A theory of financial inclusion and income inequality

Gerhard Kling[a,*], Vanesa Pesque-Cela[b], Lihui Tian[c], Deming Luo[d]

[a] *University of Aberdeen*
[b] *SOAS, University of London, United Kingdom*
[c] *Nankai University, China*
[d] *Zhejiang University, China*

Abstract

We develop a theory linking financial inclusion, defined as access to formal loans and financial assets, to income inequality. Initial inequality of households is modeled by a random variable determining initial endowments. These initial endowments can be used to invest instantaneously in human capital and financial assets. Human capital translates into income based on a strictly concave production function, suggesting optimal levels of investment. Financial assets earn yields which do not depend on the amount invested by individuals. Theoretical predictions are tested using the China Household Finance Survey (CHFS) for 2011 and 2013. Initial conditions modeled by a random variable are replaced by an actual distribution of income or assets to derive theoretical predictions regarding the proportion of the population that might benefit from financial inclusion. Financial inclusion does mitigate under-investment in education - but formal loans do not contribute. Income inequality worsens if households rely on formal or informal loans, whereas access to bank accounts improves households' prospects in the future income distribution. However, households below the 40th percentile of household income do benefit from informal loans.

Keywords: Financial inclusion, income inequality, theory

*Corresponding author: email: gerhard.kling@abdn.ac.uk; Acknowledgements: We are very grateful for the financial support provided by the ESRC and NSFC (Newton Fund). This paper is part of the research project entitled Research on China's Financial System towards Sustainable Growth: The Role of Innovation, Diversity and Financial Regulation (ESRC: ES/P005241/1 and the National Natural Science Foundation of China: 71661137002. In addition, this research was supported by the National Social Science Foundation of China (Project Number: 17ZDA071).

A theory of financial inclusion and income inequality

Abstract

We develop a theory linking financial inclusion, defined as access to formal loans and financial assets, to income inequality. Initial inequality of households is modeled by a random variable determining initial endowments. These initial endowments can be used to invest instantaneously in human capital and financial assets. Human capital translates into income based on a strictly concave production function, suggesting optimal levels of investment. Financial assets earn yields which do not depend on the amount invested by individuals. Theoretical predictions are tested using the China Household Finance Survey (CHFS) for 2011 and 2013. Initial conditions modeled by a random variable are replaced by an actual distribution of income or assets to derive theoretical predictions regarding the proportion of the population that might benefit from financial inclusion. Financial inclusion does mitigate under-investment in education - but formal loans do not contribute. Income inequality worsens if households rely on formal or informal loans, whereas access to bank accounts improves households' prospects in the future income distribution. However, households below the 40th percentile of household income do benefit from informal loans.

Keywords: Financial inclusion, income inequality, theory

1. Introduction

Economic theory suggests that financial exclusion can lead to persistent inequality. In the presence of credit market imperfections, individuals' initial wealth determines their ability to invest in human or physical capital, which prevents social mobility of the poor and perpetuates inequality (Aghion and Bolton, 1997; Banerjee and Newman, 1993; Galor and Zeira, 1993; Ghatak and Jiang, 2002; Galor and Moav, 2004; Mehrotra and Yetman, 2014). Seminal theoretical papers by Aghion and Bolton (1997), Banerjee and Newman (1993) and Galor and Zeira (1993) model heterogeneity of individuals using a distribution of initial wealth, which is similar to our approach. Their theories model a binary choice - to be or not to be an entrepreneur (Aghion and Bolton, 1997; Banerjee and Newman, 1993) or a skilled worker (Galor and Zeira, 1993), which requires investment. These models suggest that financial inclusion can reduce income inequality by increasing opportunities for education and entrepreneurship among the poor. However, if one relaxes the assumption of binary choices and introduces a continuous variable of investment in education, theoretical predictions differ. A 'middle class' emerges for whom investing in education by taking loans is not a value maximizing strategy. Even in the absence of credit rationing not all individuals might benefit from financial inclusion. Our theory addresses this research gap by modeling how individuals maximize their future discounted excess income using loans (financial inclusion) to invest in education. As individuals are heterogeneous, i.e. have different initial endowments, these individual investment decisions do influence income inequality. We establish two Theorems, demonstrating that not all individuals can benefit from financial inclusion. Hence, depending on parameter values financial inclusion can increase or decrease inequality.

While most cross-country studies have found a negative relationship between financial inclusion and income inequality (Honohan, 2008; Mookerjee and Kalipioni, 2010; Neaime and Gaysset, 2018; Kim, 2016; Turégano and Herrero, 2018), it is unclear whether these findings hold for Asia, in general, and China, in particular (Park and Mercado, 2018). Furthermore, these studies have not investigated the mechanisms or processes underlying the relationship between financial inclusion and inequality, nor have they examined whether the effects of financial inclusion vary across different financial services. Most importantly, many prior studies use country (Kim, 2016) or state level data (Jayaratne and Strahan, 1996; Burgess and Pande, 2005; Beck et al., 2007), which cannot reveal heterogeneity among individuals or households. This paper aims to contribute to this literature by investigating the relationship between financial inclusion and income inequality both theoretically and empirically using disaggregated data. Following Allen et al. (2012), financial inclusion is defined as the use of formal financial services. Our theoretical predictions are tested using data from a nationally representative sample of over 8,000 Chinese households collected by the China Household Finance Survey (CHFS) in 2011 and 2013. Our findings suggest that financial inclusion does mitigate under-investment in education - but formal loans do not contribute. Income inequality worsens if households rely on formal or informal loans, whereas access to bank accounts improve households position in the future income distribution. Yet, households in the bottom 40% of the household income distribution do benefit from informal loans. This empirical finding is expected as a consequence of Theorem 2, which derives an upper bound in terms of initial endowments. Hence, households that are 'too rich' are not expected to benefit from access to loans to finance education.

Our contribution to the literature is threefold. First, recent theoretical work on the impact of financial inclusion on the distribution of income has been limited, and has focused on entrepreneurship as the causal mechanism linking access to financial services and inequality (Besley et al., 2018; Dabla-Norris et al., 2020). Our theoretical model investigates the impact of financial inclusion on income inequality through its effects on education and human capital accumulation. By permitting a continuous control variable, i.e. investment in education, our theory reveals within country disparities, i.e. not all individuals are expected to benefit from financial inclusion. Second, with few exceptions (Zhang and Posso, 2019, e.g.), most empirical studies on this issue are cross-country studies that use supply-side data on financial inclusion collected from financial regulators. We investigate the relationship between financial inclusion and inequality in China, using demand-side data on financial inclusion collected from households. The choice of China as the setting for the empirical analysis is motivated by the fact that China has the world's largest unbanked population (Demirguc-Kunt et al., 2017). Moreover, China's inequality levels used to be close to Nordic countries but they are now approaching US levels (Piketty et al., 2019). Its increase in inequality is unprecedented (Naughton, 2018),[1] and unequal access to education has been one of the driving factors (Jain-Chandra et al., 2018). Finally, while a recent review of the literature on financial inclusion suggests that not all financial products are equally effective in reducing inequality (Demirguc-Kunt et al., 2017, p.18-19), existing studies have failed to provide theoretical and empirical evidence showing the potentially differential impact of account ownership (and the savings and payment services they provide) and credit on household income inequality. This paper aims to fill this gap. The

[1] Between 1983 and 2016, China's Gini coefficient increased from 0.28 to about 0.46 Naughton (2018)

rest of the paper is organized as follows. Section 2 reviews the theoretical and empirical literature on the relationship between financial inclusion and inequality. Section 3 derives our theoretical model. In Sections 4 and 5, we discuss our empirical strategy and data. Section 6 presents our results, and Section 7 concludes.

2. Prior research

Economic theory provides conflicting predictions about the relationship between financial inclusion and inequality. The theoretical models of Galor and Zeira (1993) and Banerjee and Newman (1993) suggest a negative linear relationship between access to finance and income inequality. In contrast, the model of Greenwood and Jovanovic (1990) predicts a non-linear, inverted U-shaped relationship between access to financial services and inequality, which depends on the level of economic development. Work by Claessens and Perotti (2007) contends that causality may run in the opposite direction, from an unequal distribution of income to an unequal distribution of access to finance.

As noted by (Aslan et al., 2017, p.6), while existing theories on the relationship between finance and inequality refer explicitly to a link between financial access (or inclusion) and income inequality, most of the empirical literature has initially examined the link between financial development and income distribution.[2] Cross-country evidence on the distributional effects of financial development is mixed. Evidence from Beck et al. (2007) indicates that financial development reduces income inequality. This is consistent with more recent findings by Zhang and Naceur (2019). However, research by Kim and Lin

[2] See recent reviews of this literature by Claessens and Perotti (2007); Demirgüç-Kunt and Levine (2009); De Haan and Sturm (2017); Cihak et al. (2020).

(2011) and Law et al. (2014) suggests that the relationship between financial development and inequality is non-linear and depends on a country's level of financial development or institutional quality. In contrast to these studies, research by Dabla-Norris et al. (2015) and De Haan and Sturm (2017) shows that financial development increases inequality. Evidence from China is equally mixed, with some studies finding a negative relationship between finance and inequality Jalil and Feridun (2011); others an inverted-U relationship Zhang and Chen (2015); and yet others a positive relationship Koh et al. (2020).

A more recent literature has investigated the relationship between financial inclusion and income inequality. This literature has thus far been dominated by cross-country studies. Most of them point to a negative relationship between financial inclusion and inequality (Honohan, 2008; Mookerjee and Kalipioni, 2010; Kim, 2016; Neaime and Gaysset, 2018; Turégano and Herrero, 2018) but their findings differ across regions and countries (Park and Mercado, 2018; Dabla-Norris et al., 2020). For instance, Mookerjee and Kalipioni (2010), as well as Neaime and Gaysset (2018), find that increasing access to financial services, through bank branches results in a less unequal income distribution. Similarly, findings by Aslan et al. (2017) show that increasing the intensity of use of financial services[3] leads to a reduction in income inequality. However, a more recent study by Park and Mercado (2018) finds that increasing the accessibility, availability and usage of financial services[4] tends to reduce income inequality, except in developing Asia (including China

[3]The intensity of use of financial services is proxied by the share of individuals having an account at a financial institution, saving at or borrowing from a financial institution, and making or receiving digital payments (Aslan et al., 2017).

[4]These various dimensions of financial inclusion are measured by: the number of automated teller machines and commercial bank branches per 100,000 adults; the number of borrowers from, and depositors with, commercial banks per 1,000 adults; and the ratio of domestic credit to GDP (Park and Mercado, 2018).

and India).

