Online Appendix

Extrapolators and Contrarians: Forecast Bias and Individual Investor Stock Trading

Steffen Andersen
Danmarks Nationalbank, Copenhagen Business School, and CEPR
san.fi@cbs.dk

Stephen G. Dimmock
National University of Singapore
dimmock@nus.edu.sg

Kasper Meisner Nielsen
Copenhagen Business School
kmn.fi@cbs.dk

Kim Peijnenburg
Tilburg University and CEPR
k.peijnenburg@tilburguniversity.edu

This document includes additional analyses and explanations. It is not intended for publication with the main paper and includes the following contents:

- Online Appendix A: Instructions for the Experiment
- Online Appendix B: Control Variables and Alternative Forecast Bias Measures
- Online Appendix C: Additional Tables

Online Appendix A: Instructions for the Experiment

In this Online Appendix, we present the instructions given to the participants in our experiment at Statistics Denmark.

A.1. Instruction for Part I: Forecasting task

In this task you will be asked to make a forecast of the value of an investment (e.g. a stock, a bond, or a property). First, we will show you the value of the investment in the previous 40 periods. Next, you have to forecast the value of the investment in the following 40 periods.

An example is shown on the top of the next page where the green dots show the value of the investment over time.

You are now ready to make your best forecast of the future value of the investment. You have to forecast the value of the investment two periods ahead each time. Your forecast of the value of the investment in the first period is indicated by a blue cross. Your forecast of the value of the investment in the second period is indicated by a yellow cross. In the figure the forecasts are that the investment will have the value of 80 in the first period and 100 in the second period.

Now, follow the instructions:

1. Make your forecast about the first period, using the mouse to drag the blue cross up or down.
2. Make your forecast about the second period, using the mouse to drag the yellow cross up or down.
3. When your crosses are in place, click 'Make forecast'.

Next, you will see the actual value of the investment in the first period – indicated by a new green dot. In the bottom figure the actual value of the investment in the first period is 90 whereas your forecast was 80. The star (which you can see, but not change) shows your forecast for the second period.

You now have to submit your forecast of the value of the investment by repeating item 1, 2, and 3. Proceed in this way, until you have submitted forecasts about the value of the investment for all 40 future periods.

Your chances of receiving a payment

When you have made your forecasts, you have a 1-in-10 chance of receiving a payment of DKK 2.000 for one of them. The random draw is made by rolling a 10-sided dice.

i. Roll the 10-sided dice. If the outcome is the number 1, you have a chance of receiving a payment for one of your 40 forecasts. If the outcome is any other number, you will not receive a payment – continue to Part II.

ii. If the outcome was 1: Roll a 4-sided dice and a 10-sided dice to select one of your forecasts. (For example: 3 + 9 = forecast no. 39).

The precision of the selected forecast determines your chances of receiving the DKK 2,000. Your chance of winning is calculated as 100 minus 5 percentage points for each number your forecast differs from the actual value. For example, if your forecast was 80 and the actual value of the investment is 90, you will have a 50 percent chance of receiving the payment (100 – 5 x

(90 – 80) = 50). If your forecast is the same as the actual value, you will have a 100 percent change of receiving the payment (100 – 5 x (0) = 100). Please note, that if your forecast differs from the actual value by more than 20, you are not eligible for receiving the payment.

The final outcome is determined by rolling two 10-stided dices to draw of a random number between 1 and 100. Each number is equally likely to occur.

iii. Roll the two 10-sided dices to determine if you receive the payment of DKK 2,0000. If your chance of winning is e.g. 50 percent, then any outcome between 1 and 50 will produce a payment. You will *not* receive a payment for outcomes between 51 and 100.

When you make your forecasts, you will not know which forecast may be selected for payment. You should therefore treat each forecast as if it is going to be selected for the payment. If you are to receive a payment, the money will be transferred to your personal bank account by Statistics Denmark.

A.2. Manuscript after Part I (Instructions to employees at Statistics Denmark).

0. Insert password "1".

1. Subject rolls 10-sided dice to determine if they will be receiving a payment.
   a. If they do not roll 1, press "No payment"
   b. If they roll 1, go to 2

2. Subject rolls a 4-sided dice and a 10-sided dice for choice of task (00=40) to determine the chance of receiving a payment.
   a. Enter the outcome of the 4 sided dice and the 10 sided dice in the numeric boxes and press "show change".
   b. Read out the chance of winning the DKK 2,000.
   c. If the chance of receiving a payment is 0 percent, press "No payment".
   d. If not, proceed to 4.

