Heterogeneity in Expectations, Risk Tolerance, and Household Stock Shares

John Ameriks
Vanguard Group

Gábor Kézdi
Central European University

Minjoon Lee
University of Michigan

Matthew D. Shapiro
University of Michigan and NBER

July 9, 2016

APPENDICES

Appendix A. Additional Tables and Figures

Table A1. Distribution of wealth in the VRI data (N=4414)

	Mean	Std	p10	p25	Median	p75	p90
Financial wealth	1,147,525	1,516,575	164,835	363,000	759,750	1,403,843	2,467,899
Home stock	360,782	578,045	31,500	125,000	235,000	420,000	1,060,000

Table A2. Summary statistics of the control variables (N=4414)

	Mean	Standard deviation
Single male	0.14	
Female in couple	0.17	
Single female	0.18	
Age	67.8	7.4
Age squared	4649	1023
In the employer-sponsored sample	0.21	
College degree	0.33	
MBA	0.07	
PhD	0.06	
Other higher degree	0.28	
Log(wealth)	13.4	1.09
Log(home equity)	11.5	3.37
Zero home equity	0.07	
Retired	0.60	
Log(Wage)	4.3	5.5
Log(Annuity Income)	6.5	5.3
Expected Log(Annuity Income)	4.3	5.3
Subjective probability of needing long-term care	0.43	0.30
Subjective probability of survival to target age	0.75	0.23

Notes.
Log variables are set to zero if the levels of the variables are zero. Zero home equity equals 1 (0) if home equity is zero (positive). Annuity income is the sum of Social Security income, defined benefit pension income and immediate annuity income, for retired households. It is set to zero for non-retired households. Expected annuity income is the sum of expected values of Social Security income, defined benefit pension income and immediate annuity income, for non-retired households. It is set to zero for retired households. Subjective probability of needing long-term care is the subjective probability chance that the respondent would need long-term care service at least for one year during her remaining life. The target age in subjective probability of survival question is set to 75 if the respondent is younger than 70, to 85 if the respondent is younger than 80, and to 95 if the respondent is younger than 90.

Table A3. The Risk Tolerance Strategic Survey Questions in VRI survey 2

Set up	Suppose you are 80 years old. Suppose, further, that for the next year: • You live alone, rent your home, and pay all your own bills. • You are in good health and will remain in good health. • You will have no medical bills or other unexpected expenses. • You do not work.
Hypothetical financial products	• Plan A guarantees that you will have $W for spending next year. • Plan B will possibly provide you with more money, but is less certain. There is a 50% chance that Plan B would double your money, leaving you with $2W, and a 50% chance that it would cut it by x%, leaving you with $ $(1-0.01\times x)W$.
Rules	• You have no other assets or income, and so the only money you have available for all your spending next year is from either Plan A or Plan B. • Any money that is not spent at the end of next year cannot be saved for the future. • You cannot give any money away or leave it as a bequest. • If you need anything next year, you have to pay for it. No one else can buy anything for you. • At the end of next year you will be offered the same choice with another $W for following year.
Parameters asked	W =100,000 and 50,000.
Question	Would you choose Plan A or Plan B?

Table A4: The stock market expectation questions in VRI survey wave 3.

Variable name	Survey question
Question Order *p-m*	
p0	What do you think is the percent chance that the stock market will be higher in twelve months than it is today? Think of a stock market index such as the Dow Jones Industrial Average and do not adjust for inflation.
p20	And what do you think is the percent chance that it will be at least 20% higher in twelve months than it is today? [If answer is greater than the p0 answer: "Please enter a response that is less than or equal to you previous response or change your previous response."]
m	Instead of probabilities, we are now interested in your expectation. By what percentage do you expect the stock market to increase or decrease in the next twelve months? Please enter a positive number for increase and negative number for decrease.
Question order *m-p*	
m	By what percentage do you expect the stock market to increase or decrease in the next twelve months? Think of a stock market index such as the Dow Jones Industrial Average and do not adjust for inflation. Please enter a positive number for increase and negative number for decrease.
p0	And what do you think is the percent chance that the stock market will be higher in twelve months than it is today?
p20	What do you think is the percent chance that it will be at least 20% higher in twelve months than it is today? [If answer is greater than the p0 answer: "Please enter a response that is less than or equal to you previous response or change your previous response."]

Note: The question orders are randomized in the survey instrument. The distributions of responses are slightly different depending on which sequence is used.

Table A5. Detailed results of the structural estimation model without covariates. (N=4,414)

	Preference	Beliefs		Bias in *p0*
	γ	μ	σ	ψ
constant	-1.148***	0.055***	0.118***	-0.539***
	(0.027)	(0.002)	(0.002)	(0.017)
Heterogeneity				
σ_u	0.704***	0.063***	0.032***	n.a.
	(0.011)	(0.001)	(0.001)	n.a.
Correlation across latent variables				
γ		0.011**	-0.004	
		(0.003)	(0.002)	
$\rho_{\mu\sigma}$		0.062**		
		(0.021)		
Measurement error				
$\sigma_{e\gamma 1}$		0.812***		
		(0.015)		
$\sigma_{e\gamma 2}$		0.544***		
		(0.016)		
σ_{em}		0.079***		
		(0.001)		
σ_{ep}		0.487***		
		(0.008)		
Log-likelihood		-48006		

Notes.
The third line reports how the latent risk tolerance parameter affects means of the belief parameter distributions. Statistics reported in Table 4 are calculated based on these parameters, where the means of belief parameter distributions are adjusted using the mean of the risk tolerance parameter multiplied with the numbers reported in the third row.
Standard errors in parentheses.
*, **, and *** implies significance at 5%, 1% and 0.1% level, respectively.