This finding is in contrast to those of China-specific studies (Zhang and Posso, 2019; Zhang et al., 2018; Huang and Zhang, 2019). While the empirical evidence on the potential inequality-reducing effects of financial inclusion in China is limited, it generally suggests that financial inclusion can contribute to reducing inequality, by facilitating entrepreneurship among the rural poor. Focusing on rural China, Zhang and Posso (2019) find that poor households benefit more from financial inclusion than rich ones. Using household data from the 2011 China Household Finance Survey, they construct a multidimensional index of financial inclusion, which includes measures of account ownership, savings, credit, and insurance, and investigate its impact on household income. Their findings show that financial inclusion has a positive effect on income, and that this effect is larger for households at the lower quantiles of the income distribution, indicating that it reduces income inequality within rural areas. Financial inclusion has also been found to reduce rural-urban income inequality[5] (Huang and Zhang, 2019). Findings by Zhang et al. (2018) indicate that digital financial inclusion is positively associated with household income, and this positive association is only significant for rural households. Similarly, Zhang et al. (2019) find that Fintech-driven financial inclusion has a positive effect on household income, and this positive effect is larger for rural than for urban households. These studies suggest that one of the mechanisms underlying the inequality-reducing effect of (digital) financial inclusion is entrepreneurship by low-income rural households (Zhang et al., 2018). Our paper contributes to this literature by investigating theoretically and empirically: (i) whether fi-

[5]Specifically, they find that financial inclusion reduces rural-urban income inequality in the long run, but can increase it in the short term, due to initial disparities between rural and urban areas in terms of access to financial infrastructure and education (Huang and Zhang, 2019).

nancial inclusion, understood as the use of formal financial services, can reduce income inequality by facilitating investment in human capital; and (ii) whether different types of financial services are equally effective in reducing inequality.

3. Theoretical considerations

The theory aims to understand the process of financial inclusion, i.e. the demand for formal financial services. These services include debt denoted D and financial assets (e.g. savings) A. Debt incurs costs in the form of interest payments with a rate r_D, while financial assets yield r_A. Banks, which are not modeled explicitly, act as intermediaries taking deposits A and providing loans D. To cover their transaction costs, there needs to be a positive net interest margin $\nu = r_D - r_A > 0$ even in the absence of default risk. This interest margin could be regarded as a measure of the quality of financial services.

Demand for finance arises due to the need of households to invest in human capital (K), i.e. education. This investment yields income (Y), i.e. there is a trade-off between consumption now and consumption later, governed by inter-temporal preferences. These preferences are captured in the discount rate r. The link between income and human capital is modeled using a strictly concave production function $f(K)$, hence, $f'(K) > 0$ and $f''(K) < 0$. This implies that there exists an optimal capital stock K^* where marginal benefits equal marginal costs. Individuals maximize their excess consumption denoted C, i.e. basic needs such as food are assumed to be covered. They can choose to invest an amount I into human capital K or consume any excess income.

The most important feature of financial inclusion is the fact that households exhibit heterogeneity with respect to access to finance. The model, hence, captures this hetero-

geneity in the initial condition. Individuals enter the stage with an initial endowment W_0, which is a realization of the random process W with probability density ρ and support on $\mathbb{R}^+$. At time $t = 0$ (the start of the model) individual i's endowment is drawn and public knowledge. The latter assumption can be relaxed, leading to information asymmetry (adverse selection). Moreover, households are risk-neutral, which simplifies the model without loss of generality in the absence of uncertainty after $t = 0$. We assume that individuals do not receive an initial income $Y(0) = 0$, which suggests the following budget constraint at $t = 0$.

$$W_0 = K(0^+) + A(0^+) - D(0^+), \quad K(0) = A(0) = D(0) = 0 \tag{1}$$

Equation (1) states that initial endowment (e.g. wealth or initial human capital) can be invested into human capital $K(0^+)$ and the financial asset $A(0^+)$. If the initial endowment is insufficient, debt $D(0^+)$ can be used to reach desired levels of human capital. Note that $\nu = r_D - r_A > 0$ implies that debt is not used to acquire the financial asset.

Individuals maximize their future discounted excess consumption C, where human capital (K), financial assets (A) and debt (D) are state variables. Individuals have an infinite time horizon. We ignore taxes. Note that investment refers to $I = \dot{K}$ and repayments of loans $S = \dot{D}$. In most applications, we consider the case where $r_D > r > r_A$, i.e. after the initial adjustment at $t = 0$ governed by (1) individuals do not increase their financial assets. Continuous discounting is applied. Capital does not depreciate.

$$C = \int_0^\infty \left(f(K) + r_A A - r_D D - S - I\right) e^{-rt} dt$$
$$\dot{K} = I,\ \dot{D} = S,\ \text{for } t > 0^+ \tag{2}$$

Note that using the initial endowment as described in the budget constraint at $t = 0$ in (1) suggests an impulse control problem because the state variables debt, assets and human capital can change instantaneously as suggested by (1). After the initial impulse, continuous processes captured in the two differential equations in (2) apply. This setup refers to an impulse control problem. In addition, control constraints apply as $I \in [0, f(K) - r_D D]$; hence, the state variables K and D enter the control constraint. To solve this problem analytically, we apply Theorem 1 in Kling (2020). A detailed mathematical discussion of this type of impulse control problem can be found in Kling (2020). The proof of Theorem 1 in Kling (2020) shows that the optimal time for an impulse is at $t = 0$ as suggested by (1). Hence, relaxing the assumption that the initial endowment W_0 is invested instantaneously (i.e. at n points in time denoted t_j with $j = 1, 2, \ldots, n$) does not affect optimal time paths of states (e.g. human capital). Furthermore, maximizing the value of the future discounted excess consumption C can be written as follows.

Theorem 1. *The optimization problem* (2) *can be rewritten as follows, where* $\kappa(D(0^+))$ *is the time into steady state and* K^* *is the steady state capital stock.*

$$\max_{\{I\}_0^\infty, \{D(t_j)\}_1^n} C = \max_{D(0^+)} \left(\frac{f(K^*)}{r} e^{-r\kappa(D(0^+))} \right),\ W_0 < K^*, r_D > r,\ \left.\frac{df}{dK}\right|_{K=K^*} = r$$

Theorem 1 refers to the case $W_0 < K^*$; hence, the individual does not have sufficient endowment to 'jump' into the preferred state of human capital K^*. Therefore, the individual can take a loan $D(0^+)$ at $t = 0$, which constitutes an impulse. As $r_D > r$, this impulse is used to move human capital from zero (the assumed initial value) to K^*_{poor}, which is determined $f_{K^*_{\text{poor}}} = r_D$. Hence, $K^*_{\text{poor}} < K^*$ and it constitutes a transient steady-state of human capital for poor individuals. The initial optimal debt level is then $D(0^+) = K^*_{\text{poor}} - W_0$.

Hence, being 'poor' in our theory means that the initial endowment W_0 is not sufficient to instantaneously reach the optimal level of human capital K^*. The initial endowment W_0 is exogenously drawn from the probability density ρ, i.e. 'rolling the dice at birth'. However, the optimal level of human capital K^* depends on the interplay between marginal benefits of human capital and the cost of debt as derived in Theorem 2. Consequently, poverty is relative and partly endogenous in our theory.

As $r_D > r$, a poor individual has an incentive to reduce the debt level and uses all resources to do so until debt is repaid. The time when debt is repaid is labeled κ_D. The repayment time can be found from the following inhomogeneous first order differential equation using the condition $D(\kappa_D) = 0$. The integrating factor is $\exp(-\int r_D dt) = \exp(-r_D t)$, and C is an arbitrary constant.

$$\begin{aligned} \dot{D} - r_D D &= -f\left(K^*_{\text{poor}}\right) \\ De^{-r_D t} &= -\int e^{-r_D t} f\left(K^*_{\text{poor}}\right) dt \\ D &= Ce^{r_D t} - \frac{f\left(K^*_{\text{poor}}\right)}{r_D} \end{aligned} \tag{3}$$

The constant C follows from the initial condition $D(0^+) = K^*_{\text{poor}} - W_0$. Note that there

is no initial cash flow from human capital as $Y(0) = 0$, i.e. $\dot{D} - r_D D = 0$ at $t = 0$.

$$C = K^*_{\text{poor}} - W_0 \tag{4}$$

Finally, we obtain κ_D setting $D(\kappa_D) = 0$.

$$\begin{aligned} 0 &= \left[K^*_{\text{poor}} - W_0\right] e^{r_D \kappa_D} - \frac{f\left(K^*_{\text{poor}}\right)}{r_D} \\ \kappa_D &= \frac{1}{r_D} \ln\left[\frac{\frac{f\left(K^*_{\text{poor}}\right)}{r_D}}{K^*_{\text{poor}} - W_0}\right] = \frac{1}{r_D} \ln\left[\frac{f\left(K^*_{\text{poor}}\right)}{\left(K^*_{\text{poor}} - W_0\right) r_D}\right] \end{aligned} \tag{5}$$

Obviously, the cash flow from human capital in the transient steady state exceeds interest payments on the initial loan. Therefore, $\kappa_D > 0$ as required based on (5).

Once all debt is repaid at time $t = \kappa_D$, a poor individual has an incentive to invest using income from human capital until K^*, the steady-state human capital stock, is reached. The time it takes to move to the final capital stock starting at time $t = \kappa_D$ follows from the following first-order differential equation.

$$\begin{aligned} \dot{K} &= f(K), \\ K(\kappa_D) &= K^*_{\text{poor}},\ K(\kappa - \kappa_D) = K^* \end{aligned} \tag{6}$$

(6) can be solved analytically depending on the production function $f(K)$. After the steady-state is reached at $t = \kappa$, any income goes into excess consumption assuming $r > r_A$, i.e. no financial asset will be accumulated. The future discounted value of excess consumption refers to the perpetuity cash flow discounted back from the time of the steady-

state $t = \kappa$ to $t = 0$ as shown in Theorem 1.

A similar analysis can be conducted for different types of individuals depending on the initial endowment W_0. Thus far, we explored CASE 3 in detail. Considering the initial condition and interest rates, the following cases emerge.

- CASE 1: $W_0 \geq K^*_{\text{rich}}$, i.e. the individual has sufficient endowment to invest optimally. Optimal investment is determined by $f_{K^*_{\text{rich}}} = \max(r, r_A)$, where f_K denotes $df(K)/dK$. Any unused endowment is invested in the financial asset $A(0) = W_0 - K^*_{\text{rich}}$ if $r_A > r$ or consumed.

