-4.85)	
Medium low wealth [100K-500K) (2,428)			-0.410***	-0.098	-0.510***	-0.120	-1.023***	-0.221
			(-2.76)		(-3.16)		(-4.25)	
Middle wealth [500K-1M) (1,930)					-0.100	-0.022	-0.613***	-0.124
					(-0.64)		(-2.62)	
Medium high wealth [1M-3M) (1,863)							-0.513**	-0.101
							(-2.30)	
(A3) Seqlogit-3: Bank's credit granting								
Low wealth (<100K) (504)	-0.667	-0.143	-0.669	-0.143	-1.193***	-0.227	-1.394***	-0.253
	(-1.61)		(-1.57)		(-2.63)		(-2.62)	
Medium low wealth [100K-500K) (2,428)			-0.002	0.000	-0.526*	-0.084	-0.728*	-0.110
			(-0.01)		(-1.92)		(-1.87)	
Middle wealth [500K-1M) (1,930)					-0.524**	-0.084	-0.725*	-0.109
					(-2.03)		(-1.93)	
Medium high wealth [1M-3M) (1,863)							-0.202	-0.025
							(-0.54)	

Notes: This table presents the estimated wealth group differences ("Parameter") based on the sequential logit model (Eq. (1)), using group dummies for income, wealth, and age. Households are categorized into five wealth groups, with each group's wealth range and number of observations indicated in brackets following the group name. "Predicted Pr." reports the predicted probability differences between the two groups, holding all other variables at their medium values while varying wealth. For further details regarding methodology and interpretation, please refer to the notes under Table 5.

5.3.3 Age group heterogeneity

Table 7 delves into age-based differences, presenting estimated differences between four age groups with the column group as the reference group. Predicted probability differences are reported alongside estimated parameters. These differences reveal how varying household head age impacts outcomes while holding all other variables constant at their medium values with income and wealth set to the middle ranges.

Unlike the baseline model (Table 4) which showed no clear connection between age and credit access, Table 7 reveals a more nuanced pattern: a nonmonotonic relationship between age and both credit demand and supply. On the credit demand side, middle-aged households (in their 40s, accounting for about one third of the whole sample) are more likely to need credit compared to both the young and old groups. This trend continues into the application stage, where middle-aged households are significantly more likely to apply for loans than their young counterparts. On the supply side, the middle-aged are less likely to get loan approval compared to both the young and the old groups. This is an interesting pattern: the middle-aged group is facing a vulnerable situation with limited access to credit despite their higher loan demand. This suggests that beyond controlled financial and demographic characteristics, factors like future earning potential or how banks perceive those young and old borrowers' business risk profiles may influence loan decisions. Further research is needed to understand this trend in more detail.

Table 7: Estimated age group heterogeneity for the self-employed

	Middle young age		Middle age		Old [50-60] (2,105)	
	Parameter	Predited Pr.	Parameter	Predited Pr.	Parameter	Predited Pr.
(A1) Seqlogit-1: Household's demand						
Young [18-29] (709)	-0.247*	-0.015	-0.317**	-0.020	-0.130	-0.008
	(-1.92)		(-2.48)		(-0.97)	
Middle young age [30-39] (1,838)			-0.070	-0.005	0.117	0.008
			(-0.87)		(1.30)	
Middle age [40-49] (2,809)					0.187**	0.013
					(2.35)	
(A2) Seqlogit-2: Household's application						
Young [18-29] (709)	-0.391*	-0.095	-0.475**	-0.114	-0.366	-0.089
	(-1.76)		(-2.16)		(-1.57)	
Middle young age [30-39] (1,838)			-0.084	-0.019	0.025	0.006
			(-0.60)		(0.16)	
Middle age [40-49] (2,809)					0.109	0.025
					(0.77)	
(A3) Seqlogit-3: Bank's credit granting						
Young [18-29] (709)	0.584	0.074	0.918*	0.131	0.531	0.066
	(1.22)		(1.95)		(1.08)	
Middle young age [30-39] (1,838)			0.334	0.057	-0.053	-0.008
			(1.48)		(-0.20)	
Middle age [40-49] (2,809)					-0.387*	-0.065
					(-1.67)	

Notes: This table presents the estimated age group differences ("Parameter") based on the sequential logit model (Eq. (1)), using group dummies for income, wealth, and age. Households are categorized into four age groups according to the household head's age, with each group's age range and number of observations indicated in brackets following the group name. "Predicted Pr." reports the predicted probability differences between the two groups, holding all other variables at their medium values while varying household head's age. For further details regarding methodology and interpretation, please refer to the notes under Table 5.