Table A6. Detailed results of the structural estimation model with covariates. (N=4,414)

	Preference	Beliefs		Bias in *p0*
	γ	μ	σ	*ψ*
Constant	-1.415***	0.071*	0.181***	-0.373
	(0.412)	(0.031)	(0.024)	(0.833)
Single male	0.038	0.004	0.002	-0.019
	(0.042)	(0.004)	(0.003)	(0.041)
Female in couple	-0.207***	0.004	0.001	-0.171***
	(0.040)	(0.004)	(0.003)	(0.039)
Single female	-0.191***	0.011**	0.000	-0.294***
	(0.041)	(0.004)	(0.003)	(0.039)
Age	-0.020	0.000	-0.002***	-0.029
	(0.015)	(0.000)	(0.000)	(0.023)
Age sq.	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Employer-	-0.003	0.015***	-0.005*	-0.167***
sponsored	(0.039)	(0.003)	(0.002)	(0.037)
College degree	0.039	-0.009**	0.007***	0.294***
	(0.035)	(0.003)	(0.002)	(0.035)
MBA	0.116	-0.004	0.004	0.222***
	(0.060)	(0.006)	(0.004)	(0.064)
PhD	0.042	-0.019**	0.023***	0.465***
	(0.061)	(0.007)	(0.006)	(0.068)
Other higher degree	0.079*	-0.010**	0.014***	0.354***
	(0.038)	(0.004)	(0.002)	(0.038)
log(wealth)	0.037**	-0.007***	0.005***	0.131***
	(0.014)	(0.001)	(0.001)	(0.014)
log(home equity)	0.029	-0.002	0.000	-0.008
	(0.015)	(0.001)	(0.001)	(0.015)
No home equity	0.338	-0.018	-0.001	-0.155
	(0.191)	(0.014)	(0.010)	(0.179)
Retired	0.386	-0.039	-0.011	-0.016
	(0.410)	(0.030)	(0.024)	(0.375)
Log(Wage)	-0.004	0.001	0.000	-0.007
	(0.011)	(0.001)	(0.001)	(0.011)
Log(Annuity	-0.030	0.008***	-0.003**	-0.048*
Income)	(0.029)	(0.001)	(0.001)	(0.022)
Expected Log(Annuity	0.015	0.004	-0.003	-0.046
Income)	(0.030)	(0.003)	(0.002)	(0.029)
LTC probability	0.009	-0.016***	0.002	0.184***
	(0.045)	(0.004)	(0.003)	(0.044)

Longevity	0.191**	0.028***	-0.004	0.034
probability	(0.063)	(0.006)	(0.004)	(0.062)
Heterogeneity				
σ_u	0.688***	0.063***	0.030***	n.a.
	(0.011)	(0.001)	(0.001)	n.a.
Correlation across latent variables				
γ		0.011**	-0.003	
		(0.004)	(0.002)	
$\rho_{\mu\sigma}$		0.009		
		(0.024)		
Measurement error				
$\sigma_{e\gamma 1}$		0.810***		
		(0.015)		
$\sigma_{e\gamma 2}$		0.557***		
		(0.016)		
σ_{em}		0.078***		
		(0.001)		
σ_{ep}		0.455***		
		(0.008)		
Log-likelihood		-47656		

Note: Standard errors in parentheses.
*, **, and *** implies significance at 5%, 1% and 0.1% level, respectively.
Reference categories are male in couple, individual client sample, not having a college degree. See notes to Table A2 for detailed description of the right hand side variables.

Table A7. Stock share regressions with raw survey answers on the right hand side (with *m* as a proxy for beliefs of mean returns μ)

	Dependent variable: survey measure of stock share		Dependent variable: administrative measure of stock share	
m	0.126**	0.153***	0.180***	0.192***
	(0.037)	(0.037)	(0.038)	(0.038)
p0-p20	0.107***	0.085***	0.098***	0.091***
	(0.016)	(0.016)	(0.017)	(0.017)
SSQ1 cat=2	0.026*	0.016	0.016	0.008
	(0.011)	(0.011)	(0.011)	(0.011)
SSQ1 cat=3	0.047***	0.035**	0.038**	0.028*
	(0.011)	(0.011)	(0.011)	(0.011)
SSQ1 cat=4	0.054***	0.044***	0.057***	0.049***
	(0.013)	(0.013)	(0.013)	(0.013)
SSQ1 cat=5	0.083***	0.073***	0.080***	0.075***
	(0.014)	(0.014)	(0.015)	(0.015)
SSQ1 cat=6	0.053	0.045	-0.023	-0.026
	(0.031)	(0.031)	(0.032)	(0.031)
Single male		0.045		0.013
		(0.031)		(0.012)
Female in couple		0.016		0.021
		(0.012)		(0.011)
Single female		-0.007		0.019
		(0.011)		(0.012)
Age		-0.007		-0.014
		(0.012)		(0.009)
Age sq.		0.000		0.000
		(0.001)		(0.000)
Employer-sponsored		-0.053***		-0.042**
		(0.011)		(0.011)
College degree		0.018		0.023*
		(0.010)		(0.010)
MBA		0.033		0.022
		(0.017)		(0.018)
PhD		0.009		0.068***
		(0.017)		(0.018)
Other higher degree		0.015		0.029**
		(0.011)		(0.011)
log(wealth)		0.017***		-0.001
		(0.004)		(0.004)
log(home equity)		0.004		0.008
		(0.004)		(0.004)

No home equity		0.031		0.080
		(0.054)		(0.055)
Retired		-0.254*		-0.318**
		(0.116)		(0.119)
Log(Wage)		0.005		-0.001
		(0.003)		(0.003)
Log(Annuity		0.002		0.023**
Income)		(0.008)		(0.008)
Expected Log(Annuity		-0.023**		-0.002
Income)		(0.008)		(0.008)
LTC probability		-0.027*		-0.035**
		(0.013)		(0.013)
Longevity		0.042*		0.034
probability		(0.018)		(0.019)
Constant		0.371***		1.028***
		(0.111)		(0.319)
R2	0.023	0.040	0.023	0.043
Observations	4414	4414	4414	4414

Note: Standard errors in parentheses.
*, **, and *** implies significance at 5%, 1% and 0.1% level, respectively.
Reference categories are male in couple, individual client sample, not having a college degree. See notes to Table A2 for detailed description of the right hand side variables.