- CASE 2: $K^*_{\text{poor}} \leq W_0 \leq K^*_{\text{rich}}$, i.e. individuals in the 'middle' do not have an incentive to take debt to invest in human capital. They will use income to grow to the optimal level of human capital K^*_{rich} assuming $r > r_A$.

- CASE 3: $W_0 < K^*_{\text{poor}}$, i.e. this individual lacks endowment to invest optimally. Optimal investment is determined by $f_{K^*_{\text{poor}}} = \max(r, r_D)$. Demand for finance refers to $D(0) = K^*_{\text{poor}} - W(0)$. If $r_D > r$ the transient steady state lasts until $t = \kappa_D$ when all debt is repaid. From $t = \kappa_D$ to $t = \kappa$ the individual funds further investment into human capital using income until K^* is reached.

As the net interest margin $\nu = r_D - r_A$ is positive, it follows that $K^*_{\text{rich}} \geq K^*_{\text{poor}}$. All households have the same function $f(K)$ transforming capital stock into income. Thus, heterogeneity stems from initial endowments.

In the absence of uncertainty after $t = 0$, information asymmetry and permitting an open economy (investments can exceed savings), households 'jump' at $t = 0$ into their steady state capital stock. Note that (2) does not permit an increase in debt after $t = 0$,

which is not optimal due to the fact that $f(K)$ is strictly concave. This also suggests that debt must be sustainable. These assumptions can be relaxed. For instance, if debt and financial assets have to be balanced then r_D and r_A could be derived endogenously. These extensions go beyond the scope of this paper.

Figure 1 illustrates the expected patterns for rich and poor individuals. Poor individuals have a demand for debt to reach their optimal capital stock $K^*_{\text{poor}} = 6$, and rich households 'jump' into $K^*_{\text{rich}} = 7.5$ instantaneously. Any excess endowment creates financial assets if $r_A > r$ or is consumed. Interestingly, Figure 1 demonstrates that there is a third group for whom debt is not attractive as r_D is too high. This 'middle class' will grow using their income to reach $K^*_{\text{rich}} = 7.5$ over time.

[Insert Figure 1]

Figure 2 illustrates optimal paths for rich individuals in red and poor individuals in green. Asset accumulation only occurs if $r_A > r$, otherwise any excess is consumed.

[Insert Figure 2]

Theorem 2 establishes the impact of financial inclusion on individuals in different parts of the initial endowment distribution. It is apparent that the distributional effects are ambiguous. Hence, the claim that financial inclusion even in ideal settings reduces inequality cannot be easily established. To obtain closed-form solutions, we choose the production function $f(K) = K^\alpha$ with $\alpha \in (0, 1)$, which implies that $f'(K) > 0$ and $f''(K) < 0$, i.e. the production function is strictly concave.

Theorem 2. *Under ideal settings (no frictions, no uncertainty after $t = 0^+$, no capital controls, $r_D > r > r_A$) and given $f(K) = K^\alpha$ with $\alpha \in (0, 1)$, financial inclusion benefits*

poor individuals where $W_0 < \left(\frac{\alpha}{r_D}\right)^{\frac{1}{1-\alpha}}$. *Individuals with* $K^*_{poor} \leq W_0 \leq K^*$ *do not benefit from financial inclusion. Rich individuals do benefit from financial assets.*

Proof. We set $f(K) = K^\alpha$ with $\alpha \in (0,1)$. As $r_D > r > r_A$, $K^*_{\text{rich}} = K^* = \left(\frac{\alpha}{r}\right)^{\frac{1}{1-\alpha}}$, whereas $K^*_{\text{poor}} = \left(\frac{\alpha}{r_D}\right)^{\frac{1}{1-\alpha}}$. If $W_0 < K^*_{\text{poor}}$, then CASE 3 applies, implying the following optimal time path of human capital. First, at $t = 0^+$ poor individuals jump to $K^*_{\text{poor}} = \left(\frac{\alpha}{r_D}\right)^{\frac{1}{1-\alpha}}$ using debt $D(0^+) = \left(\frac{\alpha}{r_D}\right)^{\frac{1}{1-\alpha}} - W_0$. Then the time until debt is repaid κ_D follows from (5).

$$\kappa_D = \frac{1}{r_D} \ln \left[\frac{\left(\frac{\alpha}{r_D}\right)^{\frac{\alpha}{1-\alpha}}}{\left(\left(\frac{\alpha}{r_D}\right)^{\frac{1}{1-\alpha}} - W_0\right) r_D} \right]$$

Then from (6), we obtain the first-order differential equation, where C is an arbitrary constant.

$$\begin{aligned} \dot{K} &= K^\alpha, \quad \kappa_D \leq t \leq \kappa \\ \int K^{-\alpha} &= \int 1 dt \\ \frac{1}{1-\alpha} K^{1-\alpha} &= C + t \end{aligned}$$

The constant C gives.

$$C = \frac{\alpha}{1-\alpha} \frac{1}{r_D} - \kappa_D$$

Finally, the time into the stead-state κ follows.

$$\begin{aligned}\kappa &= \frac{\alpha}{1-\alpha}\frac{1}{r} - \frac{\alpha}{1-\alpha}\frac{1}{r_D} + \kappa_D \\ &= \frac{\alpha}{1-\alpha}\left(\frac{1}{r} - \frac{1}{r_D}\right) + \frac{1}{r_D}\ln\left[\frac{\left(\frac{\alpha}{r_D}\right)^{\frac{\alpha}{1-\alpha}}}{\left(\left(\frac{\alpha}{r_D}\right)^{\frac{1}{1-\alpha}} - W_0\right)r_D}\right]\end{aligned}$$

Then from Theorem 1.

$$C_{\text{poor}} = \frac{1}{r}\left(\frac{\alpha}{r}\right)^{\frac{\alpha}{1-\alpha}} e^{-r\kappa}$$

A rich individual (CASE 1), i.e. $W_0 > K^*$ can reach K^* instantaneously and save excess initial endowment, which generates interest income in perpetuity.

$$C_{\text{rich}} = \frac{1}{r}\left(\frac{\alpha}{r}\right)^{\frac{\alpha}{1-\alpha}} + \left(W_0 - \left(\frac{\alpha}{r}\right)^{\frac{1}{1-\alpha}}\right)\frac{r_A}{r}$$

Finally, CASE 2, i.e. $K^*_{\text{poor}} \leq W_0 \leq K^*$, suggests that individuals in the middle category do not take debt and invest their initial endowment in human capital, i.e. $K(0^+) = W_0$. Income is invested in human capital until K^* is reached at time $t = \tau$.

$$\begin{aligned}\dot{K} &= K^\alpha, \quad 0^+ \leq t \leq \tau \\ \int K^{-\alpha} &= \int 1dt \\ \frac{1}{1-\alpha}K^{1-\alpha} &= C + t\end{aligned}$$

The constant C is determined by $K(0^+) = W_0$.

$$C = \frac{1}{1-\alpha} W_0^{1-\alpha}$$

Then the time into the steady-state follows.

$$\begin{aligned}\tau &= \frac{\alpha}{1-\alpha}\frac{1}{r} - \frac{1}{1-\alpha} W_0^{1-\alpha} \\ &= \frac{1}{1-\alpha}\left(\frac{\alpha}{r} - W_0^{1-\alpha}\right)\end{aligned}$$

Then from Theorem 1.

$$C_{\text{middle}} = \frac{1}{r}\left(\frac{\alpha}{r}\right)^{\frac{\alpha}{1-\alpha}} e^{-r\tau}$$

□

To illustrate Theorem 2, we write a Python code that calculates the proportion of the poor that benefits from financial inclusion. The calculation assumes that individual's endowment is drawn from a normal distribution. As outlined in Section 5, our sample refers to the China Household Finance Survey (CHFS), which is used to derive parameter values to compute Figure 3. For instance, cost of debt has a lower bound of 0.73%, which refers to the bottom 10% of our sample, and an upper bound of 12.00%, which is the top 10% of our sample. Figure 3 plots the proportion of the poor that benefits from financial inclusion as a function of cost of debt r_D and the production parameter α. Access to finance must be cheap to ensure that a large section of society benefits from financial inclusion.

[Insert Figure 3]

4. Empirical strategy

To test theoretical predictions, information on individuals' income (I), financial assets (A), debt (D), and human capital (K) is required. To understand the impact of financial inclusion on income inequality, it is ideal to explore a country such as China, which has undergone a rapid expansion of financial services, economic and social transition. Hence, we use data from the China Household Finance Survey (CHFS), which provides information on individuals and households. Control variables such as age, gender and location are available.

First, the empirical approach tries to estimate the link between human capital and income using various measures of education. Theoretically, in line with endogenous growth theory one would expect that investment in education K increases income I. The theory assumes an unknown functional form $I = f(K)$. It is plausible to assume (and also leads to a steady-state) that $f(K)$ in concave, i.e. $f'(K) > 0$ and $f''(K) < 0$. One could follow White (1980), which models non-linear relationships - but only works for marginal changes in variables as it basically refers to a Taylor approximation. We follow another approach using a Cobb-Douglas type function of various assets that generate a broad measure of income. Second, we require marginal costs of financing based on formal and informal loans, i.e. using the cost of debt r_D, optimal education K^* follows from $f'(K^*) = r$. Third, comparing marginal costs and benefits of investment determines under-investment in education. This might be due to financial constraints, i.e. lack of access to formal bank loans, which can be tested empirically. Fourth, we try to explain under-investment in education to understand underlying drivers including controls (e.g. gender, age), regional-effects and measures of financial inclusion. Finally, we try to establish whether financial inclu-

sion contributes to more or less income inequality comparing households' position in the income distribution over time.

5. Data and variables

The paper uses data from the China Household Finance Survey (CHFS), a nationally representative data set collected by Southwestern University of Finance and Economics. The data set includes information about household demographics, (non-financial and financial) assets, liabilities (including education loans); income and expenditures; social security and commercial insurance. The CHFS conducted in 2011 and 2013 covers 127,230 individuals in 29,733 households in the year 2011 or 2013. For some variables such as gender (*FEMALE*) or location (*RURAL*) data for all 127,230 cases are available; however, due to missing data some variables are only observed for a subset of cases as shown in the second column of Table 2. For instance, income (*INC*) is only available for 80,778 individuals.

