

NBER WORKING PAPER SERIES

SUBJECTIVE INCOME EXPECTATIONS AND HOUSEHOLD DEBT CYCLES

Francesco D'Acunto
Michael Weber
Xiao Yin

Working Paper 32715
<http://www.nber.org/papers/w32715>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2024

For very helpful comments and suggestions, we thank Elena Afanasayeva, Deniz Aydin, Michael Gelman, Yuriy Gorodnichenko, Sasha Indarte, Emi Nakamura, Peter Maxted, Alessandro Previtero, Sean Myers, Søren Leth-Petersen, Stijn Van Nieuwerburgh, Mirko Wiederholt as well as seminar and conference participants at UC Berkeley, UCL, 2022 SED Meeting, 2023 AEA Annual Meeting, 2024 Duke-UNC Asset Pricing Conference, 2024 SFS Finance Cavalcade, FRB Conference on the Macroeconomic Seeds of Financial Imbalances, LSE-IFS-UCL-CEPR-Imperial Business School workshop on Household Finance, and 2024 CES-ifo Venice Summer Institute on Expectations Formation. All errors are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2024 by Francesco D'Acunto, Michael Weber, and Xiao Yin. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Subjective Income Expectations and Household Debt Cycles
Francesco D'Acunto, Michael Weber, and Xiao Yin
NBER Working Paper No. 32715
July 2024
JEL No. D14,D84,E21,E71,G51

ABSTRACT

Matched transaction-level, credit-registry, and survey-based data reveal that consumers on average form excessively high (low) income expectations relative to ex-post realizations after unexpected positive (negative) income shocks. These extrapolative income expectations lead consumers to increase current spending, accumulate more debt, and face more defaults when lower-than-expected incomes are realized. We assess the aggregate implications by estimating a consumption model with defaultable unsecured debt and diagnostic Kalman filtering whereby consumers over-extrapolate income shocks when forming expectations. Extrapolative income expectations help explain state-dependent household debt cycles qualitatively and quantitatively.

Francesco D'Acunto
McDonough School of Business
Georgetown University
Washington, DC 20057
francesco.dacunto@georgetown.edu

Xiao Yin
30 Gordon St, London
Drayton House
London WC1H 0AX
United Kingdom
xiao.yin@ucl.ac.uk

Michael Weber
Booth School of Business
University of Chicago
5807 South Woodlawn Avenue
Chicago, IL 60637
and NBER
michael.weber@chicagobooth.edu

I Introduction

Understanding the causes and consequences of household debt cycles is important because spikes in household debt predict economic crises across space and over time (Reuven and Lansing, 2010; Mian and Sufi, 2014; Mian, Sufi, and Verner, 2017). Supply- and demand-side channels might drive household debt cycles (e.g., Agarwal et al. (2017); Mian et al. (2020); Bordalo et al. (2018); Bianchi et al. (2021); Chodorow-Reich et al. (2023)). Due to their intrinsic co-determination, isolating supply- and demand-side drivers using observational data is empirically challenging because unobserved shocks might affect both channels and both channels could influence each other.

In this paper, we provide an empirical test for a demand-side driver of household debt cycles—consumers’ inaccurate income expectations. Our individual-level panel data combine spending decisions, income inflows, and consumer-credit limits from bank accounts, debt choices from credit-registry data, and repeated income expectations elicited via surveys.¹ This setting enables us to study the dynamics of subjective income beliefs as well as their correlation with consumption and debt choices. In each period, we can compute unexpected subjective income shocks as the difference between realized income and numerical income expectations for that period based on prior survey waves. We can also compare expectations about future income with subsequent income realizations—a direct measure of subjective income expectations’ ex-post accuracy.

We start by documenting that the average consumer’s income expectations overreact to unexpected income shocks after both positive and negative shocks: after an unexpected shock in either direction today, expectations about future income are systematically inaccurate in the direction of the shock relative to subsequent realizations. The size of this overreaction increases proportionally with the size of the unexpected income shock. These results hold after keeping constant a rich set of individual-level demographics. At the same time, the size of the belief mistake is larger for lower-income consumers, consumers who face more volatile income flows, and younger consumers. These patterns hold when we exploit variation in unexpected income shocks and expectations errors within individuals

¹For recent reviews of survey-based research in economics, see Haaland et al. (2021) and D’Acunto and Weber (2024).

over time with individual fixed effects.

We then assess if these inaccurate subjective income expectations relate to consumption and debt choices. We find that they do: the larger the mistake in income expectations, the larger the present-day increase in spending and the additional debt consumers take. Furthermore, consumers' likelihood of default increases with the income-expectations mistake induced by unexpected income shocks.

Existing research finds evidence of asymmetric household debt cycles, whereby the buildup of aggregate spending and borrowing in good times is slower than the sudden drop in spending and deleveraging in bad times (Mian, Sufi, and Verner, 2017). Consistently, at the micro level, we find that the consumption response to unexpected individual negative income shocks is stronger relative to same-size positive shocks. Moreover, we find that facing larger unexpected income shocks increases consumers' uncertainty about future income—the second moment of income expectations rises. This effect is also larger for negative shocks than for same-size positive shocks.

Given the panel structure of the data, we can exploit within-individual variation in income shocks of different sizes and directions across consumers who make decisions at the same point in time. Moreover, we can add time-fixed effects to absorb aggregate shocks that affect all consumers similarly, such as the aggregate effects of the COVID-19 pandemic, which falls within our sample period. Our results arise in times of aggregate expansions and contractions alike.

Overall, the first part of the paper provides direct micro-level evidence of an expectations-based channel of debt outcomes whereby the average consumer forms systematically excessive beliefs about future income when facing unexpected income shocks. When facing positive unexpected income shocks, she consumes more and raises more debt to finance higher current spending. Once subsequent income does not reach the expected level, she is more likely to default. When facing negative unexpected income shocks, she cuts her current spending swiftly and reduces her debt.

Our micro-level evidence is based on different shocks that hit consumers at the same point in time but, in the aggregate, income shocks are likely to be correlated across consumers: in times of positive GDP growth more and more consumers will face positive

income shocks, part of which might be unexpected. Due to biased income expectations, a larger fraction of consumers will increase consumption and accumulate more debt. By contrast, when GDP growth turns negative, more and more consumers will face unexpectedly negative income shocks and swiftly cut their consumption and borrowing.

Based on this intuition, in the second part of the paper we ask if the expectations-based mechanism we document empirically could be qualitatively and quantitatively relevant to understanding the aggregate dynamics of household debt cycles. For this purpose, we augment a standard consumption model with incomplete markets and heterogeneous agents with defaultable debt. Consumers learn about the permanent component of their income process on the basis of a noisy signal (their income realization) with the Kalman filter as the rational benchmark. We model extrapolative subjective income beliefs with diagnostic expectations, which are consistent with our micro-level data-driven facts and are portable to various theoretical and empirical applications in economics.²

The model matches several empirical moments of the liquidity distribution, such as the fraction of debtors as well as the average and median of wealth-income ratios. The model also matches successfully the relationship between errors in expectations and consumption, debt, and default outcomes we observe in the microdata.

We use the model to simulate a heterogeneous-agent economy to isolate the effects of inaccurate income expectations on the dynamics of aggregate consumption, debt, and defaults. Extrapolative expectations amplify the effect of positive income shocks on income expectations and consumption. For this reason, agents start accumulating more debt after positive income shocks relative to the rational-expectations benchmark. Once we remove positive shocks, default risk increases substantially. By contrast, it does not increase under rational expectations. In a last exercise, we use the model to show that extrapolative expectations can produce aggregate patterns of consumption and debt like those that were observed in the US around the 2008-2009 Financial Crisis.

²Other microfoundations of extrapolative expectations are also consistent with our empirical results. For instance, see Barberis et al. (2018), Barberis (2018), and Li et al. (2023). Other forms of learning—such as learning about the long-run mean of the income process—could be consistent with our results too.

Our structural model’s simulation implies that extrapolative subjective income expectations like the ones we document can generate aggregate dynamics consistent with facts about aggregate household debt cycles documented in earlier research. For instance, the facts that elevated consumer sentiment implies higher subsequent debt growth (López-Salido et al., 2017) and that the end of a boom usually coincides with times of high financial fragility (Greenwood et al., 2019; Maxted, 2023).

Our evidence of a demand-side channel does not imply that supply-side channels are irrelevant to explain consumption and household debt cycles. For instance, Aydin (2022) and Yin (2022) document the impact of banks’ credit choices on consumers’ debt and spending using randomized increases in credit card limits. Bornstein and Indarte (2023) find that expanding social insurance via Medicaid eligibility increases households’ financial resilience and hence credit supply, which leads to higher household debt accumulation. And, Kluender et al. (2024) show that both beliefs (demand) and constraints (supply) drive low-income workers’ consumption decisions and assess the relative importance of these channels.

Our paper contributes to at least three strands of literature. First, the literature on the drivers of credit cycles. Theoretically, two categories of explanations emerge. On the one hand, financial frictions in the corporate and household sectors can be an amplification mechanism that induces cycles in credit supply (for instance, see Kiyotaki and Moore (1997), Gertler and Kiyotaki (2010), Brunnermeier and Sannikov (2014), He and Krishnamurthy (2019), Li (2019), and Mian et al. (2020)). On the other hand, subjective beliefs can also be relevant. For instance, beliefs might change when the incentives of producing information change over time, thereby inducing swings in asset prices and macroeconomic fluctuations (Gorton and Ordoñez (2014), Dang et al. (2020)). Allowing expectations to be extrapolative also generates credit cycles (Bordalo et al. (2018), Bianchi et al. (2021), Bordalo et al. (2021), and L’Huillier et al. (2023)). Our paper provides empirical evidence at the micro level consistent with a demand channel based on consumers’ beliefs. Our findings are also in line with Kaplan et al. (2020) who find that changes in beliefs of future housing demand were the most important driver of the increase in house prices before the 2008-2009 Global Financial Crisis and the subsequent

slump.

On the empirical side, the analysis of credit-cycle fluctuations so far has mostly focused on aggregate economy-wide or regional-level data.³ Our paper contributes to this literature by providing micro-level evidence from the household sector in a panel dataset that allows us to uncover how unexpected income shocks can induce over-reaction of expectations of future income, which leads to an excessive accumulation of debt and higher subsequent default rates.

We also contribute to the rich literature on the marginal propensity to consume (MPC) out of income shocks (Parker et al., 2013; Fuster et al., 2020; Kueng, 2018; Olafsson and Pagel, 2018; Baugh et al., 2021; Fagereng et al., 2021; Colarieti et al., 2024)⁴. Our study is closely related to work on the consumption response to shocks to total income rather than one-time wealth transfers (Ganong et al., 2020; Krueger et al., 2023). Our estimates of the consumption elasticity to income shocks are close to the estimates in Ganong et al. (2020). We contribute to this line of work by providing the first evidence on income shocks affecting consumption through a biased-beliefs channel as well as the implications for consumers' borrowing and default. Our empirical findings on how consumption adjusts to income shocks are therefore informative for models describing households' consumption-saving motives. Distinguishing between these models is important to understand both the joint dynamics of income and consumption and the response of the macroeconomy to various types of shocks. In addition, it provides useful insights for policy analysis, including the optimal design of social insurance and redistribution programs, which depend on how households insure against idiosyncratic income shocks.

Lastly, this paper contributes to the literature on the role of beliefs in explaining consumers' consumption-saving choice (see DellaVigna (2009) and Benjamin (2019) for reviews). Ameriks et al. (2020) and Ameriks et al. (2020) document the role of

³See, for instance, Bordo et al. (2001), Borio and Lowe (2002), Claessens et al. (2010), Reinhart and Rogoff (2009), Borio and Lowe (2013), Jordà et al. (2013), Baron and Xiong (2017), Greenwood et al. (2020), Krishnamurthy and Muir (2017), Mian and Sufi (2018), Mian et al. (2020), and Baron et al. (2020), among others.

⁴See Attanasio and Weber (2010) and Jappelli and Pistaferri (2010) for earlier reviews and Kaplan and Violante (2022) for a recent review of MPCs in heterogeneous agent models.)

survey-elicited beliefs on retirement choices. Manski (2004), Ameriks et al. (2020), Giglio et al. (2021), and Beutel and Weber (2022) study the relationship between beliefs and stock-market investments. Bucks and Pence (2008), Bailey et al. (2019), and Kuchler et al. (2022) analyze how beliefs affect mortgage leverage choices. D’Acunto et al. (2023, 2022), Roth and Wohlfart (2020), Chopra et al. (2023) and Coibion et al. (2022) assess the effect of macroeconomic expectations on households’ consumption, saving, and borrowing choices. Kluender et al. (2024) study the role of beliefs in explaining low-income workers’ consumption decisions and their relevance relative to financial constraints. Rozsypal and Schlafmann (2023) study the consumption response to income shocks when consumers over-estimate the persistence of their income process. In contemporaneous work, Bellifemine et al. (2024) use survey data from Italy to study the heterogeneous effects of income extrapolation on saving rates. Our work studies the drivers and effects of biased income expectations on consumption, debt, and default outcomes.

II Institutional Setting and Data

We collaborate with a large Chinese commercial bank. The bank operates nationally and is among the top 10 commercial banks in the People’s Republic of China by total assets. In 2023, the bank’s total assets amounted to more than one trillion U.S. dollars with more than 70 million active account holders.

We obtained transaction-level information on a consumer population that is representative of the Chinese banked population and fielded multiple waves of a customized survey on these consumers to elicit their subjective income expectations. Based on the fielding of our surveys, the sample period for our main analysis is from 2020 to 2023. We extend it for robustness tests to confirm the relationship between income shocks and choices within individuals over longer time periods for which subjective expectations data are not available.

For each account holder, we have also obtained data from the Credit Reference Center of the People’s Bank of China (China’s official credit registry) on total outstanding debt and its composition. The Credit Reference Center aggregates information from *all*

financial institutions from which borrowers receive credit and not only the bank with which we cooperate, which allows us to observe the full size and structure of borrowers' liabilities.

A. Identifying Primary Bank Users

The credit registry does not collect spending information. To study consumption choices, we thus need to rely on the transaction-level data from the bank. Consumers might have multiple banking relationships and multiple spending accounts, though. This would be a concern for our analysis if consumers had systematically different spending habits across accounts and these habits were a function of income expectations and past income shocks.

To reduce this concern, we follow recent research using single-provider transaction-level data (e.g., see Ganong and Noel (2019)) and impose two restrictions on the accounts that enter our empirical analysis to capture consumers who are likely to use the bank as their primary banking institution. First, we only consider consumers whose bank accounts include at least 15 outflow transactions during the sample period. An outflow is any debit from a checking, saving, or credit card account, including a cash withdrawal, an electronic payment, or a card transaction. Imposing this criterion reduces the original sample by approximately 35%. Second, we only consider consumers who have their regular income deposited at our partner bank. This restriction drops about 15% more observations. Our results are not sensitive to the levels of these cutoffs and are similar if no restrictions are imposed.

B. Measuring Income, Spending, and Debt

We compute income inflows and spending outflows from transaction-level data.

For income, we follow the steps the bank uses, which identify individual income following a classification rule of regular inflows. The bank classifies income into salary or business cash flows. Salary is defined as the regular periodic inflows (income and bonuses) for consumers who declare working as employees. The bank calculates this number in one of two alternative ways. First, if income is paid as a direct deposit from

the consumers' employers into the bank account, the number is directly labeled as salary. Otherwise, the bank can identify income if the consumer's social security insurance, which is a fixed portion of the consumer's income, is paid through the bank.⁵ We calculate income from business operations as the difference between total inflows and total outflows of transactions that the bank categorizes as business operations.

Overall, 70.5% of all income in our sample is from salary and 29.5% from business cash flows. These figures are not only representative of the Chinese banked population but also accurately computed at the individual level. We can claim this because, for each account holder, we can retrieve the income they report to the Chinese tax agency. The results of this comparison are in Panel A of Figure A.1, which is a binned scatter plot comparing the income computed by the bank and the income account holders report to the Chinese tax authority. We see a strong linear relationship between these two values, with a R^2 of 84%, which corroborates the quality of our income data. For our sample of primary bank users, the bank does not seem to miss income sources systematically.

We compute monthly total spending as the sum of all purchase transactions excluding installment payments on mortgages and vehicles plus the total amount of repayment of linked credit cards' end-of-month balances between the end of the last billing cycle and the current billing cycle.⁶

For debt, we rely on data from the Chinese credit registry, as discussed above, which includes the debt issued by individuals across all banking relationships they have. We compute debt at the individual level as the sum of outstanding interest-incurring balances on all credit cards and other unsecured personal loans in the credit registry.

⁵In China, social security payments have six components: five types of insurance and a housing provident fund. The types of insurance are paid as a fixed proportion of the worker's monthly income. One such insurance is the retirement saving insurance, which is similar to retirement savings plans in the US. The monthly contribution is 8%. However, the income base for social security is usually capped at the two tails of the income distribution. The caps vary across space and are usually set between 30%–300% or 40%–400% of the previous year's average income in each location. The uncapped distribution covers most Chinese workers ($\sim 90\%$). We remove consumers in the capped regions from the final sample.

⁶During our sample, Chinese residents mostly use Alipay or Wechat pay as a payment method for spending. There are two ways to use them: 1) linking the apps with debit or credit cards; 2) storing a small amount of money on the platforms' change wallet. With transactional history, we can observe the former but not the latter. However, the former method is the most common. A supplementary survey by the bank shows that only around 7% of all transactions are based on payments via the change wallet.

C. Survey Design and Reliability of Answers

To elicit consumers' expectations, we designed a customized survey that the bank administered on our behalf. Figure 1 plots the timeline of the survey. The bank ran the survey in 2020, 2021, and 2023. Each year, bank customers received two waves of surveys, one at the beginning of January and one at the beginning of July. Each wave elicited expectations over the subsequent six-month period.⁷

To limit survey fatigue, the bank asked that the survey include at most 12 questions.⁸ We report the English version of the survey (which was administered in Mandarin) in Appendix VI. On average, survey participants spent about 4 minutes answering the survey. The average reward was 10 CNY, which is above the 95th percentile of the hourly wage rate in China and led to a high response rate (68%).

The survey starts by describing its purpose. Besides ensuring informed participation, this step aims to eliminate any potential strategic motives in answers: without explanations, respondents might incorrectly infer that their answers are used to shape the types and quantities of financial services the bank would offer them going forward. To avoid this, respondents read the following paragraph:

The data will be analyzed by third-party research scholars for scientific research purposes and will not be evaluated by this bank. We will not disclose participants' identifiable information in any respect. We will not, to any extent, change the types of financial products we provide, including credit scores, credit limits, deposit rates, etc., based on participants' answers. Therefore, please answer based on your actual opinion.

Then, respondents report their total income over the previous 12 months. Because we observe income inflows in the data and we restrict the sample to primary bank users, we can use this question to assess if respondents answer truthfully. If we were concerned that respondents answered the survey questions randomly to finish quickly and/or to provide false information on purpose, the answer about income would capture this behavior. Panel B of Figure A.1 compares reported income values in the survey with the same respondents'

⁷The bank requested six-months lags, which align with their business auditing frequency.

⁸The overall sample receives nine questions. 29% of participants with transactions in financial accounts over the six months before fielding the surveys receive three additional questions about financial market returns.

income inflows computed from the bank’s account-level administrative data. The plot documents a strong linear relationship and a regression between the two variables yields an R^2 of around 71%, which we interpret as evidence in support of the reliability of our survey-based data.

As far as expectations about individual outcomes are concerned, we follow state-of-the-art methods to elicit not only the first moment of income expectations but also measures of uncertainty (Manski, 2004). We measure the first moment by asking:

What will your total income be most likely over the next 6 months?

To elicit higher moments, asking consumers to report a full probability distribution is highly cognitively demanding and faces the concern that most consumers ignore the concept of a probability distribution. Asking about probability distributions might thus confound actual beliefs with a measure of respondents’ cognitive abilities, which in turn shape beliefs (D’Acunto et al., 2019, 2023). We therefore rely on the triangular-distribution question design, which has become increasingly popular in economics research (for instance, see Guiso et al. (2002); Christelis et al. (2020)). This two-step question asks respondents to report a point estimate of the minimum and maximum possible future realizations of a variable—in our case, own income—which allows us to compute the second moment of subjective expectations.⁹

Specifically, we ask the following two questions:

What would be the lowest possible level of total income you believe you could make over the next 6 months?

What would be the highest possible level of total income you believe you could make over the next 6 months?

