Stock market expectations and portfolio choice of American households

Work in progress

Gábor Kézdi

(Central European University)

Robert J. Willis

(University of Michigan)

November 19, 2007

- Motivation 1: the stockholding puzzle
  - Despite historically high mean and mild variance of returns
  - Households invest too little in stock-market based assets, and many hold no such assets at all
  - Campbell (2006), Mankiw and Zeldes (1991), Haliassos and Bertaut (1995)
  - Related to the equity premium puzzle
  - Mehra and Prescott (1985), Kocherlakota (1996)
- Motivation 2: Survey measures of expectations
  - Subjective probabilities
  - Manski (2004); Hurd, van Rooij, Winter (2005)
  - Preferences vs. expectations
  - Kimball, Sahm and Shapiro (2007)

- Contribution 1: expectations explain much of the puzzle
    - We show that on average, expectations have lower mean and higher variance than historical moments
    - provides empirical support to Weitzmann (2007)
    - We show substantial heterogeneity in expectations
    - provides empirical support to disagreement models (Hong and Stein, 2007)
- Contribution 2: explicit model of survey response
    - Investment decisions and answering survey questions are different situations
    - We model both explicitly
    - Model of survey response is consistent with observed noise phenomena
    - We estimate a structural measurement model derived from those models

- The measurement problem
- Data
- Descriptive analysis
    - Signal in survey measures of expectations
    - Noise in survey measures of expectations
- Behavioral models
    - Standard model of portfolio choice
    - Implications of heterogeneity in expectations
    - A model of survey response to probability questions
    - Two versions
- Estimating unconditional moments of expectations
- Estimating predictors of expectations
    - Structural measurement model
    - Jointly with stockholding
- Conclusions

- Measure people's expectations about the one-year return on stock market index
- Assume everyone believes returns are i.i.d. normal
- But mean and variance are individual-specific

Yearly returns on S&P 500, 1950 to 2005. Normal density superimposed.

- Historical mean 0.09, historical standard deviation 0.15

Data: Health and Retirement Study (HRS) 2002

- Large set of variables
  - Expectations
  - Household assets
  - Demographics
  - Cognitive scores
  - Depression scale
  - Risk tolerance estimates by Kimball, Sahm and Shapiro (2007)
- Restricted sample
  - Financial respondents age 55 to 65
  - Expectations at individual level, investments at household level
  - Still in asset accumulation phase
- Total sample size $3,715$

The stockholding puzzle

- Direct stockholder
  - Owns stock-market based assets outside retirement accounts
- Indirect stockholder
  - Owns stock-market based assets but only on retirement accounts

Households in main sample by stockholding status (%)

	2002	2004
Direct stockholder	31	30
Indirect stockholder	13	12
Not stockholder	56	58
All	100	100

The main questionnaire of HRS 2002 has two questions

1. p_0

   By next year at this time, what is the percent chance that mutual fund shares invested in blue chip stocks like those in the Dow Jones Industrial Average will be worth more than they are today?

2. p_{10}

   By next year at this time, what is the chance they will have grown by 10 percent or more?

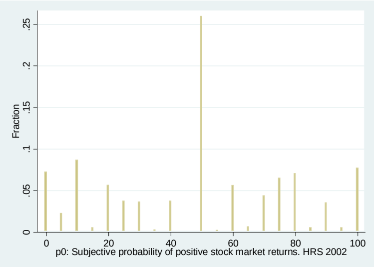
- Same questions asked from small subsample once more at the end of the survey ($n = 292$)
  - We denote module questions as p'_0 and p'_{10}
- p_0 was asked in HRS 2004 as well

Probability answers (%) by subsample. HRS 2002 & 2004

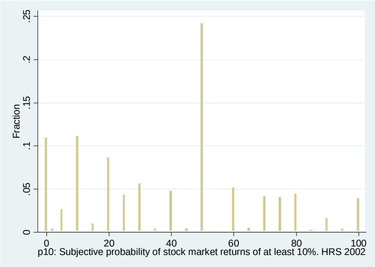
	Main sample		
	p_0	p_{10}	p_{0_2004}
Mean	49	39	52
Sd	30	28	26
Missing	0.18	0.18	0.13

	Spouses			Older respondents		
	p_0	p_{10}	p_{0_2004}	p_0	p_{10}	p_{0_2004}
Mean	46	38	49	39	31	44
Sd	28	27	26	31	28	28
Missing	0.21	0.21	0.14	0.32	0.32	0.29

