Maternal Subjective Expectations about the Technology of Skill Formation Predict Investments in Children One Year Later

By Flávio Cunha, Irma Elo, and Jennifer Culhane
June (2020)

James J. Heckman

THE UNIVERSITY OF CHICAGO

Econ 350, Winter 2021

1. Introduction

2. MSE about the Technology of Skill Formation

2.1 The Technology of Skill Formation

- Let $q_{i,0}$ and $q_{i,1}$ denote, respectively, the stocks of the human capital of child i at birth and 24 months.

- Let x_i denote maternal investments in the human capital of child i during the first two years of the child's life. Let x_i denote shocks to the development process.

- We assume that the technology of skill formation assumes the following parametric specification:

$$\ln q_{i,1} = \psi_1 \ln q_{i,0} + \psi_2 \ln x_i + \psi_3 \ln q_{i,0} \ln x_i + \nu_i. \quad (1)$$

- The translog parameterization in equation (1) is particularly convenient to make progress on the elicitation of MSE about the technology of skill formation.

- To see why, let $\mathcal{H}_i$ denote the mother's information set.

- According to the technology function denoted in (1), it follows that:

$$\mathrm{E}(\ln q_{i,1} \,|q_0, x, \mathcal{H}_i) = \mu_{i,\psi,1} \ln q_0 + \mu_{i,\psi,2} \ln x + \mu_{i,\psi,3} \ln q_0 \ln x, \quad (2)$$

2.2 Motor Social Development Scale

- For each child i and MSD item j, we define the latent variable $d^*_{i,j}$ according to the following specification:

$$d^*_{i,j} = b_{j,0} + b_{j,1}\left(\ln a_i + \frac{b_{j,2}}{b_{j,1}}\theta_i\right) - \eta_{i,j}. \tag{3}$$

- Let Φ denote the cumulative distribution function (CDF) of a normal random variable with mean 0 and variance 1.

- If we assume that $\eta_{i,j} \sim N(0,1)$ it follows that the probability that child i can perform MSD task j is equal to:

$$\Pr(d_{i,j} = 1 | a_i, \theta_i) = \Phi\left[b_{j,0} + b_{j,1}\left(\ln a_i + \frac{b_{j,2}}{b_{j,1}}\theta_i\right)\right]. \tag{4}$$

2.3 Instruments to Elicit MSE

- Figure 1 presents the survey instruments in detail.
- Panel A reproduces the elicitation items we use in the subjective probability form.
- Panel B shows the exact elicitation items in the age ranges form.
- Panel C presents the four scenarios of human capital at birth and investments.
- For the study participants, we showed a five-minute video that described the scenarios before answering any questions.

Figure 1

Panel A: Subjective-Probability Form
1. How likely is it that the baby will be able to learn to speak a partial sentence of three words or more by age two?
2. How likely is it that the baby will be able to learn to say his/her first and last names together without anyone's help by age two?
3. How likely is it that the baby will be able to learn to count three objects correctly by age two?
4. How likely is it that the baby will be able to learn his or her own age and sex by age two?
Panel B: Age-Range Form
1. What do you think are the youngest age and the oldest age a baby learns to speak a partial sentence of three words or more?
2. What do you think are the youngest age and the oldest age a baby learns to say his/her first and last names together without anyone's help?
3. What do you think are the youngest age and the oldest age a baby learns to count three objects correctly?
4. What do you think are the youngest age and the oldest age a baby learns his or her own age and sex?

Notes: This figure provides detailed information about both forms of the elicitation instrument. Panel A reproduces the elicitation items in the subjective probability form. Panel B displays the elicitation items in the age range form. Panel C describes the scenarios of human capital at birth and investments. The study participants watched a short video describing these scenarios.

Figure 1, Cont'd

Panel C: Definition of the Scenarios				
	Human capital at birth		Investments	
Scenario number	Short description	Full description	Short description	Full description
1	"High"	Pregnancy lasted 9 months, birth weight is 8 pounds, birth length is 20 inches.	"High"	6 hours per day doing HOME Scale activities.
2	"Low"	Pregnancy lasted 7 months, birth weight is 5 pounds, birth length is 18 inches.	"High"	6 hours per day doing HOME Scale activities.
3	"High"	Pregnancy lasted 9 months, birth weight is 8 pounds, birth length is 20 inches.	"Low"	2 hours per day doing HOME Scale activities.
4	"Low"	Pregnancy lasted 7 months, birth weight is 5 pounds, birth length is 18 inches.	"Low"	2 hours per day doing HOME Scale activities.

Notes: This figure provides detailed information about both forms of the elicitation instrument. Panel A reproduces the elicitation items in the subjective probability form. Panel B displays the elicitation items in the age range form. Panel C describes the scenarios of human capital at birth and investments. The study participants watched a short video describing these scenarios.

2.4 Estimation of Expectation of Child Development

- We explore the IRT model to derive an error-ridden measure of maternal expectation of the natural log of development at age 24 months, $\ln q^L_{i,j,k}$, from the reported probability $p^L_{i,j,k}$.