Table A8. Stock share regressions with raw survey answers on the right hand side (with $\left(p_0 + p_{20}\right)/2$ as a proxy for beliefs of mean returns μ)

	Dependent variable: survey measure of stock share		Dependent variable: administrative measure of stock share	
(p0+p20)/2	0.115***	0.118***	0.097***	0.089***
	(0.024)	(0.024)	(0.025)	(0.025)
p0-p20	0.076***	0.056**	0.075***	0.074***
	(0.018)	(0.018)	(0.018)	(0.018)
SSQ1 cat=2	0.023*	0.013	0.012	0.005
	(0.011)	(0.011)	(0.011)	(0.011)
SSQ1 cat=3	0.043***	0.031**	0.033**	0.024*
	(0.011)	(0.011)	(0.011)	(0.011)
SSQ1 cat=4	0.049***	0.040**	0.052***	0.045**
	(0.013)	(0.013)	(0.013)	(0.013)
SSQ1 cat=5	0.079***	0.069***	0.076***	0.071***
	(0.014)	(0.014)	(0.015)	(0.015)
SSQ1 cat=6	0.051	0.043	-0.024	-0.027
	(0.031)	(0.031)	(0.032)	(0.032)
Single male	0.051	0.043		0.014
	(0.031)	(0.031)		(0.012)
Female in couple		0.017		0.023*
		(0.012)		(0.011)
Single female		-0.006		0.023
		(0.011)		(0.012)
Age		-0.004		-0.014
		(0.012)		(0.009)
Age sq.		0.001		0.000
		(0.001)		(0.000)
Employer-sponsored		-0.052***		-0.041***
		(0.011)		(0.011)
College degree		0.014		0.020
		(0.010)		(0.010)
MBA		0.029		0.019
		(0.017)		(0.018)
PhD		0.004		0.064***
		(0.017)		(0.018)
Other higher degree		0.011		0.025*
		(0.011)		(0.011)
log(wealth)		0.017***		-0.001
		(0.004)		(0.004)
log(home equity)		0.004		0.008
		(0.004)		(0.004)

No home equity		0.033		0.080
		(0.054)		(0.055)
Retired		-0.256*		-0.321**
		(0.116)		(0.120)
Log(Wage)		0.005		-0.001
		(0.003)		(0.004)
Log(Annuity Income)		0.003		0.024**
		(0.008)		(0.008)
Expected Log(Annuity Income)		-0.022**		-0.001
		(0.008)		(0.008)
LTC probability		-0.028*		-0.037**
		(0.013)		(0.013)
Longevity probability		0.039*		0.034
		(0.018)		(0.019)
Constant		0.340**		1.010***
		(0.111)		(0.319)
R2	0.025	0.041	0.022	0.040
Observations	4414	4414	4414	4414

Note: Standard errors in parentheses.
*, **, and *** implies significance at 5%, 1% and 0.1% level, respectively.
Reference categories are male in couple, individual client sample, not having a college degree. See notes to Table A2 for detailed description of the right hand side variables.

Table A9. Stock share and preference and belief proxies. Detailed results corresponding to Table 5.

	Dependent variable: survey measure of stock share		Dependent variable: administrative measure of stock share	
$\hat{\mu}_i$	0.058***	0.055***	0.052***	0.048***
	(0.010)	(0.009)	(0.008)	(0.008)
$\hat{\sigma}_i$	-0.093*	-0.083	-0.068	-0.083*
	(0.046)	(0.051)	(0.040)	(0.038)
$\hat{\gamma}_i$	0.034***	0.033***	0.012	0.013
	(0.009)	(0.010)	(0.010)	(0.010)
Single male		0.027		0.022
		(0.022)		(0.019)
Female in couple		-0.025		0.023
		(0.021)		(0.018)
Single female		-0.031		0.013
		(0.020)		(0.019)
Age		-0.042*		-0.027
		(0.017)		(0.015)
Age sq.		0.000**		0.000
		(0.000)		(0.000)
Employer-sponsored		-0.115***		-0.081***
		(0.020)		(0.018)
College degree		0.048*		0.051**
		(0.021)		(0.017)
MBA		0.072**		0.048
		(0.027)		(0.032)
PhD		0.057		0.143***
		(0.032)		(0.025)
Other higher degree		0.053*		0.069***
		(0.022)		(0.019)
log(wealth)		0.044***		0.011
		(0.009)		(0.007)
log(home equity)		0.008		0.013*
		(0.009)		(0.006)
No home equity		0.052		0.120
		(0.118)		(0.079)
Retired		-0.448		-0.496**
		(0.244)		(0.196)
Log(Wage)		0.007		-0.002
		(0.005)		(0.005)
Log(Annuity Income)		-0.002		0.032*
		(0.016)		(0.015)

Expected Log(Annuity Income)		-0.045*		-0.006
		(0.019)		(0.012)
LTC probability		-0.032		-0.041
		(0.028)		(0.021)
Longevity probability		0.084*		0.050
		(0.033)		(0.032)
Constant	-0.001	1.136	-0.001	0.803
	(0.007)	(0.649)	(0.007)	(0.519)
R2	0.019	0.045	0.013	0.038
Observations	4414	4414	4414	4414

Standard errors in parentheses.
*, **, and *** implies significance at 5%, 1% and 0.1% level, respectively.
Reference categories are male in couple, individual client sample, not having a college degree. See notes to Table A2 for detailed description of the right hand side variables.

Table A10. Stock Shares Versus Error-Ridden Cardinal Measures of Preferences and Beliefs. Estimation without taking care of measurement error in the cardinal proxies.

	LHS variable: survey measure of stock share in total financial wealth		LHS variable: administrative measure of stock share in Vanguard	
	(1)	(2)	(3)	(4)
m_i	0.017***	0.020***	0.020***	0.021***
	(0.004)	(0.004)	(0.004)	(0.004)
$\tilde{\sigma}_i$	-0.029***	-0.019**	-0.025***	-0.020**
	(0.006)	(0.007)	(0.006)	(0.006)
$\tilde{\gamma}_i$	0.021***	0.020***	0.013**	0.013**
	(0.005)	(0.005)	(0.004)	(0.004)
Single male		0.027		0.021
		(0.022)		(0.019)
Female in couple		-0.024		0.025
		(0.020)		(0.018)
Single female		-0.026		0.019
		(0.021)		(0.019)
Age		-0.040*		-0.025
		(0.016)		(0.014)
Age sq.		0.000*		0.000
		(0.000)		(0.000)
Employer-sponsored		-0.099***		-0.067***
		(0.019)		(0.017)
College degree		0.037*		0.040*
		(0.019)		(0.017)
MBA		0.066*		0.041
		(0.031)		(0.028)
PhD		0.025		0.113***
		(0.032)		(0.028)
Other higher degree		0.036		0.052**
		(0.020)		(0.018)
log(wealth)		0.034***		0.002
		(0.008)		(0.007)
log(home equity)		0.008		0.013
		(0.008)		(0.007)
No home equity		0.054		0.118
		(0.098)		(0.088)
Retired		-0.454*		-0.497**
		(0.212)		(0.190)
Log(Wage)		0.007		-0.002
		(0.005)		(-0.005)
Log(Annuity Income)		0.004		0.037**
		(0.015)		(0.013)