In our analysis, we ensure that the same individuals have completed at least three waves of surveys. Consequently, we can exploit variation in income expectation errors, income changes, and changes in economic choices within individuals over multiple periods.

⁹Eliciting also a subjective probability that the outcome falls above the mode or midpoint typically results in identical first and second moments, see, Weber et al. (2022); Coibion et al. (2024).

This design allows us to absorb systematic time-invariant unobserved characteristics across individuals that might confound the relationship between income expectations and consumption/debt choices, such as cognitive abilities and financial literacy.

Moreover, by observing cross-sections of respondents across multiple time periods, we can assess our baseline results within time periods, which absorbs the common effects of aggregate economic shocks. This feature is important in our setting because the early sample spans the times before and after the start of the COVID-19 pandemic.

D. Summary Statistics

Table I provides summary statistics for our working sample. All local currency values are converted to U.S. dollars¹⁰ for ease of interpretation and all continuous variables are winsorized at 1-99% levels to reduce the influence of outliers.

In terms of demographics, the age distribution is symmetric around its mean (about 39 years old) and most consumers are in their active working age—the interquantile range is between 29 and 48 years. Moreover, the sample includes 52% women and 48% men, which dismisses the common concern that the transaction-based data of banks that have online operations tend to oversample men and young consumers. This feature is important for the external validity of our analysis because men and women differ systematically in the extent to which their expectations depart from subsequent realizations and how they update their expectations to economic news (e.g., see D’Acunto et al. (2021); Coibion et al. (2022)).

Moving on to transactions, the average monthly income is \$1,930, whereas the average monthly spending is \$1,245. Both distributions are right-skewed: the median income is about \$1,245 and the median spending is about \$762. Consumers have accumulated \$17,818 in savings on average at the time of the first round of surveys. Even here a fat right tail emerges—the median consumer has only \$4,261 in savings. The average stock of interest-incurring unsecured debt is around \$1,015. This figure masks substantial heterogeneity: the median consumer has no interest-bearing unsecured debt, whereas, conditional on holding any debt, the average is about \$2,300. About 43% of the consumers

¹⁰Values are divided by 6.4, which is roughly the exchange rate at the end of 2021.

in the sample have positive credit card debt, which aligns with the ranges found in the prior literature using US data (40%–60%) (Gross and Souleles, 2002; Zinman, 2009; Fulford, 2015). Average accumulated debt is lower than the credit limits consumers are assigned by all the banks with which they have relationships (about \$12,500), that is, most consumers have untapped debt capacity.

The bottom part of Table I reports summary statistics about expectations. Expected incomes are calculated as the reported most likely values of total income over the following six months. The results show that income expectations are on average higher than ex-post realized incomes. The distribution of the difference between individual income expectations and the subsequent realized values stresses this point: the average and median values are \$213 and \$390. Expected credit limits are also higher than the ex-post realized credit limits, which suggests that on average consumers expect to have access to more credit in the medium term than they actually end up having.¹¹

The sample of consumers we survey is similar to the overall bank’s customer population (see Table A.1), even though they are a slightly younger (39 vs. 41) and have 10% lower incomes than the average bank customer.

III Unexpected Income Shocks and Subjective Income Expectations

Matched observational transaction-level data with survey-based beliefs data allows us to relate income shocks to subjective income forecast errors at the individual level.

We first measure subjective forecast errors as the difference between consumers’ quantitative expectations of future income and ex-post realizations ($E_C[Y_{t+1}] - Y_{t+1}$). We then measure income changes between the previous period and the current period (ΔY_t) as well as subjective income shocks from the previous survey wave six months earlier as the difference between a consumer’s actual average income over the subsequent

¹¹Although salaries are expected to be mostly stable, total income that include components in addition to salaries could have large variations that keep consumers from perfectly forecasting the future values. For example, Ganong et al. (2024) using administrative data in the US document large monthly variations in total earnings.

six months and her expected average income over the subsequent six months ($Y_t - E_C[Y_t]$).

Subjective income shocks and subjective forecast errors might be mechanically correlated due to serial correlation in expectations errors. For this reason, we also construct a measure of objective income shocks that does not rely on expectations data. We estimate objective income shocks in period $t+1$ as the residual $\epsilon_{i,t+1}$ from the following specification:

$$Y_{i,t+1} = \rho_{j,k,a} Y_{i,t} + \Gamma X_{i,t} + \epsilon_{i,t+1}. \quad (1)$$

The strategy is similar to the estimation of expected and unexpected tax refunds in Baugh et al. (2021). In equation (1), $Y_{i,t+1}$ is consumer i 's income in period $t+1$, $X_{i,t}$ is a set of consumer demographics that includes age, age-squared, educational attainment, gender, the log of savings in the previous period, the log of the credit limit in the previous period, city fixed effects, and industry \times period fixed effects. The period is defined as a half year to be consistent with the survey design. $\rho_{j,k,a}$ is the persistence of income at the industry-city-age quintile level. Equation (1) is estimated with a random 5% of all customers in the bank's database from 2014 to 2019.

Figure 2 is a binned scatter plot of the objective income shocks derived from equation (1) against the subjective ones. Panel A plots raw measures and Panel B residuals relative to consumer demographics. The plots show a linear relationship between the two measures. For the unresidualized measure, the R^2 is 0.25, indicating a positive but far from perfect correlation between the two measures. From Panel B, residualization absorbs a certain amount of noise in the measures, yielding a slightly larger R^2 .

Although consumers appear to form expectations about income that are on average positively correlated with actual future income shocks, the correlation is far from 1 and there is scope for systematic deviations of expectations from ex-post realized outcomes, which we investigate in the next subsection.

A. Extrapolative Subjective Income Expectations

The measures of realized income, subjective income expectations, and objective income dynamics allow us to assess directly the relationship between unexpected current income

shocks and the ex-post accuracy of income expectations.

We start by presenting motivating evidence about systematic mistakes in consumers' income expectations (*income misbeliefs*). Panels A of Figure 3 depicts a binned scatter plot of ex-post realized incomes against ex-ante income expectations and Panel B plots the period-by-period changes of the variables against each other. We calculate $\Delta Y_{t+1} = Y_{t+1} - Y_t$ and $E_C[\Delta Y_{t+1}] = E_C[Y_{t+1}] - Y_t$.¹² The correlation is positive: consumers' forecasts go on average in the same direction as realized values (see also Caplin et al. (2023)). However, the cross-sectional variation of forecast errors is wide. Moreover, forecasts are on average biased upwards: Panel C of Figure 3 shows that the majority of the distribution of forecast errors lies in the positive domain. In Panel D, the distribution of forecast errors remains wide even after residualizing by demographics.

We continue by assessing the relationship between current-period unexpected income shocks and misbeliefs about subsequent-period income. Figure 4 reports binned scatter plots of forecast errors for income against unexpected income shocks. In each of the four panels, the y-axes report subjective forecast errors measured using subsequent-period income realizations. In Panel A and Panel B, the x-axes report the objective income shocks in the current period from equation (1). In Panel C and Panel D, the x-axes report the subjective unexpected income shocks in the current period, that is, realized income over the last six months minus expected income from six months earlier. Income shocks and income forecast errors are positively related irrespective of whether income shocks are measured objectively based on the specification in equation (1) or subjectively based on survey answers.¹³

To assess this relationship in a multivariate setting, we estimate the following specification:

$$E_C[Y_{i,t+1}] - Y_{i,t+1} = \beta(Y_{i,t} - E[Y_{i,t}]) + X'_{i,t}\delta + \boldsymbol{\eta} + \nu_{i,t}^Y, \quad (2)$$

where X is a vector of individual-level characteristics that include age and its square,

¹²The label $E_C[Y_{t+1}]$ indicates that the expectation is subjective and is measured from the perspective of the consumers.

¹³Bellifemine et al. (2024) find a similar pattern using Italian survey data.

educational attainment, female indicator, the log number of weekly hours worked in period t , the logarithms of monthly income and credit-card limits in period $t - 1$, which proxies for consumer’s debt capacity, consumer expected income changes from period $t - 1$ to t , and different sets of fixed effects ($\boldsymbol{\eta}$).

The left-hand-side variable in equation (2), $E_C[Y_{i,t+1}] - Y_{i,t+1}$, measures the ex-post realized subjective expectation error at time $t + 1$. On the right-hand-side, $Y_{i,t} - E[Y_{i,t}] = \epsilon_{i,t}$ measures the income shocks at time t . Our coefficient of interest is β , which estimates the marginal relationship between the current-period income shock and subsequent-period’s forecast error: β should be zero for an agent with rational expectations.

Table II reports the results for estimating equation (2). Column (1) only includes the unexpected income changes in period t as the explanatory variable. Column (2) controls for city \times year and industry of occupation fixed effects. Column (3) further adds individual fixed effects. Across columns (1)–(3), $\hat{\beta}$ is significantly larger than zero economically, different from zero statistically, and quite stable regardless of the characteristics we absorb and how we restrict the variation that identifies the coefficient. The positive value of $\hat{\beta}$ implies that the subjective income expectations of the average consumer are extrapolative.

Focusing on column (3), the inclusion of city \times year fixed effects absorbs local shocks that might induce a structural change in the income processes of all consumers in the same location. At the same time, the inclusion of individual fixed effects absorbs unobserved individual-level time-invariant characteristics that might induce negative auto-correlation of subjective expectations across periods. As a result, our specification estimates how larger vs. smaller income shocks relate to the accuracy of the same individual’s expectations. $\hat{\beta}$ in column (3) is 0.405, which means that for a one-dollar unexpected income shock in the current period, the average consumer over-estimates her income in the next period by about 40 cents.

Our setting allows us to assess the relationship between income shocks and forecast errors under different states of the economy. In columns (4)–(6) of Table II, we regress forecast errors on income shocks separately for the three years in our sample. Two findings seem relevant. On the one hand, in both normal times and in uncertain times, income

shocks are positively related to forecast errors. On the other hand, the effects of income shocks on forecast errors are slightly larger in 2020 than in 2023, that is, the degree of extrapolation appears to be slightly larger when aggregate uncertainty is higher.

Table II faces the concern that income shocks measured with (1) are negatively correlated: positive income shocks could predict negative surprises even under rational expectations. To assess this concern, in Online Appendix Table A.2, we regress objective income shocks in the subsequent period on objective and subjective income shocks in the current period. Neither objective nor subjective income shocks predict objective income shocks in the following period, which dismisses a role for negative serial correlation to explain the positive effect of income shocks on forecast errors we document.

Note that in equation (2), we regress future forecast errors on current income innovations. Because we observe subjective income errors for the same consumers over more than two periods, we can also study the relationship between future forecast errors and current subjective income shocks. We report the results, which are quite similar to our baseline findings, in Table A.3 in the Online Appendix.¹⁴

We then assess the heterogeneity of the relationship between income shocks and forecast errors. In particular, recent research in microeconomics finds that individuals' tendency to overreact to signals when forming subjective beliefs is higher when facing noisier and more volatile signals (Augenblick et al., 2021; Ba et al., 2022). We thus test if consumers who face more volatile incomes overreact more than others to unexpected income shocks. We consider four proxies for consumers' income volatility—whether the consumer belongs to the bottom half of the income distribution (Ferland et al., 2023); the standard deviation of the logarithm of expected income growth; age, because incomes tend to be more volatile among younger individuals; and educational attainment, because incomes tend to be more volatile among non-college-educated individuals.

Columns (1)–(4) of Table III report the results for estimating equation (2) once we add interactions with each of the four proxies for income volatility. Overreaction is systematically higher for consumers whose incomes are more volatile: consumers below the median of the income distribution overreact more than others by about half;

¹⁴In Table A.4, we report results where $\epsilon_{i,t+1}$ is retrieved based on $\log Y_{i,t+1}$. The findings are nearly unchanged.

those whose implied expected income growth volatility, as calculated based on assuming subjective income growth follows a triangular distribution, is higher overreact by more; older consumers extrapolate less; and, the interaction term with college education is negative although not statistically significant.

To further assess the role of income volatility, in column (5) of Table III, we consider the relationship between forecast errors and income shocks for agents who work in low-income-volatility industries and others. Specifically, we interact income shocks with a dummy variable indicating if the agent works in a government-related industry. And, indeed, we find that agents who have very stable incomes barely extrapolate at all. Because we observe several occupation industries, in Figure A.2 of the Online Appendix we dig deeper into this result by plotting forecast errors against income shocks separately for agents in each of the industry classifications we observe. Consistent with Table III, the relationship between forecast errors and income shocks is larger for consumers who work in industries with more volatile incomes (e.g., household services and business services), whereas both the size of the shocks and the steepness of the relationship are lower in industries with less volatile income (e.g., government and education).

In column (6) of Table III, we test if the degree of extrapolation differs based on the sign of income shocks. We cannot reject the null of no difference either economically or statistically when comparing extrapolation after positive or negative shocks.

All the results so far refer to the first moment of subjective income expectations. We move on to assess if unexpected income shocks also help explain variation in the second moment of subjective income expectations, that is, subjective income volatility. We do so by estimating versions of equation (2) in which the outcome variable is the standard deviation of consumers' expected income growth, which we compute based on our survey questions under the assumption that income expectations follow a simple triangular distribution.¹⁵ Because we are considering the second moment of the distribution, we use the absolute value of unexpected income shocks as the explanatory variable.

¹⁵We use the answer about the most likely earning divided by the previous-period income as the expectation, denoted as z . With the minimum and maximum divided by the previous-period income denoted as a and b , we calculate the variance as $(a^2 + b^2 + c^2 - ab - ac - bc)/18$, where $c = 3z - a - b$ is the mode. In Table A.5, we show that our results are robust to use the most likely value as the expected value, or the mid-point of a and b as the mode.

In Table IV we find that, once we absorb fixed effects, as the size of unexpected income shocks increases, irrespective of the sign, the standard deviation of expected future income growth increases as well. Moreover, in this case, the sign of unexpected income shocks matters. Shocks of the same size but different signs relate to expectations uncertainty differently. When we interact the absolute value of the unexpected income shock with a dummy for whether the shock is negative in column (4), the association between unexpected income shocks and income-expectations volatility is higher after negative shocks than after positive shocks of the same size. For each \$1,000 higher income shock, subjective income volatility is 18% larger after positive shocks but about 35.6% higher after negative shocks.

IV Extrapolative Income Expectations and Consumption, Debt, Defaults

Do the facts about subjective income expectations we documented so far have any real effects? If agents acted based on the beliefs they stated when surveyed, we would expect higher (lower) current consumption induced by overly optimistic (pessimistic) income expectations based on the perceived income effect.

We estimate variations of the following linear specification:

$$\Delta C_{i,t} = \gamma(E_C[Y_{i,t+1}] - Y_{i,t+1}) + X'_{i,t}\delta + \boldsymbol{\eta} + \nu_{i,t}^C, \quad (3)$$

where $\Delta C_{i,t}$ is the change in consumption at time t relative to the previous period ($t - 1$). $E_C[Y_{i,t+1}] - Y_{i,t+1}$ is the subjective expectations error in $t + 1$ from period t . In each year, t is the six-month period from January to June and $t + 1$ from July to December. Meanwhile, $t - 1$ refers to the six-month period from July to December of the previous year. For example, in 2020, $\Delta C_{i,t}$ is the difference between consumer i 's total spending between January 2020 and June 2020 and her spending between July 2019 and December 2019. $E_C[Y_{i,t+1}] - Y_{i,t+1}$ is the difference between subjective income expectations elicited in July 2020 for the subsequent six months and the realized total income between July

2020 and December 2020. All other variables are defined as in equation (2).

We report the estimates in Table V. Column (1) considers a baseline univariate specification: consumers whose income expectations are \$1,000 higher than the future realizations increase their current consumption by about \$206 dollars more than other consumers. In column (2), we keep constant the characteristics we observe and restrict the variation within city \times year and within individuals. Even though we explain about 50 percentage points more of the variation of current consumption choices, the estimated association between subjective income expectations and the change in spending barely varies economically or statistically.

Because forecast errors increase current consumption but not future realized income, positive forecast errors induce consumption to deviate from the optimal path. The positive response in current consumption paired with negative surprises in future income suggests that debt should increase when expectations errors are positive for agents who need to access external credit to finance current consumption. We thus assess if extrapolative income expectations predict higher borrowing. For this test, we estimate a version of equation (3) in which the left-hand-side variable is $\Delta B_{i,t+1}$. For example, in 2020, $\Delta B_{i,t+1}$ is the difference between the total outstanding interest-bearing unsecured debt at the end of December 2020 and at the end of June 2020.

Table V reveals that the larger is the difference between income expectations and ex-post income realizations, the higher is the increase in the unsecured debt the average consumer raises. The size of this estimate is similar in a univariate specification (column (3)) and when we include the full set of observables and fixed effects we observe (column (4)). The estimate indicates that, for each dollar higher misbelief in the average monthly income over the following period, average monthly unsecured debt increases by about 7.4 cents. This amount is sizable considering that unconstrained consumers do not raise debt to finance their higher spending and hence the relationship is zero for them.

The third variable we analyze is debt outcomes: because realized incomes are lower than expected unless consumers can tap into savings, they might not be able to repay their debt in full. We estimate equation (2) using the default indicator as the left-hand-side

variable.¹⁶ By construction, we only observe outcomes up to time $t + 1$ but defaults might happen at any (unobserved) subsequent time until the loan is due. For this reason, our estimates likely represent a lower bound of the actual relationship between subjective income forecast errors and the likelihood of default.

Columns (5)–(6) of Table V document that a higher difference between consumers’ income expectations and ex-post realizations is associated with a higher probability of default: for each \$1,000 higher income misbelief default increases by about 0.915-percentage points (column (6)). This magnitude is large because the average default rate in our sample is 2.31%, that is, a \$1,000 higher income misbelief leads to a 39.6% higher likelihood of default relative to the sample mean. These findings are consistent with aggregate dynamics such as credit-market sentiment triggering financial fragility (López-Salido et al., 2017)

As we discussed above, our debt data covers all sources monitored in the Chinese credit registry. By contrast, spending data is based on the accounts at the bank with which we cooperate. To assess concerns of systematic mismeasurement above and beyond the screening filters to select primary account users, we study the relationship between expectations errors and the changes in net transfers between our bank and external bank accounts (see Online Appendix Table A.6). We find no relationship between expectations errors and cash withdrawals or net transfers to external accounts.

A. Dynamic Effects

Our results so far have considered six-month horizons. However, in the long run, consumers face income realizations and become aware of their forecast errors. Consumption and debt choices should correct and the speed of this reaction is an open empirical question. To assess these dynamics, we estimate the relationship between cumulative spending and income forecast errors at various horizons:

$$C_{i,\tau+k} - C_{i,\tau-1} = \alpha + \beta_k(E_C[Y_{i,t+1}] - Y_{i,t+1}) + X'_{i,t}\delta + \boldsymbol{\eta} + \nu_{i,\tau+k}^C. \quad (4)$$

¹⁶Take 2020 as an example. The default indicator we observe is a 90-day delinquency indicator from October 2020 to March 2021. For ease of interpretation, we multiply default by 100 and divide forecast errors by 1,000.

In equation (4), τ is the first quarter of period t . Because each period includes six months, period t refers to quarters τ and $\tau + 1$ and period $t + 1$ to $\tau + 2$ and $\tau + 3$. $C_{i,\tau-1}$ is the average monthly spending in the quarter before the first survey and $E_C[Y_{i,t+1}]$ is the expected average monthly income during quarters $\tau + 2$ and $\tau + 3$.¹⁷ We fit 12 regressions based on equation (3) for k ranging from -4 to 8 excluding $k = -1$, which we use as benchmark. Equation (4) thus measures the relationship between expectations errors at time $t + 1$ and cumulative spending from quarter $\tau - 1$ to quarter $\tau + k$.

We plot the estimates in Panel A of Figure 5. The red solid lines depict positive expectation errors, and the blue dashed lines negative expectation errors. The shaded areas represent two standard error bands. Consistent with overreaction, consumers first increase their consumption during the two quarters before the time we measure expectations errors (quarters 2 and 3) as well as during the two quarters when expectation errors are measured. However, consumption starts to revert back afterwards and almost fully reaches the level of three quarters before the time during which we measure expectations. Panel B documents very similar patterns for borrowing choices.

B. Asymmetric Effects: Positive and Negative Unexpected Income Shocks

In aggregate data, household debt cycles and their correlations with households' spending have been shown to build up slowly in times of positive income growth but drop quickly in times of negative income growth. We therefore test whether unexpected income shocks in the negative domain affect consumption and debt choices more than unexpected income shocks of the same size in the positive domain.