Variables derived from the CHFS refer to income measures, education, assets, financial inclusion, motives for taking loans, individual controls and region-specific variables. To link individuals to households, we use a unique identifier (*HHID*). The *YEAR* of the survey is indicated. We use a broad measure of income (*INC*) that includes income from employment, second jobs, bonuses, income from farming and businesses. The dummy *FARM INC* indicates whether individuals receive income from farming. The dummy *IND INC* flags if individuals rely on income from businesses, which mostly refers to small businesses and self-employment.

Education variables refer to dummies that take value one if the person has no formal

education (*NO EDU*), primary education (*PRIM EDU*), secondary education (*SEC EDU*), and whether they attended college or university (*HIGH EDU*). In addition, the dummy *FOREIGN EDU* takes value one if the person studied abroad. We quantify assets as fixed assets (*FIX A*), which includes business assets, property (*HOUSE VAL*) and the value of cars (*CAR VAL*). The purchase price of properties is recorded (*HOUSE PAY*) but does not deviate substantially from current values on average. We indicate whether assets include business assets as this signals ownership of usually small companies (*BUS A*). We use a broad definition of financial assets (*FIN A*) including checking accounts, deposits, stocks, bonds, futures, warrants and wealth management products. Cash holding of households (*CASH*), and lending from households (*LEND*) are recorded.

To quantify financial inclusion, dummies signal whether individual have checking account and time deposits (*ACCOUNT*) and whether they take formal *LOAN* or informal loans *LOAN INF*. The amount of outstanding formal (*DEBT*) and informal debt (*DEBT INF*) is measured. We obtain annual interest rates on formal loans (*INTER*) and informal loans (*INTER INF*). To understand the motives of individuals for taking loans, we use a set of dummy variables. These dummies signal whether formal loans are used for financial investments (*INVEST*), consumption including buying cars (*CONSUME*), or education (*EDU*). Informal borrowing used to finance education is flagged (*INF EDU*) and the amount is measured (*INFDEBT EDU*). Finally, we use a dummy to indicate whether the main purpose of time deposits is to finance education (*DEPOSIT EDU*).

The *AGE* of individuals and their gender (*FEMALE*) serve as control variables. Location is controlled by the province and a dummy to capture rural environment (*RURAL*). Table 1 provides variable names and their definitions. Not all variables listed are used in

multivariate models; however, they are used to derive additional variables. These variables are introduced in subsequent sections.

[Insert Table 1]

Table 2 shows descriptive statistics for all variables in the sample including the number of observations *N*, mean, median, percentiles and the range. As there are two points in time for which we obtained survey data, namely 2011 and 2013, descriptive statistics reported in Table 2 refer to pooled data. Dummy variables such as no education (NO EDU) take the value one if a condition is met or zero otherwise. By construction, the mean of a dummy variable refers to the proportion of the population that has a certain characteristic. For instance, Table 2 states that the mean of the dummy variable for no education is 0.086; hence, 8.6% of the individuals in our sample have no formal education.

[Insert Table 2]

6. Empirical analysis

6.1. Descriptive findings

To obtain initial insights into patterns of income, education and other forms of investment, we provide some descriptive findings. As age is an important factor, individuals are classified into ten birth cohorts of equal size defined based on age deciles. Means of medians of variables in each birth cohort are calculated to reveal birth cohort related patterns.

Figure 4 plots average income and education for ten birth cohorts. Income peaks at median age, revealing an inverted U-shaped relationship between birth cohorts and income. On the right-hand side of Figure 4, categories of education are depicted, where

the vertical axis refers to the proportion of individuals in the respective birth cohort that exhibit the respective level of education. For instance, a value of 0.4 means that 40% have the respective characteristic. Education has undergone a profound shift exhibiting a strong birth cohort effect. The older generation at most attended primary school, whereas younger generations undergo vocational or university education.

[Insert Figure 4]

We define a broad measure of investing in education labeled *INVEST EDU*, which combines investment using time deposits (savings), formal and informal loans taken to fund education. Investment into education differs between birth cohorts, where the parents' generation and the young invest in education. Hence, there is a bipolar distribution of investment into education as shown in Figure 5. The vertical axes in both panel refer to the proportion of individuals with a respective characteristic, e.g. 0.05 means that 5% of a respective birth cohort use savings to invest in education. Savings are the predominant source of investment into education followed by informal and formal loans. Accordingly, financial inclusion might enhance investment in education by providing access to bank accounts including time deposits and formal bank loans. The following sections explore multivariate models to establish the relationships between income, investment in education and other asset classes as well as the link between education, financial inclusion and inequality suggested by our theoretical considerations.

[Insert Figure 5]

6.2. The impact of education on income

Theoretically, we expect that education should increase individuals' income. To quantify the partial impact of education on income, we run several regression models. Table 3 shows regression models that explain log income using measures of education and controls. Model [$A3$] controls for the year of the survey, and specifications [$A4$] and [$A5$] account for year and province level effects using a set of dummies. These three models exhibit higher explanatory power compared to [$A1$] and [$A2$] measured by adjusted R-squared and information criteria (AIC and BIC).

[Insert Table 3]

Model [$A1$] focuses on educational achievements using dummy variables for primary, secondary, and higher education, where the latter category includes vocational training at colleges and degree level education at universities. Most degrees refer to undergraduate degrees since postgraduate and doctorates are uncommon. No education serves as a reference category. Compared to no education, having secondary education provides the highest increase in expected earnings. Attending college or university has a slightly less pronounced effect. Obtaining a degree from a foreign university provides an additional boost to income. Specification [$A2$] accounts for the non-linear impact of age suggested by analyzing income and education patterns across birth cohorts (see Figure 4). Again an inverted U-shaped relationship prevails. Females and individuals living in rural settings are disadvantaged in the labor market, i.e. similar education leads to lower levels of income. Controlling for year and province level effects in the remaining three columns of Table 3 does not alter these results. In summary, education increases income and the

effect diminishes with higher levels of education. Location and gender matter, resulting in significantly lower earnings potential, which might undermine incentives to invest in education.

6.3. Optimal investment in education

To derive optimal investment in education, fixed and financial assets, we estimate the following Cobb-Douglas type function that determines income (*INC*), where A_{jt} refers to province j and year t effects (captured using dummies) and B_i are individual control variables including gender, age and rural settings.

$$\text{INC}_i = A_{jt}B_i(\text{INVEST EDU}_i)^{\alpha}(\text{FIN A}_i)^{\beta}(\text{FIX A}_i)^{\gamma} \tag{7}$$

Taking logs of (7) leads to a linear model, which is estimated using OLS and the Huber-White sandwich estimator for robust standard errors. The coefficients sum up to less than one, i.e. $\alpha+\beta+\gamma < 1$, (see Table 4) suggesting diminishing returns to scale. This confirms our theoretical consideration suggesting that the function linking human capital to income is concave. Table 4 demonstrates that investing in education, fixed and financial assets has diminishing returns as the three coefficients add up to less than one. Hence, doubling all three asset classes increases income by less than 100%. Yet again, there is evidence of a gender disparity and an urban-rural divide. However, women benefit from investing in education as their returns are the same indicated by an insignificant interaction term (*FEMxEDU*). In contrast, in rural settings investing in education does not yield returns in terms of higher income, demonstrated by the interaction term *RURxEDU*.

[Insert Table 4]

Using specification $[C4]$ as reference model, marginal returns of investing in education, fixed and financial assets can be calculated from the estimated coefficients $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ by taking partial derivatives of income with respect to assets (7).

$$\frac{\partial \text{INC}_i}{\partial \text{INVEST EDU}_i} = \hat{\alpha} A_{ji} B_i \frac{1}{(\text{INVEST EDU}_i)^{1-\hat{\alpha}}} (\text{FIN A}_i)^{\hat{\beta}} (\text{FIX A}_i)^{\hat{\gamma}} \tag{8}$$

From (8), it is obvious that $\frac{\partial \text{INC}_i}{\partial \text{INVEST EDU}_i} > 0$ and $\frac{\partial^2 \text{INC}_i}{\partial \text{INVEST EDU}_i^2} < 0$ suggesting an optimal level of investment in education. Similarly, marginal return of investing in fixed and financial assets can be derived. Comparing these marginal returns of investment with marginal costs such as interest rates on formal and informal loans establishes whether individuals would benefit from more investment.

Table 5 shows the 25th percentile, the median and the 75th percentile of estimated returns from education and interest rates in different cohorts. Cohorts are either defined by quantiles based on income or fixed assets. The latter is closer to our theoretical concept of endowment as fixed assets includes a wide range of assets such as property. A certain proportion of individuals in each class should benefit from investing in education; however, we observe a tendency that richer individuals, both in terms of income and assets, exhibit higher returns from education. This empirical finding is due to inherent disparities in the labor market as lower income or asset categories have a higher share of females and individuals living in rural settings. Accordingly, a further multivariate analysis is needed to disentangle these interrelationships.

[Insert Table 5]

Figure 6 illustrates that investing in education yields on average lower returns than the average return from investing in financial assets for both women and men in urban settings. As females exhibit lower income levels, they also experience lower marginal benefits from investing in education or financial assets. Interest rates in the informal sector exceed for some birth cohorts the marginal benefits of investing in education. As returns from education are diminishing with the current level of education, individuals with low levels of educational attainment do benefit more from a marginal increase of their education. Accordingly, we calculated marginal returns of investing education for each individual.

[Insert Figure 6]

Based on our reference model [$C4$] in Table 4 that derives the production function of income, we determine the marginal product of investing in education as in (8). Comparing the marginal benefit of education and interest rates on loans, we establish whether individuals under-invest in education (*UNDER*). Using logistic regressions with robust standard errors and controlling for provinces, years and birth cohorts, Table 6 provides five model specifications to reveal the determinants of under-investment in education.

[Insert Table 6]

Specification [$C1$] considers income, gender, rural settings, and dummies for educational achievements. As in previous results, females are less likely to under-invest in education, which is due to lower returns from education caused by the gender pay gap. Individuals in rural settings experience a similar problem as education leads to smaller benefits. The level of education has a positive and significant impact of being under-invested. In

addition, the impact diminishes from no formal education to secondary education. Model [$C2$] tests for financial inclusion and shows that having a bank account reduces the risk of under-investment, whereas access to loans does not exhibit a significant effect. Model [$C3$] indicates that informal loans do reduce the likelihood of under-investment, albeit to a lesser extent than having access to a bank account. Model [$C4$] considers assets such as the value of houses (*HOUSE VAL*), the value of cars (*CAR VAL*) and cash holding (*CASH*). These variables do not explain under-investment; however, one needs to consider that the sample size drops considerably due to missing data. Finally, specification [$C5$] demonstrates that households with financial assets are less likely to under-invest in education, whereas families that run small businesses tends to under-invest (*IND INC*).

