Imputing Risk Tolerance From Survey Responses

Miles S. Kimball
University of Michigan

Claudia R. Sahm
Federal Reserve Board

Matthew D. Shapiro
University of Michigan

Discussion by: Christopher Tonetti

Paper Summary

- Quantitatively measure heterogeneity in risk tolerance
- Base measure on a statistical model and responses from a large scale survey (HRS Panel)
- Ask respondents to choose between a job with a certain lifetime income vs. a job with random, but higher mean, lifetime income

Complications

- For a given person, survey responses suggest a range for the risk preference parameter, not a point value
- Survey responses are likely subject to measurement error
- Non-trivial to use this estimated risk preference parameter as an explanatory variable in regressions: a non-standard errors-in-variables problem

Survey Questions

- Original Question (1992, 1994)
  - "Suppose the chances were 50–50 that a new job would double your family income and 50–50 that it would cut it by 20%. Would you take the new job?"
- Status Quo Bias Free (1998, 2000, 2002)
  - "Suppose that you are the only income earner in the family. Your doctor recommends that you move because of allergies, and you have to choose between two possible jobs. The first would guarantee your current total family income for life. The second is possibly better paying, but the income is also less certain. There is a 50–50 chance the second job would double your total lifetime income and a 50–50 chance that it would cut it by a third. Which job would you take–the first job or the second job?"

Survey Responses

Repeat the question with varying amounts of downside risk: (1/10, 1/5, 1/3, 1/2, 3/4). Sort agents by largest rejected risk.

Table 2. Distribution of risk tolerance responses

Response category	% by HRS wave 1992	1994	1998	2000	2002
1	64.6	43.4	37.9	46.3	44.8
2		18.1	19.0	18.4	18.6
3	11.6	13.5	17.0	14.4	15.3
4	10.9	14.5	10.8	8.1	9.6
5	12.9	6.3	8.0	7.5	6.1
6		4.2	7.3	5.3	5.6
Responses	11,592	717	796	884	3,591

Category 1 rejected a job with a risky 1/10 reduction in income. Category 2 accepted the job with a risky 1/10 reduction in income, but rejected the risk of a 1/5 reduction.

Expected Utility and Parameter Bounds

- Expected Utility theory suggests a risk acceptance rule:

$$0.5U(2W) + 0.5U((1-\pi))W) \geq U(W)$$

- Assume $U(W) = \frac{W^{1-\frac{1}{\theta}}}{1-\frac{1}{\theta}}$
- Bounds on θ by category. E.g., lower bound on category 3:

$$0.5\frac{2^{1-\frac{1}{\underline{\theta}_3}}}{1-\frac{1}{\underline{\theta}_3}} + 0.5\frac{(4/5)^{1-\frac{1}{\underline{\theta}_3}}}{1-\frac{1}{\underline{\theta}_3}} = \frac{1^{1-\frac{1}{\underline{\theta}_3}}}{1-\frac{1}{\underline{\theta}_3}}$$

$$\underline{\theta}_3 = 0.27$$

similarly, $\overline{\theta}_3 = 0.50$

Estimating the Risk Proxy

- Assume: $log\theta =: x \sim N(\mu, \sigma_x^2)$

$$\begin{aligned} P(c=j) &= P(log\underline{\theta}_j < x < log\overline{\theta}_j) \\ &= \Phi((log\overline{\theta}_j - \mu)/\sigma_x) - \Phi((log\underline{\theta}_j - \mu)/\sigma_x) \end{aligned}$$

- MLE estimates of (μ, σ_x) = $(-1.98, 1.76)$
- MGF of LN dist yields $\mathbb{E}(\theta|c=j) = f(\mu, \sigma_x, \underline{\theta}_j, \overline{\theta}_j)$
- Risk proxy $h := \mathbb{E}(\theta|c) \in \{.083, .367, .706, 3.687\}$

Statistical Model of Response Error

- Recall true risk tolerance: $log\theta =: x \sim N(\mu, \sigma_x^2)$
- Let survey response error in wave w to question type q be a normal disturbance ε_{qw}
  - $q \in \{\text{original}, \text{no bias}\}$
- Respondent chooses category corresponding to $\xi_{qw} = x + \varepsilon_{qw}$