Table VI reports the results of this test. Note that in this case, the main covariate of interest is the unexpected income shock at time t instead of the subjective income forecast error. For this reason, columns (1) and (3) of Table VI report a baseline multivariate specification in which shocks of different signs are treated equally.

Columns (2) and (4) of Table VI reveal that, consistent with the literature on

¹⁷Equation (3) can be written in the form of equation (4) if we replace the left-hand-side variable with $C_{i,\tau+1} + C_{i,\tau+2} - C_{i,\tau-1} - C_{i,\tau-2}$.

aggregate household debt cycles, the micro-level estimated response of both consumption and debt to unexpected income shocks is stronger for negative shocks than positive shocks of the same size.¹⁸

C. Extending the Time Series

Our survey-based panel data includes a relatively short time-series component. Even if we show that our results are similar independent of whether we consider variation across or within consumers, income innovations are unlikely to average out to zero within consumers across a few periods (Chamberlain, 1982; Keane and Runkle, 1998; Souleles, 2004).

To tackle this concern, we exploit the fact that our observational account-level data spans a substantially longer time series than the survey-based data. We can therefore study the relationship between income innovations and spending decisions for the consumers who participated in our surveys for a longer time series. We use the longest time horizon possible based on the data the bank has been willing to share with us, which is on average 5.2 years per consumer, leading to a panel including ten or more periods for each consumer.

In Online Appendix Table A.7, we compare the results when extending the time series (even columns) relative to the baseline setting (odd columns). The conditional associations between income innovations and consumption choices, debt choices, and defaults are similar across samples.

V From Reduced Form to Quantitative Analysis: A Structural Estimation

Our analysis so far has produced reduced-form evidence that consumers form extrapolative subjective income beliefs after unexpected income shocks and that such beliefs predict their consumption and debt accumulation decisions as well as subsequent defaults. But can this micro-level demand-side channel help us understand the dynamics of aggregate

¹⁸Our estimates of the consumption elasticity to income shocks, which is between 0.15 to 0.2, see Table A.9 in the Online Appendix, are consistent with the estimates in Ganong et al. (2020).

household debt cycles?

To tackle this question, we introduce extrapolative income expectations and defaultable debt in a standard consumption model with incomplete markets and heterogeneous agents. We use the model to compare quantitatively the relationship between unexpected income shocks, income forecast errors, and consumption and debt choices for consumers with and without extrapolative expectations.

A. Income Process and Expectations Formation

In the model, time is discrete and denoted by $t = 1, 2, \dots$. A unit mass of consumers are subject to idiosyncratic income risk. For each individual i , income $y_{i,t}$ follows (as in Blundell and Preston (1998) and Carroll (1997)):

$$\begin{aligned}\log y_{i,t} &= \alpha + z_{i,t} + \epsilon_{i,t} \\ z_{i,t} &= \rho z_{i,t-1} + \eta_{i,t},\end{aligned}\tag{5}$$

where $\epsilon_{i,t}$ and $\eta_{i,t}$ are i.i.d. normal shocks with $\mathbb{E}[e^{\epsilon_{i,t}}] = 1$ and $\mathbb{E}[e^{\eta_{i,t}}] = 1$. The variances of $\epsilon_{i,t}$ and $\eta_{i,t}$ are σ_ϵ^2 and σ_η^2 , respectively. α is the life-cycle component, which we assume is constant and common knowledge.

Consumers do not know the true value of $z_{i,t}$ and need to make inferences based on Bayesian learning. The Kalman-filtering problem with respect to the persistent component of $\log y_{i,t}$ follows Guvenen (2007) and Bordalo et al. (2019). We formulate extrapolative expectations as diagnostic expectations because diagnostic expectations are portable across many economic settings (Bordalo et al., 2019; Chodorow-Reich et al., 2023; Maxted, 2023).

In each period, consumers observe $y_{i,t}$ and update their beliefs about $z_{i,t}$. Under rational expectations, the current forecast for $z_{i,t}$ is normally distributed with variance σ_z^2 and mean given by:

$$\hat{z}_{i,t} = \rho \hat{z}_{i,t-1} + \kappa [y_{i,t} - \alpha - \rho \hat{z}_{i,t-1}],\tag{6}$$

where κ is the Kalman gain of the learning process. Given an infinite horizon, we follow the common assumption that a sufficient number of periods have passed such that consumers are in a learning steady state, that is, consumers' Kalman gain is constant each period. In this case,

$$\begin{aligned}\kappa &= \frac{\rho^2 \sigma_z^2 + \sigma_\eta^2}{\rho^2 \sigma_z^2 + \sigma_\eta^2 + \sigma_\epsilon^2}, \\ \sigma_z^2 &= \frac{(1 - \kappa) \sigma_\eta^2}{1 - (1 - \kappa) \rho^2}.\end{aligned}\tag{7}$$

In contrast, under diagnostic expectations, consumers overreact to surprises in income realizations. The posterior average of $z_{i,t}$ becomes:

$$\hat{z}_{i,t}^\theta = \hat{z}_{i,t} + \theta \kappa [y_{i,t} - \alpha - \rho \hat{z}_{i,t-1}],\tag{8}$$

where $\hat{z}_{i,t}^\theta$ is the expectation of $z_{i,t}$ under diagnostic expectations, and θ is the degree of representativeness. With $\theta = 0$, the model reduces to rational learning. When $\theta > 0$, consumers overweigh representative states, their beliefs exaggerate the signal-to-noise ratio relative to the standard Kalman filter, which inflates the persistent component of income upon receiving good news and deflates it when receiving bad news. Overreaction to news increases in θ .

B. Preferences and Optimality Conditions

In this subsection, we first introduce consumer preferences and then discuss the optimality conditions.

B.1 Preferences

Household preferences follow the literature on consumer credit and default (e.g. Chatterjee et al. (2007) and Livshits et al. (2007)). Consumers at time t maximize their

expected lifetime utility with period- s utility flow of:

$$E_{i,t}^{\theta} \left[\beta^{s-t} \frac{c_{i,s}^{1-\gamma} - 1}{1-\gamma} - \chi d_{i,s} \right],$$

where the superscript θ indicates that the expectation is taken according to diagnostic expectations with degree of representativeness θ . β is the per-period discount rate, γ is the coefficient of relative risk aversion, and $d_{i,t} = 1$ if consumer i chooses to default at the end of period t .

When default occurs, consumer i incurs a fixed non-pecuniary utility cost (“stigma”) $\chi > 0$. In addition, consumers receive a pair of additively separable i.i.d. shocks, $\xi = \{\xi^D, \xi^N\}$, which are attached to the options to default or repay and are drawn from a type one extreme value distribution with scale parameter of 1. These shocks capture the fact that many defaults are associated with events such as marital disruptions and medical expenses, which we do not model explicitly. With these shocks, the model generates a positive probability of default across the whole range of borrowing. In addition, as suggested in Dempsey and Ionescu (2023), the introduction of utility shocks associated with defaulting smooths out individuals’ repayment probability functions, which eases the computation of the model.

The budget constraint in period t is

$$a_{i,t+1} = \begin{cases} (1 + r_{i,t})(a_{i,t} - c_{i,t}) + y_{i,t+1} & \text{if } d_{i,t} = 0 \\ (1 - \nu)y_{i,t+1} & \text{if } d_{i,t} = 1 \end{cases}$$

$$a_{i,t} \geq -l_{i,t}, \tag{9}$$

where $a_{i,t}$ is the total amount of available resources. $l_{i,t}$ is the credit limit, and $\nu \in [0, 1]$ is the marginal rate of garnishment. Equation (9) states that when consumers do not default, their wealth in the next period is the sum of their income and gross savings. When consumers default, their savings become zero; at the same time, they need to pay a garnishment cost equal to ν times their income in the following period. For simplicity, we assume that consumers’ borrowing capacity does not change upon default, which allows

us to discard one additional state variable.¹⁹ The interest rates differ for saving and borrowing and take the values

$$r_{i,t} = \begin{cases} r_b & \text{if } a_{i,t} < 0 \\ r_s & \text{if } a_{i,t} \geq 0. \end{cases}$$

B.2 Optimality Conditions

Consumers' problem is characterized by a set of four state variables $\Theta_{i,t} = (a_{i,t}, \hat{z}_{i,t}, z_{i,t}, \epsilon_{i,t})$. Given the overall state $\theta_{i,t}$, consumer i 's value function at time t is

$$V(\Theta_{i,t}) = \max \{V_D(\Theta_{i,t}), V_N(\Theta_{i,t})\}. \quad (10)$$

The continuation value from defaulting is

$$V_D(\Theta_{i,t}) = \max_{c_{i,t}} \frac{c_{i,t}^{1-\gamma} - 1}{1-\gamma} - \chi + \beta \mathbb{E}_{i,t}^\theta [V((1-\nu) \times y_{i,t+1}, \hat{z}_{i,t+1}, z_{i,t+1}, \epsilon_{i,t+1})] + \xi_{i,t}^D. \quad (11)$$

The continuation value from not defaulting is

$$V_N(\Theta_{i,t}) = \max_{c_{i,t}} \frac{c_{i,t}^{1-\gamma} - 1}{1-\gamma} + \beta \mathbb{E}_{i,t}^\theta [V(a_{i,t+1}, \hat{z}_{i,t+1}, z_{i,t+1}, \epsilon_{i,t+1}) | I_{i,t}] + \xi_{i,t}^N. \quad (12)$$

Given that ξ follows a type one extreme value distribution, the probability of default is

$$pd(\Theta_{i,t}) = [1 + \exp\{V_N(\Theta_{i,t}) - V_D(\Theta_{i,t})\}]^{-1}. \quad (13)$$

C. Calibration

The estimation consists of two stages. In the first stage, we rely on our income and beliefs microdata. We pin down the parameters associated with the income process such as the degree of extrapolation θ and the marginal rate of garnishment. In the second stage, we use the simulated method of moments (SMM) to get the estimates of consumers'

¹⁹Some studies assume that defaults go hand in hand with a temporary inability to borrow, i.e. $l_{i,t} = 0$ (Chatterjee et al., 2007; Livshits et al., 2007; Dempsey and Ionescu, 2023), but Livshits et al. (2007) show that the costs of default from changing borrowing capacities are quantitatively small.

coefficient of risk aversion γ and non-pecuniary costs of default χ .

First Stage: The first-stage estimation requires the use of income and survey data. We first set the time frequency t to six months to be consistent with the survey frequency. For the parameters governing the income process, α , ρ , σ_ϵ^2 , and σ_η^2 , we residualize all individual incomes by age quintiles, year, education, occupation industry, city, and gender fixed effects, and estimate the process in (5) using the method of moments. After residualization, α is set to 0. When solving the model, we measure consumption and saving with respect to the level of average income.

We report the estimation results in Panel A of Table VII. Log income is highly serially correlated (AR(1) coefficient of 0.97). The average annualized interest rate on deposits in the data is around 2.8%. We therefore set $r_s = 1.4\%$. We use various interest rates and fees on the credit card borrowing from the bank to determine r_b . The average daily interest rate on credit card debt is 5 basis points. After accounting for a 45 days interest-free period, a 2.5% reward rate, and 1% of debt-related fees, we set r_b to 5.5% per six months. We set a uniform credit limit for all consumers ($l_{i,t} = 1.4$) so that the average credit limit to average income in the model matches the average total credit limit to annual income in the data (73%).²⁰ We compute the marginal garnishment rate ν from the bank’s database—it is about 5%, which corresponds roughly to one third of the average monthly income in the data.²¹ We set the discount rate $\beta = 0.95^{0.5}$, so that the annual rate is 0.95. Finally, we calculate θ to match the regression coefficient of subjective beliefs errors on income shocks, which yields $\theta = 1.68$.

Second Stage: We use SMM to estimate γ and χ . The targeted moments are the average wealth-consumption ratio and average default rate. The intuition behind the relationship between targeted moments and estimated parameters is as follows. The risk-aversion parameter γ captures the curvature of the utility function. Higher risk aversion

²⁰The level of credit limit to income is larger than the 20%-35% range in many previous papers using data from the Survey of Consumer Finances in the US (Kaplan and Violante, 2014; Lee and Maxted, 2023). However, it is very close to the number in a recent report by Experian using administrative data. The report documents that the average credit card limit in 2019 for an average American was around \$31,400, which is around 60% of the average individual income of \$54,129 in 2019, and should be a lower bound of the ratio between total limits over all lines of credit and income. See [here](#) and [here](#) for the report.

²¹This number is slightly smaller than the estimate of around 9% from DeFusco et al. (2024) in the US, conditional on being garnished for delinquent debt.

increases consumers' willingness to save, thereby increasing the wealth-to-consumption ratio. The stigma cost χ directly affects consumers' willingness to default. A higher χ indicates a higher cost of default and hence a lower default rate.²²

D. Reproducing the Empirical Results

We now discuss the results of the structural model.

D.1 Goodness of Fit

Table VII reveals that the model does a good job in matching the targeted moments as well as non-targeted moments. In Panels B and C, the estimation fits the empirical moments closely. The average wealth to half-year total consumption is 0.846. The default rate is around 2.31%, which yields a γ of 2.51 and a χ of 24.45.

Panel D shows the results for a set of non-targeted moments. The estimation also fits these empirical moments. The average (median) wealth to six-month income ratio is 1.351 (0.731) in the data and 1.263 (0.780) in the model. Because liquidity affects the MPC in the sample, the stationary distribution of savings should be consistent with the empirical one. Panel D also shows that the model can match several moments of the empirical distribution of liquid savings. For example, in the data, 31.38% of total liquid assets are held by consumers in the top 5% of asset holdings. In the model, this number is 31.07%. In addition, 29.89% of individuals have negative net liquid wealth in the data; in the model, the corresponding number is 31.89%. Therefore, the model is capable of matching both first and higher moments.

D.2 Extrapolation of Income Shocks

We use the model to study the effects of extrapolative income expectations on economic choices. We first compare the relationship between objective income shocks and subjective forecast errors under diagnostic expectations and rational expectations. We simulate 100,000 periods of income data following the process in (5) and construct subjective and objective income expectations based on (8) and (7). We then regress subjective forecast

²²A detailed description of the model solution and estimation is in the Online Appendix.

errors at $t + 1$ on the objective income shock at t . When performing the analysis, we drop the first 100 periods, which serve as the burn-in period in the simulation.

We report the model results in columns (1) of Table VIII. Panel B refers to diagnostic expectations and Panel C to rational expectations. For comparison, Panel A reports the empirical counterparts in the data. Given that θ is calibrated to match the relationship between future subjective belief errors and current income shocks, the regression coefficient of subjective forecast errors on objective income shocks under diagnostic expectations matches the empirical counterpart perfectly: when $\theta = 1.68$, each dollar higher objective income shock leads to a 40-cent income forecast error. The estimate in column (1) of Panel C sheds light on the importance that θ be positive to recover the empirical relationship: when θ is set to zero, the relationship between objective income shocks and forecast errors becomes economically insignificant. Each dollar higher objective income shock leads to a *negative* 4-cent income forecast error.²³

D.3 Consumption, Borrowing, and Defaults

We continue the quantitative exercise and explore whether the model can generate the empirical relationship between subjective forecast errors and economic choices. To perform the analysis, we simulate the model for 20,000 individuals with 1,000 periods, after a burn-in period of 100. We then drop simulated data with savings to average income ratio larger than 8.13, which is the highest value in the data. After obtaining the simulated sample, we run regressions of changes in total consumption at t , total debt at the beginning of period $t + 1$, and the default indicator at the end of $t + 1$ on the forecast errors for period $t + 1$. Debt at the beginning of $t + 1$ is defined as the negative of $a_{i,t}$, conditional on $a_{i,t}$ being negative. To be consistent with the analysis in Table V, we control for log income at $t - 1$, expected changes in income at t , and individual fixed effects.

We report the results in columns (2)–(4) of Table VIII. For comparison, Panel A reports the empirical counterparts. The relationship between forecast errors and economic

²³The statistical significance is a result of high precision of simulated data with a large number of data points.

decisions implied by the model are close to those in the data. Specifically, a one-dollar higher income forecast error at $t+1$ relates to 20.6 cents higher current consumption in the data and leads to 25.1 cents higher consumption in the model. Meanwhile, a one-dollar higher income forecast error at $t+1$ relates to 7.5 cents higher debt in the data and leads to 7.1 cents higher debt in the model. For default, since the units are different, we standardize the forecast errors so that the coefficients measure the association between the default probability for a one standard deviation higher forecast error. In the data, each standard deviation higher forecast error increases default by 91.8 basis points and the model-implied quantity is 93.1 basis points.

In Panel C of Table VIII we see that, without extrapolative expectations (θ equal to zero), the relationships between forecast errors and consumption, debt, and default become much smaller. Forecast errors do not have an economically significant impact on current consumption. And, although negative income surprises lead mechanically to more debt and defaults, with rational expectations, the effects are about 60% smaller for debt and 75% for defaults.

Overall, the version of the model that includes extrapolative expectations is able to produce the patterns we observe in the data, whereas the same model with rational expectations is unable to do so.

***E.* Aggregate Household Debt Cycles**

After verifying that the model with extrapolative expectations can reproduce our reduced-form empirical results, we move on to assess if the same model is also able to produce an aggregate household credit cycle à la Minsky-Kindleberger. In a Minsky-Kindleberger credit cycle, boom-bust patterns in household leverage start with household over-optimism after positive shocks to fundamentals. That is, after positive shocks, households exaggerate the informativeness of good news about future growth and take excessive leverage, which builds up financial fragility in the economy. When the effects of positive shocks diminish, households face a negative income surprise and their ability to repay debt is low. Consequently, at the end of expansions, both the demand for (additional) debt and financial fragility are elevated.

Following Maxted (2023), we use impulse response functions (IRF) to study consumer responses to a series of positive transitory income shocks starting from the stochastic steady state. For a baseline, we start by simulating 20,000 diagnostic-expectations individuals ($\theta = 1.68$) for 1,000 periods and a half-year frequency without introducing additional shocks. To derive the IRFs, we use the same simulation but introduce a three-year sequence of positive income shocks, which results in a three-standard-deviation cumulative shock over the three years from periods 895 to 900. The IRFs are the (percentage) differences between the sample averages of the two simulations. Then, we repeat the same exercise but replace diagnostic expectations with rational learning— θ is set to zero. Comparing the IRFs across the two calibrations reveals the effects of extrapolative expectations on economic outcomes.

We plot the IRFs in Figure 6. The top left panel plots the transitory income shocks and the top right panel the updates in expected log income, $o_{i,t} = \kappa(1+\theta)(y_{i,t} - \alpha - \hat{z}_{i,t-1})$. The bottom four panels are the percentage differences of average income expectations, consumption, borrowing, and default rate relative to the no-shock simulation. The red solid lines plot the results under diagnostic expectations ($\theta = 1.68$) and the blue dashed lines under rational learning ($\theta = 0$).

Figure 6 shows a strong amplification effect of diagnostic expectations on income beliefs. Panel B and Panel C show that, initially, after facing positive shocks, the average diagnostic-expectations agent (D) increases her income expectations more than twice as much as a rational-learning agent (R). Given a much higher expected income, consumption increases more for D . In terms of debt, Panel A and Panel D show that when D faces the first positive shock, realized income increases by around 20% and consumption increases by around 14%. Because we only have one asset in the model, an MPC smaller than one indicates an initial drop of debt, which is consistent with Agarwal et al. (2007) and Coibion et al. (2020). In Panel F, relative to R , the initial positive income shock leads to a lower default rate for D . This result arises because higher expectations about future income induce a higher perceived garnishment cost associated with default.

As D continues to face positive income shocks, income expectations, and consumption continue to be higher than for R . At the same time, the trajectories of debt and the

default rate keep diverging. For R , income and income expectations increase smoothly and R continues to de-lever. An increase in current-period assets and expectations about future income decrease the default rate. In contrast, because of higher future income expectations and smaller income surprises, D starts to accumulate more debt than R . Panel F shows that a lower level of current-period assets tends to increase the motive for defaulting. Ultimately, this channel outweighs the marginal garnishment channel and default rates start to increase for D .

In period 1, the series of positive shocks are removed. As a result, income expectations decline. Panel B and Panel C show that this reduction is larger for D than R due to over-extrapolation. Consequently, consumption drops more for D . The larger negative surprise induces D to have a higher need to smooth the negative shock but also results in lower income expectations, thus creating a spike of debt. Relative to R , lower current-period asset holding and lower expected future earnings increase D 's default rate substantially. As time elapses, income expectations converge. Both D 's debt and default probabilities start to decrease to R 's level.

