Online Appendix:

Microfounding Household Debt Cycles with Extrapolative Expectations

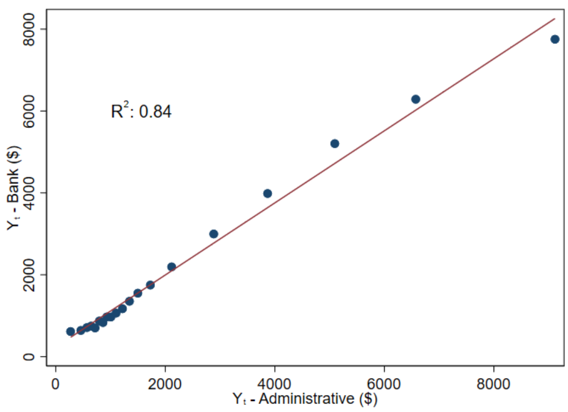
Francesco D'Acunto, Michael Weber, Xiao Yin

Not for Publication

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: Comparisions of Income Measures



Panel B: Survey Answers and Bank Measures

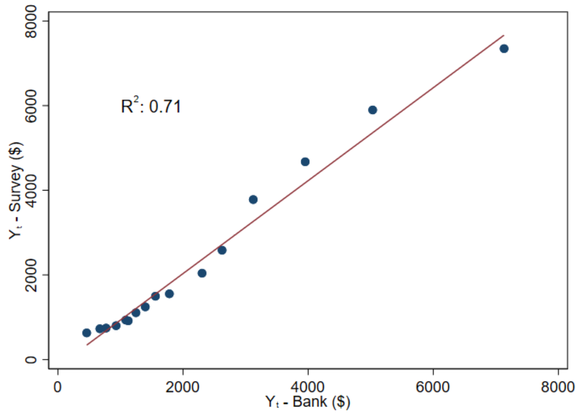


Figure A.2. Equilibrium Distribution of Saving

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.

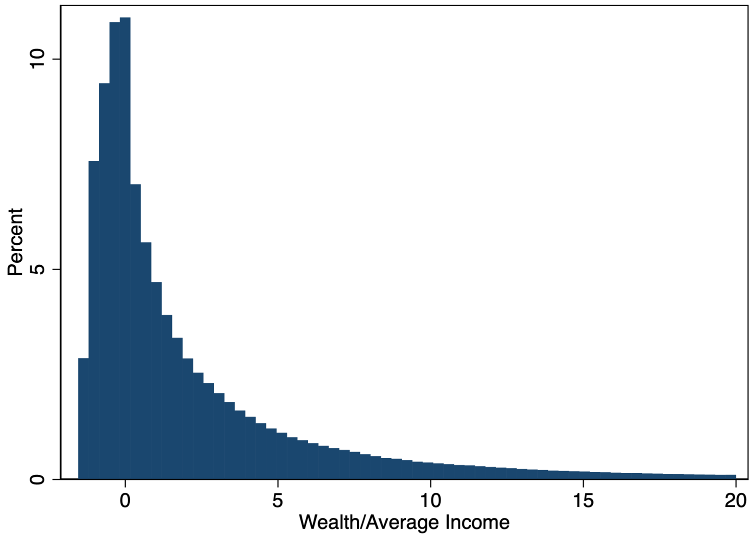


Figure A.3. Impulse Responses after Transitory Income Shocks

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 $\theta = 0$.