Expected Log(Annuity Income)		-0.041**		-0.002
		(0.015)		(0.013)
LTC probability		-0.043		-0.050
		(0.023)		(0.021)
Longevity probability		0.106**		0.065*
		(0.033)		(0.030)
constant	-0.001	1.120*	-0.000	0.781
	(0.007)	(0.565)	(0.006)	(0.507)
R^2	0.013	0.039	0.012	0.038
N	4414	4414	4414	4414

Notes. In these regressions the cardinal proxies $\hat{\mu}_i, \hat{\sigma}_i, \hat{\gamma}_i$ are replaced with $m_i, \tilde{\sigma}_i, \tilde{\gamma}_i$, respectively, where m_i is the raw answer to the expected stock returns question, $\tilde{\sigma}_i = 0.2 / \left(\Phi^{-1}(p_{0i}) - \Phi^{-1}(p_{20i})\right)$ (the denominator replaced with 0.2 if zero), and $\tilde{\gamma}_i$ is the median valie of the CRRA risk tolerance parameter that corresponds to the answers to the first set of the risk tolerance questions (κ set to zero).
Standard errors in parentheses.
*, **, and *** implies significance at 5%, 1% and 0.1% level, respectively.
Reference categories are male in couple, individual client sample, not having a college degree. See notes to Table A2 for detailed description of the right hand side variables.

Table A11. Stock Shares Versus Cardinal Proxies for Preferences and Beliefs. Employer-sponsored subsample

	LHS variable: survey measure of stock share in total financial wealth		LHS variable: administrative measure of stock share in Vanguard	
	(1)	(2)	(3)	(4)
$\hat{\mu}_i$	0.067***	0.062**	0.083***	0.080***
	(0.018)	(0.019)	(0.014)	(0.015)
$\hat{\sigma}_i$	-0.122	-0.037	-0.014	0.055
	(0.097)	(0.107)	(0.088)	(0.087)
$\hat{\gamma}_i$	0.070**	0.068*	0.016	-0.007
	(0.029)	(0.031)	(0.032)	(0.040)
constant	-0.074***	1.930	-0.030	3.388
	(0.017)	(1.896)	(0.015)	(1.836)
control variables	N	Y	N	Y
R^2	0.026	0.040	0.033	0.079
N	923	923	923	923

Notes.
Employer-sponsored sample are those who only have 401(k) type accounts at Vanguard.

Table A12. Stock Shares Versus Cardinal Proxies for Preferences and Beliefs. Individual-client subsample

	LHS variable: survey measure of stock share in total financial wealth		LHS variable: administrative measure of stock share in Vanguard	
	(1)	(2)	(3)	(4)
$\hat{\mu}_i$	0.059***	0.055***	0.041***	0.036***
	(0.010)	(0.012)	(0.011)	(0.009)
$\hat{\sigma}_i$	-0.075	-0.089	-0.091	-0.112*
	(0.051)	(0.055)	(0.051)	(0.046)
$\hat{\gamma}_i$	0.027**	0.024*	0.012	0.013
	(0.010)	(0.010)	(0.010)	(0.011)
constant	0.024**	1.099*	0.011	0.765
	(0.009)	(0.525)	(0.007)	(0.570)
control variables	N	Y	N	Y
R^2	0.016	0.032	0.008	0.028
N	3491	3491	3491	3491

Notes.
Individual-client sample is the complement of Employer-sponsored sample.

Table A13. Stock Shares Versus Cardinal Proxies for Preferences and Beliefs. Share of wealth at Vanguard at least 50 percent

	LHS variable: survey measure of stock share in total financial wealth		LHS variable: administrative measure of stock share in Vanguard	
	(1)	(2)	(3)	(4)
$\hat{\mu}_i$	0.057***	0.053***	0.045***	0.044***
	(0.015)	(0.013)	(0.011)	(0.011)
$\hat{\sigma}_i$	-0.139*	-0.131	-0.008	-0.018
	(0.067)	(0.076)	(0.055)	(0.053)
$\hat{\gamma}_i$	0.035**	0.038*	0.029*	0.029**
	(0.012)	(0.015)	(0.012)	(0.014)
constant	0.005	0.776	-0.032***	1.193
	(0.009)	(0.870)	(0.007)	(0.756)
control variables	N	Y	N	Y
R^2	0.020	0.034	0.018	0.042
N	1909	1909	1909	1909

Table A14. Stock Shares Versus Cardinal Proxies for Preferences and Beliefs. Share of wealth at Vanguard at least 70 percent

	LHS variable: survey measure of stock share in total financial wealth		LHS variable: administrative measure of stock share in Vanguard	
	(1)	(2)	(3)	(4)
$\hat{\mu}_i$	0.058***	0.054**	0.060***	0.058***
	(0.016)	(0.017)	(0.013)	(0.013)
$\hat{\sigma}_i$	-0.127	-0.107	-0.018	-0.008
	(0.084)	(0.075)	(0.061)	(0.067)
$\hat{\gamma}_i$	0.041**	0.045**	0.036**	0.039**
	(0.013)	(0.015)	(0.012)	(0.014)
constant	0.004	0.470	-0.046***	0.698
	(0.015)	(1.225)	(0.012)	(1.032)
control variables	N	Y	N	Y
R^2	0.019	0.036	0.003	0.061
N	1241	1241	1241	1241

Table A15. Stock Shares Versus Cardinal Proxies for Preferences and Beliefs. Households with directly held stocks