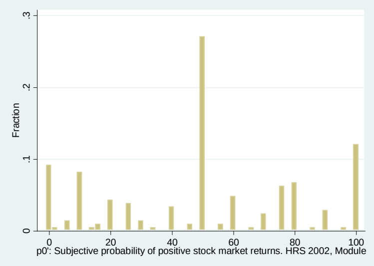
- Probabilities implied by historical distribution
  - $p_0^{hist} = 73$
  - $p_{10}^{hist} = 47$


Fraction
0
.05
.1
.15
.2
.25
0
20
40
60
80
100
p0: Subjective probability of positive stock market returns. HRS 2002

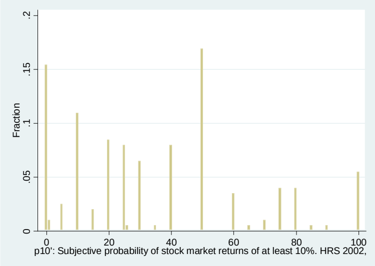
p0


Fraction
0
.05
.1
.15
.2
.25
0
20
40
60
80
100
p10: Subjective probability of stock market returns of at least 10%. HRS 2002

p10


Fraction
0
.1
.2
.3
0
20
40
60
80
100
p0': Subjective probability of positive stock market returns. HRS 2002, Module

p0'


Fraction
0
.05
.1
.15
.2
0
20
40
60
80
100
p10': Subjective probability of stock market returns of at least 10%. HRS 2002,

p10'

Reported probabilities (%) by stockholder status

	2002			2004
	$\bar{p}_0$	$\bar{p}_{10}$	$\bar{p}_0 - \bar{p}_{10}$	$\bar{p}_0$
Direct stockholder	55	44	11	59
Indirect stockholder	52	42	10	56
Not a stockholder	41	34	7	46
All	48	39	9	100

- Probabilities implied by historical distribution (%)
  - $p_0^{hist} = 73$
  - $p_{10}^{hist} = 47$
  - $p_0^{hist} - p_{10}^{hist} = 25$

Four types of noise in the probability answers.

- Missing answers
  - 18 per cent of respondents
- Round answers and focal answers at 0, 50, 100
  - 25-30 per cent of all answers are focal at 50
- Zero and negative probability mass between p_0 and p_{10}
  - For 43% of responses, $p_0 = p_{10}$;
  - For 14% of responses, $p_0 < p_{10}$
- Test-retest noise
  - Core and module answers are similarly distributed
  - Large differences

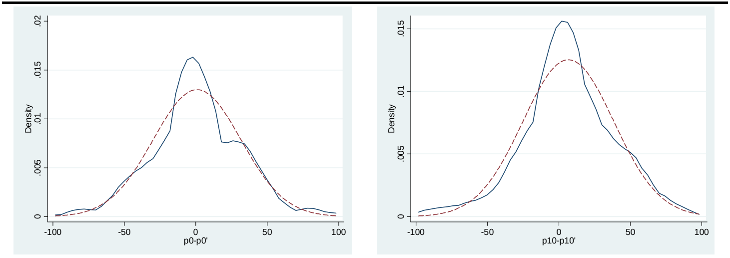
$$\begin{aligned} Corr\left(p_{0i}, p'_{0i}\right) &= 0.48 \\ Corr\left(p_{10i}, p'_{10i}\right) &= 0.35 \end{aligned}$$

	Missing p0		p0 focal at 0 or 100		p0 focal at 50		p0 other round	
Stockholding	-0.22		-0.01		-0.02		0.04	
	[0.01]**		[0.01]		[0.02]		[0.02]*	
Education		-0.01		0		-0.01		0.01
		[0.00]**		[0.00]		[0.00]*		[0.00]
Word recall		-0.02		0		-0.03		0.02
		[0.01]*		[0.01]		[0.01]**		[0.01]*
Counting back		-0.06		-0.02		0.03		0
		[0.01]**		[0.01]**		[0.01]**		[0.01]
Female		0.06		-0.02		0.08		-0.04
		[0.01]**		[0.02]		[0.02]**		[0.02]
Black		0.03		0		-0.04		0.03
		[0.02]		[0.02]		[0.02]		[0.03]
Hispanic		0.18		-0.01		-0.04		0.1
		[0.03]**		[0.03]		[0.04]		[0.04]*
Log wealth		-0.01		0.00		0.00		0.00
		[0.00]**		[0.00]		[0.00]		[0.00]
Born in US		-0.07		-0.05		0		0.06
		[0.03]**		[0.03]		[0.03]		[0.04]
Ever military		0.01		0.05		0		-0.03
		[0.01]		[0.02]**		[0.02]		[0.03]
Constant	0.27	0.44	0.16	0.24	0.27	0.28	0.5	0.37
	[0.01]**	[0.05]**	[0.01]**	[0.05]**	[0.01]**	[0.06]**	[0.01]**	[0.07]**
Observations	3715	3715	3049	3049	3049	3049	3049	3049
R-squared	0.08	0.17	0.00	0.01	0.00	0.01	0.00	0.01