- To do so, we invert equation (4) and solve for $\theta^L_{i,j,k}$:

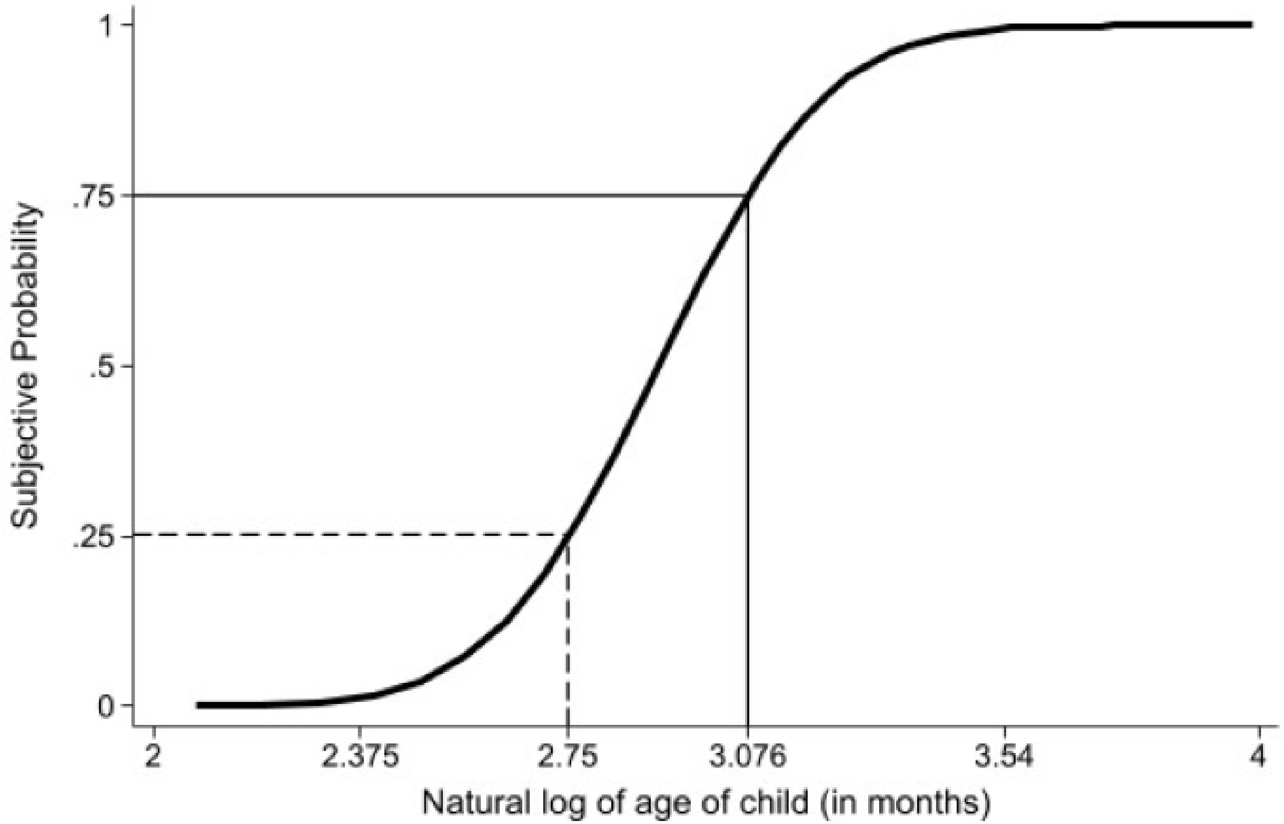
$$\ln q^L_{i,j,k} \equiv \left(\ln 24 + \frac{b_{2,j}}{b_{1,j}}\theta^L_{i,j,k}\right) = \left[\frac{\Phi^{-1}(p^L_{i,j,k}) - b_{j,0}}{b_{j,1}}\right]. \quad (5)$$

- The thick solid curve in Figure 2 shows the prediction from the IRT model for the MSD item, "speak a partial sentence of three words or more."

- The horizontal axis in Figure 2 shows the natural logarithm of the child's age (in months), while the vertical axis shows the maternal subjective probability that a child will "speak a partial sentence of three words or more."

- We use the IRT model to transform the subjective probability that the mother reports into the corresponding natural log of age—the scale that we use for human capital at age two.

Figure 2



Notes: This figure shows how we use the IRT model to relocate and rescale maternal subjective probability reports (shown in the vertical axis) to error-ridden measures of the expectation of the natural log of human capital at age two years (shown in the horizontal axis) for two scenarios of investments ("high" vs. "low") when human capital at birth is "high." When the investment is "low," the mother reports that there is a 25% chance that the child will learn how to speak a partial sentence with three words or more by age 24 months. When the investment is "high," the mother reports that the probability is 75% by age 24 months. These probabilities correspond to $2.75 = ln\ 16$ and $3.076 = ln\ 22$, respectively. According to the IRT model, 25% of the 16-month-old children and 75% of the 22-month-old children speak a partial sentence with three words or more.

- To infer the respondent's subjective probability that the child will learn how to speak partial sentences by age 24 months, we need to construct how the probability varies with age.

- In the analysis of the age-range data, we use the mother's answer to estimate a mother-scenario specific IRT model along with the parameterization used in equation (3).

- Indeed, let $d_{i,j,k}^{*}$ denote the latent variable that is determined according to:

$$d_{i,j,k}^{*} = \tilde{b}_{i,j,k,0} + \tilde{b}_{i,j,k,1} \ln a_{i,j,k} - \tilde{\eta}_{i,j,k}, \tag{6}$$

where the shock $\tilde{\eta}_{i,j,k}$ is normally distributed with mean zero and variance one.