3. Subject rolls two 10-sided dice to determine payment (00=100).
   a. Enter outcome of the two 10-sided dices in the numeric box
   b. Press continue, and verify the prize outcome

Online Appendix B: Control Variables and Alternative Forecast Bias Measures

Risk Aversion

In our experiment, we presented 40 lottery pairs to participants in a random order. The design of these pairs is based on the work of Wakker, Erev, and Weber (1994). Specifically, we selected four sets of their main lottery pairs, as detailed in Wakker, Erev, and Weber (1994), resulting in a total of 16 lottery pairs. We made significant adjustments to the prize values in these pairs. We additionally employ the same lotteries in the loss domain (16 lottery pairs), and endow subjects in each lottery with the maximum amount potentially lost. Furthermore, we took two of the sets, comprising eight lottery pairs, and endowed subjects with the maximum potential loss. However, we added half of this endowment to the prizes, creating lotteries in a mixed frame. This completes 40 lottery pairs. In this study, we use the fraction of safe choices as a measure of risk aversion.

Below are the instructions for the lottery task.

Instructions for Part II: Lottery task

In this task we will present you with 40 problems. In each of the problems you should choose between Option A and Option B and pick your preferred option. After selecting your preferred options, a random draw will determine if you can receive a payment, based on one of the problems. Each number is equally likely to occur. An example is shown on the next page.

In the example, the payment for Option A is:

DKK 400	if the random number is between 1 and 50	(50 pct. chance)
DKK 0	if the random number is between 51 and 100	(50 pct. chance)

In the example, the payment for Option B is:

DKK 600	if the random number is between 1 and 40	(40 pct. chance)
DKK 0	if the random number is between 41 and 100	(60 pct. chance)

The is an equal probability to draw all numbers.

Your chances of receiving a payment depends on your choices

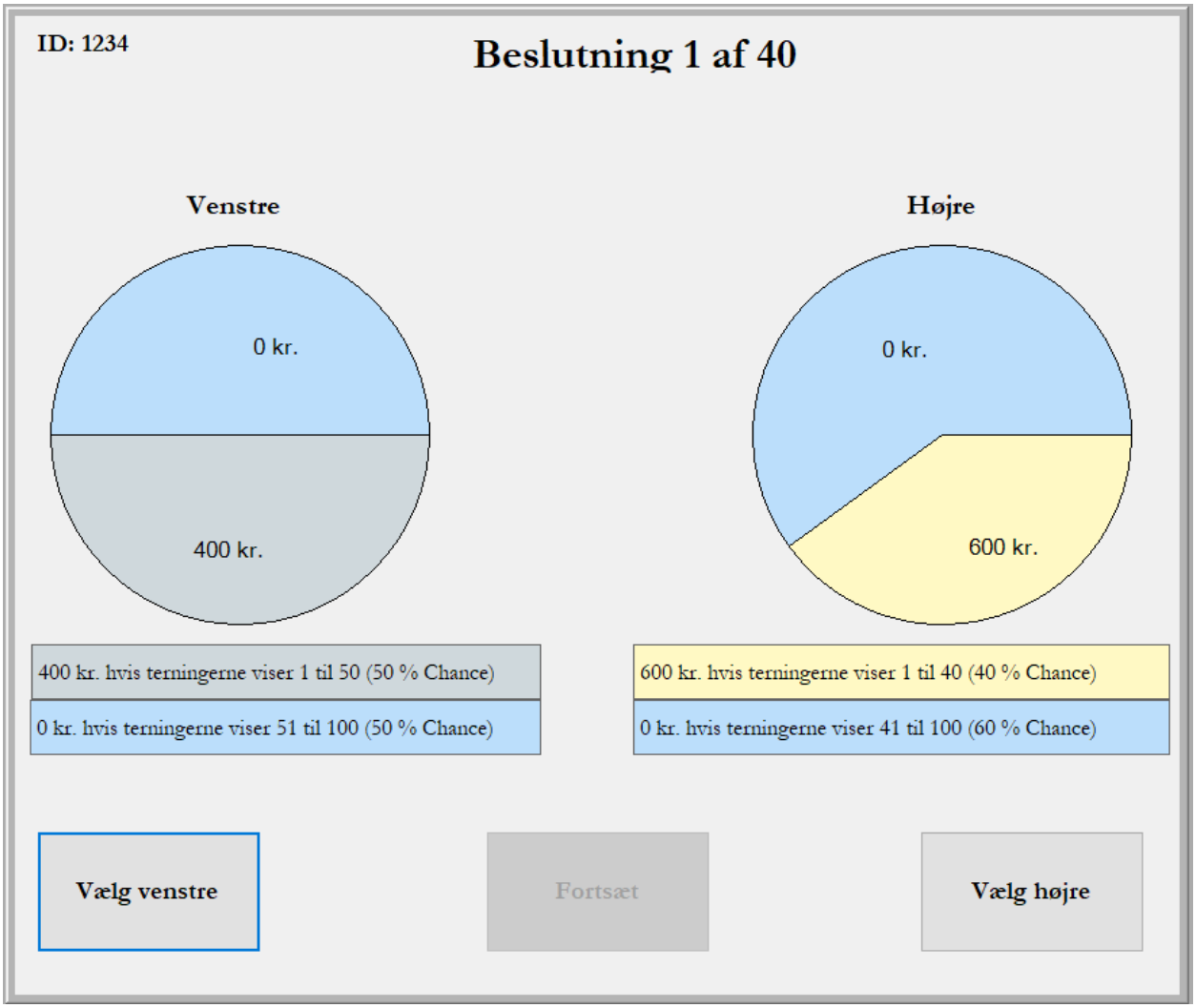
When you have solved all 40 tasks, you have to determine whether you are eligible to receive a payment on one of the problems. This is determined by rolling a 10-sided die.