	p0 = p10		p0 < p10		|p0 – p0_mod|	
Stockholding	-0.07		-0.04		-1.83	
	[0.02]**		[0.01]**		[3.12]	
Education		-0.01		-0.01		-0.91
		[0.00]*		[0.00]*		[0.76]
Word recall		0		-0.01		-0.78
		[0.01]		[0.01]		[1.79]
Counting back		-0.01		0		2.8
		[0.01]		[0.01]		[2.32]
Female		0.07		0.03		7.65
		[0.02]**		[0.02]*		[3.91]
Black		0.04		0.01		-11.18
		[0.03]		[0.02]		[4.13]**
Hispanic		-0.08		0.06		-0.38
		[0.04]		[0.03]		[7.49]
Log wealth		-0.01		0.00		-0.63
		[0.00]*		[0.00]		[0.47]
Born in US		-0.05		0.02		-1.61
		[0.04]		[0.03]		[5.39]
Ever military		0.03		0.01		4.49
		[0.03]		[0.02]		[4.32]
Constant	0.46	0.6	0.16	0.18	23.25	37.95
	[0.01]**	[0.07]**	[0.01]**	[0.05]**	[2.31]**	[13.53]**
Observations	3049	3049	3049	3049	185	185
R-squared	0.00	0.02	0.00	0.01	0.00	0.05

Do noise features contain relevant information?

- Missing answers
  - yes
  - likely to reflect genuine ignorance (extreme uncertainty)
- Round answers and focal answers at 0, 50, 100
  - not much (50 may)
- Zero and negative probability mass between p_0 and p_{10}
  - some information but weak
- Test-retest noise
  - Looks like pure survey noise
  - Looks pretty close to normal



Off-the shelf classical model of Merton (1969)

- Setup:
  - Given wealth at $t = 0$, max expected utility of wealth at $t = T$
  - $u(W_t) = \frac{W_t^{1-1/\alpha}}{1-1/\alpha}$
  - Safe asset has yearly return r
  - Value of risky asset follows random walk w/drift, cont's time: $\frac{dS}{S} = \mu \; dt + \sigma \; dz$
  - Returns are i.i.d. normal with mean μ and standard deviation σ
- Solution: optimal fraction of wealth in risky asset
  - $s_t^* = s^* = \alpha \frac{\mu - r}{\sigma^2}$
  - Constant in time, determined only by risk aversion and returns (mean and variance of risky asset's returns)

If a model of demand, all parameters may be heterogeneous:

$$s_i^* = \alpha_i \frac{\mu_i - r_i}{\sigma_i^2}$$

- Heterogeneity unrestricted if interpreted as a model of demand
  - Equilibrium questions not asked
- We focus on heterogeneity in expectations $(\mu_i,\ \sigma_i^2)$
  - May be the consequence of learning, from stock market events (Brennan, 1998) or other
  - May be "erroneous" or "biased"
- We do not focus on heterogeneity in risk aversion
  - Preferences (estimated for HRS respondents by Kimball, Sahm, Shapiro, 2007)
  - May be endogenous (Dohmen, Falk, Huffman & Sunde, 2007)

At investment decision at time t, individual i retrieves

$$\begin{aligned} R_{i(t+1)} &= \mu_{it} + \eta_{it} \\ \eta_{it} | \mu_{it} &\sim N\left(0, \sigma_{it}^2\right) \end{aligned}$$

- $R_{i(t+1)}$: returns as i sees they could be one year from time t
- μ_{it}: mean yearly returns as seen by i at t
- η_{it}: possible random deviation from mean as seen by i at t