- If we combine the model in equation (6) with age ranges provided by the respondent, we conclude that, according to respondent i, the probability that the child will learn how to do MSD task j in scenario k when $a_{i,j,k} = a_{i,j,k}$ is:

$$\Delta_0 = \Phi\left[\tilde{b}_{i,j,k,0} + \tilde{b}_{i,j,k,1} \ln \underline{a}_{i,j,k}\right]. \tag{7}$$

- Analogously, the probability that the child will learn how to do MSD task j in scenario k when $a_{i,j,k} = \bar{a}_{i,j,k}$ is

$$\Delta_1 = \Phi\left[\tilde{b}_{i,j,k,0} + \tilde{b}_{i,j,k,1} \ln \overline{a}_{i,j,k}\right]. \tag{8}$$

- If we manipulate the system in equations (7) and (8), we conclude that, for arbitrary j and k, the following equalities hold:

$$\tilde{b}_{i,j,k,1} = \frac{\Phi^{-1}(\Delta_1) - \Phi^{-1}(\Delta_0)}{\ln \overline{a}_{i,j,k} - \ln \underline{a}_{i,j,k}}, \quad (9)$$

and

$$\tilde{b}_{i,j,k,0} = \frac{\Phi^{-1}(\Delta_0) \ln \overline{a}_{i,j,k} - \Phi^{-1}(\Delta_1) \ln \underline{a}_{i,j,k}}{\ln \overline{a}_{i,j,k} - \ln \underline{a}_{i,j,k}}. \quad (10)$$

- The individual-specific IRT model states that this probability is:

$$p_{i,j,k}^{A} = \Phi\left[\tilde{b}_{i,j,k,0} + \tilde{b}_{i,j,k,1} \ln 24\right]. \tag{11}$$

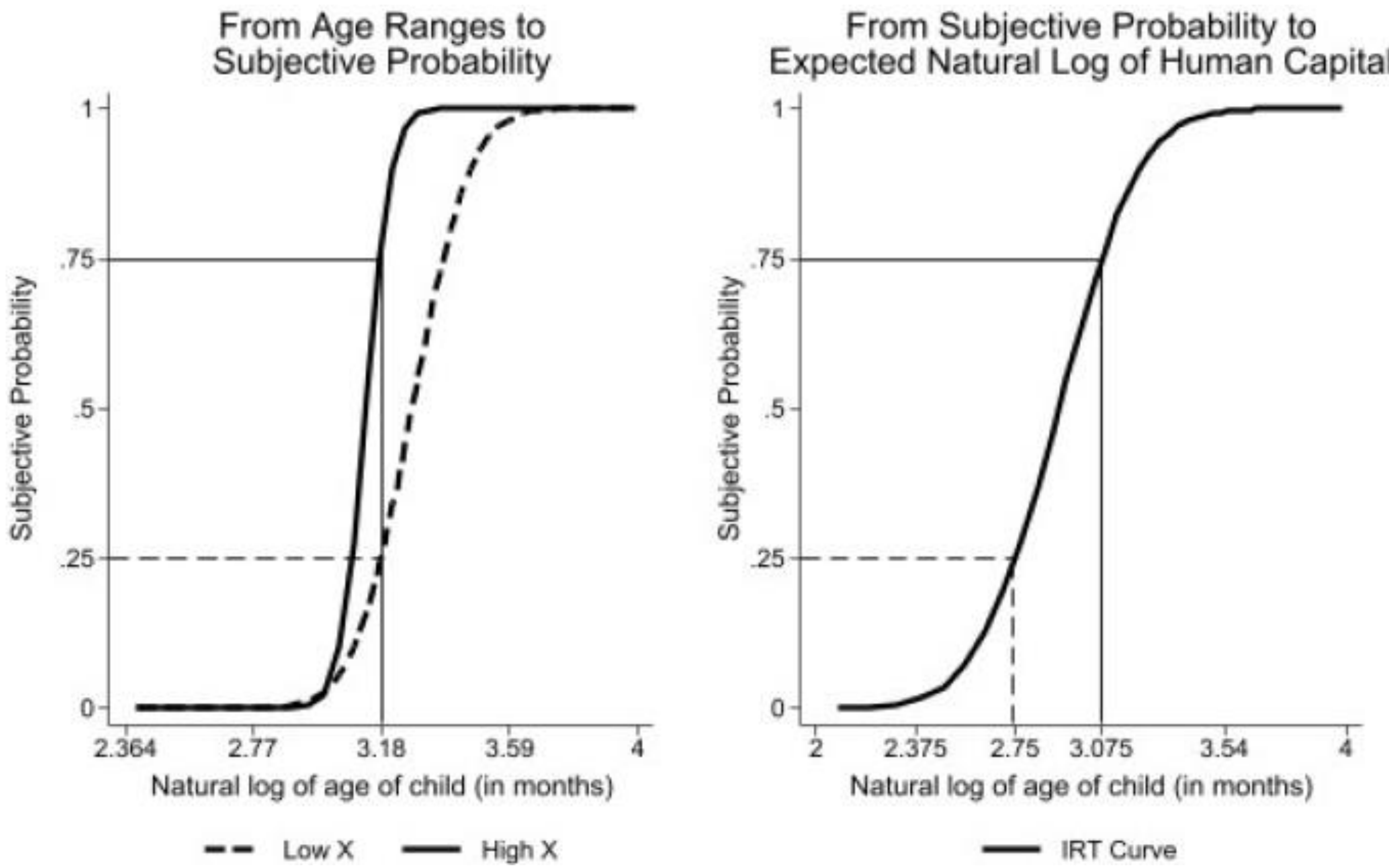
- We use equation (11), together with the IRT probability in equation (4), to derive an error-ridden measure of maternal expectation of the natural log of development at age 24 months, $\ln q_{i,j,k}^{A}$, from the implied probability $p_{i,j,k}^{A}$

- To do so, we invert (4) and solve for $\ln q_{i,j,k}^{A}$:

$$\ln q_{i,j,k}^{A} = \left(\ln 24 + \frac{b_{2,j}}{b_{1,j}} \theta_{i,j,k}^{A}\right) = \left[\frac{\Phi^{-1}\left(p_{i,j,k}^{A}\right) - b_{j,0}}{b_{j,1}}\right]. \tag{12}$$

- We illustrate the algorithm with Figure 3.
- The left panel in Figure 3 captures the first step of the process, which is described by equations (9), (10), and (11).
- To construct this example, we assume that the study participant states that the age ranges are 21 and 25 for the scenario in which both human capital and investment are "high."