i. Throw the 10-sided dice. If the outcome is the number 1, you have a chance of receiving a payment for one of your decisions. If the outcome is any other number, you do not have this chance – proceed to Part III.

ii. If the outcome is 1: Throw one 4-sided dice and one 10-sided dice to select the decision that may generate a payment. For example: 2 + 7 = decision no. 27.

iii. Throw two 10-sided dices to determine the payment – depending on the choice you made. (For example: If you chose Option A in decision above. If the outcome of rolling the two 10-sided dices is '49', then you win DKK 400).

When you make your choice between Option A and Option B in the 40 decision, you will not know which decision that might be selected for payment. You should therefore treat each decision as if it is going to be selected for the payment. If you are to receive a payment, the money will be transferred to your personal bank account by Statistics Denmark.



Financial Literacy

The measure of financial literacy is the number of correct responses to four financial literacy questions from Lusardi and Mitchell (2007):

> Suppose you had 1000 kroner in a savings account and the interest rate was 2% per year. After 5 years, how much do you think you would have in the account if you left the money to grow?
>
> 1) More than 1020 kroner
> 2) Exactly 1020 kroner
> 3) Less than 1020 kroner
> 4) Do not know
>
> Imagine that the interest rate on your savings account was 1% per year and inflation was 2% per year. After 1 year, would you be able to buy more than, exactly the same as, or less than today with the money in this account?
>
> 1) More than today
> 2) Exactly the same as today
> 3) Less than today
> 4) Do not know
>
> Please tell us whether this statement is true or false. Buying a single company stock usually provides a safer return than a stock mutual fund.
>
> 1) True
> 2) False
> 3) Don't know
>
> If the interest rate falls, what should happen to bond prices?
>
> 1) Rise
> 2) Fall
> 3) Stay the same
> 4) None of the above
> 5) Don’t know

Overconfidence

Following Moore and Healy (2008), we measure overconfidence as the difference between the number of financial literacy questions the subject believes are answered correctly minus the actual number of correct answers.

> How many of the previous four questions do you have correct?
>
> 1) Zero correct
> 2) One correct
> 3) Two correct
> 4) Three correct
> 5) Four correct
> 6) Don’t know

Numeracy

Numeracy is the number of correct responses to three numerical problems. The numeracy questions we use are based on three questions in the Health and Retirement Study and the English Longitudinal Study of Ageing:

> If the chance of getting a disease is 10 percent, how many people out of 1,000 would be expected to get the disease?
>
> 1) About 1 person
> 2) About 10 people
> 3) About 100 people
> 4) About 1000 people
> 5) Don't know
>
> If 5 people all have the winning numbers in the lottery and the prize is two million kroners, how much will each of them get?
>
> 1) 200,000 kroner
> 2) 400,000 kroner
> 3) 1,000,000 kroner
> 4) 2,000,000 kroner
> 5) Don't know
>
> A second-hand car dealer is selling a car for 60,000 kroner. This is two-thirds of what it cost new. How much did the car cost new?
>
> 1) 70,000 kroner
> 2) 90,000 kroner
> 3) 120,000 kroner
> 4) 180,000 kroner
> 5) Don't know

Trust

The trust question we use is based on Guiso, Sapienza, and Zingales (2008): "Generally speaking, would you say that most people can be trusted or that you have to be very careful when dealing with people?