6.4. Financial inclusion and inequality

To test whether investing in education and financial inclusion improves a household's position in terms of its income decile, we focus on a sub-sample of households surveyed in 2011 and 2013. 38.9% of households in the 2013 survey were also included in the 2011 survey. All variables are aggregated on the household level. *HH INC* refers to the combined household income, *HH EDU* and *HH FIN* represent investment in education and financial assets on the household level, respectively. We take the average age of members of the household (*HH AGE*) to account for cohort effects. We use indicators for account ownership (*HH ACCOUNT*), having a formal loan (*HH LOAN*), and having an informal loan (*HH LOAN INF*). If one member of the household has an account, a formal or informal loan, the household is assigned the value one in the respective indicator variable. Finally, we also include whether households have a time deposit for the purpose of financing education *HH DEPOSIT EDU*, whether households take informal loans to fund

education *HH INF EDU* and if so the amount of loans taken *HH INFDEBT EDU*. The position of households in income deciles is determined in 2011 and 2013, resulting in an ordinal variable from 1 (poorest) to 10 (richest) household. To account for the ordinal nature of the dependent variable, we estimate ordered logistic regressions reported in Table 7.

Holding the sample composition constant permits the use of lagged explanatory variables on the household level. Accordingly, we can establish the causal order or relationships, which goes beyond the previous empirical models Kling et al. (2017). All explanatory variables are lagged, i.e. they refer to 2011, whereas the dependent variable is the income decile of the household in 2013. The previous income decile in 2011 is a predictor of the future outcome, and as expected exhibits a positive and significant partial impact.

[Insert Table 7]

All specifications in Table 7 use robust standard errors and account for provinces using dummies. Model [$D1$] considers the impact of the household's previous position *INC DEC*, indicating that having a better starting position enhances the chances to stay ahead in the income distribution. Moreover, a dummy for rural settings is added, reconfirming that households in rural settings struggle to climb the ladder. Specification [$D1$] also controls for educational investment *HH EDU* and the average age of household members *HH AGE*. Education has a positive albeit small effect on enhancing a household's position in the income distribution. Models [$D2$] and [$D3$] test for financial inclusion. Financial inclusion is both, a blessing and a curse, as having a bank account improves household's position *HH ACCOUNT* - but having formal *HH LOAN* or informal loans *HH LOAN INF* diminishes prospects.

As shown in [*D*4], time deposits for education *HH DEPOSIT EDU* does have a significant and positive impact as well as the amount of informal loans for education *HH INFDEBT EDU*. The latter, however, exhibits a small magnitude of impact. After incorporating all variables in model [*D*5], the significance of time deposits for education *HH DEPOSIT EDU* vanishes. In general, the findings of model [*D*3] remain unchanged in [*D*5].

In line with Theorem 2, we rerun the analyses for different income deciles, which reveals a very different pattern if we consider the bottom 40% of the income distribution.[6] Table 8 restricts the sample to households below the 40th income percentile, indicating that informal loans *HH LOAN INF* do lift households in the bottom 40% to higher income deciles. In addition, poorer households seem to be better off in rural than urban settings. Accordingly, splitting the sample in line with theoretical suggestions provides further insides.

[Insert Table 8]

7. Conclusion

Our theory relaxes the assumption of binary choices made in prior studies (Aghion and Bolton, 1997; Banerjee and Newman, 1993; Galor and Zeira, 1993) permitting continuous investment decisions. Intuitively, increasing financial inclusion should lead to more investment in education, building human capital, which translates into higher earnings. However, depending on the initial endowment some households might under-invest

[6] Splitting the sample into the bottom 40% of households and the top 60% in terms of household income is in line with (Demirguc-Kunt et al., 2017, p.18).

as they face higher interest rates on loans. Financial inclusion also constitutes access to financial assets. If returns from financial assets exceed returns from investing in human capital, inequality can increase. Theorem 2 demonstrates that not all individuals might benefit from financial inclusion as a 'middle class' might find it unattractive to invest in education using expensive loans compared to their time preference.

Using household data from China, we show that education increases incomes; however, marginal benefits of higher attainments fade. In addition, females and individuals in rural settings receive lower earnings, diminishing their returns from investing in education. This in turn leads to lower optimal levels of investment in education as predicted by theoretical considerations. Quantifying under-investment in education reveals that a segment of the population should invest more as their marginal benefits outweigh marginal costs in terms of interest rates on formal and informal loans. Financial inclusion does mitigate the investment gap in education; however, this effect is predominately driven by having access to savings accounts and not formal bank loans. In addition, informal loans have a positive but smaller effect on closing the gap in education. Finally, we demonstrate that households can move to higher income deciles if they have access to bank accounts. Yet, having formal or informal loans diminishes the prospects of households. Hence, the impact of financial inclusion on inequality is mixed, suggesting that increasing access to loans might not result in better outcomes.

As suggested by Theorem 2 and in line with (Demirguc-Kunt et al., 2017, p.18), we split the sample into the bottom 40% of households and the top 60% in terms of household income. Our theoretical prediction that only poorer households benefit from loans is confirmed empirically. This is in line with recent research focusing on rural China that poor

households benefit more from financial inclusion (Zhang and Posso, 2019). Moreover, we show that informal loans are beneficial, while formal loans do not help households to climb the income ladder. This finding underlines that formal loans are not yet as effective as informal channels of finance, requiring further policy responses. These findings are consistent with the observation made by (Demirguc-Kunt et al., 2017, p.18-19) that "not all financial products are equally effective in reaching development goals, such as reductions in poverty and inequality".

While more research is needed, recent studies and our own findings suggest that the largest impacts come from access to bank accounts and savings - but also informal loans in the case of poorer households Demirguc-Kunt et al. (2017). In fact, findings from recent experimental studies conducted in various developing countries indicate that the impact of payment, (micro-)savings, and (micro-)insurance services on the income of poor and low-income households is consistently and significantly more positive than that of (micro-) credit (Banerjee et al., 2015b,a; Brune et al., 2011; Demirguc-Kunt et al., 2017; Dupas and Robinson, 2013; Karlan et al., 2014). More research is needed to gain a better understanding of how different dimensions of financial inclusion promote inclusive development.

References

Aghion, P., Bolton, P., 1997. A theory of trickle-down growth and development. The Review of Economic Studies 64 (2), 151–172.

Allen, F., Demirguc-Kunt, A., Klapper, L., Peria, M. S. M., 2012. The foundations of financial inclusion: Understanding ownership and use of formal accounts. The World Bank.

Aslan, G., Deléchat, C., Newiak, M., Yang, F., 2017. Inequality in financial inclusion, gender gaps, and income inequality. International Monetary Fund Working Papers 17 (236), 1–31.

Banerjee, A., Duflo, E., Glennerster, R., Kinnan, C., 2015a. The miracle of microfinance? Evidence from a randomized evaluation. American Economic Journal: Applied Economics 7 (1), 22–53.

Banerjee, A., Karlan, D., Zinman, J., 2015b. Six randomized evaluations of microcredit: Introduction and further steps. American Economic Journal: Applied Economics 7 (1), 1–21.

Banerjee, A., Newman, A., 1993. Occupational choice and the process of development. Journal of Political Economy 101 (2), 274–298.

Beck, T., Levine, R., Levkov, A., 2007. Big bad banks? The impact of US branch deregulation on income distribution. The World Bank.

Besley, T., Burchardi, K., Ghatak, M., 2018. The gains from financial inclusion: Theory and a quantitative assessment.

Brune, L., Giné, X., Goldberg, J., Yang, D., 2011. Commitments to save: A field experiment in rural Malawi. The World Bank.

Burgess, R., Pande, R., 2005. Do rural banks matter? Evidence from the Indian social banking experiment. American Economic Review 95 (3), 780–795.

Cihak, M., Sahay, R., et al., 2020. Finance and inequality. Tech. rep., International Monetary Fund.

Claessens, S., Perotti, E., 2007. Finance and inequality: Channels and evidence. Journal of comparative Economics 35 (4), 748–773.

Dabla-Norris, E., Ji, Y., Townsend, R., Unsal, D., 2015. Identifying constraints to financial inclusion and their impact on GDP and inequality: A structural framework for policy. No. 15-22. International Monetary Fund.

Dabla-Norris, E., Ji, Y., Townsend, R. M., Unsal, D. F., 2020. Distinguishing constraints on financial inclusion and their impact on GDP, TFP, and the distribution of income. Journal of Monetary Economics.

De Haan, J., Sturm, J.-E., 2017. Finance and income inequality: A review and new evidence. European Journal of Political Economy 50, 171–195.

Demirguc-Kunt, A., Klapper, L., Singer, D., 2017. Financial inclusion and inclusive growth: A review of recent empirical evidence. The World Bank.

Demirgüç-Kunt, A., Levine, R., 2009. Finance and inequality: Theory and evidence. Annu. Rev. Financ. Econ. 1 (1), 287–318.

Dupas, P., Robinson, J., 2013. Savings constraints and microenterprise development: Evidence from a field experiment in Kenya. American Economic Journal: Applied Economics 5 (1), 163–92.

Galor, O., Moav, O., 2004. From physical to human capital accumulation: Inequality and the process of development. The Review of Economic Studies 71 (4), 1001–1026.

Galor, O., Zeira, J., 1993. Income distribution and macroeconomics. The Review of Economic Studies 60 (1), 35–52.

Ghatak, M., Jiang, N. N.-H., 2002. A simple model of inequality, occupational choice, and development. Journal of development Economics 69 (1), 205–226.

Greenwood, J., Jovanovic, B., 1990. Financial development, growth, and the distribution of income. Journal of political Economy 98 (5, Part 1), 1076–1107.

Honohan, P., 2008. Cross-country variation in household access to financial services. Journal of Banking & Finance 32 (11), 2493–2500.

Huang, Y., Zhang, Y., 2019. Financial inclusion and urban–rural income inequality: Long-run and short-run relationships. Emerging Markets Finance and Trade, 1–15.

Jain-Chandra, S., Khor, N., Mano, R., Schauer, J., Wingender, P., Zhuang, J., 2018. Inequality in China: Trends, drivers and policy remedies. International Monetary Fund.

Jalil, A., Feridun, M., 2011. Long-run relationship between income inequality and financial development in China. Journal of the Asia Pacific Economy 16 (2), 202–214.