Figure 3



Notes: This figure shows the two steps involved in transforming age ranges to error-ridden measures of expectation of the natural log of human capital at age 24 months. In the first step, which we show in the left panel, we use maternal reports of the age ranges and assumptions about the interpolating function and the Parameters Δ_0 and Δ_1. For this figure, we assume that the interpolating function is the normal CDF, $\Delta_0 = 10\%$ and $\Delta_1 = 90\%$. We show the transformation from age ranges to subjective probability for two scenarios of investments ("high" vs. "low") when human capital at birth is "high."

2.5 Identification

- Next, we shed light on the source of identification for MSE.
- In essence, the identifying information comes from three different types of moments from the raw data.
- First, to identify $\mu_{i,\psi,3}$ in equation (2), the coefficient on the interaction between human capital at birth and investments, we rely on the following differences-in-differences:

$$\mathbf{M}_{i,1} \equiv \underbrace{\frac{\Delta \mathrm{E}\left(\ln q_{i,1} \,|\overline{q}_0, \mathcal{H}_i\right)}{\Delta \ln x \times \Delta \ln q_0}}_{\substack{\textit{Difference between} \\ \textit{Scenarios } 1 \text{ and } 3.}} - \underbrace{\frac{\Delta \mathrm{E}\left(\ln q_{i,1} \,|\underline{q}_0, \mathcal{H}_i\right)}{\Delta \ln x \times \Delta \ln q_0}}_{\substack{\textit{Difference between} \\ \textit{Scenarios 2 and 4.}}} = \mu_{i,\psi,3}, \tag{13}$$

where $\Delta \ln x = \ln \overline{x} - \ln \underline{x}$, $\Delta \ln q_0 = \ln \overline{q}_0 - \ln \underline{q}_0$ and

$$\Delta \mathrm{E}\left(\ln q_{i,1} \,|q_0, \mathcal{H}_i\right) = \mathrm{E}\left(\ln q_{i,1} \,|(q_0, \overline{x}), \mathcal{H}_i\right) - \mathrm{E}\left(\ln q_{i,1} \,|(q_0, \underline{x}), \mathcal{H}_i\right).$$

- We can use the following moment to identify $\mu_{i,\psi,2}$ in equation (2):

$$\mathbf{M}_{i,2} \equiv \frac{\ln \overline{q}_0 \, \Delta \mathrm{E}\left(\ln q_{i,1} \left| \underline{q}_0, \mathcal{H}_i\right.\right)}{\Delta \ln x \times \Delta \ln q_0} - \frac{\ln \underline{q}_0 \, \Delta \mathrm{E}\left(\ln q_{i,1} \left| \overline{q}_0, \mathcal{H}_i\right.\right)}{\Delta \ln x \times \Delta \ln q_0} = \mu_{i,\psi,2}. \quad (14)$$

- Note that equation (14) is just a "weighted" difference-in-difference moment in which the weights are the scenarios for human capital at birth.

- Finally, we can derive a moment that identifies $\mu_{i,\psi,1}$ in equation (2):

$$\mathbf{M}_{i,3} \equiv \frac{\ln \overline{x} \, \Delta \mathrm{E}\left(\ln q_{i,1} \left| \underline{x}, \mathcal{H}_i\right.\right)}{\Delta \ln x \times \Delta \ln q_0} - \frac{\ln \underline{x} \, \Delta \mathrm{E}\left(\ln q_{i,1} \left| \overline{x}, \mathcal{H}_i\right.\right)}{\Delta \ln x \times \Delta \ln q_0} = \mu_{i,\psi,1}. \quad (15)$$

2.6 Estimation of MSE

Note that $\ln q^L_{i,j,k}$ defined in equation (5) and $\ln q^A_{i,j,k}$ defined in equation (12) are two error-ridden measures of maternal expectations about the natural log of child development. Let $\epsilon^f_{i,j,k}$ denote the measurement error in form f, scenario k, MSD item j, and participant i. We define $\epsilon_i = \left(\epsilon^L_{i,1,1}, \epsilon^L_{i,1,2}, \ldots, \epsilon^L_{i,j,k}, \ldots, \epsilon^L_{i,J,4}, \epsilon^A_{i,1,1}, \ldots \epsilon^A_{i,j,k}, \ldots, \epsilon^A_{i,J,4}\right)'$. We define in a similar fashion the vectors $\mathrm{E}(\ln q_i \,|Z_i, \mathcal{H}_i)$, $\ln q_i$ and Z_i, where $Z_{i,j,k} = \left(\ln q_{0,i,j,k}\,, \ln x_{i,j,k}\,, \ln q_{0,i,j,k} \ln x_{i,j,k}\right)$. Finally, let $\mu_{\psi,i} = \left(\mu_{\psi,i,1}, \mu_{\psi,i,2}, \mu_{\psi,i,3}\right)$. Therefore:

$$\ln q_{i,1} = \mathrm{E}(\ln q_i \,|Z_i, \mathcal{H}_i) + \epsilon_i = Z_i \mu_{\psi,i} + \epsilon_i. \qquad (16)$$

- Although the number of scenarios by itself is sufficient to estimate the individual MSE parameters, it is not enough, without additional restrictions, to estimate maternal subjective uncertainty (MSU) about the parameters of the technology of skill formation (1).