> 1. Most people can be trusted
> 2.
> 3.
> 4.
> 5.
> 6.
> 7. One has to be very careful with other people
> 8. Don't know

Optimism

Optimism is proxied by the difference between the subjects self-perceived life-expectancy and their life-expectancy according to gender-specific survival tables (Puri and Robinson, 2007). The question we use is "About how long do you think you will live?"

Until about age _____

Do not know

Alternative measures of Forecast Bias

The specifications of the alternative measures of forecast bias are the same as in Afrouzi et al. (2023).

Diagnostic expectations

This model for forecast bias is introduced by Bordalo, Gennaioli, and Shleifer (2018):

$$F_{i,t}(x_{i,t+1}) - E_{i,t}(x_{i,t+1}) = b_i \cdot \left(E_{i,t}(x_{i,t+1}) - E_{i,t-1}(x_{i,t+1})\right) + \varepsilon_{i,t}. \quad (1)$$

Sticky expectations

An often-used model of expectations formation is sticky expectations (Woodford, 2003). A common specification of this model is:

$$F_{i,t}(x_{i,t+1}) - E_{i,t}(x_{i,t+1}) = b_i \cdot \left(F_{i,t-1}(x_{i,t+1}) - E_{i,t}(x_{i,t+1})\right) + \varepsilon_{i,t} \quad (2)$$

where $F_{i,t}(x_{i,t+1})$ indicates subject *i*'s forecast of next period's realization $x_{i,t+1}$, $F_{i,t-1}(x_{i,t+1})$ indicates subject *i*'s forecast in the previous period of next period's realization $x_{i,t+1}$, and $E_{i,t}(x_{i,t+1})$ is the rational forecast. The parameter b_i measures forecast bias.

Extrapolative expectations

In this model of expectations formation (Metzler, 1941), the expectations are influenced by the current outcome and the recent trend. We follow Afrouzi et al. (2023) to specify extrapolative expectations:

$$F_{i,t}(x_{i,t+1}) - x_{i,t} = b_i \cdot (x_{i,t} - x_{i,t-1}) + \varepsilon_{i,t}. \quad (3)$$

Adaptive expectations

A common specification for adaptive expectations (Cagan, 1956) is:

$$F_{i,t}(x_{i,t+1}) - F_{i,t-1}(x_{i,t}) = b_i \cdot \left(x_{i,t} - F_{i,t-1}(x_{i,t})\right) + \varepsilon_{i,t}. \quad (4)$$

Online Appendix C: Additional Tables

Table C.1: Comparison of participants and non-participants

This table provides a comparison of the individual characteristics and portfolio characteristics of individuals that participated in our experiments ("Participants"), and individuals that declined our invitation to participate in our experiments ("Non-participants"). Individual and portfolio characteristics are defined in Appendix Table A1. Column 3 reports statistical inference for test of equal means between participants and non-participants. The symbols *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

	Participants	Non-participants	Significance
	(1)	(2)	(3)
Age	50.46	49.45	***
Male	0.69	0.56	***
Married	0.64	0.69	***
Children	0.81	0.86	***
Education	16.46	16.02	***
Financial assets (000's)	2,012.79	1,857.59	
Income (000's)	747.62	819.04	*
Housing assets (000's)	1,939.25	2,067.84	
Number of buys	36.41	29.30	*
Number of sells	28.75	23.01	
Prior annual ret. (buys)	22.49%	26.41%	*
Prior annual ret. (sales)	25.99%	25.03%	
N	958	23,852	

Table C.2: Correlations between measures of forecast bias

This table reports the correlation matrix for our main measure of *Forecast Bias* and the alternative measures of forecast bias used in Table 5. In column (1), *Forecast Bias* is estimated using equation (2) in the paper. In column (2), *Forecast Bias Residual* is generated using an investor-specific estimated persistence parameter and standard deviation of the error term based on the 80 realizations of the stochastic process (see Section 1.2 in the paper for details). In column (3), *Forecast Bias Limited Information* is based on a subject-specific rational benchmark that is updated every round of the elicitation procedure using the realizations that subject has observed until that point in the experiment (see Section 1.2 for details). In column (4), *Forecast Bias Rank* is the rank transformation of *Forecast Bias*. In column (5), *Diagnostic Expectations* is estimated using the diagnostic expectations function (Online Appendix B, eq. 1). In column (6), *Sticky Expectations* is estimated using the sticky expectations function (Online Appendix B, eq. 2) and is multiplied by -1 to be directionally consistent with the other measures. In column (7), *Extrapolative Expectations* is estimated using the extrapolative expectations function (Online Appendix B, eq. 3). In column (8), *Adaptive Expectations* is estimated using the adaptive expectations function (Online Appendix B, eq. 4).