Jayaratne, J., Strahan, P. E., 1996. The finance-growth nexus: Evidence from bank branch deregulation. The Quarterly Journal of Economics 111 (3), 639–670.

Karlan, D., Ratan, A., Zinman, J., 2014. Savings by and for the poor: A research review and agenda. Review of Income and Wealth 60 (1), 36–78.

Kim, D.-H., Lin, S.-C., 2011. Nonlinearity in the financial development–income inequality nexus. Journal of Comparative Economics 39 (3), 310–325.

Kim, J.-H., 2016. A study on the effect of financial inclusion on the relationship between income inequality and economic growth. Emerging Markets Finance and Trade 52 (2), 498–512.

Kling, G., 2020. Measuring financial exclusion of firms. Finance Research Letters, forthcoming.

Kling, G., Harvey, C., Maclean, M., 2017. Establishing causal order in longitudinal studies combining binary and continuous dependent variables. Organizational Research Methods 20 (4), 770–799.

Koh, S. G., Lee, G. H., Bomhoff, E. J., 2020. The income inequality, financial depth and economic growth nexus in China. The World Economy.

Law, S. H., Tan, H. B., Azman-Saini, W., 2014. Financial development and income inequality at different levels of institutional quality. Emerging Markets Finance and Trade 50 (sup1), 21–33.

Mehrotra, A., Yetman, J., 2014. Financial inclusion and optimal monetary policy. BIS Working Papers No 476.

Mookerjee, R., Kalipioni, P., 2010. Availability of financial services and income inequality: The evidence from many countries. Emerging Markets Review 11 (4), 404–408.

Naughton, B., 2018. The Chinese economy: Adaptation and growth. MIT Press.

Neaime, S., Gaysset, I., 2018. Financial inclusion and stability in MENA: Evidence from poverty and inequality. Finance Research Letters 24, 230–237.

Park, C.-Y., Mercado, J., 2018. Financial inclusion, poverty, and income inequality. The Singapore Economic Review 63 (01), 185–206.

Piketty, T., Yang, L., Zucman, G., 2019. Capital accumulation, private property, and rising inequality in China, 1978–2015. American Economic Review 109 (7), 2469–96.

Turégano, D., Herrero, A., 2018. Financial inclusion, rather than size, is the key to tackling income inequality. The Singapore Economic Review 63 (01), 167–184.

White, H., 1980. Using least squares to approximate unknown regression functions. International Economic Review 21, 149–170.

Zhang, Q., Chen, R., 2015. Financial development and income inequality in China: An application of SVAR approach. Procedia Computer Science 55, 774–781.

Zhang, Q., Posso, A., 2019. Thinking inside the box: A closer look at financial inclusion and household income. The Journal of Development Studies 55 (7), 1616–1631.

Zhang, R., Naceur, S. B., 2019. Financial development, inequality, and poverty: Some international evidence. International Review of Economics & Finance 61, 1–16.

Zhang, X., Zhang, J., He, Z., 2018. Is Fintech inclusive? Evidence from Chinas household survey data.

Zhang, X., Zhang, J., Wan, G., Luo, Z., 2019. Fintech, growth, and inequality: Evidence from China's Household Survey Data. The Singapore Economic Review.

Table 1: Definition of variables

Variable	Definition
HHID	Unique identifier of households which is used to link individuals to households.
YEAR	Year of the survey is either 2011 or 2013.
INCOME MEASURES	
INC	Individual income, including second jobs and bonuses.
FARM INC	Dummy for farming income.
IND INC	Dummy for business income which mostly refers to small businesses.
EDUCATION MEASURES	
NO EDU	Dummy takes value one if person has no formal education.
PRIM EDU	Dummy takes value one if person has primary education.
SEC EDU	Dummy takes value one if person has secondary education, which includes junior high, high school, secondary/vocational schools.
HIGH EDU	Dummy takes value one if person attended college or university (BSc, MSc, PhD).
FOREIGN EDU	Dummy takes value one if person studied abroad.
ASSETS	
FIX A	Fixed assets from business, property and cars.
BUS A	Dummy for business assets.
HOUSE PAY	Purchase price of house.
HOUSE VAL	Current value of house.
CAR VAL	Current value of car.
FIN A	Financial assets including checking accounts, deposits, stocks, bonds, futures, warrants and wealth management products.
CASH	Cash in household.
LEND	Lending from household.
FINANCIAL INCLUSION	
ACCOUNT	Dummy if individual has checking account and time deposits.
LOAN	Dummy for taking formal loan.
LOAN INF	Dummy for taking informal loan.
DEBT	Amount of formal debt from banks.
DEBT INF	Amount of informal debt.
INTER	Annual interest rates on formal loans.
INTER INF	Annual interest rates on informal loans.
REASONS FOR LOANS	
INVEST	Dummy if the loan is used for investing in financial assets.
CONSUME	Dummy if the loan is used for consumption and cars.
EDU	Dummy if the loan is used loan for education.
DEPOSIT EDU	Dummy if the main purpose of time deposits is to finance education.
INF EDU	Dummy if informal loans are used to finance education.
INFDEBT EDU	The amount of informal loans used to fund education.
INDIVIDUAL CONTROLS	
AGE	Age in years.
FEMALE	Dummy takes one for females.
LOCATION	
PROVINCE	Province.

Table 2: Descriptive statistics

	N	mean	sd	min	p25	p50	p75	max
NO EDU	127230	0.086	0.280	0.000	0.000	0.000	0.000	1.000
PRIM EDU	127230	0.173	0.379	0.000	0.000	0.000	0.000	1.000
SEC EDU	127230	0.437	0.496	0.000	0.000	0.000	1.000	1.000
HIGH EDU	127230	0.304	0.460	0.000	0.000	0.000	1.000	1.000
FOREIGN EDU	127230	0.008	0.091	0.000	0.000	0.000	0.000	1.000
INC	80778	24852.851	1.85e+05	0.000	1.000	200.000	21250.000	8.00e+06
FARM INC	127230	0.314	0.464	0.000	0.000	0.000	1.000	1.000
IND INC	127230	0.133	0.339	0.000	0.000	0.000	0.000	1.000
FIX A	102498	12103.717	2.10e+05	0.000	6.500	20.000	55.000	1.00e+07
BUS A	127230	0.143	0.350	0.000	0.000	0.000	0.000	1.000
HOUSE PAY	96166	125.985	6342.897	0.000	1.300	5.000	14.000	7.74e+05
HOUSE VAL	96233	42.720	72.595	0.001	6.250	20.000	45.000	800.000
CAR VAL	20374	8.420	10.540	-1.000	3.000	5.500	10.000	200.000
ACCOUNT	127230	0.612	0.487	0.000	0.000	1.000	1.000	1.000
FIN A	75205	81315.065	2.61e+05	0.000	2500.000	16000.000	60000.000	7.51e+06
CASH	127230	4800.921	37909.442	0.000	500.000	1000.000	3000.000	4.00e+06
LEND	127230	5702.248	69091.700	0.000	0.000	0.000	0.000	5.00e+06
LOAN	127230	0.041	0.199	0.000	0.000	0.000	0.000	1.000
LOAN INF	127230	0.107	0.309	0.000	0.000	0.000	0.000	1.000
DEBT	5440	40033.282	3.05e+05	0.000	2.000	8.000	120.000	5.00e+06
DEBT INF	9657	37023.208	8.94e+05	0.000	2.000	15.000	5000.000	5.00e+07
INTER	4291	0.068	0.050	0.000	0.023	0.073	0.096	0.750
INTER INF	9634	0.106	2.518	0.000	0.000	0.000	0.025	117.647
INVEST	127230	0.021	0.142	0.000	0.000	0.000	0.000	1.000
CONSUME	127230	0.016	0.124	0.000	0.000	0.000	0.000	1.000
EDU	127230	0.020	0.140	0.000	0.000	0.000	0.000	1.000
DEPOSIT EDU	127230	0.027	0.161	0.000	0.000	0.000	0.000	1.000
INF EDU	127230	0.070	0.255	0.000	0.000	0.000	0.000	1.000
INFDEBT EDU	127230	854.835	12329.933	0.000	0.000	0.000	0.000	1.20e+06
AGE	127215	39.008	21.019	0.000	23.000	39.000	55.000	113.000
FEMALE	127230	0.491	0.500	0.000	0.000	0.000	1.000	1.000
RURAL	127230	0.381	0.486	0.000	0.000	0.000	1.000	1.000

Table 3: The impact of education on income

	[*A*1]	[*A*2]	[*A*3]	[*A*4]	[*A*5]
PRIM_EDU	1.296***	0.116	0.109	0.091	0.079
SEC_EDU	3.311***	1.556***	1.565***	1.484***	0.960***
HIGH_EDU	3.282***	3.022***	3.128***	2.890***	1.466***
FOREIGN_EDU	2.641***	0.704***	0.994***	0.881***	0.711***
AGE		0.288***	0.294***	0.280***	0.159***
AGE_2		-0.003***	-0.003***	-0.003***	-0.002***
FEMALE		-0.600***	-0.630***	-0.627***	-0.465***
RURAL		-1.739***	-2.166***	-2.013***	-0.210***
FARM_INC					-5.026***
IND_INC					-3.062***
ll	-2.19e+05	-2.13e+05	-2.01e+05	-1.99e+05	-1.89e+05
aic	4.37e+05	4.26e+05	4.02e+05	3.99e+05	3.77e+05
bic	4.37e+05	4.26e+05	4.02e+05	3.99e+05	3.77e+05
r2_a	0.054	0.188	0.415	0.440	0.581
N	73500	73494	73494	73494	73494

Note: All models refer to OLS regressions with robust standard errors.
Model A3 account for the year of the survey, and specifications A4
and A5 control for provinces and years.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: The impact of education on income

	[B1]	[B2]	[B3]	[B4]	B5
INVEST_EDU	0.407***	0.311***	0.220***	0.167***	0.313***
FIX_A		0.346***	0.334***	0.260***	0.258***
FIN_A			0.266***	0.206***	0.201***
AGE				0.213***	0.214***
AGE_2				-0.003***	-0.003***
FEMALE				-0.447***	-0.815
FEMxEDU					0.037
RURAL				-1.249***	2.714***
RURxEDU					-0.411***
ll	-1.54e+04	-1.24e+04	-6947.445	-6770.697	-6752.324
aic	30739.835	24833.219	13902.889	13557.395	13524.649
bic	30753.099	24852.572	13926.379	13604.375	13583.373
r2_a	0.361	0.432	0.402	0.476	0.483
N	5610	4680	2624	2624	2624

Note: All models refer to OLS regressions with robust standard errors.