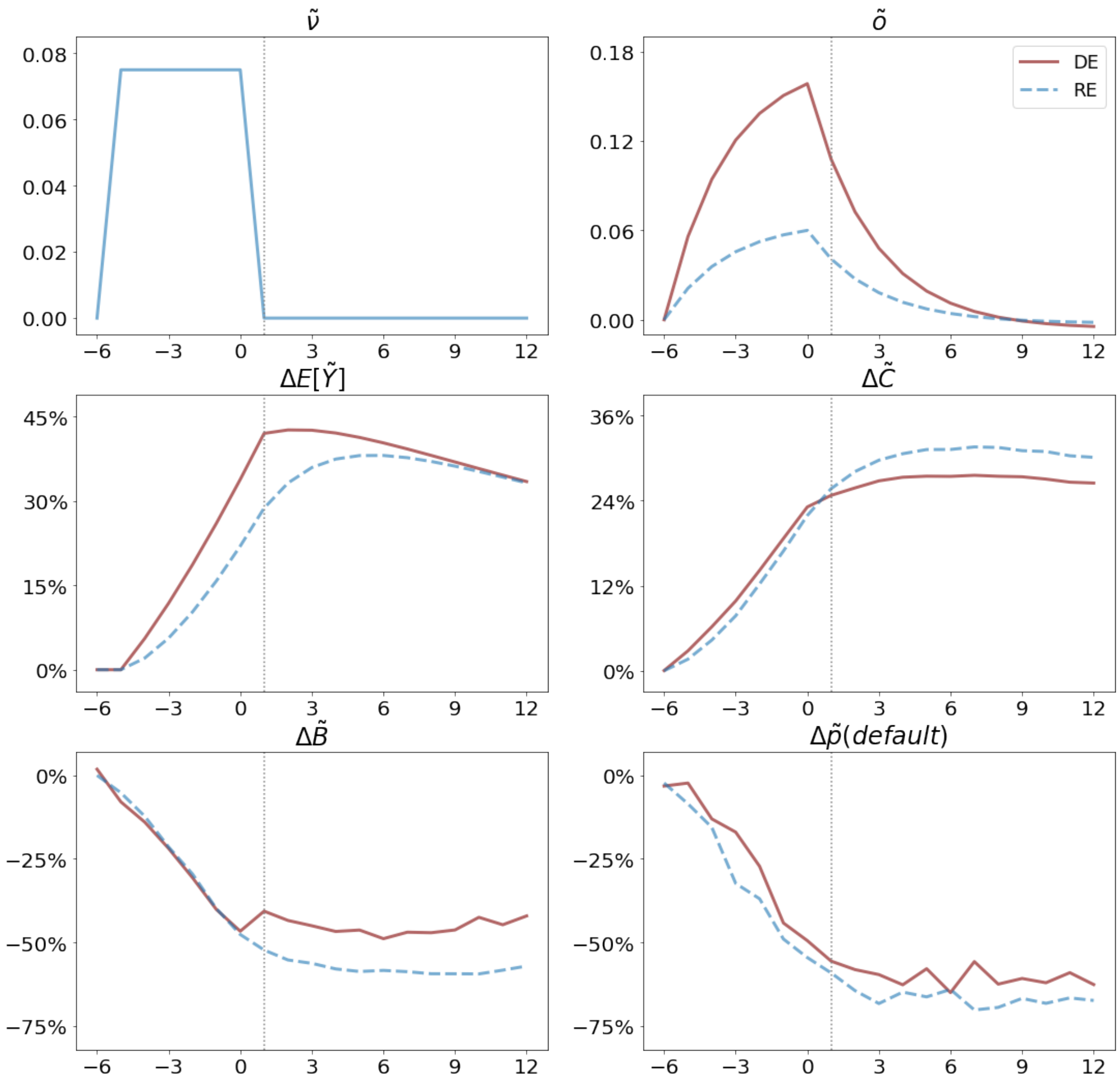


TABLE A.1. Income Shocks and Misbeliefs – Subjective Income Shocks

$E_C[Y_{t+1}]$ is the expected level of income ($ thousands) in period $t+1$ based on survey question 8 sent in period t. $E_C[\Delta Y_t]$ is the expected changes ($ thousands) in 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 6, 7, and 8 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) $E_C[Y_{t+1}] - Y_{t+1}$	(2) $E_C[Y_{t+1}] - Y_{t+1}$	(3) ΔC_t	(4) ΔC_t	(5) ΔB_{t+1}	(6) ΔB_{t+1}
$Y_t - E_C[Y_t]$	0.361***	0.364***	0.165***	0.372***	0.231*	0.081
	(0.071)	(0.074)	(0.024)	(0.083)	(0.137)	(0.049)
$E_C[\Delta Y_t]$					0.065	-0.021
					(0.057)	(0.035)
$1_{\{Y_t - E[Y_t]<0\}}$					0.391*	0.163**
					(0.210)	(0.080)
$(Y_t - E[Y_t]) \times 1_{\{Y_t - E[Y_t]<0\}}$		0.019	-0.016	0.053	0.055	-0.014
		(0.081)	(0.028)	(0.085)	(0.085)	(0.028)
$\log Hours$		-0.116	-0.134***	-0.164	-0.171	-0.133***
		(0.099)	(0.050)	(0.148)	(0.145)	(0.048)
$SD(E_C[\Delta \log Y_{t+1}])$		0.000	-0.011	0.010	0.010	-0.011
		(0.015)	(0.012)	(0.031)	(0.031)	(0.012)
$\log Y_{t-1}$		-0.285**	-0.051	-0.201	-0.193	-0.049
		(0.122)	(0.048)	(0.130)	(0.129)	(0.048)
$\log L_{t-1}$		0.024	0.012	-0.029	-0.029	0.012
		(0.028)	(0.012)	(0.026)	(0.026)	(0.012)
N	10,467	10,467	10,467	10,467	10,467	10,467
City × 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×Year Level in Parentheses
* p <0.10 ** p <0.05 *** p <0.01