	Forecast Bias (1)	*Residual* (2)	*Limited Info* (3)	*Rank* (4)	*Diagnostic* (5)	*Sticky* (6)	*Extrapolative* (7)	*Adaptive* (8)
Forecast Bias	1							
Residual	0.98	1						
Limited Info	0.92	0.96	1					
Rank	0.96	0.94	0.88	1				
Diagnostic Expectations	0.95	0.93	0.88	0.92	1			
Sticky Expectations	0.74	0.73	0.70	0.70	0.76	1		
Extrapolative Expectations	0.81	0.79	0.76	0.81	0.92	0.73	1	
Adaptive Expectations	0.88	0.87	0.83	0.89	0.91	0.85	0.87	1

Table C.3: Forecast bias and individual characteristics

This table provides a comparison of the individual characteristics of *Contrarians*, *No Bias*, and *Extrapolators*. *Contrarians* is defined as *Forecast Bias* < -0.5, *No Bias* is defined as -0.5 ≤ *Forecast Bias* ≤ 0.5, and *Extrapolators* is defined as *Forecast Bias* > 0.5. The final three columns report statistical inference for test of equal means between *Contrarians*, *No Bias*, and *Extrapolators*. Appendix Table A1 defines the variables. The symbols *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

	Contrarians	*No bias*	*Extrapolators*	(1) vs.	(1) vs.	(2) vs.
	(1)	(2)	(3)	(2)	(3)	(3)
Age	51.45	50.35	49.91	*	**	
Male	0.68	0.69	0.69			
Married	0.65	0.61	0.67			
Children	0.81	0.80	0.81			
Education	16.19	16.43	16.65		**	
Financial assets (000's)	1,308.47	3,193.80	1,368.57			
Income (000's)	727.14	787.88	724.03			
Housing assets (000's)	1,880.45	1,980.19	1,939.90			
Post-Covid experiment	0.28	0.37	0.34	**		
Risk aversion	0.47	0.50	0.50	**	*	
Financial literacy	3.29	3.44	3.43	**	**	
Numeracy	2.75	2.84	2.87	**	***	
Optimism	4.60	4.39	4.74			
Overconfidence	0.19	0.13	0.23			
Trust	4.05	4.25	4.37		**	

Table C.4: Forecast bias and stock purchases: Different horizons

This table reports results of weighted-least squares regressions of the relation between stock purchases and *Forecast Bias* for different past return horizons. The dependent variable equals 100 if the stock is purchased and zero otherwise. The key independent variable is *Forecast Bias x Performance measure*, where the performance measure is *Lagged return*. In columns (1) through (4), *Performance measure* is the lagged return over the prior three months, six months, one year, and three years, respectively. The return is winsorized at the 1st and 99th percentiles. The unit of observation is investor-stock-day over the period 2011-2021. The observations are weighted such that each investor has equal weight in the regressions. *Forecast Bias* is adjusted to have a standard deviation of one. Both columns include an indicator for a stock previously held by the subject, an indicator for a stock currently owned by the subject, the portfolio weight of a current holding (zero for stocks not currently held), and the fraction of all investor-stock purchases for the full Danish population in the past month that were in the stock. The specification includes investor-day fixed effects. Standard errors are clustered at the individual-level and appear in parentheses. The symbols *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Excess return horizon	Prior 3 months	Prior 6 months	Prior 1 year	Prior 3 years
	(1)	(2)	(3)	(4)
Forecast Bias x Perform.	0.007	0.005*	0.004***	0.003***
	(0.004)	(0.003)	(0.001)	(0.001)
Performance measure	-0.020***	-0.012***	-0.008***	-0.0004***
	(0.005)	(0.003)	(0.002)	(0.001)
Investor-Day Fixed Effect	Yes	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes
N	34,570,055	34,271,247	33,675,400	30,982,379