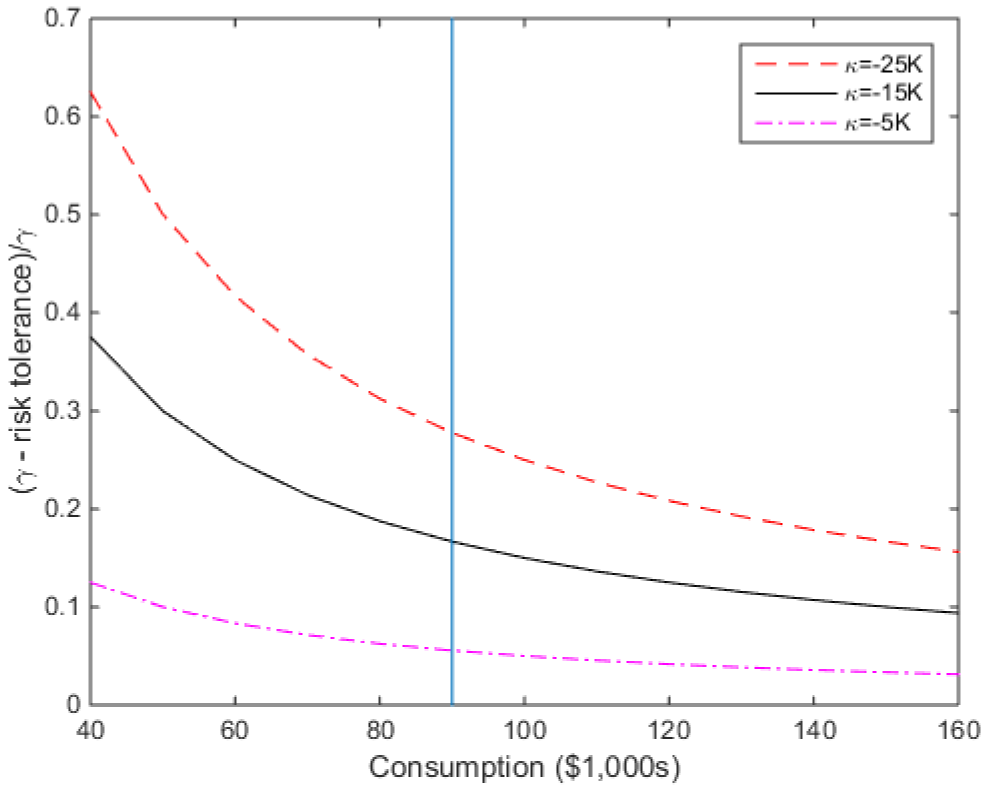
	LHS variable: survey measure of stock share in total financial wealth		LHS variable: administrative measure of stock share in Vanguard	
	(1)	(2)	(3)	(4)
$\hat{\mu}_i$	0.051*	0.067**	0.045*	0.039*
	(0.024)	(0.023)	(0.020)	(0.019)
$\hat{\sigma}_i$	-0.147	-0.136	-0.169	-0.095
	(0.156)	(0.126)	(0.149)	(0.107)
$\hat{\gamma}_i$	0.023	0.022	-0.001	0.011
	(0.017)	(0.024)	(0.030)	(0.036)
constant	0.070***	1.321	0.045*	3.797**
	(0.018)	(1.771)	(0.018)	(1.600)
control variables	N	Y	N	Y
R^2	0.013	0.026	0.011	0.042
N	639	639	639	639

Table A16. Observed stock shares and theoretically optimal stock shares. Attenuation to belief heterogeneity.

	Survey Stock Share		Administrative Stock Share	
	(1)	(2)	(3)	(4)
λ	0.063***	0.059***	0.052***	0.049***
	(0.010)	(0.010)	(0.008)	(0.008)
control variables	N	Y	N	Y
R^2	0.015	0.042	0.012	0.038
N	4414	4414	4414	4414
p-value of Wald test on restriction	0.605	0.520	0.386	0.644

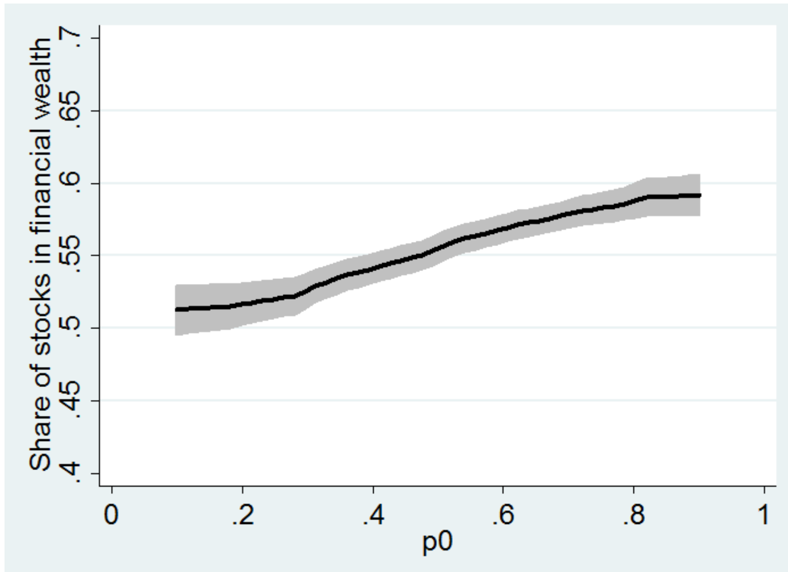
Notes. Regression results from equation (12), imposing $\beta_1^0 = 1$, $\beta_2^0 = -2$, and omitting risk tolerance term ($\lambda\beta_3^0 \frac{\gamma_i - \overline{\gamma}}{\overline{\gamma}}$). The correlation between the distribution of the risk tolerance parameter and that of belief parameters is estimated to be negligible (see Table A5 and A6), so omitting risk tolerance term does not affect inferences on the effect of belief heterogeneity. λ in this exercise can be interpreted as the attenuation factor to belief heterogeneity only. Bootstrap standard errors in parentheses. *, **, and *** implies significance at 5%, 1% and 0.1% level, respectively.

Figure A1. Difference between relative risk tolerance and γ (as a fraction of γ) over different levels of consumption and κ.