For some of the analysis we decompose η

$$\begin{aligned} \eta_i &= \delta_{it} + \varepsilon_{t+1} \\ (\delta_{it}, \varepsilon_{t+1}) &\sim N\left(\mathbf{0}, \left\langle \sigma_{\delta it}^2, \sigma_{\varepsilon}^2 \right\rangle\right) \end{aligned}$$

- ε_{t+1}: risk, faced by everybody
- δ_{it}: uncertainty specific to individual i

- We refer to R as
  - *fundamental subjective returns*, or
  - *fundamental expectations about the returns*
- We refer to heterogeneity in R_i as *relevant heterogeneity*

Investment situations are different from answering survey questions

- At investment: strong motivation, enough time, access to help
- At survey: no motivation, little time, no access to help

- When making an investment decision
  - People form fundamental expectations $R_{i(t+1)}$
  - Calculate statistics relevant for the decision $\left(\mu_i, \sigma_i^2\right)$
- When answering the HRS probability questions
  - People retrieve a noisy version of fundamental expectations,

$$\tilde{R}_{i(t+1)}$$

  - Calculate probabilities in a fast-and-frugal way

    Note: probabilities are statistics that are not directly relevant for the investment decision

When confronted with probability question j at survey t, individual i retrieves

$$\tilde{R}_{i(t+1)j} = \mu_{it} + \eta_{it} + v_{itj},$$

v_{ijt}: question-specific survey noise.

- We refer to $\tilde{R}_{ij}$ as
  - *noisy subjective returns*, or
  - *noisy expectations about the returns*
- We refer to heterogeneity in $\tilde{R}$ as *measured heterogeneity*

Survey noise v_j is classical (in R)

- Independent of relevant heterogeneity (both measured and unmeasured)
- Core draws of v are independent of module draws
- Adjacent draws $(p_0$ versus $p_{10})$ may be correlated

$$\begin{aligned} v_{ij}|\mu_i, \eta_i &\sim N\left(0, \sigma_v^2\right) \\ Corr\left(v_0, v_{0'}\right) &= Corr\left(v_{10}, v_{10'}\right) = 0 \\ Corr\left(v_0, v_{10}\right) &= Corr\left(v_{0'}, v_{10'}\right) = \rho_v \end{aligned}$$

- $\rho_v < 1$ if there is (random) lack of attention
  - a way to accommodate $p_0 \leq p_{10}$ (and $p'_0 \leq p'_{10}$)

- Given $\tilde{R}_{ij}$, people have to answer probability question p_{ij}
  - Omit survey index t and t+1 for simplicity

Two alternative models

- A simple but ad-hoc model
  - People give answers that are in the same interval as a precisely calculated probability could be
- A more structural model with modal response
  - Following Hill, Perry and Willis (2006).
  - People form a distribution of possible answers
  - Of that distribution they pick the mode for an answer
- Both model is consistent with
  - less time and incentives in a survey situation than at investment
  - observed noise features of rounding and focal answers

The benchmark model

- The precise probability from $\tilde{R}$ at question j with threshold τ_j

$$
\begin{aligned}
p_{ij}^* &= \Pr\left(\tilde{R}_{ij} > \tau_j | \mu_i, v_{ij}\right) = \Pr\left(\frac{\eta_i}{\sigma_i} > \frac{\tau_j - \mu_i - v_{ij}}{\sigma_i}\right) \\
&= \Phi\left(\frac{\mu_i + v_{ij} - \tau_j}{\sigma_i}\right)
\end{aligned}
$$

- More uncertainty leads pushes answers to the middle in a monotonic fashion
  - as $\sigma_i \to \infty$, $p_{ij}^* \to 0.5$

Assume interval response

$$
p_{ij} \in [\underline{b}, \bar{b}] \Leftrightarrow p_{ij}^* \in [\underline{b}, \bar{b}] \Leftrightarrow \underline{b} \leq \Phi\left(\frac{\mu_i + v_{ij} - \tau_j}{\sigma_i}\right) < \bar{b}
$$