The patterns in Figure 6 are consistent with several recent findings in macroeconomics. For example, López-Salido et al. (2017) show that elevated credit-market sentiment is associated with a higher credit growth in subsequent years. Consistently, Panel E shows that as initial income shocks create more optimistic beliefs, the debt growth rate becomes positive for D , whereas it remains negative for R . In addition, the elevated default rate after the expansion (Panel F) is consistent with the findings in Greenwood et al. (2019) and Maxted (2023) that financial fragility arises at the end of economic expansions.

***F.* Simulating the Global Financial Crisis**

Economists' renewed interest in the drivers of aggregate household debt cycles over the last few years was arguably motivated by the Global Financial Crisis. Our last exercise thus assesses whether our model can also generate patterns of unsecured borrowing and default risk consistent with those observed in the US during the Global Financial Crisis. Because our parameter values are estimated with Chinese data, we focus on the patterns'

dynamics rather than their quantitative levels.

Figure 7 shows the results. The top panel plots the selected shocks. In the middle and bottom panels, the red solid lines and blue dotted lines present the simulation when $\theta = 1.68$ and $\theta = 0$, respectively. Again, the model frequency is half a year; the black dashed lines represent the data. In the middle panel, $\overline{B}/\overline{Y}$ is the ratio between consumer loans, credit cards, and other revolving plans and income. In the bottom panel, default probability is the delinquency rate on consumer loans.²⁴

Two results stand out in the middle panel. First, the aggregate debt level at the equilibrium (before 2004) is larger with diagnostic expectations. This result arises because, at the micro level, diagnostic expectations widen the cross-sectional dispersion of income expectations. In response to the same distribution of income shocks, more people have more optimistic expectations and more people have more pessimistic expectations. Because debt is bounded at zero, the share of people with more optimistic expectations ends up accumulating more debt, whereas more pessimistic people cannot de-accumulate debt to negative levels. Expectations errors do not wash out at the macro level causing more aggregate debt in equilibrium.

Second, under diagnostic expectations over income surprises, the model can reproduce the boom-bust cycles of the average unsecured-debt to average income ratio. The pattern aligns closely with the average credit-card debt to GDP ratio in the US. If diagnostic expectations are removed, the cyclical behavior becomes weaker: the 2009 peak never surpasses the 2004 equilibrium level.

The bottom panel reveals that these shocks, which are reverse-engineered to match the actual debt dynamics in the data, can replicate the cyclical patterns in default rates: at the end of the expansion, default rates shoot up and then slowly decline. However, when removing diagnostic expectations, default rates stay nearly constant from 2004 to 2012.

²⁴Since we are studying average debt to personal income in the model, we multiply GDP by the labor share to get total income. The resulting series trends down after 2000, so we detrend it over our sample period. For the delinquency rate, the model-implied default probability is based on all consumers, but the delinquency rate is conditional on debtholders. To make the series comparable, we adjust the delinquency rate by dividing it by the proportion of debtors in the economy. Hence, we get a debt delinquency rate for all consumers rather than only for debtor.

Overall, we find that the model can reproduce an aggregate boom-bust debt cycle like the one observed during the Global Financial Crisis when income expectations are extrapolative but not under rational learning about income.

VI Conclusions

Our results suggest that extrapolative expectations can be a demand-side driver of aggregate household debt cycles. First, combining survey-based subjective income expectations with transaction-level and credit-registry data, we show that consumers form extrapolative expectations after unexpected income shocks. Positive subjective forecast errors relate to higher current consumption, debt accumulation, and subsequent defaults. Moreover, positive and negative unexpected shocks have asymmetric effects on consumption and debt accumulation, which is consistent with the aggregate dynamics of household debt cycles documented in the literature. In a quantitative exercise, we show that extrapolative income expectations can produce Minsky-Kindleberger credit cycles after positive income shocks as well as patterns of unsecured debt and default risk that are consistent with those observed in the US during the Global Financial Crisis. The same model with rational learning cannot produce these results.

Future research should dig deeper into both the reduced-form results and structural analysis of this paper. On the reduced-form side, deepening our understanding of the income expectations-formation process within and across consumers is important. On the structural side, the agents in our setting have the same degree of extrapolation but the degree of extrapolation is likely heterogeneous across consumers.²⁵ Studying more advanced macroeconomic models that allow for this heterogeneity and assessing its quantitative implications is an interesting avenue for future research.

²⁵For subjective macroeconomic expectations, cross-sectional variation has been documented based on cognition (D’Acunto et al., 2019, 2021), socioeconomic status (Kuhnen and Miu, 2017; Das et al., 2020), and local experiences (Kuchler and Zafar, 2019), among others.

References

- Agarwal, S., S. Chomsisengphet, N. Mahoney, and J. Stroebel (2017, 07). Do Banks Pass through Credit Expansions to Consumers Who want to Borrow? *The Quarterly Journal of Economics* 133(1), 129–190.
- Agarwal, S., C. Liu, and N. Souleles (2007). The reaction of consumer spending and debt to tax rebates-evidence from consumer credit data. *Journal of Political Economy* 115(6), 986–1019.
- Ameriks, J., J. Briggs, A. Caplin, M. Lee, M. D. Shapiro, and C. Tonetti (2020, January). Older americans would work longer if jobs were flexible. *American Economic Journal: Macroeconomics* 12(1), 174–209.
- Ameriks, J., J. Briggs, A. Caplin, M. D. Shapiro, and C. Tonetti (2020). Long-term-care utility and late-in-life saving. *Journal of Political Economy* 128(6), 2375–2451.
- Ameriks, J., G. Kézdi, M. Lee, and M. D. Shapiro (2020). Heterogeneity in expectations, risk tolerance, and household stock shares: The attenuation puzzle. *Journal of Business & Economic Statistics* 38(3), 633–646.
- Attanasio, O. P. and G. Weber (2010, September). Consumption and saving: Models of intertemporal allocation and their implications for public policy. *Journal of Economic Literature* 48(3), 693–751.
- Augenblick, N., E. Lazarus, and M. Thaler (2021). Overinference from weak signals and underinference from strong signals. *arXiv preprint arXiv:2109.09871*.
- Aydin, D. (2022, January). Consumption response to credit expansions: Evidence from experimental assignment of 45,307 credit lines. *American Economic Review* 112(1), 1–40.
- Ba, C., J. A. Bohren, and A. Imas (2022). Over-and underreaction to information. *Available at SSRN 4274617*.
- Bailey, M., E. Dávila, T. Kuchler, and J. Stroebel (2019). House price beliefs and mortgage leverage choice. *The Review of Economic Studies* 86(6), 2403–2452.
- Barberis, N. (2018). Psychology-based models of asset prices and trading volume. In *Handbook of behavioral economics: applications and foundations 1*, Volume 1, pp. 79–175. Elsevier.
- Barberis, N., R. Greenwood, L. Jin, and A. Shleifer (2018). Extrapolation and bubbles. *Journal of Financial Economics* 129(2), 203–227.
- Baron, M., E. Verner, and W. Xiong (2020, 10). Banking Crises Without Panics*. *The Quarterly Journal of Economics* 136(1), 51–113.
- Baron, M. and W. Xiong (2017, 01). Credit Expansion and Neglected Crash Risk*. *The Quarterly Journal of Economics* 132(2), 713–764.
- Baugh, B., I. Ben-David, H. Park, and J. A. Parker (2021, January). Asymmetric consumption smoothing. *American Economic Review* 111(1), 192–230.
- Bellifemine, M., A. Couturier, and S. H. Maasoum (2024). A distributional theory of household sentiment. *Technical Report*.
- Benjamin, D. J. (2019). Chapter 2 - errors in probabilistic reasoning and judgment biases. In B. D. Bernheim, S. DellaVigna, and D. Laibson (Eds.), *Handbook of Behavioral*

- Economics - Foundations and Applications 2*, Volume 2 of *Handbook of Behavioral Economics: Applications and Foundations 1*, pp. 69–186. North-Holland.
- Beutel, J. and M. Weber (2022). Beliefs and portfolios: Causal evidence. *Chicago Booth Research Paper* (22-08).
- Bianchi, F., C. Ilut, and H. Saijo (2021). Diagnostic business cycles. *NBER Working Paper* 28604.
- Blundell, R. and I. Preston (1998, 05). Consumption Inequality and Income Uncertainty*. *The Quarterly Journal of Economics* 113(2), 603–640.
- Bordalo, P., N. Gennaioli, R. L. Porta, and A. Shleifer (2019). Diagnostic expectations and stock returns. *The Journal of Finance* 74(6), 2839–2874.
- Bordalo, P., N. Gennaioli, and A. Shleifer (2018). Diagnostic expectations and credit cycles. *The Journal of Finance* 73(1), 199–227.
- Bordalo, P., N. Gennaioli, A. Shleifer, and S. J. Terry (2021, January). Real credit cycles. Working Paper 28416, National Bureau of Economic Research.
- Bordo, M., B. Eichengreen, D. Klingebiel, and M. S. Martinez-Peria (2001). Is the crisis problem growing more severe? *Economic Policy* 16(32), 52–82.
- Borio, C. and P. Lowe (2002, 07). Asset Prices, Financial and Monetary Stability: Exploring the Nexus. *BIS Working Papers* (114).
- Borio, C. and P. Lowe (2013). Systemic Banking Crises Database. *IMF Economic Review* 61(114), 225–270.
- Bornstein, G. and S. Indarte (2023). The impact of social insurance on household debt. Available at SSRN 4205719.
- Brunnermeier, M. K. and Y. Sannikov (2014, February). A macroeconomic model with a financial sector. *American Economic Review* 104(2), 379–421.
- Bucks, B. and K. Pence (2008). Do borrowers know their mortgage terms? *Journal of Urban Economics* 64(2), 218–233.
- Caplin, A., V. Gregory, E. Lee, S. Leth-Petersen, and J. Sæverud (2023, March). Subjective earnings risk. Working Paper 31019, National Bureau of Economic Research.
- Carroll, C. D. (1997, 02). Buffer-Stock Saving and the Life Cycle/Permanent Income Hypothesis*. *The Quarterly Journal of Economics* 112(1), 1–55.
- Chamberlain, G. (1982, June). Panel data. Working Paper 913, National Bureau of Economic Research.
- Chatterjee, S., D. Corbae, M. Nakajima, and J.-V. Ríos-Rull (2007). A quantitative theory of unsecured consumer credit with risk of default. *Econometrica* 75(6), 1525–1589.
- Chodorow-Reich, G., A. M. Guren, and T. J. McQuade (2023, 04). The 2000s Housing Cycle with 2020 Hindsight: A Neo-Kindlebergerian View. *The Review of Economic Studies* 91(2), 785–816.
- Chopra, F., C. Roth, and J. Wohlfart (2023). Home price expectations and spending: Evidence from a field experiment. Available at SSRN 4452588.
- Christelis, D., D. Georgarakos, T. Jappelli, and M. van Rooij (2020). Consumption uncertainty and precautionary saving. *Review of Economics and Statistics* 102(1), 148–161.

- Claessens, S., M. Ayhan Kose, and M. E. Terrones (2010). The global financial crisis: How similar? how different? how costly? *Journal of Asian Economics* 21(3), 247–264. The Financial Crisis of 2008-09: Origins, Issues, and Prospects.
- Coibion, O., D. Georgarakos, Y. Gorodnichenko, G. Kenny, and M. Weber (2024). The effect of macroeconomic uncertainty on household spending. *American Economic Review* 114(3), 645–677.
- Coibion, O., Y. Gorodnichenko, and M. Weber (2020, August). How did u.s. consumers use their stimulus payments? Working Paper 27693, National Bureau of Economic Research.
- Coibion, O., Y. Gorodnichenko, and M. Weber (2022). Monetary policy communications and their effects on household inflation expectations. *Journal of Political Economy* 130(6), 1537–1584.
- Colarieti, R., P. Mei, and S. Stantcheva (2024, March). The how and why of household reactions to income shocks. Working Paper 32191, National Bureau of Economic Research.
- D’Acunto, F., D. Hoang, M. Paloviita, and M. Weber (2019). Cognitive abilities and inflation expectations. In *AEA Papers and Proceedings*, Volume 109, pp. 562–66.
- D’Acunto, F., D. Hoang, M. Paloviita, and M. Weber (2021). Cognitive constraints and the transmission of economic policies. Technical report, National Bureau of Economic Research.
- D’Acunto, F., D. Hoang, M. Paloviita, and M. Weber (2023). Iq, expectations, and choice. *The Review of Economic Studies* 90(5), 2292–2325.
- D’Acunto, F., D. Hoang, and M. Weber (2022). Managing households’ expectations with unconventional policies. *The Review of Financial Studies* 35(4), 1597–1642.
- D’Acunto, F., U. Malmendier, and M. Weber (2021). Gender roles produce divergent economic expectations. *Proceedings of the National Academy of Sciences* 118(21), e2008534118.
- D’Acunto, F. and M. Weber (2024). Why survey-based subjective expectations are meaningful and important. *Annual Review of Economics*.
- Dang, T. V., G. Gorton, and B. Holmström (2020). The information view of financial crises. *Annual Review of Financial Economics* 12(1), 39–65.
- Das, S., C. M. Kuhnen, and S. Nagel (2020). Socioeconomic status and macroeconomic expectations. *The Review of Financial Studies* 33(1), 395–432.
- DeFusco, A. A., B. Enriquez, and M. Yellen (2024, March). Wage garnishment in the united states: New facts from administrative payroll records. *American Economic Review: Insights* 6(1), 38–54.
- DellaVigna, S. (2009, June). Psychology and economics: Evidence from the field. *Journal of Economic Literature* 47(2), 315–72.
- Dempsey, K. and F. Ionescu (2023). Borrowing premia in unsecured credit markets. Technical report.
- Fagereng, A., M. B. Holm, and G. J. Natvik (2021, October). Mpc heterogeneity and household balance sheets. *American Economic Journal: Macroeconomics* 13(4), 1–54.
- Ferland, E., C. M. Kuhnen, G. Li, and I. Ben-David (2023). Extrap-

- relative uncertainty and household economic behavior. Available at SSRN: <https://ssrn.com/abstract=3293399> or <http://dx.doi.org/10.2139/ssrn.3293399>.
- Fulford, S. L. (2015). How important is variability in consumer credit limits? *Journal of Monetary Economics* 72, 42–63.
- Fuster, A., G. Kaplan, and B. Zafar (2020, 11). What Would You Do with \$500? Spending Responses to Gains, Losses, News, and Loans. *The Review of Economic Studies* 88(4), 1760–1795.
- Ganong, P., D. Jones, P. J. Noel, F. E. Greig, D. Farrell, and C. Wheat (2020, July). Wealth, race, and consumption smoothing of typical income shocks. Working Paper 27552, National Bureau of Economic Research.
- Ganong, P. and P. Noel (2019, July). Consumer spending during unemployment: Positive and normative implications. *American Economic Review* 109(7), 2383–2424.
- Ganong, P., P. Noel, C. Patterson, J. Vavra, and A. Weinberg (2024). Earnings instability.
- Gertler, M. and N. Kiyotaki (2010). Chapter 11 - financial intermediation and credit policy in business cycle analysis. Volume 3 of *Handbook of Monetary Economics*, pp. 547–599. Elsevier.
- Giglio, S., M. Maggiori, J. Stroebel, and S. Utkus (2021, May). Five facts about beliefs and portfolios. *American Economic Review* 111(5), 1481–1522.
- Gorton, G. and G. Ordoñez (2014, February). Collateral crises. *American Economic Review* 104(2), 343–78.
- Greenwood, R., S. G. Hanson, and L. J. Jin (2019, April). Reflexivity in credit markets. Working Paper 25747, National Bureau of Economic Research.
- Greenwood, R., S. G. Hanson, A. Shleifer, and J. A. Sørensen (2020, June). Predictable financial crises. Working Paper 27396, National Bureau of Economic Research.
- Gross, D. B. and N. S. Souleles (2002, 02). Do Liquidity Constraints and Interest Rates Matter for Consumer Behavior? Evidence from Credit Card Data. *The Quarterly Journal of Economics* 117(1), 149–185.
- Guiso, L., T. Jappelli, and L. Pistaferri (2002). An empirical analysis of earnings and employment risk. *Journal of Business & Economic Statistics* 20(2), 241–253.
- Guvenen, F. (2007, June). Learning your earning: Are labor income shocks really very persistent? *American Economic Review* 97(3), 687–712.
- Haaland, I., C. Roth, and J. Wohlfart (2021). Designing information provision experiments. *Journal of Economic Literature*.
- He, Z. and A. Krishnamurthy (2019, October). A macroeconomic framework for quantifying systemic risk. *American Economic Journal: Macroeconomics* 11(4), 1–37.
- Jappelli, T. and L. Pistaferri (2010). The consumption response to income changes. *Annual Review of Economics* 2(1), 479–506.
- Jordà, O., M. Schularick, and A. M. Taylor (2013). When credit bites back. *Journal of Money, Credit and Banking* 45(s2), 3–28.
- Kaplan, G., K. Mitman, and G. L. Violante (2020). The housing boom and bust: Model meets evidence. *Journal of Political Economy* 128(9), 3285–3345.
- Kaplan, G. and G. L. Violante (2014). A model of the consumption response to fiscal stimulus payments. *Econometrica* 82(4), 1199–1239.

- Kaplan, G. and G. L. Violante (2022). The marginal propensity to consume in heterogeneous agent models. *Annual Review of Economics* 14, 747–775.
- Keane, M. P. and D. E. Runkle (1998). Are financial analysts’ forecasts of corporate profits rational? *Journal of Political Economy* 106(4), 768–805.
- Kiyotaki, N. and J. Moore (1997). Credit cycles. *Journal of Political Economy* 105(2), 211–248.
- Kluender, R., S. Indarte, U. Malmendier, and M. Stepner (2024). The roles of beliefs versus constraints in consumption decisions of low-income workers. *Working Paper*.
- Krishnamurthy, A. and T. Muir (2017, September). How credit cycles across a financial crisis. Working Paper 23850, National Bureau of Economic Research.
- Krueger, D., E. Malkov, and F. Perri (2023, November). How do households respond to income shocks? Working Paper 31894, National Bureau of Economic Research.
- Kuchler, T., M. Piazzesi, and J. Stroebel (2022, April). Housing market expectations. Working Paper 29909, National Bureau of Economic Research.
- Kuchler, T. and B. Zafar (2019). Personal experiences and expectations about aggregate outcomes. *The Journal of Finance* 74(5), 2491–2542.
- Kueng, L. (2018, 06). Excess Sensitivity of High-Income Consumers*. *The Quarterly Journal of Economics* 133(4), 1693–1751.
- Kuhnen, C. M. and A. C. Miu (2017). Socioeconomic status and learning from financial information. *Journal of Financial Economics* 124(2), 349–372.
- Lee, S. C. and P. Maxted (2023, May). Household borrowing and mpc’s in heterogeneous-agent models. Technical report.
- L’Huillier, J.-P., S. R. Singh, and D. Yoo (2023). Incorporating diagnostic expectations into the new keynesian framework. *Review of Economic Studies* (forthcoming).
- Li, W. (2019). Public liquidity and financial crises. Available at SSRN: <https://ssrn.com/abstract=3175101>.
- Li, Z., S. Van Nieuwerburgh, and W. Renxuan (2023, July). Understanding rationality and disagreement in house price expectations. Working Paper 31516, National Bureau of Economic Research.
- Livshits, I., J. MacGee, and M. Tertilt (2007, March). Consumer bankruptcy: A fresh start. *American Economic Review* 97(1), 402–418.
- López-Salido, D., J. C. Stein, and E. Zakrajšek (2017, 05). Credit-Market Sentiment and the Business Cycle*. *The Quarterly Journal of Economics* 132(3), 1373–1426.
- Manski, C. F. (2004). Measuring expectations. *Econometrica* 72(5), 1329–1376.
- Maxted, P. (2023, 03). A Macro-Finance Model with Sentiment. *The Review of Economic Studies*, rdad023.
- Mian, A. and A. Sufi (2014). What explains the 2007-2009 drop in employment? *Econometrica* 82(6), 2197–2223.
- Mian, A. and A. Sufi (2018, August). Finance and business cycles: The credit-driven household demand channel. *Journal of Economic Perspectives* 32(3), 31–58.
- Mian, A., A. Sufi, and E. Verner (2017). Household debt and business cycles worldwide. *The Quarterly Journal of Economics* 132(4), 1755–1817.
- Mian, A., A. Sufi, and E. Verner (2020). How does credit supply expansion affect the

- real economy? the productive capacity and household demand channels. *The Journal of Finance* 75(2), 949–994.
- Olafsson, A. and M. Pagel (2018, 04). The Liquid Hand-to-Mouth: Evidence from Personal Finance Management Software. *The Review of Financial Studies* 31(11), 4398–4446.
- Parker, J. A., N. S. Souleles, D. S. Johnson, and R. McClelland (2013, October). Consumer spending and the economic stimulus payments of 2008. *American Economic Review* 103(6), 2530–53.
- Pistaferri, L. (2001, 08). Superior Information, Income Shocks, and the Permanent Income Hypothesis. *The Review of Economics and Statistics* 83(3), 465–476.
- Reinhart, C. M. and K. S. Rogoff (2009, May). The aftermath of financial crises. *American Economic Review* 99(2), 466–72.
- Reuven, G. and K. J. Lansing (2010). Global household leverage, house prices, and consumption. *FRBSF Economic Letter*.
- Roth, C. and J. Wohlfart (2020). How do expectations about the macroeconomy affect personal expectations and behavior? *Review of Economics and Statistics* 102(4), 731–748.
- Rozsypal, F. and K. Schlafmann (2023, October). Overpersistence bias in individual income expectations and its aggregate implications. *American Economic Journal: Macroeconomics* 15(4), 331–71.
- Souleles, N. (2004). Expectations, heterogeneous forecast errors, and consumption: Micro evidence from the michigan consumer sentiment surveys. *Journal of Money, Credit and Banking* 36(1), 39–72.
- Weber, M., F. D’Acunto, Y. Gorodnichenko, and O. Coibion (2022). The subjective inflation expectations of households and firms: Measurement, determinants, and implications. *Journal of Economic Perspectives* 36(3), 157–184.
- Yin, X. (2022, October). Learning in the Limit: Income Inference from Credit Extension. Technical report.
- Zinman, J. (2009). Debit or credit? *Journal of Banking & Finance* 33(2), 358–366.