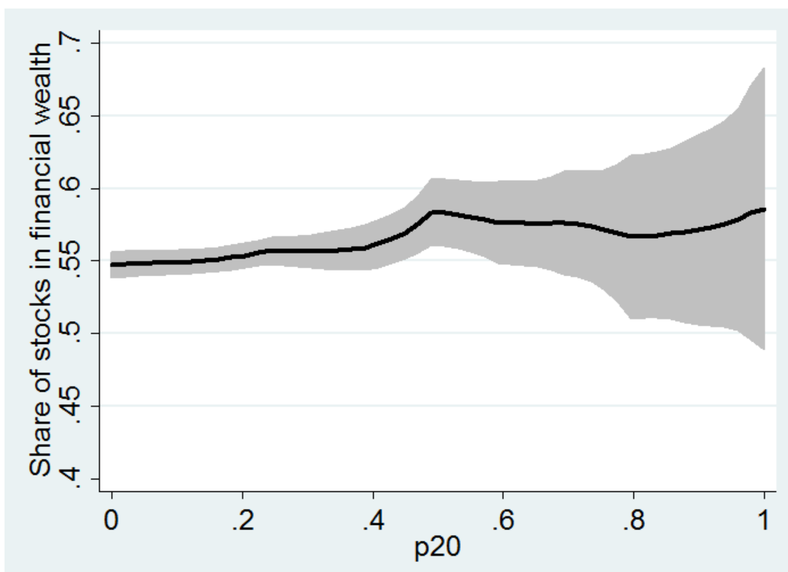


Notes.
The vertical line shows the mean level of household income in the VRI (before retirement), to approximate the average level of household consumption.

Figure A2. Bi-variate non-parametric regression of stock share in total financial wealth on each probability questions on stock market expectation



p0



p20

Appendix B. Details on Structural Estimation Procedure

The distributions of the true latent variables are assumed as (8), (9) and (10) in the main text:

$$\begin{matrix} \mu_i = \bar{\mu} + u_{\mu i} \\ \sigma_i = \bar{\sigma} + u_{\sigma i} \end{matrix}, \begin{pmatrix} u_{\mu i} \\ u_{\sigma i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u\mu}^2 & \rho_{\mu\sigma}\sigma_{u\mu}\sigma_{u\sigma} \\ . & \sigma_{u\sigma}^2 \end{pmatrix} \right) \quad \text{(B.1)}$$

$$\log(\gamma_i) = \bar{\gamma} + u_{\gamma i},\ u_{\gamma i} \sim N(0, \sigma_{u\gamma}^2) \quad \text{(B.2)}$$

We allow the beliefs about returns to depend on risk preference, so the covariates of $\bar{\mu}$ and $\bar{\sigma}$ include the latent γ_i.

These latent variables are related to observed survey responses in the following way.

$$\begin{matrix} \log(\tilde{\gamma}_{ij}) = \log(\gamma_i) + \varepsilon_{\gamma ij} \qquad \text{for}\ j = 1,2 \\ \varepsilon_{\gamma ij} \sim N(0, \sigma_{\varepsilon\gamma j}^2) \end{matrix} \quad \text{(B.3)}$$

$$\frac{(W+\kappa)^{1-1/\tilde{\gamma}_i}}{1-1/\tilde{\gamma}_i} \ \text{vs.}\ 0.5\frac{(2W+\kappa)^{1-1/\tilde{\gamma}_i}}{1-1/\tilde{\gamma}_i} + 0.5\frac{((1-x)W+\kappa)^{1-1/\tilde{\gamma}_i}}{1-1/\tilde{\gamma}_i} \quad \text{(B.4)}$$

$$\tilde{m}_i = \mu_i + \varepsilon_{mi},\ \varepsilon_{mi} \sim N(0, \sigma_{\varepsilon m}^2) \quad \text{(B.5)}$$

$$\tilde{p}_{0i} = \Phi(\frac{\mu_i}{\sigma_i} + \varepsilon_{0i}),\ \varepsilon_{0i} \sim N(\psi, \sigma_{\varepsilon p}^2) \quad \text{(B.6)}$$

$$\tilde{p}_{20i} = \Phi(\frac{\mu_i - 0.2}{\sigma_i} + \varepsilon_{20i}),\ \varepsilon_{20i} \sim N(0, \sigma_{\varepsilon p}^2) \quad \text{(B.7)}$$

The variables $\tilde{m}_i$, $\tilde{p}_{0i}$, and $\tilde{p}_{20i}$ are before rounding. Actual survey response m_i is a rounded version of $\tilde{m}_i$ as m_i is restricted to take an integer value. Survey responses p_{0i} and p_{20i} are to take a value from the set {0,5,10,15,25,35,…,75,85,90,95,100}, we assume that $\tilde{p}_{0i}$ and $\tilde{p}_{20i}$ are rounded to the closest values allowed for each response. Also note that the survey does not allow for p_{20i} to be larger than p_{0i}. Hence when we observe $p_{20i} = p_{0i}$, we consider the possibility that the survey response error actually generated $\tilde{p}_{20i} > \tilde{p}_{0i}$ but after imposing the constraint we observe the equality in the actual responses. Together with interval responses, these formulae tell the range of survey response error terms that generate the responses of individual i that we observe, given μ_i, σ_i, and γ_i. The parameter values governing the distribution of the survey response errors allow us to calculate the conditional probability of the joint responses.

The parameters to be estimated are $\bar{\gamma}, \bar{\mu}, \bar{\sigma}, \psi, \rho_{\mu\sigma}, \sigma^2_{u\mu}, \sigma^2_{u\sigma}, \sigma^2_{u\gamma}, \sigma^2_{\varepsilon m}, \sigma^2_{\varepsilon p}, \sigma^2_{\varepsilon\gamma 1}, \sigma^2_{\varepsilon\gamma 2}$. We allow for $\bar{\gamma}, \bar{\mu}, \bar{\sigma}$, and ψ to vary with covariates.

Algorithm of likelihood function calculation

We use the Gaussian quadrature approximation of the normal distribution to numerically integrate the density functions over multiple dimensions. Let θ be the vector of parameters. Given a fixed θ_0 the likelihood function is calculated through the following algorithm:

(1) Based on the parameter values that govern the true belief and preference parameter distributions in θ_0, and using Gaussian Quadrature approximation, generate K nodes $\{\mu_k, \sigma_k, \gamma_k\}_{k=1}^{K}$ of belief and preference parameters, with corresponding probabilities $\{\pi_k\}_{k=1}^{K}$ such that $\sum_{k=1}^{K} \pi_k = 1$.