6. Conclusions

This paper examines the challenges of accessing bank loans for self-employed businesses in China. Building on two CHFS surveys, we construct a three-stage sequential logit model to illuminate three crucial phases: household loan demand, their application decisions, and banks' approval choices. The analysis yields several key insights. First, credit constraints were widespread among Chinese self-employed households in 2013 and 2015. More than half of those needing loans faced credit constraints, discouraged from applying or rejected by banks. This critical barrier represses entrepreneurship and economic growth. Second, economic conditions, risk tolerance, financial literacy, and even socio-demographic characteristics all influence household borrowing behavior and loan success. This complexity underlines the need for nuanced policies that consider these diverse influencing factors. Third, significant fixed costs associated with applying for loans can disproportionately discourage households from pursuing loans. This raises concerns about equitable access to credit and suggests the need for policies like financial literacy training or simplified loan processes to reduce these barriers. Fourth, the analysis sheds light on banks' often-opaque loan assessment criteria, revealing the significant influence of socio-demographic features on their preferences. This raises questions about potential biases and the necessity for transparent lending practices. Fifth, the extended model, considering non-monotonic influences of income, wealth, and age, demonstrates substantial heterogeneities across groups in credit needs, application choices and loan approvals. This underlines the need for more granular and targeted policies based on these diverse characteristics, like tailoring financial literacy programs for different age groups.

Methodologically, the sequential logit model enables a joint examination of both the supply and demand sides of the credit market. This approach allows me to investigate whether specific population groups face more credit rationing because they demand more loans or because they are perceived as bad credit risks by banks. The advantage of this approach becomes clear when compared to a simple logit model: analyzing loan approvals without disentangling demand-side influences risks confounding results. The sequential logit model avoids this pitfall, providing a more nuanced and accurate picture of the factors at play.

Recognizing the vital role of small and micro businesses in generating jobs and reducing poverty, the Chinese government has launched the "Inclusive Small and Micro Loans" Program since 2012 (PBC, 2012). This policy has improved credit access for marginalized groups, making it easier for them to start businesses. Our research builds on this progress by offering further policy insights for designing effective self-employment bank credit programs. While our analysis doesn't find evidence of gender or age discrimination in loan approvals, it emphasizes the need for targeted policies that specifically address the challenges faced by low-income, less wealthy, and less-educated households in accessing credit. By empowering these groups to pursue self-employment, we can unlock their entrepreneurial potential and drive economic growth, job creation and poverty alleviation.

References:

In 2012, the People's Bank of China (PBOC) issued a guideline on 普惠小微贷款. The guideline called for financial institutions to provide more loans to small businesses and individuals with limited access to traditional credit.

Arora, Raksha, "Borrowing Behavior in China," *Gallup Poll*, March 29, 2005.

Balboni, C., Bandiera, O., Burgess, R., Ghatak, M., & Heil, A. (2021). Why Do People Stay Poor? *The Quarterly Journal of Economics, 137*(2), 785-844. doi:10.1093/qje/qjab045

Chivakul, M., & Chen, K. (2008). What Drives Household Borrowing and Credit Constraints? Evidence from Bosnia & Herzegovina. *IMF Working Paper, WP/08/202*.

Crook, J. (2001). The Demand for Household Debt in the USA: Evidence from the 1995 Survey of Consumer Finance. *Applied Financial Economics, 11*(1), 83-91.

Gan, L., Yin, Z., Jia, N., Xu, S., Ma, S., & Zheng, L. (2014). *Data You Need to Know About China: Research Report of China Household Finance Survey 2012*. Berlin: Springer.

Herkenhoff, K., Phillips, G. M., & Cohen-Cole, E. (2021). The Impact of Consumer Credit Access on Self-Employment and Entrepreneurship. *Journal of Financial Economics, 141*(1), 345-371. doi:https://doi.org/10.1016/j.jfineco.2021.03.004

Herreño, J., & Ocampo, S. (2023). The Macroeconomic Consequences of Subsistence Self-Employment. *Journal of monetary economics, 136*, 91-106. doi:https://doi.org/10.1016/j.jmoneco.2023.02.002

Hout, M., & Rosen, H. (2000). Self-Employment, Family Background, and Race. *The Journal of Human Resources, 35*(4), 670-692. doi:10.2307/146367

Jappelli, T. (1990). Who Is Credit Constrained in the U. S. Economy? *The Quarterly Journal of Economics, 105*(1), 219-234. doi:10.2307/2937826

Magri, S. (2002). *Italian Households' Debt: Determinants of Demand and Supply*. Retrieved from

Mare, R. D. (1980). Social Background and School Continuation Decisions. *Journal of the american Statistical association, 75*(370), 295-305.

Perraudin, W. R. M., & Sorensen, B. E. (1992). The Credit-Constrained Consumer: An Empirical Study of Demand and Supply in the Loan Market. *Journal of Business & Economic Statistics, 10*(2), 179-192. doi:10.1080/07350015.1992.10509897

Wang, S.-Y. (2012). Credit Constraints, Job Mobility, and Entrepreneurship: Evidence from a Property Reform in China. *The Review of Economics and Statistics, 94*(2), 532-551. doi:10.1162/REST_a_00160

Zhang, T., Zhang, M., & Zhang, L. (2021). Self-employment of Chinese rural labor force: Subsistence or opportunity?—An empirical study based on nationally representative micro-survey data. *Journal of Asian Economics, 77*, 101397. doi:https://doi.org/10.1016/j.asieco.2021.101397

Zhu, C., Lee, C.-C., Hong, J., & Shi, X. (2023). Multidimensional cultural distance and self-employment of internal migrants in China. *International Review of Economics & Finance, 86*, 58-81. doi:https://doi.org/10.1016/j.iref.2023.03.005