Figure 1. Survey Design

This figure describes the timeline of the survey waves run in each of three year (2020, 2021, 2023).

For each year of 2020, 2021, 2023

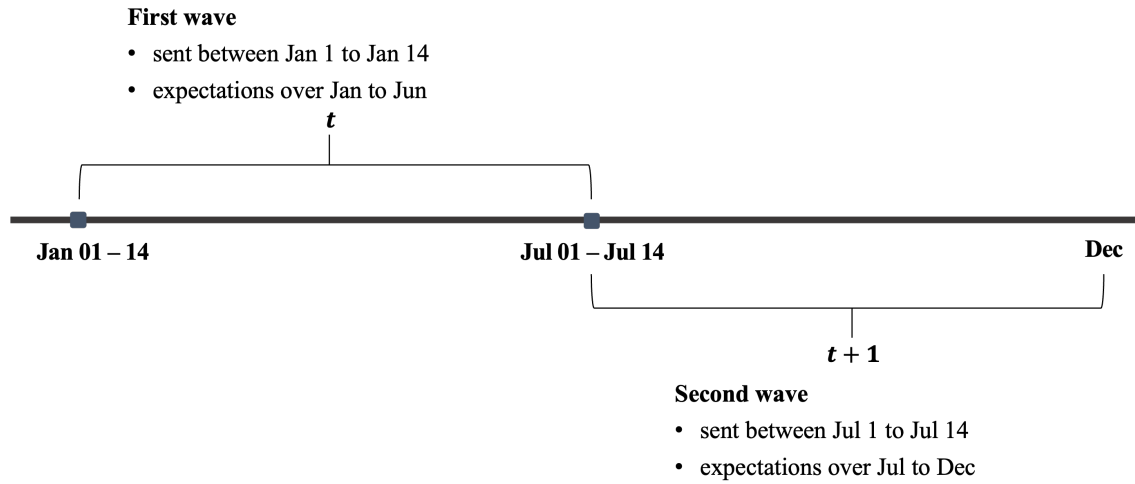


Figure 2. Subjective and Objective Income Shocks

This figure plots binned-scatter plots of subjective income shocks (measured from the surveys) against objective income innovations (measured based on equation (1)) at the individual level. Panel A plots the raw values. In Panel B, variables are residualized by income changes in period t , age, age-squared, education degree, expected log standard deviation of expected income growth in period $t + 1$, number of hours worked per week, log income in period t , occupation industry fixed effects, and city \times year fixed effects. All variables are winsorized at 1% level within each wave.

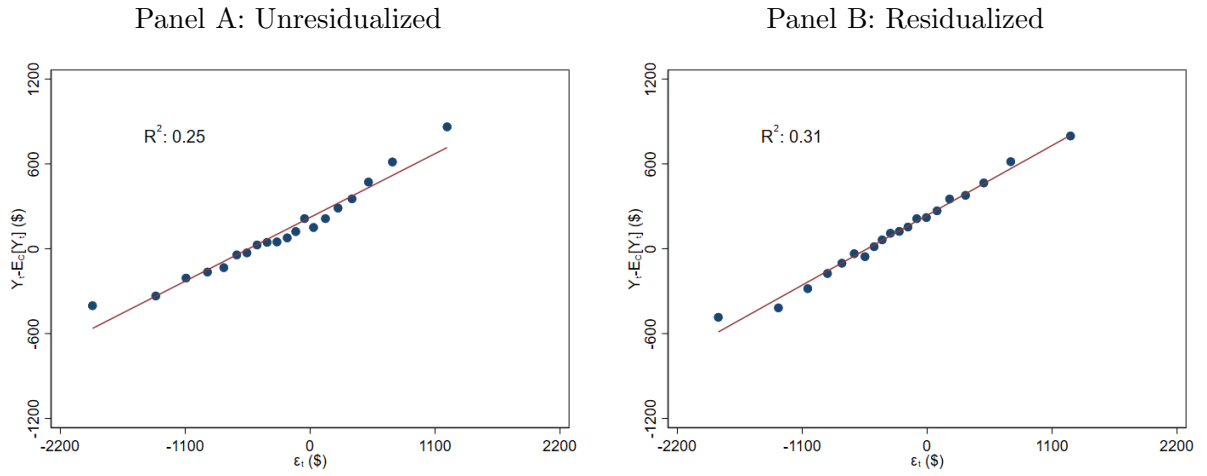


Figure 3. Mistakes in Subjective Income Expectations

Panel A is a binned scatter plot of consumers' ex-post income realizations against ex-ante income expectations. Y_{t+1} is consumers' realized income in period $t + 1$, measured in US dollars. $E_C[Y_{t+1}]$ is the income level consumers expect to realize in period $t + 1$ based on survey question 8, measured in US dollars. Panel B plots within-individual income changes against changes in income expectations. Panels C and D are histograms of the individual-level difference between income expectations and ex-post income realizations. All variables are winsorized at the 1% level within each wave.

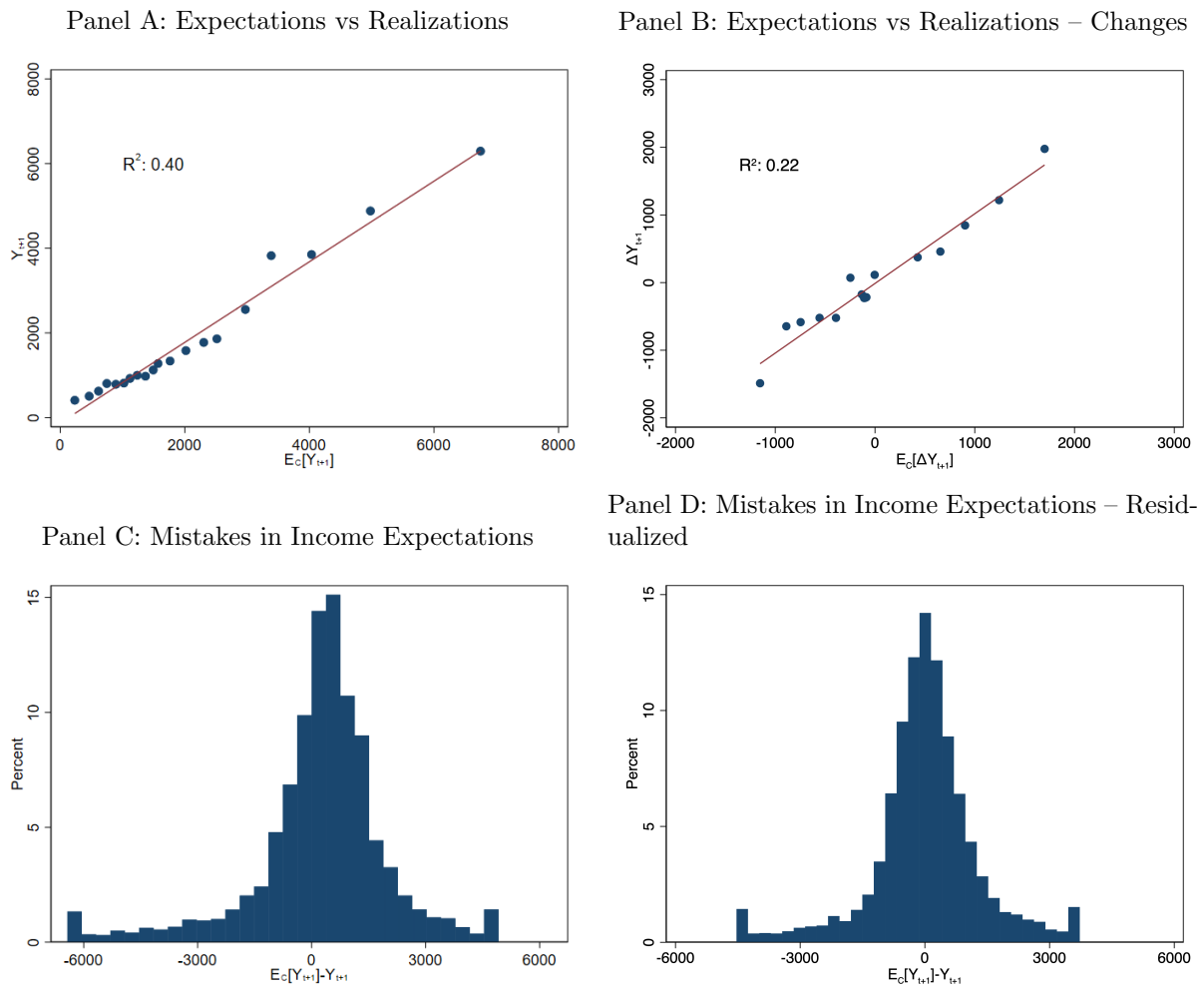


Figure 4. Income Shocks and Mistakes in Subjective Income Expectations

This figure plots binned scatter plots of mistakes in subjective income expectations against income shocks. Y_t is consumers' realized income in period t measured in US dollars. $E_C[Y_t]$ is the income level consumers expect to realize in period t based on survey question 8 answered in period $t - 1$ measured in US dollars. $\epsilon_{i,t}$ is consumers' income innovation measured based on equation (1). Panels A and C plot the raw values. In panels B and D, variables are residualized by income changes in period t , age, age-squared, education degree, expected log standard deviation of expected income growth in period $t + 1$, number of hours worked per week, log income in period t , occupation industry fixed effects, and city \times year fixed effects. All variables are winsorized at 1% level within each wave.

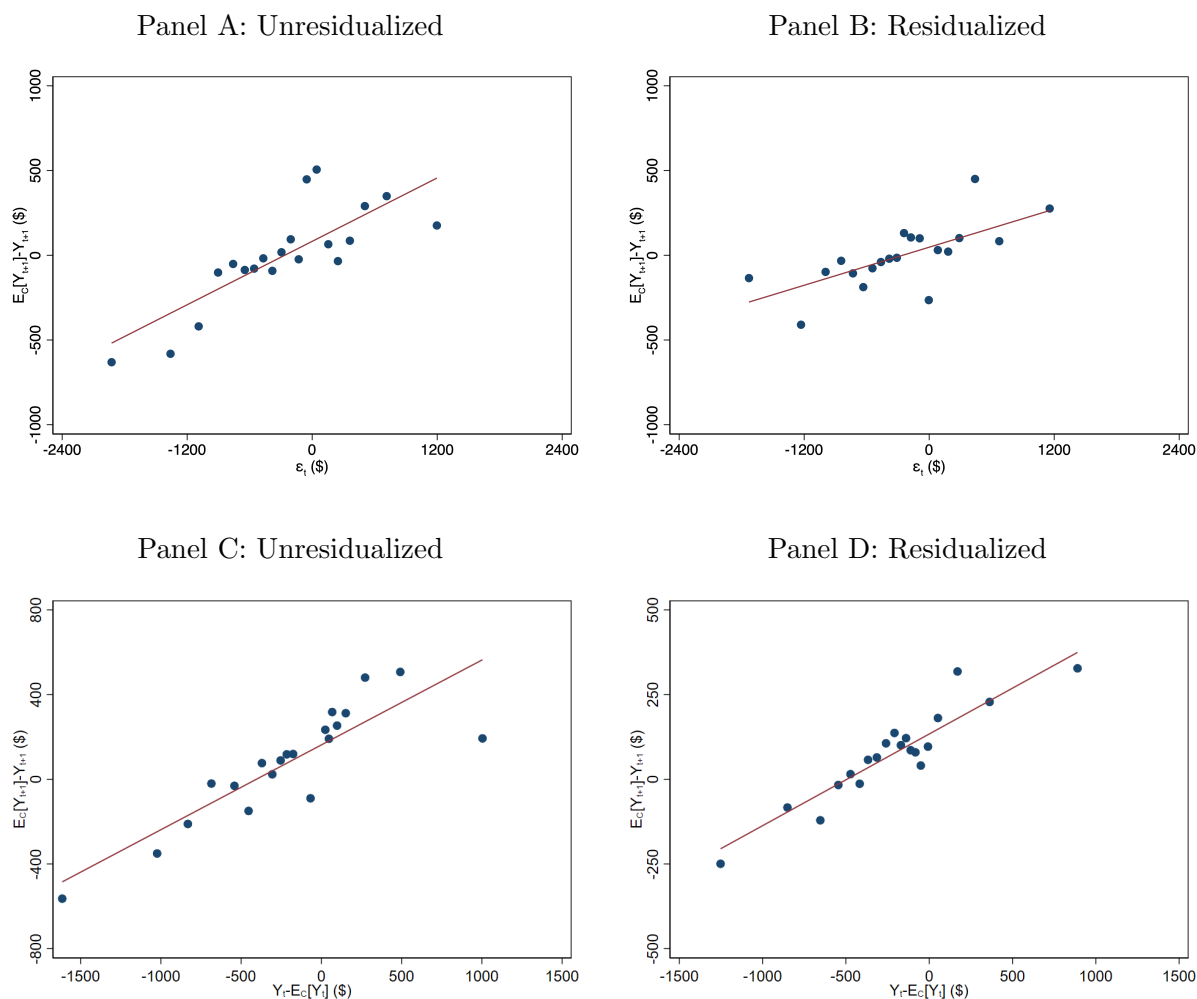


Figure 5. Dynamics of Consumption and Debt Accumulation by Mistakes in Subjective Income Expectations

This figure reports the coefficient estimates attached to mistakes in subjective income expectations on cumulative spending and debt choices. Estimation is based on the specification in equation (4). In Panel A, the outcome variable is total non-durable spending. In Panel B, it is the stock of unsecured debt. The red solid lines report the estimates for consumers with positive expectation errors and the blue dashed lines for consumers with negative expectation errors. The width of the shaded regions is twice the size of estimated standard errors.

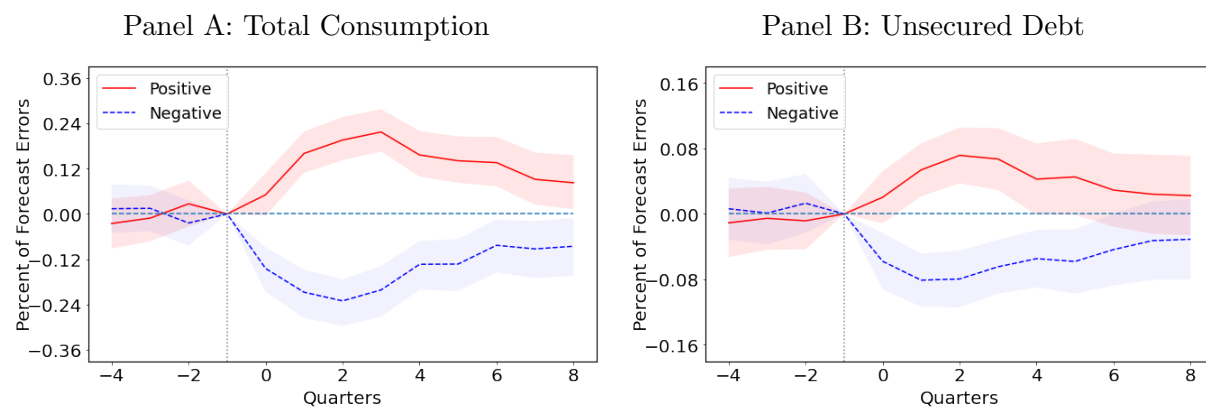


Figure 6. Structural Model: Impulse Responses after Transitory Income Shocks

This figure plots the IRFs based on the versions of our model with either diagnostic expectations (DE) or rational expectations (RE) after a series of transitory income shocks at the half-year frequency. The simulation is based on 20,000 individuals initially at the stochastic steady state. Starting at the stochastic steady state, each individual receives a 3-year sequence of positive shocks that result in a 3-standard-deviation cumulative shock over three years. The top left panel plots the income shocks. The top right panel plots the updates in expected log income under both forms of expectations, where $o_{i,t} = \kappa(1 + \theta)(y_{i,t} - \alpha - \hat{z}_{i,t-1})$. The bottom four panels are the percentage difference in income expectations, consumption, borrowing, and default rate relative to when no shocks are introduced. In all panels, red solid lines plot the results when $\theta = 1.65$ (DE) and blue dashed lines when $\theta = 0$ (RE).

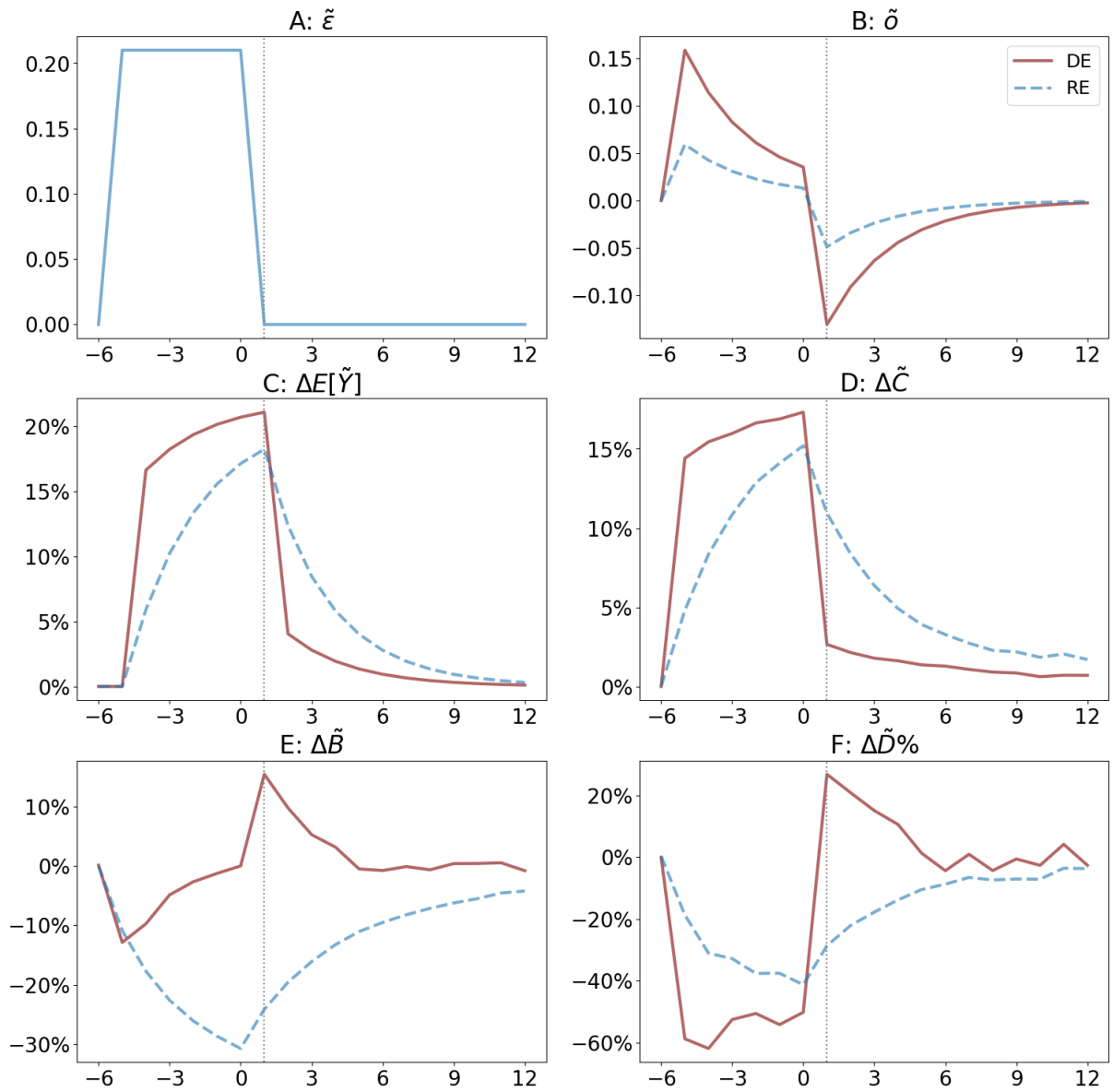


Figure 7. Simulating the Global Financial Crisis in the US

This figure simulates the 2008-2009 Global Financial Crisis in the US based on a series of transitory income shocks at the half-year frequency. The simulation is based on 20,000 individuals. Starting at the stochastic steady state, each individual receives a series of transitory shocks as indicated in the top panel. In the bottom two panels, the red solid lines present the simulation when $\theta = 1.65$ (diagnostic expectations, DE); the blue dotted lines plot the results when $\theta = 0$ (rational expectations, RE); the black dashed lines plots the results in the data. In the middle panel, \bar{B}/\bar{Y} is the ratio of total debt from credit cards and other credit lines on GDP, divided by the labor share. \bar{B}/\bar{Y} is detrended over 2003–2012. In the bottom panel, $\bar{p}(\text{default})$ is consumers' debt delinquency rate multiplied by the proportion of consumers who hold positive debt.