- Let $\Sigma_i = Var(\psi | q_0, x, \mathcal{H}_i), \sigma_{v,i}^2 = Var(v_i | q_0, x, \mathcal{H}_i)$ and note that, given parameterization (1),

$$\operatorname{Var}(\ln q_{i,1} | q_0, x, \mathcal{H}_i) = Z_i' \Sigma_i Z_i + \sigma_{v,i}^2 \tag{17}$$

2.7 A Factor Model Approach

- Then, the combination of parameterization (2) with Assumptions 1, 2, and 3 produces the following latent MSE model:

$$
\begin{aligned}
M^f_{i,j,1} &\equiv \frac{\Delta q^f_{i,j,\overline{q}_0} - \Delta q^f_{i,j,\underline{q}_0}}{\Delta \ln x \times \Delta \ln q_0} = \mu_{i,\psi,3} + \frac{\Delta \epsilon^f_{i,j,\overline{q}_0} - \Delta \epsilon^f_{i,j,\underline{q}_0}}{\Delta \ln x \times \Delta \ln q_0} \\
M^f_{i,j,2} &\equiv \frac{\ln \overline{q}_0 \left(\Delta q^f_{i,j,\underline{q}_0}\right) - \ln \underline{q}_0 \left(\Delta q^f_{i,j,\overline{q}_0}\right)}{\Delta \ln x \times \Delta \ln q_0} = \mu_{i,\psi,2} + \frac{\ln \underline{q}_0 \left(\Delta \epsilon^f_{i,j,\overline{q}_0}\right) - \ln \overline{q}_0 \left(\Delta \epsilon^f_{i,j,\underline{q}_0}\right)}{\Delta \ln x \times \Delta \ln q_0} \qquad (18) \\
M^f_{i,j,3} &\equiv \frac{\ln \overline{x} \left(\Delta q^f_{i,j,\underline{x}}\right) - \ln \underline{x} \left(\Delta q^f_{i,j,\overline{x}}\right)}{\Delta \ln x \times \Delta \ln q_0} = \mu_{i,\psi,1} + \frac{\ln \overline{x} \left(\Delta \epsilon^f_{i,j,\overline{x}}\right) - \ln \underline{x} \left(\Delta \epsilon^f_{i,j,\underline{x}}\right)}{\Delta \ln x \times \Delta \ln q_0},
\end{aligned}
$$

where

$$
\Delta q^f_{i,j,q_0} = \ln q^f_{i,j,(q_0,\overline{x})} - \ln q^f_{i,j,(q_0,\underline{x})}, \qquad \text{for } q_0 \in \left\{\underline{q}_0, \overline{q}_0\right\},
$$

$$
\Delta q^f_{i,j,x} = \ln q^f_{i,j,(\overline{q}_0,x)} - \ln q^f_{i,j,(\underline{q}_0 x)}, \qquad \text{for } x \in \{\underline{x}, \overline{x}\}.
$$

- However, the latent MSE variables need not be orthogonal.

- Further, if the factors are highly correlated, the three latent MSE variables may be summarized by fewer than three factors.

- To illustrate the issue, we return to the ideal moments (13), (14), and (15). Some trivial algebra leads us to the following equations:

$$\mathbf{M}_{i,3} + \ln \overline{x}\, \mathbf{M}_{i,1} = \Delta \ln x \left[\Delta \mathrm{E}\left(\ln q_{i,1} \,\middle|\, \overline{x}, \mathcal{H}_i \right) \right]$$
$$\mathbf{M}_{i,3} + \ln \underline{x}\, \mathbf{M}_{i,1} = \Delta \ln x \left[\Delta \mathrm{E}\left(\ln q_{i,1} \,\middle|\, \underline{x}, \mathcal{H}_i \right) \right]. \tag{20}$$

- These conclusions also are valid for the relationship between $\mu_{i,\psi,2}$ and $\mu_{i,\psi,3}$.
- Indeed, if we condition on scenarios for human capital at birth, then we can derive a parallel set of cross-moment relationships:

$$\mathbf{M}_{i,2} + \ln \overline{q}_0 \mathbf{M}_{i,1} = \Delta \ln q_0 \left[\Delta \mathrm{E}\left(\ln q_{i,1} \middle| \overline{q}_0, \mathcal{H}_i\right)\right]$$
$$\mathbf{M}_{i,2} + \ln \underline{q}_0 \mathbf{M}_{i,1} = \Delta \ln q_0 \left[\Delta \mathrm{E}\left(\ln q_{i,1} \middle| \underline{q}_0, \mathcal{H}_i\right)\right] \quad (21)$$

3. Data

4. Results

4.1 Sample Characteristics from the PHD Study

- Table 1 presents the demographic characteristics of the PHD study sample.
- The participants are relatively young: 63.25% of the mothers enrolled in the study were born between 1988 and 1997.
- The majority of these mothers are Non-Hispanic Black (around 54%).
- At the time of recruitment into the study, 60% of the participants were single, 30% were married, and 10% were cohabiting.
- The participants tended to have lower educational attainment than do national representative samples: 42% were high-school dropouts; 42% had a high school degree or some post-secondary schooling, but only 16% had a four-year college degree or more education.

Table 1: Demographic Characteristics of PHD Study Participants

Characteristic	Percentage
Year of Birth	
Mother born between 1968 and 1977	3.54%
Mother born between 1978 and 1987	33.21%
Mother born between 1988 and 1987	63.25%
Race and Ethnicity	
Mother is Hispanic	13.02%
Mother is Non-Hispanic Black	53.77%
Mother is Non-Hispanic White	26.64%
Other	6.57%
Educational Attainment	
Less than high school diploma	42.09%
High school or some college	41.36%
Four-year college diploma or higher	16.55%
Marital Status	
Single[a]	60.71%
Cohabiting	9.49%
Married	29.81%

Notes: a. In the single category, we include one participant who reported being separated and two participants who reported being divorced at the time of enrollment in the study. The remaining individuals in this category (496 out of 499) reported being single and never married at the time of enrollment into the study. b. We conducted the first wave when the mothers were in the second trimester of their first pregnancy. c. We conducted the second wave when the children were 9–12 months old.