Intervals exogenously given at as $[0, 5)$, $[5, 15)$, ..., $[95, 100]$

The modal response model

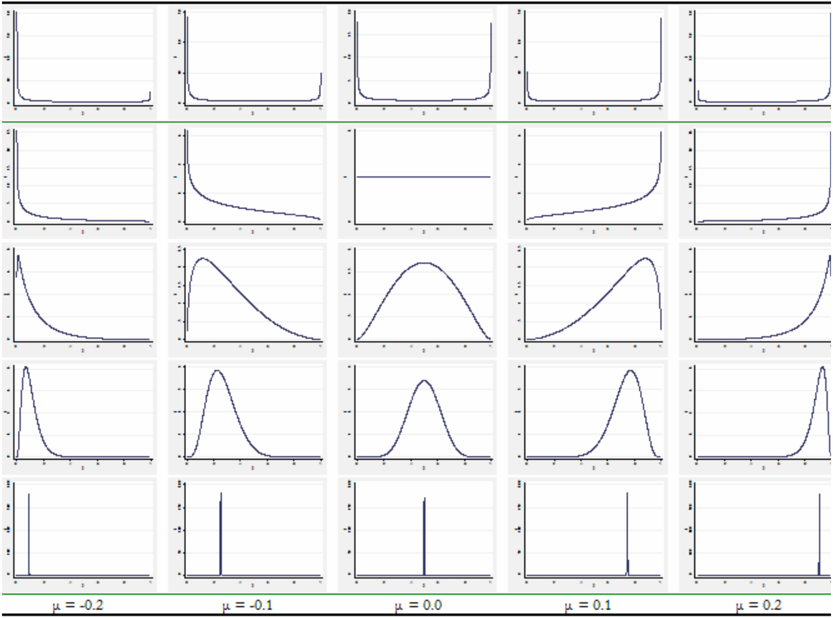
- People make a guess
- The guess is modeled as picking a possible answer from the distribution of potential answers

$$q_{ij} = \Pr\left(\tilde{R}_{ij} > \tau_j | \mu_i, \delta_i, v_{ij}\right) = \int_{-\infty}^{\tau_j} \left(\mu_i + v_{ij} + \delta_i + \varepsilon\right) dF\left(\varepsilon\right)$$

  - $\tau_0 = 0, \quad \tau_{10} = 0.1$

- From its distribution with c.d.f. G and p.d.f. g

$$\begin{aligned} G_i\left(q_{ij}\right) &\equiv \Pr\left[\Pr\left(\tilde{R}_{ij} > \tau_j | \mu_i, \delta_i, v_{ij}\right) < q_{ij} | \mu_i, v_{ij}\right] \\ g_i\left(q_{ij}\right) &\equiv \frac{\partial G_i\left(q_{ij}\right)}{\partial q_{ij}} = \frac{\sigma_\varepsilon}{\sigma_{\delta i}} \frac{\phi\left[\frac{\tau_j - \mu_i - v_{ij} + \sigma_\varepsilon \Phi^{-1}\left(p_j\right)}{\sigma_{\delta i}}\right]}{\phi\left(\Phi^{-1}\left(p_j\right)\right)} \end{aligned}$$


μ = -0.2
μ = -0.1
μ = 0.0
μ = 0.1
μ = 0.2

- We assume that people pick the mode: $p_{ij} = \mathrm{mod}\,(q_{ij})$
  - a likely "true" q, easy to locate given mental image of g
- The mode is the maximum of g. The first-order condition is

$$p_{ij}^{*} \equiv p_{ij\_FOC} = \Phi\left[(\tau_j - \mu_i - v_{ij})\,\frac{\sigma_\varepsilon}{\sigma_{\delta i}^2 - \sigma_\varepsilon^2}\right]$$

- The maximum is p^* if g concave. Otherwise more complicated

$$\begin{aligned}
p_{ij} &= p_{ij}^{*} \quad \mathit{if} \quad \sigma_{\delta i}^2 < \sigma_\varepsilon^2 \quad \& \quad 0.005 \leq p_{ij}^{*} < 0.995 \\
&= 0 \quad \mathit{if} \quad \left[\sigma_{\delta i}^2 < \sigma_\varepsilon^2 \quad \& \quad p_{ij}^{*} < 0.005\right] \\
&\qquad \mathit{OR} \quad \left[\sigma_{\delta i}^2 \geq \sigma_\varepsilon^2 \quad \& \quad p_{ij}^{*} \geq 0.995\right] \\
&= 1 \quad \mathit{if} \quad \left[\sigma_{\delta i}^2 < \sigma_\varepsilon^2 \quad \& \quad p_{ij}^{*} \geq 0.995\right] \\
&\qquad \mathit{OR} \quad \left[\sigma_{\delta i}^2 \geq \sigma_\varepsilon^2 \quad \& \quad p_{ij}^{*} < 0.005\right] \\
&= 0.5 \quad \mathit{if} \quad \sigma_{\delta i}^2 \geq \sigma_\varepsilon^2 \quad \& \quad 0.005 \leq p_{ij}^{*} < 0.995
\end{aligned}$$