All specifications control for provinces and years.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: Returns and interest rates in income or asset groups

	p25	p50	p75	p25	p50	p75
	Returns			Interest rates		
1 (INC)	0.025	0.053	0.133	0.015	0.070	0.096
2 (INC)	0.026	0.054	0.144	0.015	0.076	0.096
3 (INC)	0.032	0.078	0.188	0.030	0.080	0.099
4 (INC)	0.045	0.086	0.200	0.036	0.072	0.095
1 (ASS)	0.021	0.043	0.091	0.015	0.076	0.098
2 (ASS)	0.034	0.067	0.145	0.030	0.080	0.100
3 (ASS)	0.052	0.097	0.198	0.015	0.070	0.090
4 (ASS)	0.057	0.106	0.312	0.030	0.070	0.090

The table shows the bottom 25%, the median, and the top 25% of returns from investing in education and interest rates. The first four rows refer to quantiles defined by income (INC) starting from the lowest 1 to the highest 4. The last four rows refer to quantiles defined by fixed assets (ASS) starting from the lowest 1 to the highest 4.

Table 6: Determinants of underinvestment in education

	[C1]	[C2]	[C3]	[C4]	[C5]
INC	0.000**	0.000***	0.000***	0.000	0.000***
FEMALE	-0.487***	-0.456***	-0.456***	-0.622***	-0.437***
RURAL	-0.619***	-0.928***	-0.915***	-0.606**	-0.928***
NO_EDU	1.348***	0.849***	0.881***	1.981**	0.635***
PRIM_EDU	0.986***	0.701***	0.723***	1.590***	0.537***
SEC_EDU	0.649***	0.551***	0.563***	1.356***	0.466***
FOREIGN_EDU	-0.169	-0.144	-0.137	-0.207	-0.148
LOAN		0.097	0.181		
ACCOUNT		-5.940***	-5.964***		
LOAN_INF			-0.318***		
CASH				-0.000	
HOUSE_VAL				0.001	
CAR_VAL				0.004	
FIN_A					-0.136***
FARM_INC					-0.157
IND_INC					0.218**
ll	-6317.404	-5606.295	-5598.203	-872.457	-5344.259
aic	12772.807	11354.589	11340.406	1854.914	10832.518
bic	13413.999	12014.366	12009.475	2249.042	11453.678
r2_p	0.061	0.167	0.168	0.136	0.085
N	80228	80228	80228	9566	41242

Note: All models refer to logistic regressions with robust standard errors.

All specifications control for provinces, years and birth cohorts.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: Determinants of income inequality

	[D1]	[D2]	[D3]	[D4]	[D5]
L_INC_DEC	0.134***	0.131***	0.143***	0.134***	0.143***
L_HH_EDU	0.004*	0.004*	0.005**	-0.004	-0.002
L_HH_AGE	-0.037***	-0.037***	-0.037***	-0.037***	-0.037***
RURAL	-0.390***	-0.353***	-0.313***	-0.377***	-0.307***
L_HH_ACCOUNT		0.158***	0.129**		0.122**
L_HH_LOAN			-0.334**		-0.317**
L_HH_LOAN_INF			-0.432***		-0.428***
L_HH_DEPOSIT_EDU				0.530**	0.390
L_HH_INF_EDU				0.088	0.111
L_HH_INFDEBT_EDU				0.000**	0.000**
ll	-1.38e+04	-1.38e+04	-1.38e+04	-1.38e+04	-1.38e+04
aic	27752.872	27743.060	27689.163	27744.303	27684.898
bic	27998.803	27995.823	27955.588	28010.728	27971.817
r2_p	0.056	0.057	0.059	0.057	0.059
N	6846	6846	6846	6846	6846

Note: All models refer to ordered logistic regressions with robust standard errors.

All specifications control for provinces and years.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 8: Determinants of income inequality

	[E1]	[E2]	[E3]	[E4]	[E5]
L_INC_DEC	0.276***	0.277***	0.269***	0.278***	0.270***
L_HH_EDU	0.000	0.000	-0.001	-0.008	-0.009
L_HH_AGE	-0.019***	-0.019***	-0.018***	-0.019***	-0.018***
RURAL	0.977***	0.971***	0.948***	0.979***	0.952***
L_HH_ACCOUNT		-0.023	-0.006		-0.005
L_HH_LOAN			0.218		0.233
L_HH_LOAN_INF			0.361**		0.364**
L_HH_DEPOSIT_EDU				0.327	0.423
L_HH_INF_EDU				0.245	0.189
L_HH_INFDEBT_EDU				0.000	0.000
ll	-2141.830	-2141.797	-2136.118	-2141.313	-2135.521
aic	4343.660	4345.593	4338.236	4348.626	4343.043
bic	4521.241	4529.094	4533.575	4543.965	4556.140
r2_p	0.211	0.211	0.213	0.211	0.213
N	2750	2750	2750	2750	2750

Note: All models refer to ordered logistic regressions with robust standard errors.

All specifications control for provinces and years.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure 1: Optimal investment, debt and financial assets

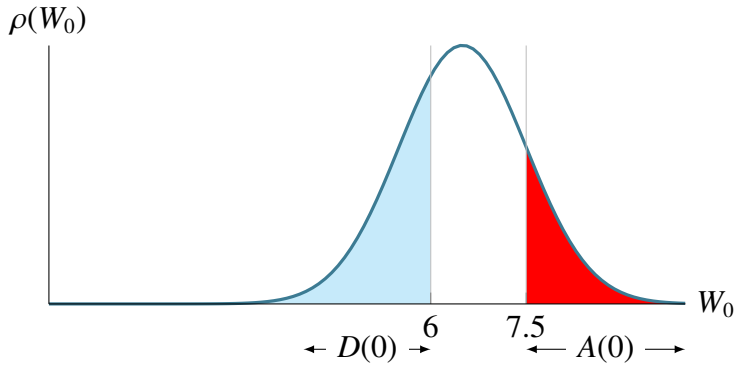


Figure 1 plots the initial distribution of endowment, where $\rho(W_0)$ refers to the probability density. The poor with initial endowment below optimal levels of human capital $K^* = 7.5$ can take loans $D(0)$ to reach the transient steady-state $K_{poor} = 6$. The rich with $W_0 > K^* = 7.5$ can invest in financial assets $A(0)$. Individuals with initial endowment from $K_{poor} = 6$ to $K^* = 7.5$ cannot benefit from financial inclusion.

Figure 2: Optimal paths of investment and debt

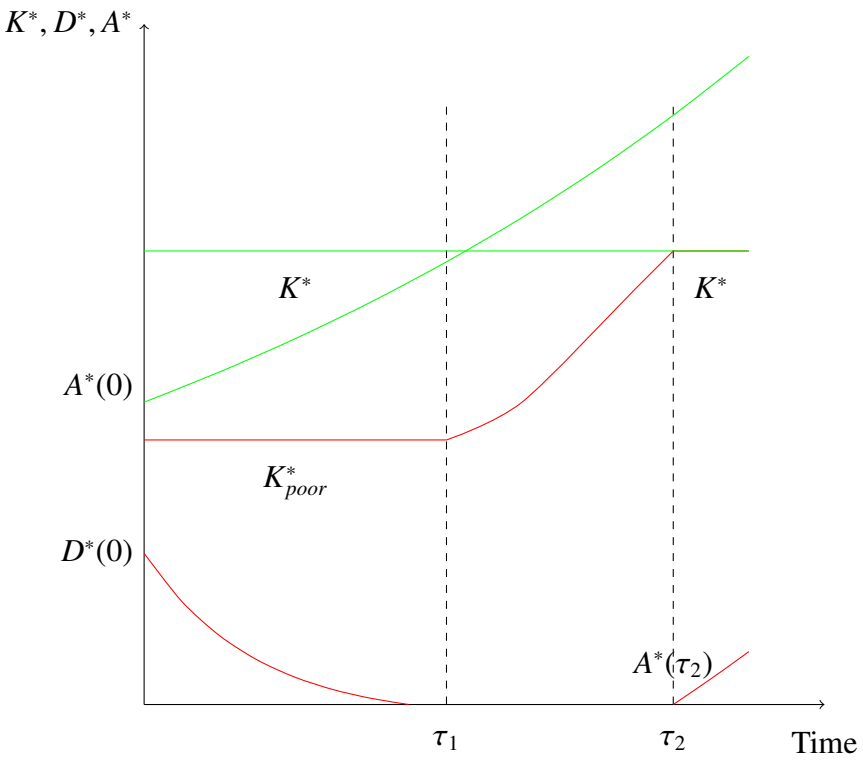
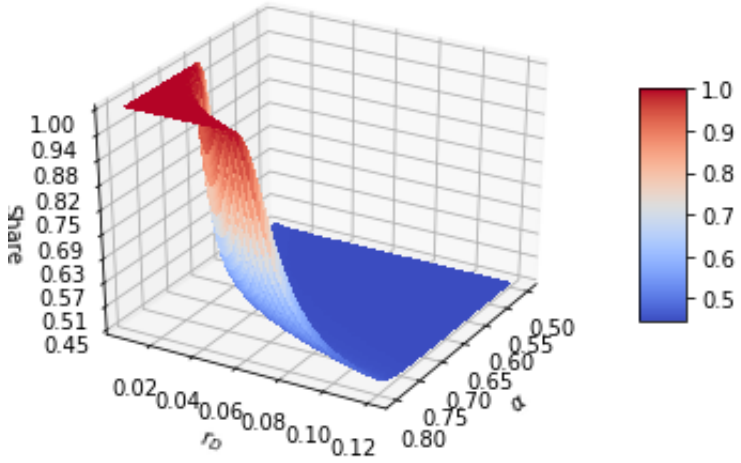