Table 1: Demographic Characteristics of PHD Study Participants, Cont'd

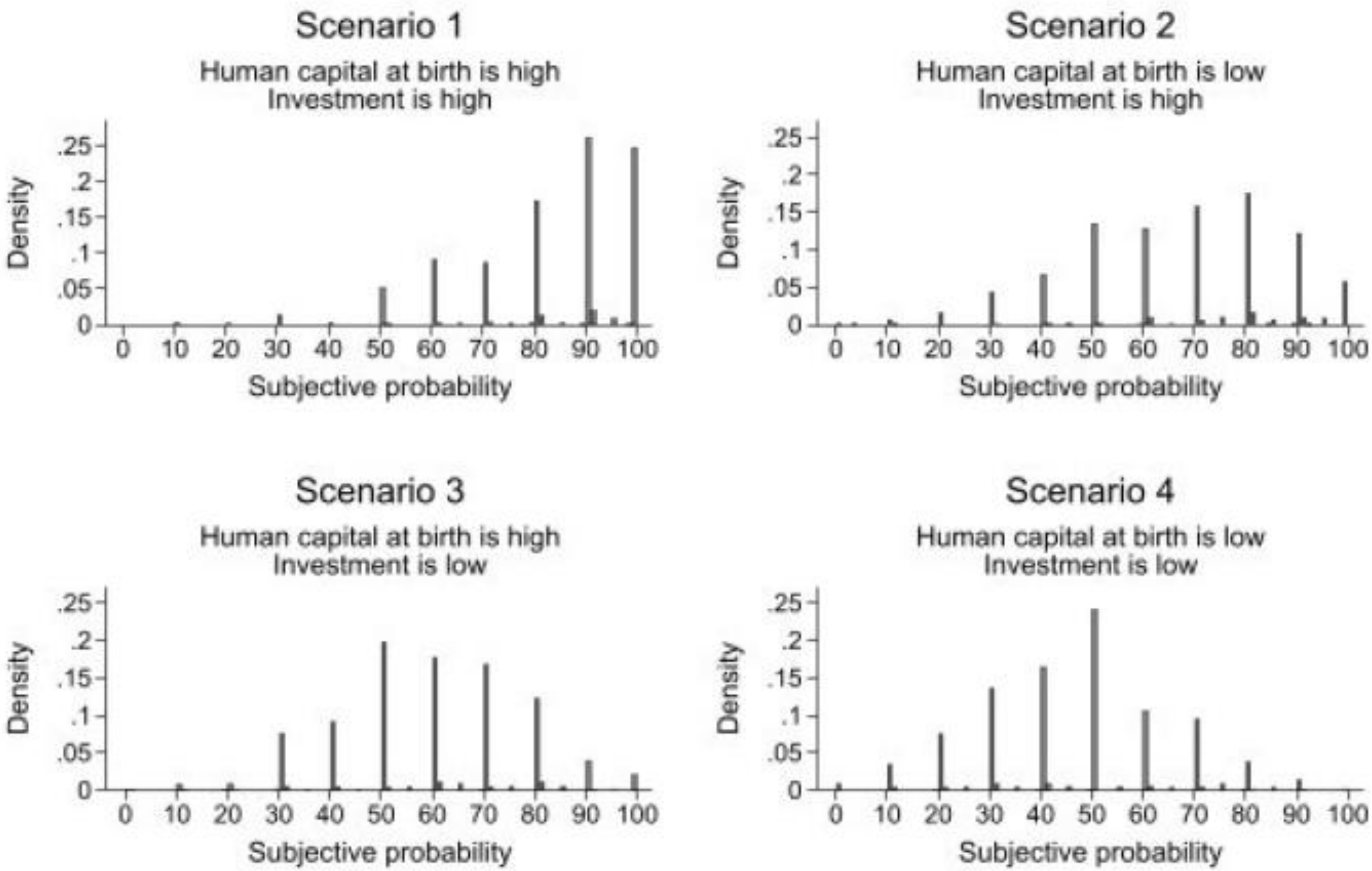
Characteristic	Percentage
Center for Epidemiological Studies Depression Scale	
The score is greater than or equal to 16	29.32%
Household Income Per Year (y)	
$y < \$25,000$	44.77%
$\$25,000 \leq y < \$55,0000$	20.56%
$\$55,000 \leq y < \$105,0000$	16.06%
$y \geq \$105,000$	18.61%
Sample Size	
First wave[b]	822
Second wave[c]	687

Notes: a. In the single category, we include one participant who reported being separated and two participants who reported being divorced at the time of enrollment in the study. The remaining individuals in this category (496 out of 499) reported being single and never married at the time of enrollment into the study. b. We conducted the first wave when the mothers were in the second trimester of their first pregnancy. c. We conducted the second wave when the children were 9–12 months old.

4.2 Analysis of the Elicitation Data

- Figure 4 presents the histograms of answers to that form.
- In the paper, we focus our attention on the first MSD item in both forms ("speaks a partial sentence with three words or more").
- The most noticeable pattern in responses is heaping on round numbers.
- Although respondents could choose any integer between 0 and 100, the raw data show that respondents overwhelmingly chose multiples of 10.
- The heaping that we observe for "speaking a partial sentence with three words or more" is also present for the other three MSD items we use in this study.