Survey nonresponse

- Recall that 18 per cent of respondents do not answer the stock market expectation questions
- Nonresponse is likely to reflect genuine ignorance
- Morel here simply as

$$\sigma_{it} \longrightarrow \infty$$

- So that

$$s^* \longrightarrow 0$$

- Won't focus more on nonresponse formally

Estimate unconditional moments of relevant heterogeneity and noise by GMM (intuition from MM)

$$E(\mu_i), \quad E(\sigma_i), V(\mu_i), \quad V(\sigma_i)$$
$$\sigma_v^2, \quad \rho$$

- Use moments of all four probability answers $p_0, p_{10}, p_{0'}, p_{10'}$
- Use minimal assumptions
  - Normality of $\tilde{R}$ conditional on μ_i and σ_i
  - Normality of v_{ij}, correlation assumptions as above
  - Independence of μ_i and σ_i
- Two goals
  - Explain average stockholding
  - Obtain noise parameters for future calibration

- Expected values identified from expected values of levels and differences of p_{ij}

$$
\begin{aligned}
E\left[\frac{\mu_i}{\sigma_i}\right] &= E\left[\Phi^{-1}\left(p_{ij}\right)\right] + E\left[\frac{\tau_j}{\sigma_i}\right] \\
E\left[\sigma_i\right] &= \frac{0.1}{E\left[\Phi^{-1}\left(p_{i0}\right) - \Phi^{-1}\left(p_{i10}\right)\right]}
\end{aligned}
$$

- Need to assume independence of μ_i and σ_i in order to get

$$
E\left[\mu_i\right] = E\left[\frac{\mu_i}{\sigma_i}\right] E\left[\sigma_i\right]
$$

- Expected values of squares identified from expected values of squares of levels and differences
- But we also need $\Pr\left[p_{i0} < p_{i10}\right]$

$$
\begin{aligned}
E\left[\sigma_i^2\right] &= \frac{2\left(1-\rho\right)\sigma_v^2 + 0.01}{E\left[\left\{\Phi^{-1}\left(p_{i0}\right) - \Phi^{-1}\left(p_{i10}\right)\right\}^2\right]} \\
\sigma_v^2 &= \frac{1}{2}E\left[\sigma_i^2\right]E\left[\left\{\Phi^{-1}\left(p_{i0}\right) - \Phi^{-1}\left(p_{i0'}\right)\right\}^2\right] \\
\rho &= 1 + \frac{0.05}{\sigma_v^2\Phi^{-1}\left[\Pr\left(p_{i0} < p_{i10}\right)\right]}
\end{aligned}
$$

- so that

$$
V\left(\sigma_i\right) = E\left(\sigma_i^2\right) - \left[E\left(\sigma_i\right)\right]^2
$$

Together with independence of μ_i and σ_i we get

$$
\begin{aligned}
E\left[\frac{\mu_i^2}{\sigma_i^2}\right] &= E\left[\left\{\Phi^{-1}\left(p_{i0}\right)\right\}^2\right] - \frac{\sigma_v}{E\left[\sigma_i^2\right]} \\
E\left[\mu_i^2\right] &= E\left[\mu_i^2/\sigma_i^2\right] E\left[\sigma_i^2\right] \\
V\left(\mu_i\right) &= E\left(\mu_i^2\right) - \left[E\left(\mu_i\right)\right]^2
\end{aligned}
$$

Need distributional assumptions in order to calibrate

$$\begin{aligned} E\left[s_i^*\right] &= E\left[\alpha \frac{\mu_i - r}{\sigma_i^2}\right] \\ \Pr\left[s_i^* > 0\right] &= \Pr\left[\alpha \frac{\mu_i - r}{\sigma_i^2} > 0\right] \end{aligned}$$

Assume

$$\begin{aligned} \mu_i &\sim N\left(E_\mu, V_\mu\right) \\ \log \sigma_i &\sim N\left(E_{\log \sigma}, V_{\log \sigma}\right) \end{aligned}$$

where $E_{\log \sigma} = \log E\left[\sigma_i\right] - \frac{1}{2} \log \left(1 + \frac{V[\sigma_i]}{E[\sigma_i]^2}\right)$ and $V_{\log \sigma} = \log \left(1 + \frac{V[\sigma_i]}{E[\sigma_i]^2}\right)$