Table C.5: Forecast bias and stock sales - Robustness

This table reports results of a weighted-least squares regression of the relation between stock sales and *Forecast Bias*. The dependent variable equals 100 if the stock is sold and zero otherwise. The key independent variables are *Forecast Bias x Performance measure*, where the performance measure is *Capital gain*. Capital gains are winsorized at the 1st and 99th percentiles. The unit of observation is investor-stock-day over the period 2013-2021. The observations are weighted such that each investor has equal weight in the regressions. *Forecast Bias* is adjusted to have a standard deviation of one. The specification includes monthly holding length fixed effects, investor-day fixed effects, as well as controls for the portfolio weight of a current holding, the fraction of all investor-stock sales for the full Danish population in the past month that were in the stock, and additional performance terms. Standard errors are clustered at the individual-level and appear in parentheses. The symbols *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

	(1)
Forecast Bias × Perform.	-0.904**
	(0.406)
Performance	-11.306***
	(2.260)
I(Performance>0)	9.135***
	(1.593)
Performance squared	4.138***
	(0.910)
Performance cubed	-0.364***
	(0.089)
Performance highest	6.724***
	(2.207)
Performance lowest	2.886
	(2.180)
Portfolio weight	16.894***
	(4.584)
Stock sales fraction	1.449***
	(0.036)
Investor-Day Fixed Effect	Yes
Length of Holding Fixed Effect	Yes
N	176,451

Table C.6: Net flows, trading, and conditional net flows - Robustness

This table reports the results of conditional logit regressions of trading activity during the period 2012-2021. In columns (1), (2), and (3), the dependent variables are indicators equal to 100 if the absolute value of the subject's net flows is greater than 1%, the value of net flows is greater than 1%, and value of net flows is less than -1%, respectively. The sample includes all investor-months in which the subject owns individual stocks. In all columns, the observations are weighted such that each investor has equal weight in the regressions. Lag market return is the return on the Danish stock market index over the prior year. Lag excess return is the subject's stock return over the prior year less the lag market return. All columns include controls for the value of beginning of month stock holdings, financial assets, housing assets, income, education, children, and marital status, as well as individual and year-month fixed effects. Appendix Table A1 defines the variables. The number of observations differs from that reported in Table 7 because the conditional logit model drops groups with no variation in the dependent variable. Standard errors are clustered at the individual-level and appear in parentheses. The symbols *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

	Trade month	Buy month	Sell month
	(1)	(2)	(3)
Forecast Bias × Lag market return	-0.061	-0.008	-0.070
	(0.130)	(0.146)	(0.141)
Forecast Bias × Lag excess return	0.054	0.041	0.041
	(0.056)	(0.055)	(0.061)
Lag excess return	0.293***	0.178**	0.248***
	(0.069)	(0.072)	(0.072)
Individual Fixed Effect	Yes	Yes	Yes
Year-Month Fixed Effect	Yes	Yes	Yes
Control variables	Yes	Yes	Yes
N	73,643	60,700	66,221

References

Afrouzi, H., S.Y. Kwon, A. Landier, Y. Ma, and D. Thesmar. 2023. Overreaction in expectations: Evidence and theory. *Quarterly Journal of Economics* 138, 1713-1764.

Bordalo, P., N. Gennaioli, and A. Shleifer. 2018. Diagnostic expectations and credit cycles. *Journal of Finance* 73, 199-227.

Cagan, P. 1956. The monetary dynamics of hyperinflation. In *Studies in the Quantity Theory of Money*, M. Friedman Editor, University of Chicago Press.

Guiso, L., P. Sapienza, and L. Zingales. 2008. Trusting the stock market. *Journal of Finance* 63, 2557-2600.

Lusardi, A., and O.S. Mitchell. 2007. Baby boomer retirement security: The roles of planning, financial literacy, and housing wealth. *Journal of Monetary Economics* 54, 205-224.

Metzler, L.A. 1941. The nature and stability of inventory cycles. *Review of Economics and Statistics* 23, 113-129.

Moore, D.A., and P.J. Healy. 2008. The trouble with overconfidence. *Psychological Review* 115, 502-517.

Puri, M., and D.T. Robinson. 2007. Optimism and economic choice. *Journal of Financial Economics* 86, 71-99.

Wakker, P., I. Erev, and E.U. Weber. 1994. Comonotonic independence: The critical test between classical and rank-dependent utility theories. *Journal of Risk and Uncertainty* 9, 195-230.

Woodford, M. 2003. Imperfect common knowledge and the effects of monetary policy. In *Knowledge, Information, and Expectations in Modern Macroeconomics*, P. Aghion, R. Frydman, J.E. Stiglitz, and M. Woodford Editors, Princeton University Press.