Figure 4



Notes: This figure shows the histograms of maternal reports of subjective probability for the MSD item "child speaks a partial sentence with three words or more" for all scenarios of human capital at birth and investments. The figure shows a pattern of answers that indicates that the higher human capital at birth or investment, the higher the likelihood that the child will be able to accomplish this task by age 24 months.

- Table 2 presents the expectations of child developmental outcomes according to the subjective probability form (see Section 2.4.1).

- We present these probabilities for the four MSD items and the four different scenarios in which human capital at birth and investments are varied.

- Several remarkable results are evident in this table.

Table 2: Summary Statistics of Answers to the Subjective Probability-Elicitation Form

Brief description of the MSD item	Panel A					
	Scenario 1 Human capital at birth is "high" Investment is "high"			Scenario 2 Human capital at birth is "low" Investment is "high"		
	Mean	Median	SD	Mean	Median	SD
Speaks partial sentence	82.27	90.00	16.97	67.28	70.00	20.31
	(0.59)	(0.84)	(4.30)	(0.71)	(1.68)	(4.37)
Counts three objects correctly	79.27	80.00	19.58	65.63	70.00	21.32
	(0.68)	(0.84)	(4.74)	(0.74)	(0.84)	(4.48)
Knows own age and sex	83.89	90.00	16.81	69.80	70.00	20.04
	(0.59)	(0.84)	(4.39)	(0.70)	(0.84)	(4.24)
Says first and last names together	82.23	90.00	19.30	69.19	70.00	21.37
	(0.67)	(0.84)	(5.15)	(0.75)	(1.60)	(4.73)

Notes: This table shows the mean, median, and standard deviation (SD) of maternal reports of subjective probability for each MSD item and scenario.

Table 2: Summary Statistics of Answers to the Subjective Probability-Elicitation Form, Cont'd

	Panel B					
	Scenario 3 Human capital at birth is "high" Investment is "low"			Scenario 4 Human capital at birth is "low" Investment is "low"		
Brief description of the MSD item	Mean	Median	SD	Mean	Median	SD
Speaks partial sentence	59.18	60.00	18.43	46.23	50.00	18.76
	(0.64)	(0.92)	(3.99)	(0.65)	(0.84)	(4.02)
Counts three objects correctly	56.63	60.00	20.25	44.69	41.00	19.55
	(0.71)	(0.84)	(4.34)	(0.68)	(0.84)	(4.20)
Knows own age and sex	59.84	60.00	18.82	47.15	50.00	19.13
	(0.66)	(0.84)	(4.15)	(0.67)	(0.84)	(4.13)
Says first and last names together	59.68	60.00	20.11	47.15	50.00	20.75
	(0.70)	(1.68)	(4.33)	(0.72)	(0.84)	(4.41)

Notes: This table shows the mean, median, and standard deviation (SD) of maternal reports of subjective probability for each MSD item and scenario.

- Figure 5 presents the histograms when we aggregate maternal subjective probability reports across MSD items for the subjective probability form.

- For Scenario 1, we still have some heaping at high probability values, but heaping has been reduced for other scenarios because maternal reports across MSD items for a given scenario are correlated but far from perfect.

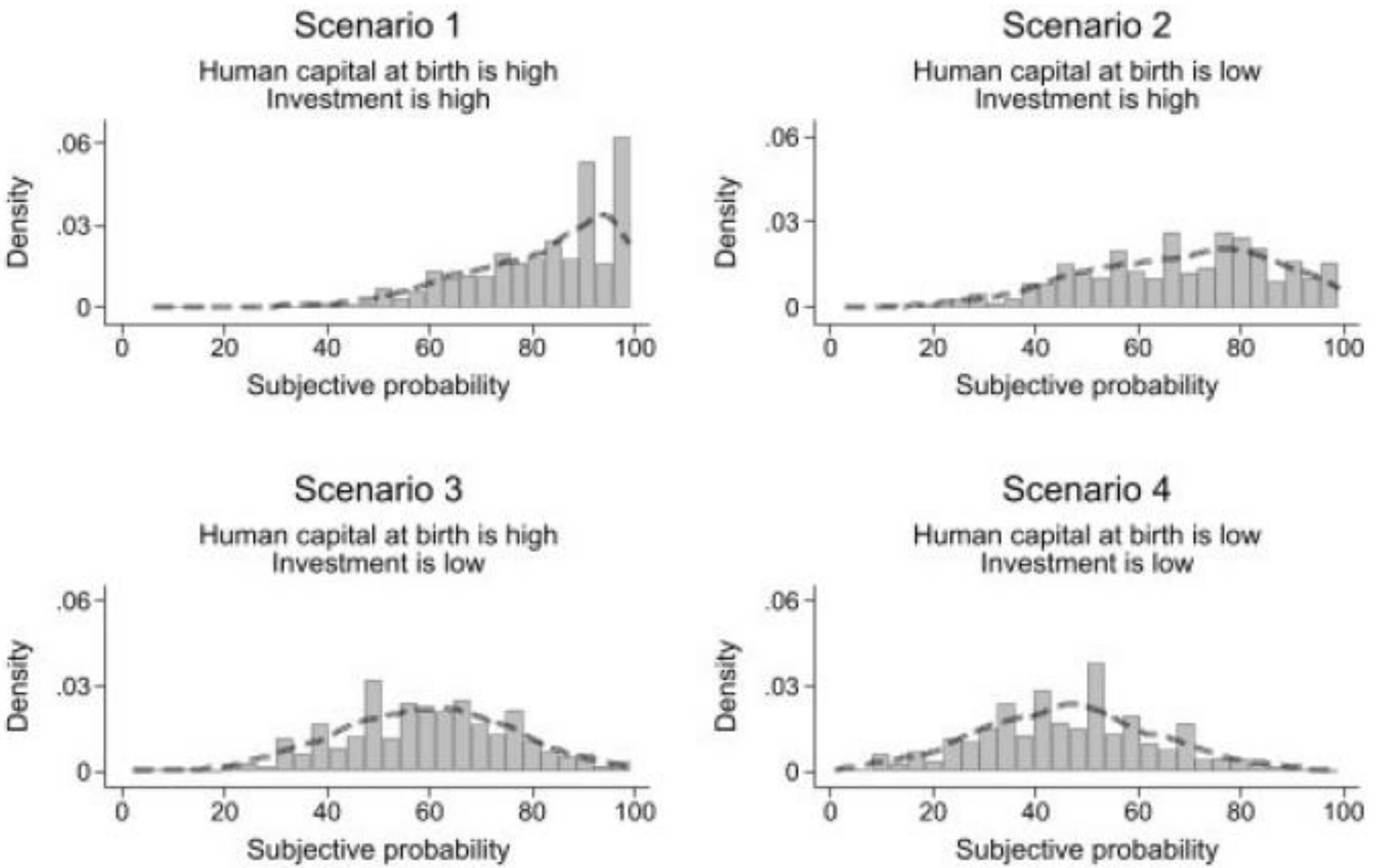
- This property of high, but imperfect, correlation is an essential result for addressing heaping and measurement error.