(2) For each $\{\mu_k, \sigma_k, \gamma_k\}$ and each individual, calculate $[\varepsilon_{mi}^{low}, \varepsilon_{mi}^{high}]$, $[\varepsilon_{0i}^{low}, \varepsilon_{0i}^{high}]$, $[\varepsilon_{20i}^{low}, \varepsilon_{20i}^{high}]$, $[\varepsilon_{\gamma 1i}^{low}, \varepsilon_{\gamma 1i}^{high}]$, $[\varepsilon_{\gamma 2i}^{low}, \varepsilon_{\gamma 2i}^{high}]$ such that survey response error terms realized in these ranges generate the observed responses after rounding and corresponding constraints.

(3) For each $\{\mu_k, \sigma_k, \gamma_k\}$ and each individual, calculate the joint likelihood of the realization of the error terms in the range found in (2), using Gaussian CDF under the parameter values governing the error term distributions in θ_0. Let π_{ki}^{ε} denote this joint likelihood.

(4) The likelihood for each individual is calculated as integration over *k* nodes as following:

$$L_i = \sum_{k=1}^{K} \pi_{ki}^{\varepsilon} \pi_k \tag{B.8}$$

Then the joint likelihood is calculated as products of L_i over individuals.

Calculation of the proxy variables

Under the estimated parameters, the proxy variables are calculated as expected values conditional on the observed responses. The individual-specific proxy variables are obtained using the econometric model outlined above. The likelihood function of the model specifies the probability distribution of the observed responses conditional on the latent beliefs and preferences. The distribution of the latent variables conditional on the observed responses can be obtained from the likelihood function using Bayes' theorem. Integrating out this function yields the individual-specific proxy variables ($\hat{\mu}_i$, $\hat{\sigma}_i$ and $\hat{\gamma}_i$) as the conditional expectations of the latent variables given the observed survey responses. We use the same numerical approximation for this calculation. Using the Bayes' Rule, $\hat{\theta}_i$ is calculated as:

$$\hat{\theta}_i \equiv E[\theta_i \mid m_i, p_{0i}, p_{20i}, SSQ_{1i} SSQ_{2i}] = \frac{1}{L_i} \sum_{k=1}^{K} \theta_k \pi_{ki}^{\varepsilon} \pi_k \ . \tag{B.9}$$

Appendix C. Details on Structural Life-Cycle Model of Portfolio Choice

Health Transition and Preferences The model starts from age 55, which is the lowest value in the VRI, and the household can live up to age 110 at most.[1] The probability of survival up to next period $(1-\pi_D)$ is a function of age. The household evaluate flow utility from the consumption using (1). It discounts next period utility by time discount factor β. When it dies, it leaves the bequest, and bequest utility is modeled as:

$$U_{Beq,i}(B) = \theta_{Beq} \frac{(B+\kappa_{Beq})^{1-1/\gamma_i}}{1-1/\gamma_i} \tag{C.2}$$

where θ_{Beq} determines the strength of the bequest motive and κ_{Beq} determines whether it is necessity or luxury, compared to its own consumption.

Labor Income Process The household retires at age 65. Until then, the labor income is exogenously determined as:

$$\log(Y_{it}) = \log(\overline{y}_i) + \nu_{it}, \;\; \nu_{it} \sim N(0, \sigma_\nu^2) \text{ for } t < 65. \tag{C.3}$$

Given that households have only 10 years until retirement in this model, we abstract from permanent income shocks. After retirement, the household receives annuity income which captures Social Security income and defined benefit pension income and hence is not exposed to any uncertainty. This annuity income is modeled as a fraction (λ) of the mean income before retirement:

$$\log(Y_{it}) = \log(\lambda) + \log(\overline{y}_i) \text{ for } t \geq 65. \tag{C.4}$$

[1] To avoid the complications arising from the joint survival process, we assume that the household dies when the head dies. Essentially, the model is looking at the single households' portfolio choice. Stock share regression using singles only give the essentially the same results as our baseline results using the full sample.

Financial Assets Households can invest in two different assets, a riskless asset and a risky asset where the latter represents stocks. The gross real return on a risk free asset is set as a constant $\bar{R}_f$. The subjective belief on distribution of the real gross return on a risky asset, R_t, is modeled as:

$$R_{t+1,i} = \mu_i + \eta_{t+1},\ \ \eta_{t+1} \sim N(0, \sigma_i^2) \tag{C.5}$$

where η_{t+1} is an i.i.d. stock return shock. Note that this subjective belief process is heterogeneous across households. We assume that the aggregate stock return shock is uncorrelated with the idiosyncratic labor income shock, following Cocco, Gomes and Maenhout (2005).

Optimization problem of the households Let W_{it} be beginning-of-period cash in hand of a household and α_{it} be share of savings of this period invested to stocks. We assume that short sales and leveraged stock holdings are not allowed.[2] Then the household solves the following optimization problem (we drop the subscripts *i* and *t*):

$$\begin{aligned} V(W,t) = \max_{C,W',\alpha}\ &\{U(C) + \beta E[(1-\pi_D(t))V(W',t+1) + \pi_D(t)U_{Beq}(W')]\} \\ s.t.\ \ &\mathrm{W}' = [(W-C)((1-\alpha)\bar{R}_f + \alpha R_s)] + y' \\ &\mathrm{C} \le W \\ &\alpha \in [0,1] \end{aligned} \tag{C.6}$$

Computation We solve for the optimal policy function numerically using backward induction. The last period (at age 110) maximization is a static one so the value function is trivially

[2] Optimal stock share could go over 100% if we allowed leveraging, since labor earnings and retirement income are close substitutes to the risk-free asset, due to zero correlation with stock return for the former and the absence of risk for the latter. In addition, when we approximate the labor income process as a discrete process, even the worst possible realization of income guarantees positive resources net of the subsistence level of consumption (as in Cocco, Gomes and Maenhout (2005)) since mean level of labor income is much higher than the subsistence level of consumption.

obtained. This value function is used as a continuation value for the maximization program of the penultimate period. We repeat this until we solve for the maximization problem at the first period. For the choice over continuous spaces, i.e. over C and α, the optimization is done using grid search. With the curvature parameters the problem is no more homogenous to the scale, so it cannot be normalized as typically done in the literature (see Cocco, Gomes and Maenhout (2005) and Pang and Warshawsky (2010) for example). This does not increase computational burden too much since we abstract from permanent income shocks.