TABLE A.2. Income Shocks and Misbeliefs – 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 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) $E_C[y_{t+1}] - y_{t+1}$	(2) $E_C[y_{t+1}] - y_{t+1}$	(3) Δc_t	(4) Δc_t	(5) $Default$	(6) $Default$
$y_t - E[y_t]$	0.412***	0.210***				
	(0.021)	(0.031)				
$E_C[y_{t+1}] - y_{t+1}$			0.200***	0.180***	1.095***	1.454***
			(0.015)	(0.020)	(0.149)	(0.246)
$\log Hours$		-0.071		-0.065		-4.153*
		(0.072)		(0.109)		(2.237)
$SD(E_C[\Delta \log Y_{t+1}])$		0.014		0.054		0.143
		(0.025)		(0.042)		(0.242)
$\log Y_{t-1}$		-0.142*		-0.063		-17.142***
		(0.076)		(0.066)		(4.146)
$\log L_{t-1}$		-0.015		-0.023		-0.094
		(0.015)		(0.023)		(0.267)
N	9,845	9,845	9,845	9,845	9,845	9,845
Industry FE	No	No	Yes	No	Yes	No
City × Round FE	No	No	Yes	Yes	Yes	Yes
Individual FE	No	No	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.3. Misbelief and Economic Behaviors – Longer Sample

ΔC_t is the differences ($ thousands) in the monthly average consumption between t and $t-1$. ΔB_{t+1} is the differences ($ 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 (4). ΔY_t is the changes ($ 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) ΔC_t	(2) ΔC_t	(3) ΔB_{t+1}	(4) ΔB_{t+1}	(5) $Default$	(6) $Default$
$Y_t - E[Y_t]$	0.356***	0.323***	0.169***	0.139***	3.657***	3.168***
	(0.078)	(0.093)	(0.040)	(0.031)	(1.160)	(0.874)
$E[\Delta Y_t]$	0.069	0.069	-0.007	-0.023	-0.061	-0.197
	(0.094)	(0.098)	(0.032)	(0.029)	(0.530)	(0.316)
$\log Y_{t-1}$	-0.222*	-0.219***	-0.057	-0.051	-16.810***	-21.33***
	(0.116)	(0.117)	(0.052)	(0.044)	(3.846)	(5.12)
$\log L_{t-1}$	-0.023	-0.021	0.015	0.007	0.091	0.184
	(0.026)	(0.019)	(0.012)	(0.009)	(0.298)	(0.117)
N	10,497	46,348	10,497	46,348	10,497	46,348
City × 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.4. Misbeliefs, 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 expected changes ($ thousands) in income from period $t-1$ to period t. $SD(E_C[\Delta \log Y_t])$ 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) $\Delta Withdraw_t$	(2) $\Delta Withdraw_t$	(3) $\Delta Transfer_t$	(4) $\Delta Transfer_t$
$E_C[Y_{t+1}] - Y_{t+1}$	0.031	0.019	-0.020	-0.024
	(0.064)	(0.057)	(0.016)	(0.027)
$E[\Delta Y_t]$		0.171**		-0.076**
		(0.068)		(0.030)
Age		0.011*		
		(0.006)		
Age^2		-0.000*		
		(0.000)		
$Female$		-0.014**		
		(0.007)		
$College$		0.019***		
		(0.005)		
$SD(E[\Delta \log Y_{t+1}])$		-0.024***		-0.031***
		(0.008)		(0.006)
$\log Hours$		0.000		-0.032***
		(0.001)		(0.008)
$\log Y_{t-1}$		0.024**		0.005
		(0.012)		(0.005)
$\log L_{t-1}$		-0.020*		-0.012**
		(0.012)		(0.006)
N	10,497	10,497	10,497	10,497
City FE × 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.5. 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

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' personal 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 the participants' personal answers. Therefore, please answer based on your true thoughts.

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

- Yes
- No

1. What was your average monthly income over the past year? ______

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. What's the lowest possible credit card limit you believe you could have in 6 months? ______

4. What's the highest possible credit card limit you believe you could have in 6 months? ______

5. What would your total credit card limit most likely be in 6 months? ______

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

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

7. What would be the highest possible level of average monthly income you believe you would get over the next 6 months?

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

8. What would your average monthly income most likely be in the next 6 months?

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

9. How many hours do you usually work per week? ______

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. Our strategies follow Pistaferri (2001). Specifically, let log income, $y_{i,t}$, follow

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

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

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

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 (4) 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.

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

$\Delta y_t - E[\Delta y_t]$ 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 (4). 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 Misbeliefs

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***	0.329***			0.329***	0.341***		
	(0.019)	(0.022)			(0.062)	(0.054)		
ξ			0.048	0.078**			0.123*	0.107*
			(0.031)	(0.034)			(0.062)	(0.061)
$\log Hours$		-0.055		-0.094		-0.085		-0.087
		(0.071)		(0.071)		(0.067)		(0.070)
$SD(E_C[\Delta \log Y_{t+1}])$		-0.001		0.081***		0.051*		0.060**
		(0.030)		(0.030)		(0.029)		(0.030)
$\log Y_{t-1}$		-0.049		-0.143		-0.189**		-0.134
		(0.080)		(0.087)		(0.084)		(0.084)
$\log L_{t-1}$		-0.016		-0.014		-0.013		-0.014
		(0.016)		(0.015)		(0.016)		(0.015)
N	9,888	9,888	9,888	9,888	9,888	9,888	9,888	9,888
City × 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

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