Results

	Upper	Middle	Lower
$E[\mu_i]$	$-.07$	$-.07$	$-.07$
$E[\sigma_i]$	.41	.41	.41
$Std[\mu_i]$	.24	.39	1.17
$Std[\sigma_i]$	.26	.68	2.30
σ_v	.28	.46	1.36
ρ	.42	.38	.36
Implied probab. stockholder	.38	.43	.48
Implied avg share in stocks			
$\alpha = 1$	.24	.33	.43
$\alpha = .2$	.12	.25	.40

Compare with

- Historical $\mu = .09$ $\sigma = .15$
- Sample stockholding probab $= .50$, avg. share in stocks $= .24$

Estimation of a structural model on individual-level data

- Based on
  - Model of portfolio choice
  - Model of survey response (two versions)
- Two goals for now
  - Identifying systematic variation in expectation
  - Contrasting implied individual heterogeneity in stockholding with observed heterogeneity
- Further goals not pursued in this version
  - Evaluate alternative models of survey response
  - Establish causality of expectations

Structural equations for three latent variables

$$
\begin{aligned}
s_i^* &= \alpha \frac{(\mu_i - r)}{\sigma_i^2} + u_{si} \\
\mu_i &= \beta_\mu' x_{\mu i} + u_{\mu i} \\
\log\left(\sigma_i^2\right) &= \beta_\sigma' x_{\sigma i} + u_{\sigma i}
\end{aligned}
$$

The three latent variables are related to three observed variables

- Observed stockholding
- Observed answers to the two probability questions in the main questionnaire

Latent stockholding versus observed stockholding

Two models

- Tobit-type specification (corner solutions)

$$s_i = \left\{ \begin{array}{lll} 0 & \textit{if} & s_i^* < 0 \\ s_i^* & \textit{if} & 0 \leq s_i^* \leq 1 \\ 1 & \textit{if} & s_i^* > 1 \end{array} \right\}$$

- Probit-type specification

$$S_i = \left\{ \begin{array}{lll} 0 & \textit{if} & s_i^* \leq 0 \\ 1 & \textit{if} & s_i^* > 0 \end{array} \right\}$$

- r is calibrated

Heterogeneity in subjective mean and variance versus observed probability answers

Two models

- Benchmark (intervals of precise probabilities)

$$\begin{aligned} p_{ij} &\in [\underline{b}, \bar{b}] \Leftrightarrow p^*_{ij} \in [\underline{b}, \bar{b}] \\ &\Leftrightarrow \underline{b} \leq \Phi\left(\frac{\mu_i + v_{ij} - \tau_j}{\sigma_i}\right) < \bar{b} \end{aligned}$$

- Modal response
  - Not estimated in this version (work in progress)

- σ_v^2 and $\rho = Corr\,(v_{i0}, v_{i10})$ are calibrated
- $Corr\,(u_{si}, u_{\mu i}) = Corr\,(u_{si}, u_{\sigma i})$ are also calibrated (to be zero)

Four versions of models for s_i^*

$$
\begin{aligned}
s_i^* &= \alpha \frac{(\mu_i - r)}{\sigma_i^2} + u_{si} && (1) \\
s_i^* &= \beta_s + \alpha \frac{(\mu_i - r)}{\sigma_i^2} + u_{si} && (2) \\
s_i^* &= \beta_\alpha a_i \frac{(\mu_i - r)}{\sigma_i^2} + u_{si} && (3) \\
s_i^* &= \beta_s + \beta_\alpha a_i \frac{(\mu_i - r)}{\sigma_i^2} + u_{si} && (4)
\end{aligned}
$$

where a_i is the Kimball-Sahm-Shapiro measure of individual risk tolerance

Identification by

- Calibration of noise parameters from unconditional moments
- Exclusion restrictions

- Both x_μ and x_σ contain
  - gender, race, education
- Included in x_μ only
  - Optimism (avg. response to the probability of survival, sunny day, income will keep up with inflation), from HRS 1992-2002
  - Psychological distress (inverse~)
  - Stock market the month before (variation from month of interview)
- Included in x_σ only
  - Fraction of focal answers on other probability questions, from HRS 1992-2002