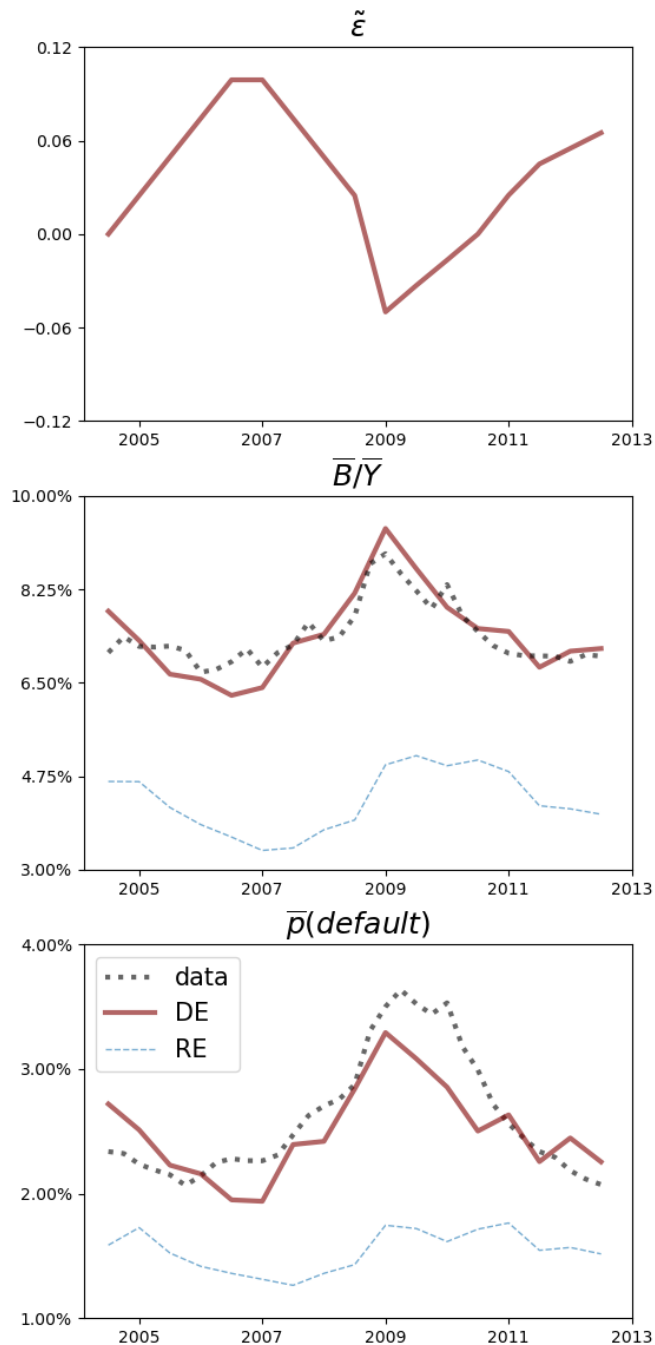


TABLE I. Summary Statistics

This table reports summary statistics for our sample. Spending is the average monthly spending from the checking accounts plus spending from credit card accounts. Income is the average monthly income inflow computed as discussed in the main text. Savings is the average level of savings. Limit is the credit card limit retrieved for each consumer from the Chinese credit registry. Debt is the average interest-incurring credit card debt for each consumer. Debt|Debt > 0 is the average interest-incurring credit card debt for those who hold any debt. Income - E[Income] is the income shock computed based on equation (1). $E_C[\text{Income}]$ is based on the answers from survey questions Q6. $SD_C[\Delta \log \text{Income}]$ is the subjective standard deviations of expected income growth. All level variables are measured in US dollars and winsorized at the 1% level within each wave.

	(1)	(2)	(3)	(4)	(5)	(6)
	Mean	SD	p25	Median	p75	N
Age	39.28	9.00	29	39	48	10497
Female	0.52	0.50	0	1	1	10497
Spending	1245.31	1588.01	292.73	761.68	1451.23	10497
Income	1929.88	2139.66	706.97	1187.36	2407.33	10497
Savings	17817.82	29855.22	1645.41	4261.38	19893.30	10497
Limit	12473.88	17380.76	2031.25	3906.25	15625.00	10497
Debt	1014.86	1655.40	0.00	0.00	1510.59	10497
Debt Debt > 0	2315.65	1830.48	651.09	1863.64	3717.34	4631
$E_C[\text{Income}]$	2011.21	1587.33	900.01	1504.89	2723.21	10497
$SD_C[\Delta \log \text{Income}]$	0.36	0.37	0.14	0.24	0.45	10497
$E_C[\text{Income}] - \text{Income}$	212.57	1730.91	-432.87	389.85	1103.88	10497

TABLE II. Income Shocks and Mistakes in Subjective Income Expectations

This table reports the results for estimating a set of ordinary least-squares regressions. $E_C[Y_{t+1}]$ is the expected level of income (measured in US dollar thousands) in period $t + 1$. $E_C[\Delta Y_t]$ is the difference between reported expected income (measured in US dollar thousands) in period t and realized income in period $t - 1$. $E[Y_t]$ is the expected income at time t estimated based on equation (1). $SD(E_C[\Delta \log Y_t])$ is the standard deviation of expected income growth in period t based on survey questions 4, 5, and 6 and assuming that income growth follows a Triangular distribution. *College* is an indicator for whether the consumers' highest degree is college and above. $\log Hours_t$ is the number of hours the consumer reports working per week in period t . $\log Y_t$ and $\log L_t$ are the log monthly income and log credit card limit in period t . Columns (1)–(3) use the whole sample. Columns (4)–(6) consider observations in 2020, 2021, and 2023 separately. Because we only have one observation per individual per year, these specifications include no individual fixed effects. All variables are winsorized at the 1% level within each wave.

	(1) All	(2) All	(3) All	(4) 2020	(5) 2021	(6) 2023
	$E_C[Y_{t+1}] - Y_{t+1}$	$E_C[Y_{t+1}] - Y_{t+1}$	$E_C[Y_{t+1}] - Y_{t+1}$	$E_C[Y_{t+1}] - Y_{t+1}$	$E_C[Y_{t+1}] - Y_{t+1}$	$E_C[Y_{t+1}] - Y_{t+1}$
$Y_t - E[Y_t]$	0.457*** (0.045)	0.397*** (0.055)	0.405*** (0.091)	0.443*** (0.062)	0.396*** (0.055)	0.334*** (0.069)
$E_C[\Delta Y_t]$		-0.028 (0.055)	0.041 (0.089)	-0.029 (0.055)	-0.011 (0.053)	-0.033 (0.061)
<i>Age</i>		-0.029*** (0.011)		-0.023* (0.011)	-0.028** (0.012)	-0.033*** (0.011)
Age^2		0.000** (0.000)		0.000** (0.000)	0.000** (0.000)	0.000** (0.000)
<i>Female</i>		-0.155*** (0.047)		-0.101** (0.047)	-0.159** (0.072)	-0.169*** (0.053)
<i>College</i>		0.068* (0.039)		0.063 (0.044)	0.093* (0.051)	0.070* (0.040)
$\log Hours_t$		-0.012 (0.072)	-0.125 (0.090)	-0.006 (0.081)	-0.013 (0.087)	-0.013 (0.079)
$SD(E_C[\Delta \log Y_t])$		-0.019* (0.009)	-0.012 (0.015)	-0.027** (0.014)	-0.023 (0.013)	-0.013 (0.014)
$\log Y_{t-1}$		-0.235*** (0.029)	-0.319** (0.129)	-0.203*** (0.035)	-0.251*** (0.048)	-0.259*** (0.041)
$\log L_{t-1}$		0.002 (0.011)	0.031 (0.031)	0.009 (0.033)	0.012 (0.023)	0.011 (0.023)
N	10,497	10,497	10,497	4,377	3,103	3,017
Industry FE	No	Yes	No	Yes	Yes	Yes
City \times Round FE	No	Yes	Yes	No	No	No
City FE	No	No	No	Yes	Yes	Yes
Individual FE	No	No	Yes	–	–	–
R^2	3.38%	12.19%	60.75%	9.76%	11.78%	16.83%

Standard Errors Clustered at City Level in Parentheses

* p<0.10 ** p<0.05 *** p<0.01

TABLE III. Income Shocks and Mistakes in Subjective Income Expectations—Heterogeneity Analysis

This table reports the results for estimating a set of ordinary least-squares regressions. $E_C[Y_{t+1}]$ is the expected level of income (measured in US dollar thousands) in period $t + 1$. $E_C[\Delta Y_t]$ is the difference between reported expected income (measured in US dollar thousands) in period t and realized income in period $t - 1$. $E[Y_t]$ is the expected income at time t estimated based on equation (1). SD_H is a dummy variable that equals 1 if the standard deviation of expected income growth in period t based on survey questions 4, 5, and 6 and assuming that income growth follows a Triangular distribution is above the median. Age_H is a dummy variable that equals 1 if the consumer's age is above 38. $College$ is an indicator for whether the consumers' highest degree is college and above. Gov is an indicator for whether the consumers works in government-related position. $1_{\{Y_t - E[Y_t] < 0\}}$ is a dummy variable indicating if the income shock is negative. Controls include log number of hours worked, log income and credit limit in period $t - 1$, subjective uncertainty about income in period t , and the variables in levels. All variables are winsorized at the 1% level within each wave.

	(1)	(2)	(3)	(4)	(5)	(6)
	$E_C[Y_{t+1}] - Y_{t+1}$	$E_C[Y_{t+1}] - Y_{t+1}$	$E_C[Y_{t+1}] - Y_{t+1}$	$E_C[Y_{t+1}] - Y_{t+1}$	$E_C[Y_{t+1}] - Y_{t+1}$	$E_C[Y_{t+1}] - Y_{t+1}$
$Y_t - E[Y_t]$	0.509*** (0.074)	0.323*** (0.095)	0.545*** (0.065)	0.549*** (0.143)	0.409*** (0.111)	0.387*** (0.097)
$Y_H \times (Y_t - E[Y_t])$	-0.233** (0.114)					
$SD_H \times (Y_t - E[Y_t])$		0.167** (0.084)				
$Age_H \times (Y_t - E[Y_t])$			-0.207* (0.113)			
$Degree_H \times (Y_t - E[Y_t])$				-0.198 (0.136)		
$Gov \times (Y_t - E[Y_t])$					-0.399*** (0.118)	
$1_{\{Y_t - E[Y_t] < 0\}} \times (Y_t - E[Y_t])$						-0.096 (0.189)
N	10,497	10,497	10,497	10,497	10,497	10,497
Controls	Yes	Yes	Yes	Yes	Yes	Yes
City \times Round FE	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	60.74%	60.77%	60.77%	60.76%	60.77%	60.76%

Standard Errors Clustered at City Level in Parentheses

* $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$

TABLE IV. Income Shocks and Subjective Income Uncertainty

This table reports the results for estimating a set of ordinary least-squares regressions. $E_C[Y_{t+1}]$ is the expected level of income (measured in US dollar thousands) in period $t + 1$. $E_C[\Delta Y_t]$ is the difference between reported expected income (measured in US dollar thousands) in period t and realized income in period $t - 1$. $E[Y_t]$ is the expected income at time t estimated based on equation (1). $abs(Y_t - E[Y_t])$ is the absolute value of income surprises in period t . $1_{\{Y_t - E[Y_t] < 0\}}$ is an indicator variable equal to 1 if the income surprise in period t is negative. $SD(E[\Delta \log Y_{t+1}])$ is the standard deviation of the income growth in period $t + 1$ based on survey questions 4, 5, and 6 assuming income growth follows a Triangular distribution. $\log Hours_t$ is the number of hours the consumer reports working per week in period t . $\log Y_{t-1}$ and $\log L_{t-1}$ are the log monthly income and log credit card limit in period $t - 1$. All variables are winsorized at the 1% level within each wave.

	(1)	(2)	(3)	(4)
	$\Delta SD(E[\Delta \log Y_{t+1}])$	$\Delta SD(E[\Delta \log Y_{t+1}])$	$\Delta SD(E[\Delta \log Y_{t+1}])$	$\Delta SD(E[\Delta \log Y_{t+1}])$
$abs(Y_t - E[Y_t])$	0.046 (0.029)	0.282*** (0.073)	0.308*** (0.071)	0.182** (0.084)
$1_{\{Y_t - E[Y_t] < 0\}}$				-0.050 (0.052)
$abs(Y_t - E[Y_t]) \times 1_{\{Y_t - E[Y_t] < 0\}}$				0.174* (0.090)
$E_C[\Delta Y_t]$			-0.094*** (0.016)	-0.101*** (0.020)
$\log Hours_t$			-0.011 (0.044)	-0.014 (0.045)
$\log Y_{t-1}$			-0.442*** (0.141)	-0.445*** (0.141)
$\log L_{t-1}$			0.002 (0.012)	0.001 (0.011)
N	10,497	10,497	10,497	10,497
Industry FE	No	Yes	Yes	Yes
City \times Round FE	No	Yes	Yes	Yes
Individual FE	No	Yes	Yes	Yes
R^2	0.05%	65.69%	66.07%	66.09%

Standard Errors Clustered at City Level in Parentheses

* $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$

TABLE V. Mistakes in Subjective Income Expectations and Choice: Consumption, Debt, and Defaults

This table reports the results for estimating a set of ordinary least-squares regressions. ΔC_t is the difference in the monthly average consumption between t and $t - 1$ (measured in US dollar thousands). ΔB_{t+1} is the difference in the average interest-incurring unsecured debt between $t + 1$ and t (measured in US dollar thousands). $E_C[Y_{t+1}]$ is the expected level of income (measured in US dollar thousands) in period $t + 1$. $E_C[\Delta Y_t]$ is the difference between reported expected income (measured in US dollar thousands) in period t and realized income in period $t - 1$. $SD(E_C[\Delta \log Y_t])$ is the standard deviation of expected income growth in period t based on survey questions 4, 5, and 6 and assuming that income growth follows a Triangular distribution. $\log Hours_t$ is the number of hours the consumer reports working per week in period t . $\log Y_{t-1}$ and $\log L_{t-1}$ are log monthly income and log credit card limit in period $t - 1$. All variables are winsorized at the 1% level within each wave.

	(1)	(2)	(3)	(4)	(5)	(6)
	ΔC_t	ΔC_t	ΔB_{t+1}	ΔB_{t+1}	$Default_{t+1}$	$Default_{t+1}$
$E_C[Y_{t+1}] - Y_{t+1}$	0.206*** (0.025)	0.201*** (0.036)	0.087*** (0.010)	0.074*** (0.014)	0.735*** (0.082)	0.915*** (0.138)
$E_C[\Delta Y_t]$		0.016 (0.085)		-0.039 (0.026)		-0.688 (0.555)
$\log Hours_t$		-0.169 (0.137)		-0.138*** (0.050)		-5.902** (2.748)
$SD(E_C[\Delta \log Y_t])$		0.011 (0.030)		-0.011 (0.011)		0.267 (0.307)
$\log Y_{t-1}$		-0.143 (0.123)		-0.022 (0.050)		-16.485*** (3.905)
$\log L_{t-1}$		-0.031 (0.025)		0.012 (0.012)		0.007 (0.291)
N	10,497	10,497	10,497	10,497	10,497	10,497
Industry FE	No	Yes	No	Yes	No	Yes
City \times Round FE	No	Yes	No	Yes	No	Yes
Individual FE	No	Yes	No	Yes	No	Yes
R^2	3.27%	54.76%	2.91%	61.00%	0.74%	53.65%

Standard Errors Clustered at City Level in Parentheses

* $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$

TABLE VI. Asymmetric Effects of Mistakes in Subjective Income Expectations

This table reports the results for estimating a set of ordinary least-squares regressions. ΔC_t is the difference in the monthly average consumption between t and $t - 1$ (measured in US dollar thousands). ΔB_{t+1} is the difference in the average interest-incurring unsecured debt between $t + 1$ and t (measured in US dollar thousands). $E_C[\Delta Y_t]$ is the difference between reported expected income (measured in US dollar thousands) in period t and realized income in period $t - 1$. $E[Y_t]$ is the expected income in time t estimated based on equation (1). $1_{\{Y_t - E[Y_t] < 0\}}$ is an indicator variable equal to 1 if the income surprise in period t is negative. $SD(E[\Delta \log Y_{t+1}])$ is the standard deviation of the income growth in period $t + 1$ based on survey questions 4, 5, and 6 assuming income growth follows a Triangular distribution. $\log Hours_t$ is the number of hours the consumer reports working per week in period t . $\log Y_{t-1}$ and $\log L_{t-1}$ are the log monthly income and log credit card limit in period $t - 1$. All variables are winsorized at the 1% level within each wave.

	(1)	(2)	(3)	(4)
	ΔC_t	ΔC_t	ΔB_{t+1}	ΔB_{t+1}
$Y_t - E[Y_t]$	0.352*** (0.078)	0.204* (0.113)	0.167*** (0.039)	0.130* (0.066)
$1_{\{Y_t - E[Y_t] < 0\}}$		-0.040 (0.109)		0.067 (0.045)
$(Y_t - E[Y_t]) \times 1_{\{Y_t - E[Y_t] < 0\}}$		0.198* (0.118)		0.118* (0.069)
$E_C[\Delta Y_t]$	0.067 (0.093)	0.067 (0.092)	-0.009 (0.033)	-0.007 (0.032)
$\log Hours_t$	-0.177 (0.143)	-0.182 (0.144)	-0.139*** (0.050)	-0.144*** (0.049)
$SD(E_C[\Delta \log Y_t])$	0.008 (0.030)	0.009 (0.030)	-0.012 (0.012)	-0.011 (0.011)
$\log Y_{t-1}$	-0.230* (0.119)	-0.231* (0.119)	-0.065 (0.052)	-0.059 (0.051)
$\log L_{t-1}$	-0.024 (0.026)	-0.024 (0.026)	0.014 (0.012)	0.014 (0.012)
N	10,497	10,497	10,497	10,497
Industry FE	Yes	Yes	Yes	Yes
City \times Round FE	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes
R^2	53.59%	53.60%	60.26%	60.28%

Standard Errors Clustered at City Level in Parentheses

* $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$

TABLE VII. Structural Model: Estimation

This table reports the estimated parameters of the structural model. Panel A presents the parameters estimated in the first stage and Panel B those estimated in the second stage based on SMM. Panel C reports the matched moments and Panel D the moments not targeted directly. w/c is the average wealth-consumption ratio. $p(\text{default})$ is the proportion of defaults. w/y is the average wealth-income ratio. $\text{median}(w/y)$ is the median of the wealth-income ratio. low liq share is the fraction of consumers with savings lower than three months of income. top 5% liq share is the fraction of savings held by the top 5% individuals in the model. Estimation of moments in the model is based on a simulation of 5,000 individuals with 500 periods, after a burning period of 1,000 periods for the distribution to reach the steady state. Estimation of moments in the data is based on a random 5% of active customers in the bank's database. Model moments are trimmed at savings to average income ratio larger than 10.