<u>Calibration</u> We solve this model for various sets of subjective belief and risk tolerance parameter values that are in the range supported by the evidence from the VRI, to understand the effects of heterogeneous belief and preference on the optimal stock share. The curvature parameter for the ordinary utility function (κ) is fixed at the value estimated from the VRI (-17K). Time discount factor (β) is set to be 0.96, a value that is typically used in the literature for annual models.

The probability of survival π_D is estimated from the HRS (1994 – 2010). For the parameters for the bequest utility function we use the median values ($\theta_{Beq} = 32,\ \kappa_{Beq} = 64K$) from Ameriks et al. (2015) who estimate heterogeneity in preferences regarding long-term care expenditure and bequests. The parameters imply that a bequest is a luxury good compared to the ordinary consumption, but once the bequest motive kicks in for wealthy households the marginal utility from leaving bequest is large. Risk free return ($\overline{R}_f$) is set to be 1.02. In the baseline model we use \$90,000 for the mean income before retirement ($\overline{y}$) and assume 0.5 for the replacement rate after retirement (λ). These values are close to means from the VRI data. The variance of transitory income shocks (σ_v^2) is set to be 0.07, which is close to the value used in Cocco et al. (2005).[3]

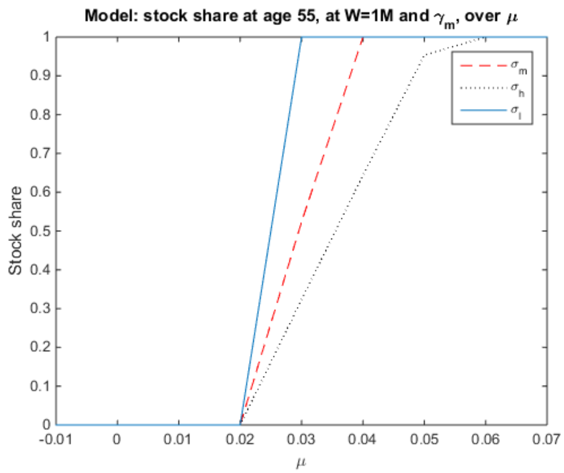
Table C1 summarizes the calibration of the parameters, and figure C1 summarizes the results.

Table C1. Calibration of Parameters for the Life-Cycle Model

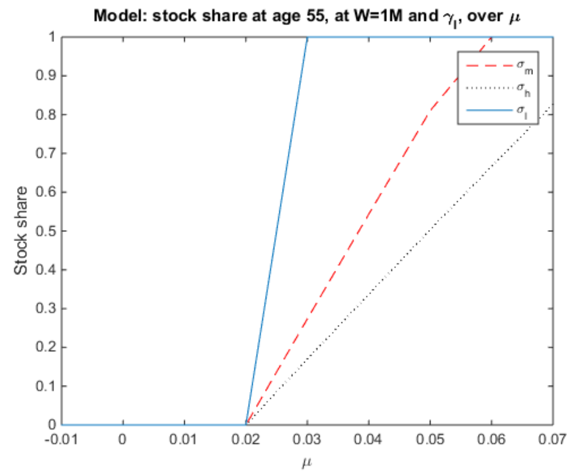
Parameters	Value	Target/Source
κ	-17K	VRI estimation
β	0.96	Standard
π_D		HRS estimation
$\overline{R}_f$	1.02	Cocco, Gomes and Maenhout (2005)
θ_{Beq}	32	Ameriks et al. (2015)
κ_{Beq}	64K	Ameriks et al. (2015)
$\overline{y}$	$80,000	VRI data
λ	0.5	VRI data
σ_v^2	0.07	Cocco, Gomes and Maenhout (2005)

[3] They estimated it to be 0.058 for college graduates. We set it slightly larger here given that our model does not have permanent income shocks.

Figure C1. Stock share and the expected value of stock returns (μ) at different levels of the standard deviation of stock returns (σ) and risk tolerance (γ). Results from the life cycle portfolio choice model.

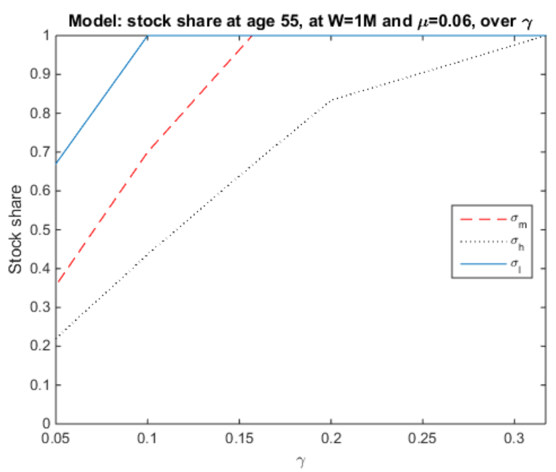


Portfolio choice model,
medium level of risk tolerance (γ = 0.32)

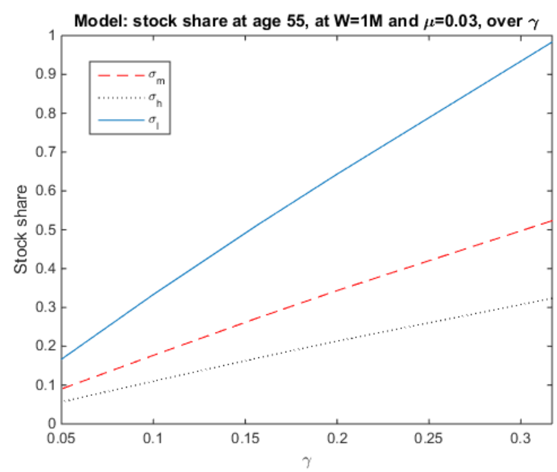


Portfolio choice model,
low level of risk tolerance (γ = 0.16)

Figure C2. Stock share and the risk tolerance (γ) at different levels of the standard deviation of stock returns (σ) and expected value of stock returns (μ). Results from the life cycle portfolio choice model.



Portfolio choice model,
medium level of expected return (μ = 0.06)



Portfolio choice model,
low level of expected return (μ = 0.03)