Predictors of μ_i

	(1)	(2)	(3)	(4)
Female	$-.130^{**}$	$-.125^{**}$	$-.135^{**}$	$-.131^{**}$
Black	$-.205^{**}$	$-.203^{**}$	$-.213^{**}$	$-.209^{**}$
Hispanic	$-.094^{*}$	$-.098^{*}$	$-.098^{**}$	$-.099^{**}$
Education	$.022^{**}$	$.022^{**}$	$.022^{**}$	$.021^{**}$
Optimism other domains	$.946^{**}$	$.922^{**}$	$.968^{**}$	$.958^{**}$
Happiness	$.020^{*}$	$.022^{*}$	$.019$	$.021^{*}$
S&P 500	$.0002^{*}$	$.0002^{*}$	$.0002^{*}$	$.0002^{*}$
Constant	$-.942^{**}$	$-.938^{**}$	$-.929^{**}$	$-.943^{**}$

Predictors of $\ln \sigma_i^2$

	(1)	(2)	(3)	(4)
Female	.414**	.407**	.410**	.405**
Black	.125	.099	.125	.107
Hispanic	−.124	−.145	−.121	−.136
Education	.018	.023	.019	.023
Uncertainty other domains	5.38**	5.31**	5.38**	5.34**
Constant	−2.00**	−2.06**	−2.01**	−2.03**

Other estimates

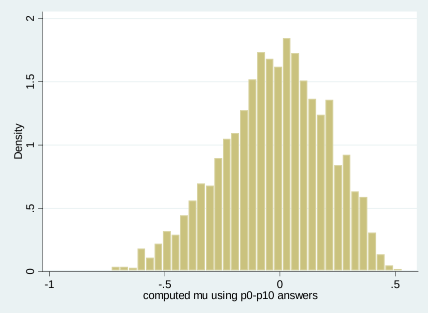
	(1)	(2)	(3)	(4)
α	$.416^{**}$	$.411^{**}$		
Coeff. on risk tolerance			1.85^{**}	1.75^{**}
Constant in s_i^*		$.237^{**}$		$.179^{**}$
$Std\left(u_\mu\right)$	$.210^{**}$	$.208^{**}$	$.232^{**}$	$.233^{**}$
$Std\left(u_\sigma\right)$	1.79^{**}	1.79^{**}	1.82^{**}	1.83^{**}
$Corr\left(u_\mu, u_\sigma\right)$	$.36^{**}$	$.40^{**}$	$.35^{**}$	$.37^{**}$
$Std\left(u_s\right)$	1.14^{**}	1.02^{**}	1.12^{**}	1.04^{**}
Implied moments				
$E\left(\mu_i\right)$	$-.00$	$-.02$	$-.00$	$-.02$
$E\left(\sigma_i^2\right)$	.60	.59	.61	.61
$E\left(s_i^*\right)$	.26	.24	.08	.24

Measures of fit

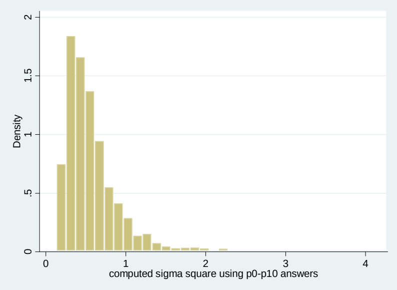
	(1)	(2)	(3)	(4)
Log likelihood	−15.492	−15.458	−15.495	−15.475
Percent correctly predicted				
Overall	.67	.62	.67	.60
Observed stockholders	.67	.98	.67	.98
Observed nonholders	.67	.21	.67	.15

Estimated relevant heterogeneity

$E\left(\mu_i | x_{\mu i}\right)$



$E\left(\sigma_i^2 | x_{\sigma i}\right)$



- Conclusion 1: expectations explain much of the puzzle
  - On average, expectations have lower mean and higher variance than historical moments
  - There is substantial heterogeneity in expectations
  - Results consistent with observed stockholding
    - both on average and at individual level
- Conclusion 2: survey answers to probability questions
  - Contain important information
  - Contain substantial noise
  - It is possible to separate the two

A to do list

- Estimate modal response model
  - Evaluate relative performance of two alternative models of survey response
- Include other predictors
  - Cognitive measures and wealth are most important
- Investigate surprising performance of risk tolerance measures
  - Multiplicative noise may be important
- Try establishing causality
  - Exclusion restrictions may be helpful