Panel A		Panel B			Panel C			Panel D		
First-Stage Parameters		Second-Stage Parameters			Targeted Moments			Not Targeted Moments		
	Estimates		Estimates	S.E.		Data	Model		Data	Model
	(1)		(2)	(3)		(4)	(5)		(6)	(7)
α	0	γ	2.511	(0.267)	w/c	0.846	0.846	w/y	1.351	1.263
ρ	0.970	χ	24.453	(1.201)	$p(\text{default})$	2.31%	2.31%	$\text{median}(w/y)$	0.731	0.780
σ_ν	0.150							top 5% liq share	31.38%	31.07%
σ_ϵ	0.420							debtor %	29.89%	31.89%
β	0.975									
ν	0.050									
r_b	0.055									
r_s	0.014									
θ	1.680									

TABLE VIII. Structural Model: Linking Income Shocks, Mistakes in Income Expectations, and Consumption-Debt Choices

In Panel A (data), $E_C[Y_{t+1}]$ is the expected level of income (measured in US dollar thousands) in period $t + 1$ based on survey question 8 answered by the consumer in period t . In Panels B and C (model), $E_C[Y_{t+1}]$ is the expected level of income at time t under diagnostic expectations with $\theta = 1.65$ and $\theta = 0$, respectively. In Panel A, $E[Y_{t+1}]$ is the expected income in time t estimated based on equation (1). In Panel B and C, $E[Y_{t+1}]$ is the expected income when $\theta = 1.68$ and $\theta = 0$, respectively. ΔC_t is the changed in total consumption, ΔB_{t+1} is the next-period changes in total liquid debt. *Default* is a dummy variable that is equal to 100 if consumer i has a 90-day delinquency, and zero otherwise. All variables in Panel A are winsorized at the 1% level within each wave. For columns (1) in Panels B and C, data is based on 100,000 periods of simulation with a burning periods of 100. For columns (2) to (4) in Panels B and C, the analysis is based on a simulation of 20,000 individuals with 1,000 periods, after a burning period of 100 periods. Simulated data are right-trimmed at a savings-to-average-income ratio larger than 8.13, which is the highest value in the data.

	(1) $E_C[Y_{t+1}] - Y_{t+1}$	(2) ΔC_t	(3) ΔB_{t+1}	(4) $Default_{t+1}$
Panel A: Data				
$Y_t - E[Y_t]$	0.405*** (0.091)			
$E_C[Y_{t+1}] - Y_{t+1}$		0.201*** (0.025)	0.074*** (0.014)	0.918*** (0.138)
N	10,497	10,497	10,497	10,497
Panel B: $\theta = 1.680$				
$Y_t - E[Y_t]$	0.404*** (0.004)			
$E_C[Y_{t+1}] - Y_{t+1}$		0.251*** (0.000)	0.071*** (0.000)	0.931*** (0.003)
N	999,998	8,433,591	8,433,591	8,433,591
Panel C: $\theta = 0$				
$Y_t - E[Y_t]$	-0.040*** (0.003)			
$E_C[Y_{t+1}] - Y_{t+1}$		-0.002*** (0.000)	0.029*** (0.000)	0.249*** (0.001)
N	999,998	8,902,876	8,902,876	8,902,876

Standard Errors in Parentheses
 * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$

Online Appendix:

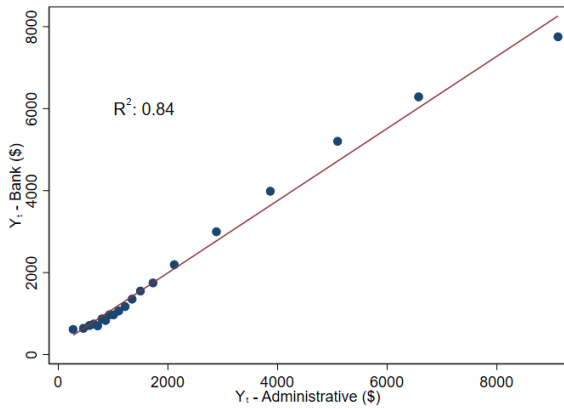
Subjective Income Expectations and Household Debt Cycles

Francesco D'Acunto, Michael Weber, Xiao Yin

Figure A.1. Comparing Computed Individual-level Income and Registry-based Income

Panel A in this figure is a binned scatter plot that compares the income values computed by the bank based on the transaction-level data the bank accesses and following the steps described in section *B*. of the paper and the registry income values reported by the same consumers in our sample to the Chinese tax authority, which can be accessed through one-to-one matching of individual tax identifiers. Panel B compares consumer answers from survey question 1 and the income from the bank at the same period.

Panel A: Comparisons of Income Measures



Panel B: Survey Answers and Bank Measures

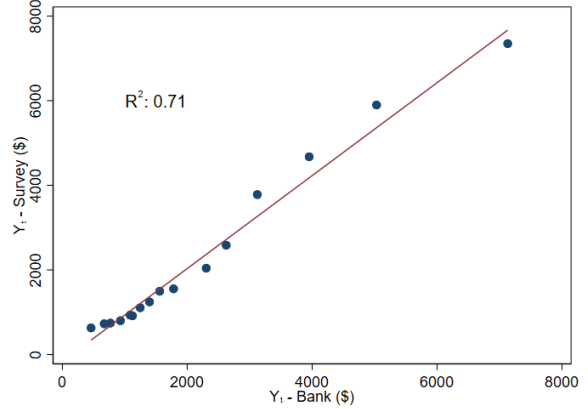


Figure A.2. Income Shocks and Mistakes in Subjective Income Expectations by Industry

This figure presents the binned-scatter plots of misbeliefs in future income vs past income shocks. The four panels respectively plots individuals working in government, education, household service, and business service.

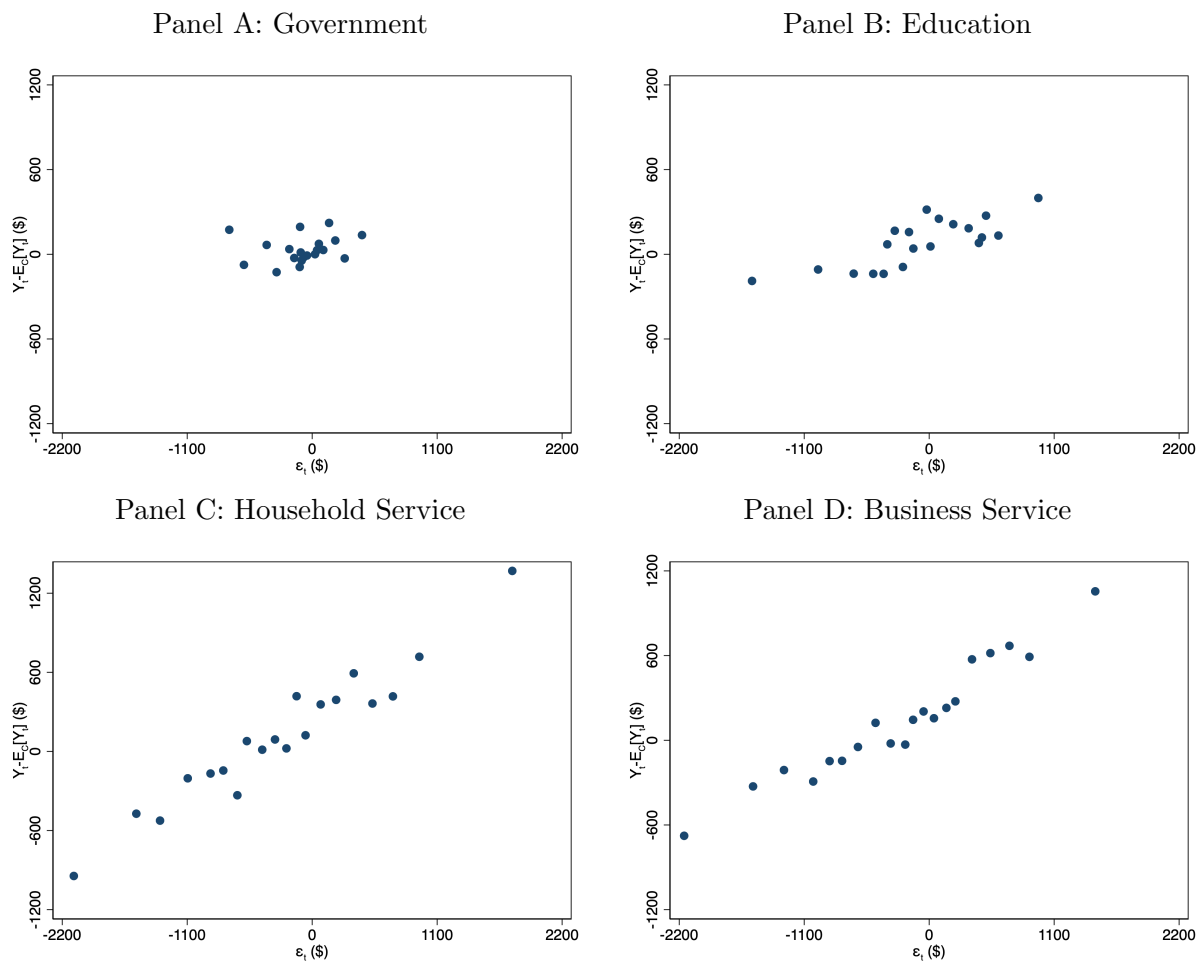


Figure A.3. Equilibrium Distribution of Saving

This figure is a histogram of the equilibrium distribution of consumer saving to average income ratio. Saving is equal to total resource available $a_{i,t}$ at the beginning of the period minus the total consumption $c_{i,t}$. The plot is based on simulated steady-state distribution based on a simulation of 5,000 individuals with 500 periods, after a burning period of 1,000 periods for the distribution to reach the steady state. The distribution is right trimmed at saving to average income ratio larger than 20.

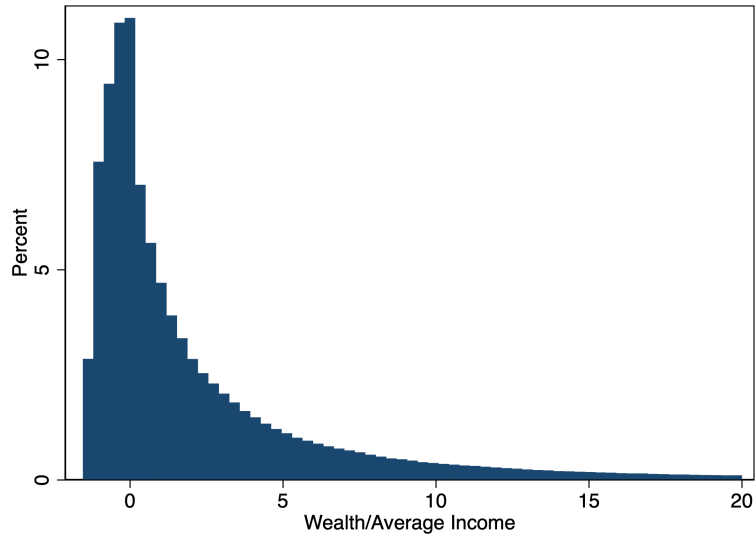


Figure A.4. Impulse Responses after Transitory Income Shocks

This figure gives the IRFs after a series of persistent income shocks at the half-year frequency. The simulation is based on 20,000 individuals initially at the stochastic steady states. Starting at the stochastic steady states, each individual receives a 3-year sequence of positive shocks that result in a 3 standard deviation cumulative shock over three years. The top left panel gives the introduced transitory income shocks. The top right panel gives the updates in expected log income, where $o_{i,t} = \kappa(1+\theta)(y_{i,t} - \alpha - \hat{z}_{i,t-1})$. The bottom four panels are respectively the percentage difference in income expectations, consumption, borrowing, and default probability relative to when shocks are not introduced. The red solid lines present the results when $\theta = 1.641$; the blue dashed lines present the results when $\theta = 0$.

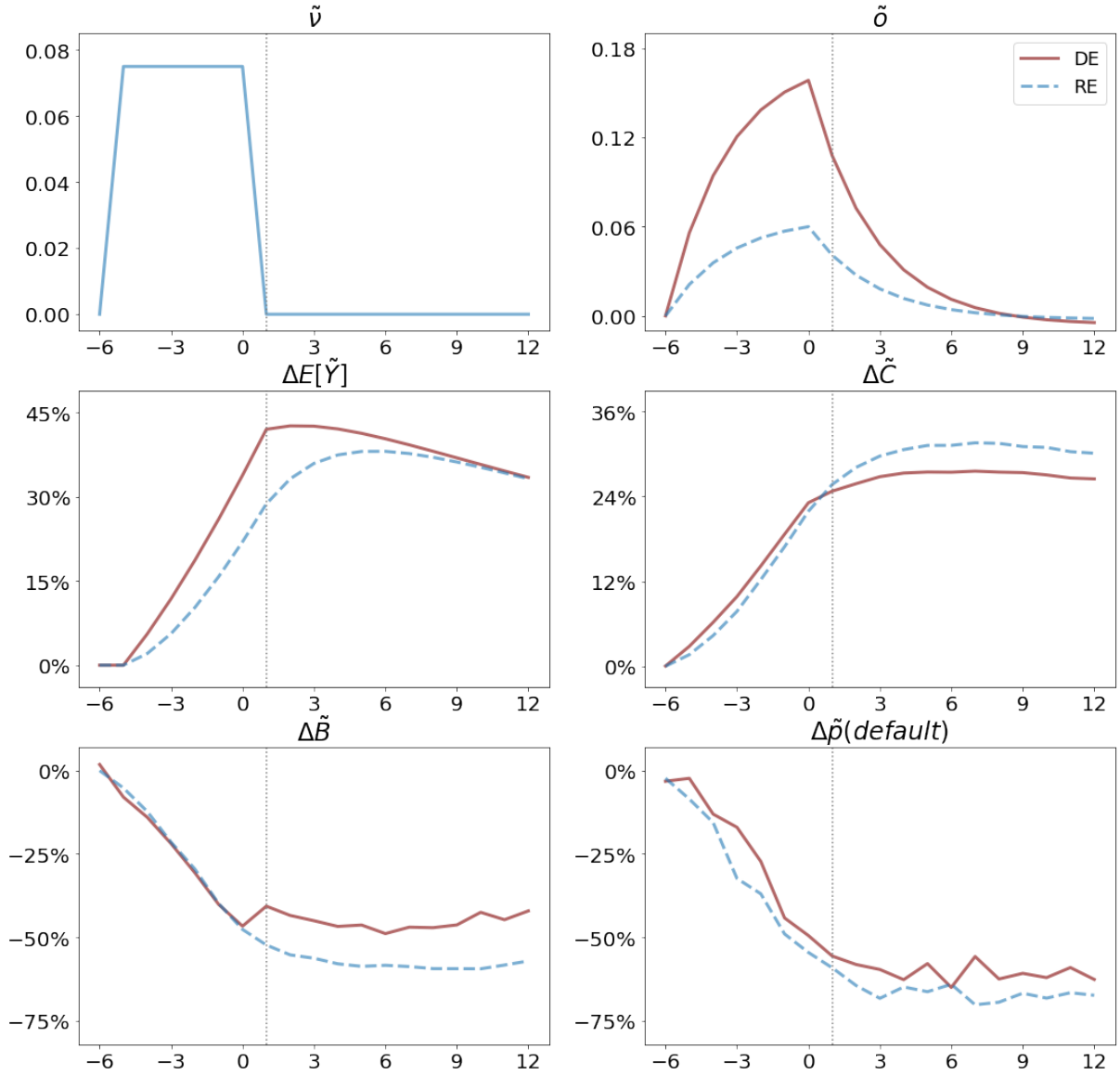


TABLE A.1. Data Comparison

Panel A gives the summary statistics of the main sample. Panel B gives the summary statistics of a 5% random sample of all consumers at the bank. All variables are winsorized at 1% level.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Panel A: Survey Sample					Panel B: Whole Sample				
	Mean	SD	p25	Median	p75	Mean	SD	p25	Median	p75
Age	39.28	9.00	29	39	48	41.23	10.23	29	41	50
Female	0.52	0.50	0	1	1	0.50	0.50	0	0	1
Spending	1245.31	1588.01	292.73	761.68	1451.23	1397.84	2103.67	347.07	846.23	1863.81
Income	2067.28	2124.95	722.30	1227.24	2459.59	2258.03	2632.17	843.32	1467.32	2866.33
Saving	17817.82	29855.22	1645.41	4261.38	19893.30	20121.23	38661.00	1787.90	7762.33	24378.90
Limit	12473.88	17380.76	2031.25	8593.75	23437.50	135671.09	23241.89	2812.50	10156.25	28125.00
Debt	1014.86	1655.40	0.00	0.00	1510.59	1003.69	2355.40	0.00	0.00	2123.66
Debt Debt > 0	2315.65	1830.48	651.09	1863.64	3717.34	2247.84	2543.23	451.75	1932.52	4021.21

TABLE A.2. Shock Correlation

$E_C[Y_{t+1}]$ is the expected level of income in period $t + 1$. $E[Y_t]$ is the expected income at time t estimated from (1). Y_t is income in period t . Δc_t is log changes in total spending from period $t - 1$ to t . $SD(E_C[\Delta \log Y_t])$ is the standard deviation of expected income growth in period t based on survey questions 4, 5, and 6 assuming income growth follows a Triangular distribution. $\log Hours_t$ is the log number of hours the customers usually work every week in period t . $\log Y_{t-1}$ and $\log L_{t-1}$ are respectively log monthly income and log credit card limit in period $t - 1$. All variables are winsorized at 1% level by each wave.

	(1) $E[Y_{t+1}] - Y_{t+1}$	(2) $E[Y_{t+1}] - Y_{t+1}$	(3) $E[Y_{t+1}] - Y_{t+1}$	(4) $E[Y_{t+1}] - Y_{t+1}$
$Y_t - E[Y_t]$	0.057 (0.096)	0.044 (0.090)		
$Y_t - E_C[Y_t]$			0.063 (0.089)	0.077 (0.098)
$E_C[\Delta Y_t]$		-0.022 (0.051)		-0.013 (0.033)
$\log Hours_t$		-0.015 (0.053)		-0.014 (0.051)
$SD(E_C[\Delta \log Y_t])$		-0.019 (0.014)		-0.024* (0.012)
$\log Y_{t-1}$		-0.105 (0.065)		-0.100* (0.055)
$\log L_{t-1}$		0.009 (0.010)		0.009 (0.012)
N	10,497	10,497	10,497	4,377
City \times Round FE	No	Yes	No	Yes
Individual FE	No	Yes	No	Yes
R^2	1.06%	29.34%	0.75%	24.34%

Standard Errors Clustered at City Level in Parentheses

* $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$

TABLE A.3. Income Shocks and Mistakes in Subjective Income Expectations—Subjective Income Shocks

$E_C[Y_{t+1}]$ is the expected level of income (\$ thousands) in period $t + 1$. $E_C[\Delta Y_t]$ is the difference between reported expected income (measured in US dollar thousands) in period t and realized income in period $t - 1$. $SD(E_C[\Delta \log Y_t])$ is the standard deviation of expected income growth in period t based on survey questions 4, 5, and 6 assuming income growth follows a Triangular distribution. *Degree* is the consumers' highest degree earned. $\log Hours_t$ is the log number of hours the customers usually work every week in period t . $\log Y_{t-1}$ and $\log L_{t-1}$ are respectively log monthly income and log credit card limit in period $t - 1$. All variables are winsorized at 1% level by each wave.