Statistical Assumptions

- $\xi_{qw} = x + \varepsilon_{qw} = x + b_q + \kappa_q + e_{qw}$
- b_q is commmon bias across individuals of question type q
  - $b_{\text{no bias}} = 0$
- $\kappa_q \sim N(0, \sigma^2_{\kappa q})$ is individual's persistent response error for question type q
- $e_{qw} \sim N(0, \sigma^2_{eq})$ is individual's transitory response error for wave w and question type q
- $\xi_{qw} \sim N(\mu + b_q, \sigma^2_x + \sigma^2_{\kappa q} + \sigma^2_{eq})$

Likelihoods

- Individuals who respond only in 1 wave:

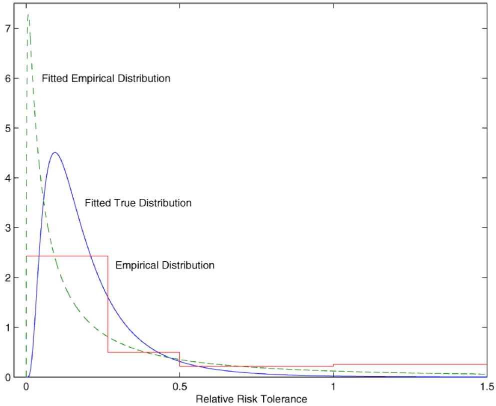
$$P(c_w = j) = \Phi((log\overline{\theta}_j - \mu - b_q)/\sigma_q) - \Phi((log\underline{\theta}_j - \mu - b_q)/\sigma_q)$$

- Individuals who respond in both waves:

$$P(c_w = j, c_{w'} = k) = \vec{\Phi}(\overline{N}_{jq}, \overline{N}_{kq'}, \rho) + \vec{\Phi}(\underline{N}_{jq}, \underline{N}_{kq'}, \rho) - \vec{\Phi}(\overline{N}_{jq}, \underline{N}_{kq'}, \rho) - \vec{\Phi}(\underline{N}_{jq}, \overline{N}_{kq'}, \rho)$$

- $\vec{\Phi}$ is bivariate normal CDF
- $\overline{N}_{jq} := (log\overline{\theta}_j - \mu - b_q)/\sigma_q$
- $\rho = corr(\xi_{qw}, \xi_{qw'})$

MLE Estimated Distribution of Risk Tolerance



Distribution of Risk Preferences

Table 4. Distribution of risk preferences

Parameter	Log risk tolerance	Risk tolerance	Risk aversion
Mean	−1.84	.206	8.2
	(.03)	(.008)	(.3)
Median	−1.84	.159	6.3
	(.03)	(.005)	(.2)
Mode	−1.84	.094	3.7
	(.03)	(.004)	(.2)
Standard deviation	.73	.172	6.8
	(.04)	(.018)	(.7)
Fractiles			
1	−1.54	.029	1.2
5	−1.32	.048	1.9
10	−1.20	.063	2.5
25	−1.01	.097	3.9
50	−.80	.159	6.3
75	−.59	.259	10.3
90	−.40	.402	16.0
95	−.28	.523	20.8
99	−.07	.858	34.1

Studying Behavior with the Proxy

- Define $u := \theta - h = \theta - \mathbb{E}(\theta|c)$
- Regression to analyze effects of risk tolerance and other regressors on choice y

$$y = \theta\delta_\theta + z\delta_z + v$$

- Assume $\mathbb{E}(v|\theta, z) = 0$
- OLS consistently estimates $\{\delta_\theta, \delta_z\}$

Non Classical Measurement Error

- If θ not observed, use proxy in regression

$$y = h\delta_\theta + z\delta_z + \eta$$
$$\eta = u\delta_\theta + v$$

- u is uncorrelated with h and correlated with θ
- Regressors z that correlate with θ will also correlate with u
- If z is correlated with u, OLS not consistent

Moment Condtions For GMM: Risk Proxy h

- Independent error assumptions and properties of condtional expectations:

$$\mathbb{E}(h\eta) = \mathbb{E}(hu)\delta_\theta + \mathbb{E}(h\nu) = 0 \tag{1}$$

Moment Condtions For GMM: Regressors z

- Assume conditional expectation of each observable $z_k \in z$ is linear in risk tolerance

$$z_k = \theta \beta_k + \zeta$$

- Assume $\mathbb{E}(\zeta | \theta) = 0$
- $\beta_k = \mathbb{E}(\theta^2)^{-1} \mathbb{E}(\theta z_k)$