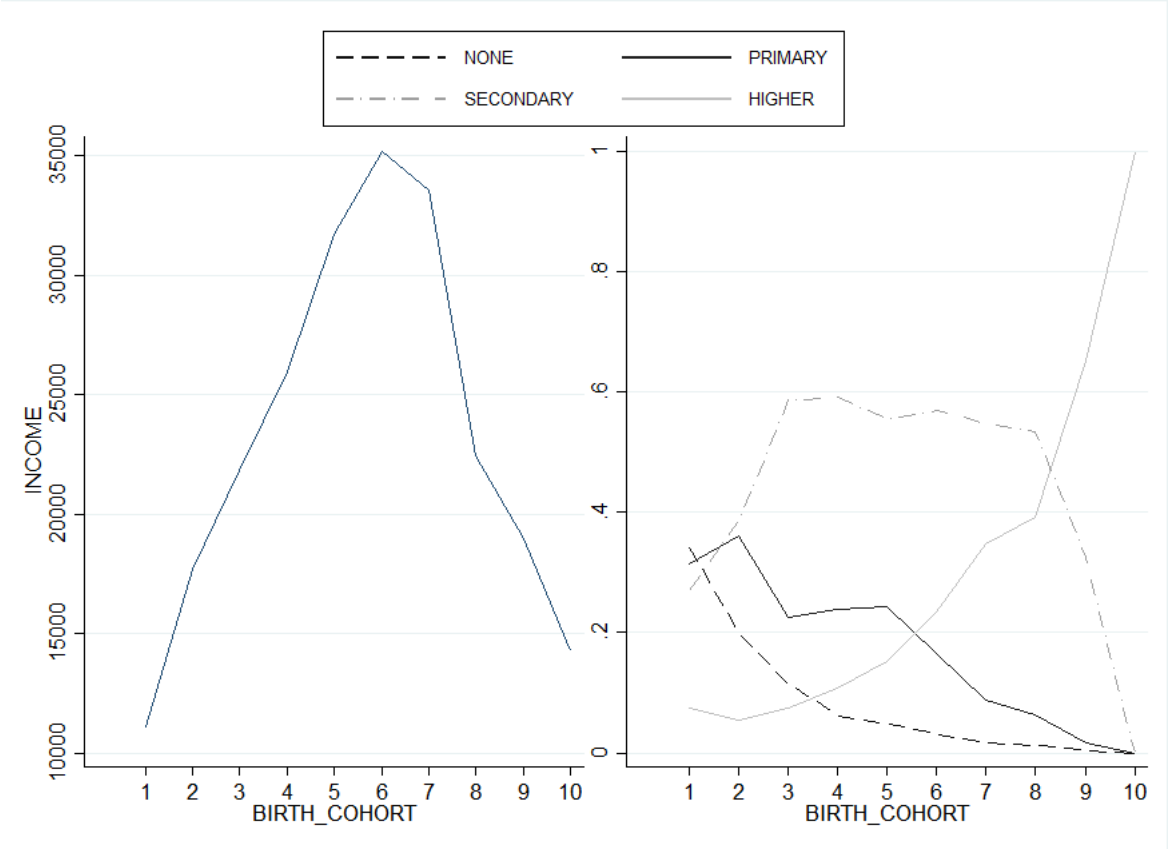
Figure 2 plots the optimal paths for the poor (in red) who use debt $D^*(0)$ to reach the transient steady-state of human capital K^*_{poor}. After repaying debt at time τ_1, they use their incomes to invest in human capital to reach the optimal capital stock K^* at time τ_2. Any additional income will be invested in the financial asset $A^*(\tau_2)$. The rich (in green) 'jump' into the optimal capital stock K^*. Any additional endowment at $t = 0$ is invested in the financial asset $A^*(0)$, which cumulates over time with income.

Figure 3: Proportion of the poor who can benefit from financial inclusion (Theorem 2)



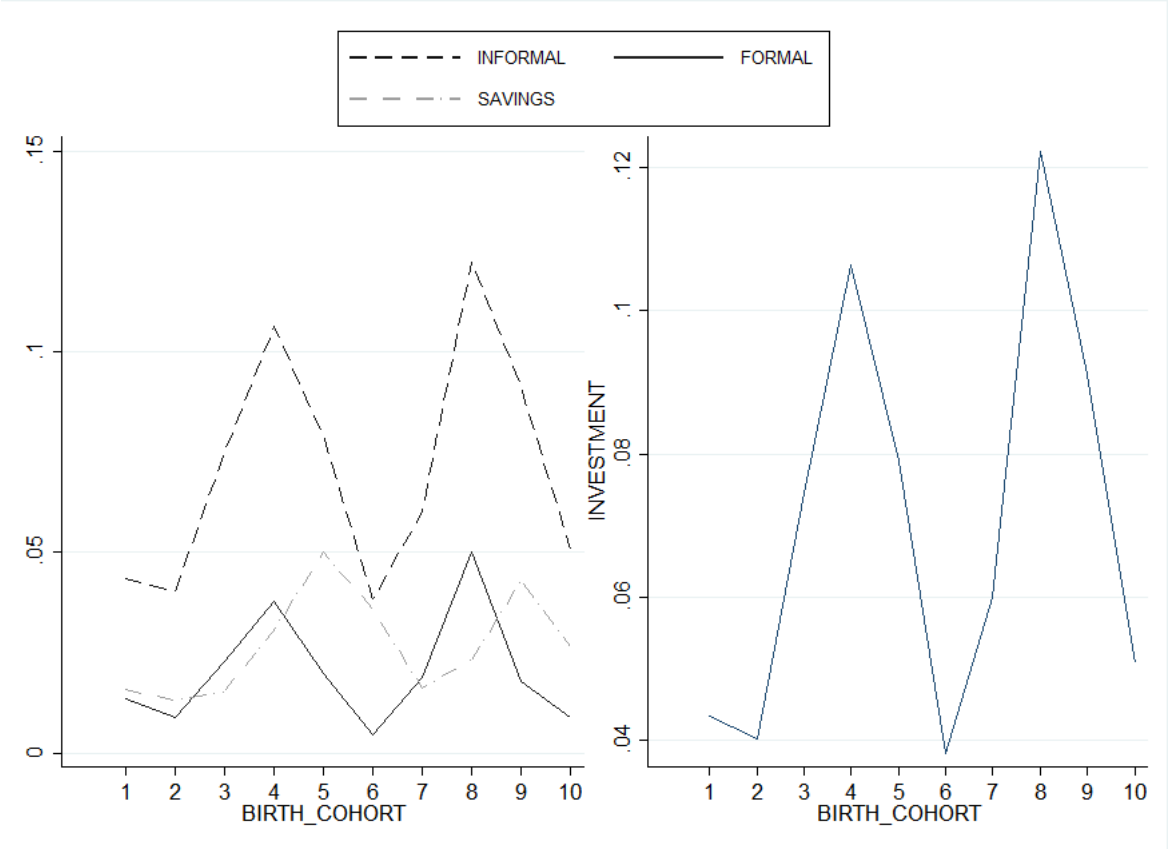
Theorem 2 states that the poor who benefit from financial inclusion have initial wealth $W_0 < (\alpha/r_D)^{1/(1-\alpha)}$, where r_D is cost of capital and α is the production parameter. Figure 3 plots the proportion of the poor that fulfills this condition. Cost of debt, r_D, is in the range 0.73% to 12% in line with descriptive statistics for China.

Figure 4: Cohort effect in income and education



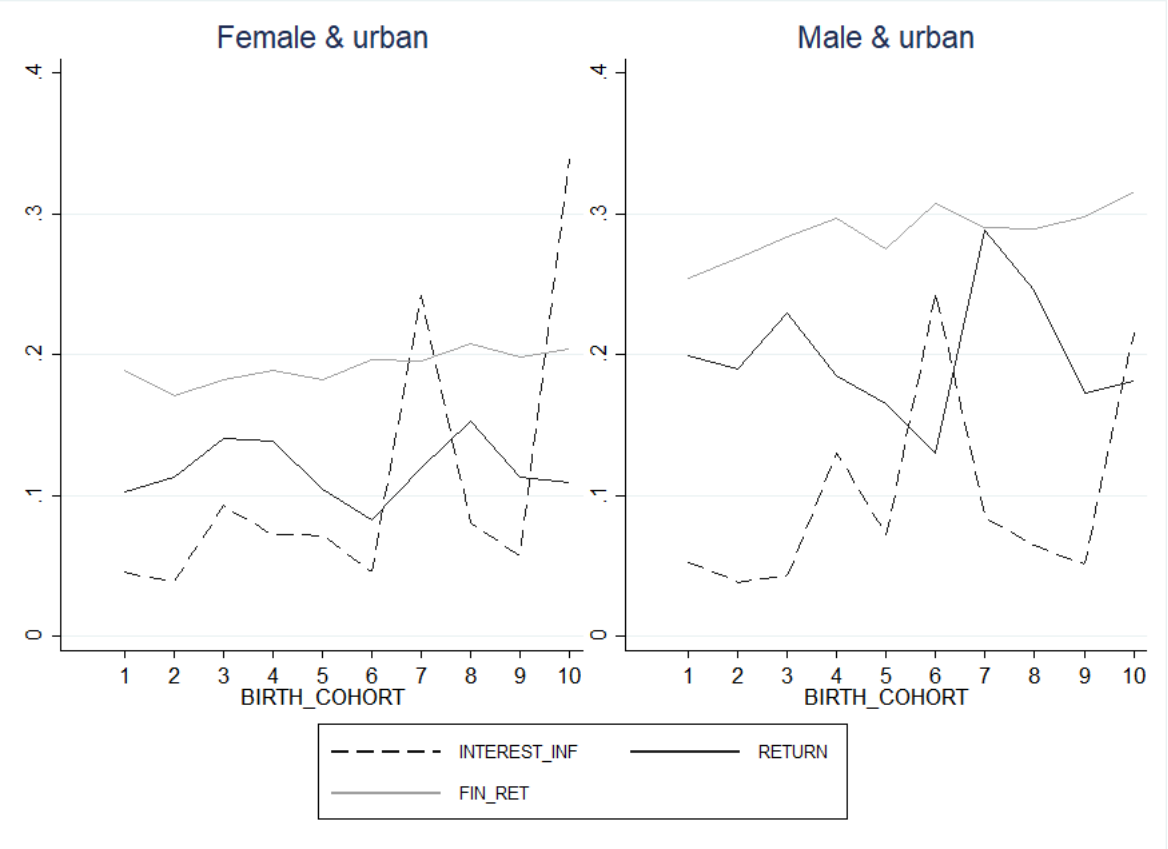
The left panel shows income levels for different birth cohorts, whereas the right panel depicts the proportion of individuals in each birth cohort that exhibit no eduction (NONE), primary, secondary or higher education.

Figure 5: Investment into education by cohort



The left panel shows the proportion of individuals in each birth cohort that use savings, formal or informal loans to invest in education. The right panel depicts the proportion of individuals in each birth cohort that invests in education.

Figure 6: Returns from investing in education and financial assets



The figure shows the returns from investing in education (RETURN) compared to returns from financial assets (FIN_RET) and interest rates on informal loans (INTEREST_INF). The left panel focuses on females in urban areas, whereas the right panel depicts results for males in urban settings.