	(1)	(2)	(3)	(4)	(5)	(6)
	$E_C[Y_{t+1}] - Y_{t+1}$	$E_C[Y_{t+1}] - Y_{t+1}$	ΔC_t	ΔC_t	ΔB_{t+1}	ΔB_{t+1}
$Y_t - E_C[Y_t]$	0.369*** (0.071)	0.368*** (0.074)	0.165*** (0.024)	0.372*** (0.083)	0.231* (0.137)	0.082* (0.048)
$E_C[\Delta Y_t]$					0.065 (0.057)	-0.023 (0.035)
$1_{\{Y_t - E[Y_t] < 0\}}$					0.391* (0.210)	0.163** (0.080)
$(Y_t - E[Y_t]) \times 1_{\{Y_t - E[Y_t] < 0\}}$		0.019 (0.081)	-0.016 (0.028)	0.053 (0.085)	0.055 (0.085)	-0.014 (0.028)
$\log Hours$		-0.116 (0.099)	-0.134*** (0.050)	-0.164 (0.148)	-0.171 (0.145)	-0.133*** (0.048)
$SD(E_C[\Delta \log Y_{t+1}])$		0.000 (0.015)	-0.011 (0.012)	0.010 (0.031)	0.010 (0.031)	-0.012 (0.013)
$\log Y_{t-1}$		-0.285** (0.122)	-0.051 (0.048)	-0.201 (0.130)	-0.193 (0.129)	-0.048 (0.047)
$\log L_{t-1}$		0.024 (0.028)	0.012 (0.012)	-0.029 (0.026)	-0.029 (0.026)	0.013 (0.012)
N	10,467	10,467	10,467	10,467	10,467	10,467
City \times Round FE	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	61.04%	61.09%	60.56%	53.90%	53.96%	60.61%

Standard Errors Clustered at City \times Year Level in Parentheses

* $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$

TABLE A.4. Income Shocks and Mistakes in Subjective Income Expectations—Alternative Specification

$E_C[Y_{t+1}]$ is the expected level of income in period $t + 1$ based on survey question 8 sent in period t . $E[Y_t]$ is the expected income at time t estimated from

$$y_{i,t+1} = \rho_{j,k,a} y_{i,t} + \Gamma X_{i,t} + \epsilon_{i,t+1},$$

where $y_{i,t+1} = \log Y_{i,t+1}$. All variables are winsorized at 1% level by each wave.

	(1) All	(2) All	(3) All	(4) 2020	(5) 2021	(6) 2023
	$E_C[Y_{t+1}] - Y_{t+1}$	$E_C[Y_{t+1}] - Y_{t+1}$	$E_C[Y_{t+1}] - Y_{t+1}$	$E_C[Y_{t+1}] - Y_{t+1}$	$E_C[Y_{t+1}] - Y_{t+1}$	$E_C[Y_{t+1}] - Y_{t+1}$
$Y_t - E[Y_t]$	0.453*** (0.042)	0.389*** (0.053)	0.382*** (0.082)	0.457*** (0.089)	0.369*** (0.084)	0.339*** (0.112)
$E_C[\Delta Y_t]$		-0.035 (0.052)	0.035 (0.089)	0.012 (0.073)	-0.053 (0.078)	-0.132 (0.121)
<i>Age</i>		-0.025*** (0.011)				
<i>Age</i> ²		0.000** (0.000)				
<i>Female</i>		-0.159*** (0.042)				
<i>College</i>		0.067* (0.038)				
$\log Hours_t$		-0.012 (0.070)	-0.133 (0.093)	-0.019 (0.072)	-0.135 (0.101)	0.253 (0.182)
$SD(E_C[\Delta \log Y_t])$		-0.022 (0.018)	0.023 (0.038)	-0.003 (0.021)	-0.049* (0.024)	0.021 (0.055)
$\log Y_{t-1}$		-0.241*** (0.032)	-0.301** (0.130)	-0.233*** (0.049)	-0.270*** (0.049)	-0.217** (0.101)
$\log L_{t-1}$		0.007 (0.014)	0.034 (0.031)	0.011 (0.015)	0.019 (0.019)	0.055 (0.042)
N	10,497	10,497	10,497	4,377	3,103	3,017
Industry FE	No	Yes	No	Yes	Yes	Yes
City × Round FE	No	Yes	Yes	Yes	Yes	Yes
Individual FE	No	No	Yes	No	No	No
R^2	3.44%	12.57%	60.61%	9.99%	12.81%	17.83%

Standard Errors Clustered at City Level in Parentheses

* $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$

TABLE A.5. Income Shocks and Subjective Uncertainty—Robustness

$E_C[Y_{t+1}]$ is the expected level of income in period $t + 1$. $E[Y_t]$ is the expected income at time t estimated from (1). Y_t is income in period t . ΔC_t is log changes in total spending from period $t - 1$ to t . $SD(E_C[\Delta \log Y_t])$ is the standard deviation of expected income growth in period t based on survey questions 4, 5, and 6 assuming income growth follows a Triangular distribution. In columns (1) - (3), SD is calculated such that the most likely values divided by the previous-period income treated as the mode of the distribution. In columns (4) - (6), SD is calculated such that the mid-points of the lowest and highest values divided by the previous-period income are treated as the mode of the distribution. All variables are winsorized at 1% level by each wave.

	(1)	(2)	(3)	(4)	(5)	(6)
	Most Likely Values as Mode			Mid Points as Mode		
	$E_C[Y_{t+1}] - Y_{t+1}$	$\Delta SD(E[\Delta \log Y_{t+1}])$	$\Delta SD(E[\Delta \log Y_{t+1}])$	$E_C[Y_{t+1}] - Y_{t+1}$	$\Delta SD(E[\Delta \log Y_{t+1}])$	$\Delta SD(E[\Delta \log Y_{t+1}])$
$Y_t - E[Y_t]$	0.325*** (0.105)			0.326*** (0.090)		
$SD_H \times (Y_t - E[Y_t])$	0.176* (0.092)			0.156** (0.077)		
$abs(Y_t - E[Y_t])$		0.293*** (0.071)	0.165* (0.089)		0.316*** (0.070)	0.184** (0.081)
$1_{\{Y_t - E[Y_t] < 0\}}$			-0.043 (0.057)			-0.053 (0.052)
$abs(Y_t - E[Y_t]) \times 1_{\{Y_t - E[Y_t] < 0\}}$			0.173* (0.091)			0.179* (0.090)
N	10,497	10,497	10,497	10,497	10,497	10,497
Controls	Yes	Yes	Yes	Yes	Yes	Yes
City \times Round FE	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	60.79%	59.91%	64.43%	60.77%	66.00%	66.02%

Standard Errors Clustered at City Level in Parentheses

* $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$

TABLE A.6. Mistakes in Subjective Income Expectations, Cash Withdrawals, and Cross-Bank Transfers

$\Delta Withdraw_t$ is the changes in cash withdrawal between t and $t - 1$, $\Delta Transfer_t$ is the changes in the net transfers from this bank to other bank accounts. $E[Y_{t+1}]$ is the expected level of income (\$ thousand) in period $t + 1$ based on survey question 8 sent in period t . $E_C[\Delta Y_t]$ is the difference between reported expected income (measured in US dollar thousands) in period t and realized income in period $t - 1$. $SD(E_C[\Delta \log Y_t])$ is the standard deviation of expected income growth in period t based on survey questions 4, 5, and 6 assuming income growth follows a Triangular distribution. $\log Hours_t$ is the log number of hours the customers usually work every week in period t . $\log Y_{t-1}$ and $\log L_{t-1}$ are respectively log monthly income and log credit card limit in period $t - 1$. All variables are winsorized at 1% level by each wave.

	(1)	(2)	(3)	(4)
	$\Delta Withdraw_t$	$\Delta Withdraw_t$	$\Delta Transfer_t$	$\Delta Transfer_t$
$E_C[Y_{t+1}] - Y_{t+1}$	0.031 (0.064)	0.013 (0.055)	-0.020 (0.016)	-0.029 (0.031)
$E[\Delta Y_t]$		0.170** (0.064)		-0.072** (0.033)
<i>Age</i>		0.012* (0.006)		
<i>Age</i> ²		-0.000* (0.000)		
<i>Female</i>		-0.015** (0.007)		
<i>College</i>		0.020*** (0.004)		
$SD(E[\Delta \log Y_{t+1}])$		-0.027*** (0.009)		-0.032*** (0.007)
$\log Hours$		0.000 (0.001)		-0.033*** (0.008)
$\log Y_{t-1}$		0.025** (0.013)		0.006 (0.005)
$\log L_{t-1}$		-0.021* (0.012)		-0.011** (0.005)
N	10,497	10,497	10,497	10,497
City FE \times Round FE	No	No	Yes	Yes
Individual FE	No	No	Yes	Yes
R^2	0.27%	17.32%	56.32%	59.99%

Standard Errors Clustered at City Level in Parentheses

* $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$

TABLE A.7. Objective Unexpected Income Shocks and Choice—Longest Time Series Available

ΔC_t is the differences (in \$ thousands) in the monthly average consumption between t and $t - 1$. ΔB_{t+1} is the differences (in \$ thousands) in the end-of-period interest-incurring unsecured debt between $t + 1$ and t . $Default_{t+1}$ is an indicator for 90-day delinquency in $t + 1$. $E[Y_t]$ is estimated based on equation (1). ΔY_t is the changes (in \$ thousands) in income between period $t - 1$ and period t . Columns (1), (3), and (5) focus on the same sample period as that in the main analysis. Columns (2), (4), and (6) use a longer sample that includes all the data available for the same survey participants. All variables are winsorized at 1% level by each wave.

	(1)	(2)	(3)	(4)	(5)	(6)
	ΔC_t	ΔC_t	ΔB_{t+1}	ΔB_{t+1}	<i>Default</i>	<i>Default</i>
$Y_t - E[Y_t]$	0.356*** (0.078)	0.323*** (0.093)	0.169*** (0.040)	0.139*** (0.031)	3.657*** (1.160)	3.168*** (0.874)
$E[\Delta Y_t]$	0.069 (0.094)	0.069 (0.098)	-0.007 (0.032)	-0.023 (0.029)	-0.061 (0.530)	-0.197 (0.316)
$\log Y_{t-1}$	-0.222* (0.116)	-0.219*** (0.117)	-0.057 (0.052)	-0.051 (0.044)	-16.810*** (3.846)	-21.33*** (5.12)
$\log L_{t-1}$	-0.023 (0.026)	-0.021 (0.019)	0.015 (0.012)	0.007 (0.009)	0.091 (0.298)	0.184 (0.117)
N	10,497	46,348	10,497	46,348	10,497	46,348
City \times Round FE	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
Longer Sample	No	Yes	No	Yes	No	Yes
R^2	51.46%	44.96%	61.33%	53.29%	56.25%	48.33%

Standard Errors Clustered at City Level in Parentheses

* $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$

TABLE A.8. Income Shocks and Mistakes in Subjective Income Expectations—log scale

$E_C[y_t]$ is the log of expected level of income (\$ thousands) in period 1 based on survey question 8 sent in period 0. Δy_t is the changes in log income between period $t - 1$ and period t . $SD(E_C[\Delta \log Y_t])$ is the standard deviation of expected income growth in period t based on survey questions 4, 5, and 6 assuming income growth follows a Triangular distribution. $\log Hours_t$ is the log number of hours the customers usually work every week in period t . $\log Y_{t-1}$ and $\log L_{t-1}$ are respectively log monthly income and log credit card limit in period $t - 1$. All variables are winsorized at 1% level by each wave.

	(1)	(2)	(3)	(4)	(5)	(6)
	$E_C[y_{t+1}] - y_{t+1}$	$E_C[y_{t+1}] - y_{t+1}$	Δc_t	Δc_t	<i>Default</i>	<i>Default</i>
$y_t - E[y_t]$	0.405*** (0.024)	0.211*** (0.039)				
$E_C[y_{t+1}] - y_{t+1}$			0.200*** (0.015)	0.178*** (0.020)	1.093*** (0.149)	1.452*** (0.246)
$\log Hours$		-0.071 (0.072)		-0.065 (0.109)		-4.153* (2.237)
$SD(E_C[\Delta \log Y_{t+1}])$		0.015 (0.023)		0.055 (0.040)		0.144 (0.241)
$\log Y_{t-1}$		-0.141* (0.075)		-0.062 (0.065)		-17.149*** (4.452)
$\log L_{t-1}$		-0.012 (0.022)		-0.023 (0.025)		-0.103 (0.277)
N	9,845	9,845	9,845	9,845	9,845	9,845
Industry FE	Yes	No	Yes	No	Yes	No
City \times Round FE	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	No	Yes	No	Yes	No	Yes
R^2	4.50%	62.94%	2.32%	54.58%	0.61%	53.80%

Standard Errors Clustered at City Level in Parentheses

* $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$

TABLE A.9. Elasticity of Total Spending to Income Shocks

y_t is log income in period t . Δc_t is log changes in total spending from period $t - 1$ to t . $SD(E_C[\Delta \log Y_t])$ is the standard deviation of expected income growth in period t based on survey questions 4, 5, and 6 assuming income growth follows a Triangular distribution. $\log Hours_t$ is the log number of hours the customers usually work every week in period t . $\log Y_{t-1}$ and $\log L_{t-1}$ are respectively log monthly income and log credit card limit in period $t - 1$. All variables are winsorized at 1% level by each wave.

	(1)	(2)	(3)	(4)
	Δc_t	Δc_t	Δc_t	Δc_t
$y_t - E[y_t]$	0.233*** (0.023)	0.236*** (0.023)	0.232*** (0.023)	0.185*** (0.022)
$E[y_t] - y_{t-1}$		-0.106*** (0.023)	-0.098*** (0.022)	-0.058 (0.042)
$\log Hours$		0.034 (0.050)	0.068 (0.054)	-0.070 (0.104)
$SD(E_C[\Delta \log Y_{t+1}])$		0.015 (0.012)	0.021* (0.013)	0.079* (0.044)
$\log Y_{t-1}$		-0.102*** (0.020)	-0.075*** (0.022)	-0.105 (0.067)
$\log L_{t-1}$		-0.002 (0.010)	0.001 (0.010)	-0.023 (0.021)
N	10,442	10,442	10,442	10,442
Industry FE	No	No	Yes	No
City \times Round FE	No	No	Yes	Yes
Individual FE	No	No	No	Yes
R^2	1.79%	2.80%	5.59%	54.16%

Standard Errors Clustered at City Level in Parentheses

* $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$

B: Household Consumption Survey

Please read the following information carefully:

To better understand the households' consumption behaviors, we selected a certain number of active users to participate in a survey. The survey is expected to take about 5 minutes. If you choose to take the survey, you will be awarded a 10-CNY Red Pocket.

The data will be analyzed by third-party research scholars for scientific research purposes and will not be evaluated by this bank. We will not disclose participants' identifiable information in any respect. We will not, to any extent, change the types of financial products we provide, including credit scores, credit limits, deposit rates, etc., based on participants' answers. Therefore, please answer based on your actual opinion.

Please confirm if you would like to participate and complete the survey.

- Yes
- No

1. What was your total income over the past 12 months? _____
2. How many banks or financial platforms do you usually use for daily transactions?
 - (a) 0
 - (b) 1
 - (c) 2
 - (d) 3 or more
3. How many hours do you usually work per week? _____

For the following three questions, we would like to ask you about your income expectations over the next six months. Please consider the case if you do not have a job change.

Note: income includes wages, salaries, bonuses, commission, etc., excluding earnings from financial investment.

4. What would be the lowest possible level of total income you believe you would get over the next 6 months?

5. What would be the highest possible level of total income you believe you would get over the next 6 months?

6. What would your total income most likely be over the next 6 months?

For the following three questions, we would like to ask you about your income expectations over the next 12 months. Please consider the case if you do not have a job change.¹

7. What would be the lowest possible level of total income you believe you would get over the next 12 months?

8. What would be the highest possible level of total income you believe you would get over the next 12 months?

9. What would your total income most likely be over the next 12 months?

¹Questions 7 - 9 were asked in 2023.

For the following three questions, please answer if you are currently investing in the stock market.

10. What would be the lowest amount of income you believe you could earn from investing in the financial market in total over the next 6 months?

Note: use negative number for a negative return. _____

11. What would be the highest amount of income you believe you could earn from investing in the financial market in total over the next 6 months?

Note: use negative number for a negative return. _____

12. What would be the total amount of income you would most likely earn from investing in the financial market in the next 6 months?

Note: use negative number for a negative return. _____

C: Transitory and Persistent Shocks

In this section, we separately study the effects of transitory and persistent income shocks on forecast errors.² Our strategies follow Pistaferri (2001). Specifically, let log income, $y_{i,t}$, follow

$$\begin{aligned} y_{i,t} &= Z_{i,t}\Theta + \Gamma_i\Pi + p_{i,t} + \epsilon_{i,t} \\ p_{i,t} &= p_{i,t-1} + \xi_{i,t}, \end{aligned}$$

where Γ_i is the time-invariant component and $Z_{i,t}$ the time-variant component. The log growth in income can be summarized as

$$\Delta y_{i,t} = \Delta Z_{i,t}\Theta + \xi_{i,t} + \Delta\epsilon_{i,t}.$$

Following Pistaferri (2001), $\Delta Z_{i,t}\Theta = \gamma_0 + \gamma_1 age_{i,t}$. Then with subjective expectation data and under the assumption of rational expectations, the transitory and persistent shocks can be retrieved with

$$\begin{aligned} \epsilon_{i,t} &= -E_{i,t}[\Delta y_{i,t+1}] + (\gamma_0 + \gamma_1 age_{i,t+1}) \\ \xi_{i,t} &= \Delta y_{i,t+1} - E_{i,t-1}[\Delta y_{i,t}] - \epsilon_{i,t} - \gamma_1. \end{aligned}$$

Note that this decomposition strategy requires the assumption that subjective expectations satisfy rational expectations, which is violated in our setting. Instead, we use both objective income expectations as retrieved from (1) and subjective income expectations as elicited from the surveys to perform the decomposition.

Table C.1 gives the variance decomposition of the persistent and transitory shocks. as shown, more than three-quarters of the variations of unexpected income growth comes from transitory shocks.

In Table C.2, we regress log forecast errors on current transitory and persistent shocks. Panels A and B respectively gives results using objective and subjective expectations. Similar to the main analysis, using objective and subjective expectations yields similar conclusions. In both panels, both transitory and permanent shocks lead to higher forecast errors in the future. However, transitory shocks in general yield larger forecast errors.

²We thank Deniz Aydin for encouraging this analysis.

Table C.1. Variance Decomposition of Persistent and Transitory Income Shocks

$\Delta y_t - E[\Delta y_t]$ is the unexpected income growth. ξ is the persistent income shock. ϵ is the transitory income shock. In Panel A, expectations are based on the objective income expectations (1). In Panel B, expectations are based on the subjective income expectations elicited from the surveys.

	(1) $Var(\Delta y_t - E[\Delta y_t])$	(2) $Var(\xi)$	(3) $Var(\epsilon)$	(4) $2 \times Cov(\xi, \epsilon)$
Panel A: Objective				
Level	0.498	0.101	0.376	0.021
Percent	100.00%	20.28%	75.50%	4.22%
Panel B: Subjective				
Level	0.596	0.133	0.504	-0.041
Percent	100.00%	22.35%	84.49%	-6.85%

Table C.2. Persistent and Transitory Income Shocks and Mistakes in Subjective Income Expectations

The RHS variables across all columns are the log forecast errors in the next period, $E_C[y_{t+1}] - y_{t+1}$. $E_C[y_t]$ is the log of expected level of income (\$ thousands) in period 1 based on survey question 8 sent in period 0. ξ is the persistent income shock. ϵ is the transitory income shock. $SD(E_C[\Delta \log Y_t])$ is the standard deviation of expected income growth in period t based on survey questions 6, 7, and 8 assuming income growth follows a Triangular distribution. $\log Hours_t$ is the log number of hours the customers usually work every week in period t . $\log Y_{t-1}$ and $\log L_{t-1}$ are respectively log monthly income and log credit card limit in period $t - 1$. All variables are winsorized at 1% level by each wave.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: Objective Expectations				Panel B: Subjective Expectations			
ϵ	0.331*** (0.019)	0.329*** (0.022)			0.329*** (0.062)	0.341*** (0.054)		
ξ			0.048 (0.031)	0.078** (0.034)			0.123* (0.062)	0.107* (0.061)
$\log Hours$		-0.055 (0.071)		-0.094 (0.071)		-0.085 (0.067)		-0.087 (0.070)
$SD(E_C[\Delta \log Y_{t+1}])$		-0.001 (0.030)		0.081*** (0.030)		0.051* (0.029)		0.060** (0.030)
$\log Y_{t-1}$		-0.049 (0.080)		-0.143 (0.087)		-0.189** (0.084)		-0.134 (0.084)
$\log L_{t-1}$		-0.016 (0.016)		-0.014 (0.015)		-0.013 (0.016)		-0.014 (0.015)
N	9,888	9,888	9,888	9,888	9,888	9,888	9,888	9,888
City \times Round FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	63.76%	63.78%	62.62%	62.71%	62.79%	62.88%	62.69%	62.76%

Standard Errors Clustered at City Level in Parentheses

* $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$