Moment Condtions For GMM: Regressors z

- Substitute in h

$$z_k = h\beta_k + u\beta_k + \zeta$$

- Assuming ζ independent of $\varepsilon \Rightarrow \mathbb{E}(h\zeta) = 0$
- Recall: $\mathbb{E}(uh) = 0$
- Thus, OLS of z_k on h consistently estimates β_k:

$$\beta_k = \mathbb{E}(\theta^2)^{-1}\mathbb{E}(\theta z_k) = \mathbb{E}(h^2)^{-1}\mathbb{E}(hz_k)$$

- Define: $\lambda = \mathbb{E}(\theta^2)/\mathbb{E}(h^2)$
- Then algebra shows:

$$\mathbb{E}(\theta z_k) = \lambda \mathbb{E}(hz_k) \qquad (2)$$

Moment Condtions For GMM

- Restate the original model

$$y = \lambda h\delta_\theta + z\delta_z + \omega$$
$$\omega = (\theta - \lambda h)\delta_\theta + \nu$$

- EQ(1) & EQ(2) yield moment conditions:

$$\mathbb{E}(h\eta) = \mathbb{E}(h(y - h\delta_\theta + z\delta_z)) = 0 \quad (3)$$
$$\mathbb{E}(z'\omega) = \mathbb{E}(z'(y - \lambda h\delta_\theta + z\delta_z)) = 0 \quad (4)$$

- Can implement GMM since we have an estimate of λ from MLE estimation

Risk Tolerance and Financial Portfolios

Table 7. Effect of risk preferences on the share of financial wealth in stocks

			Risk tolerance proxy			
Control for log risk tolerance	None	Categorical survey response	Ignoring response error	Modeling response error	Modeling response error	Including application covariates
Estimator	OLS	OLS	OLS	OLS	GMM	OLS
Category 3		−.026				
		(.010)				
Category 4		.022				
		(.012)				
Category 5–6		.025				
		(.011)				
Proxy			.008	.146	.162	.152
			(.003)	(.054)	(.060)	(.056)
Male	.024	.023	.023	.023	.014	.018
	(.007)	(.007)	(.007)	(.007)	(.008)	(.008)
Education						
>16 years	.034	.032	.032	.031	.012	.019
	(.012)	(.012)	(.012)	(.012)	(.013)	(.014)
13–16 years	.036	.035	.035	.035	.024	.029
	(.009)	(.009)	(.009)	(.009)	(.009)	(.009)
<12 years	−.023	−.024	−.023	−.023	−.026	−.024
	(.009)	(.009)	(.009)	(.009)	(.009)	(.009)
Black	−.029	−.029	−.029	−.028	−.024	−.027
	(.009)	(.009)	(.009)	(.009)	(.009)	(.009)
Hispanic	−.035	−.036	−.035	−.035	−.034	−.038
	(.012)	(.012)	(.012)	(.012)	(.013)	(.012)
Age/10	−.002	−.001	−.001	−.001	.006	.006
	(.008)	(.008)	(.008)	(.008)	(.008)	(.008)
Log income	.002	.003	.003	.002	.004	.003
	(.005)	(.005)	(.005)	(.005)	(.005)	(.005)
Log wealth	.046	.047	.046	.046	.047	.046
	(.002)	(.002)	(.002)	(.002)	(.002)	(.002)
R^2	.170	.172	.171	.171	.178	.177

MLE Estimates of Log Risk Tolerance

Parameter	Ignoring response error	Modeling response error
Log risk tolerance		
Mean μ	−1.98	−1.84
	(.03)	(.03)
Standard deviation σ_x	1.76	.73
	(.03)	(.04)
Status quo bias b_o		−.11
		(.04)
Response error standard deviation		
Original question, transitory σ_{eo}		1.39
		(.05)
Original question, persistent $\sigma_{\kappa o}$		.73
		(.10)
SQB-free question, transitory σ_{ef}		1.43
		(.03)
SQB-free question, persistent $\sigma_{\kappa f}$		.60
		(.09)
Number of individuals	11,616	11,616
Number of responses	11,616	17,580
Number of parameters	2	7
Log-likelihood	−12,073.4	−21,208.3

Imputation of Risk Preference

Table 5. Imputation of risk preference

Response category	Log risk tolerance	Risk tolerance	Risk aversion
1	−2.107	.153	10.4
2	−1.811	.203	7.6
3	−1.693	.228	6.7
4	−1.575	.257	6.0
5	−1.419	.301	5.1
6	−1.172	.387	4.0