- Figure 6 presents histograms of the expected log of human capital at age two.

- To obtain these estimates, we proceed in two steps.

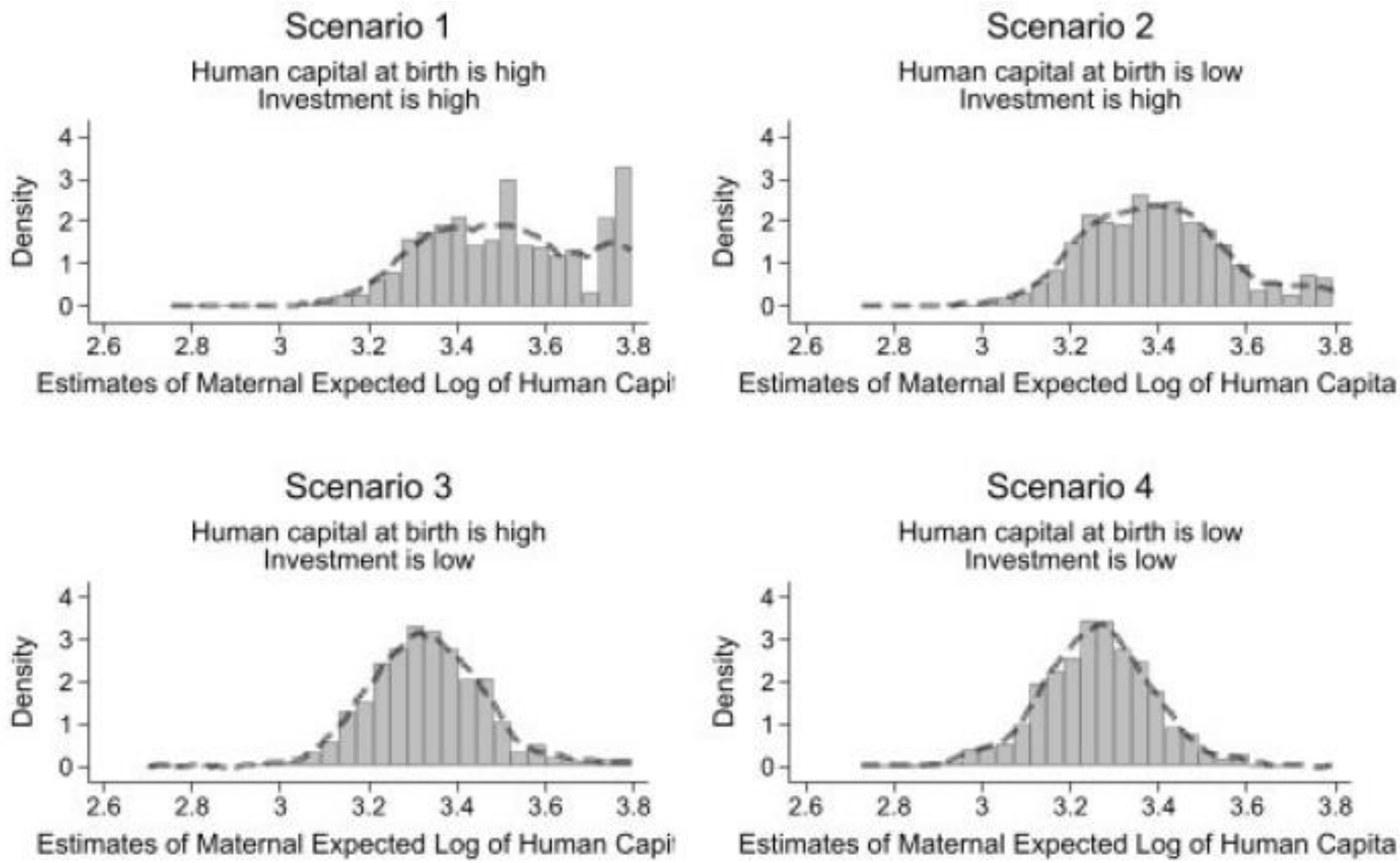
- In the first step, we invert the IRT equation (5). The input in equation (5) is the subjective probability for each scenario and MSD item.

Figure 5



Notes: This figure shows the histograms of subjective probability after we average maternal reports across the MSD items for each scenario of human capital at birth and investment. The result is that subjective probabilities are far less likely to exhibit heaping that we observed in Figure 4.

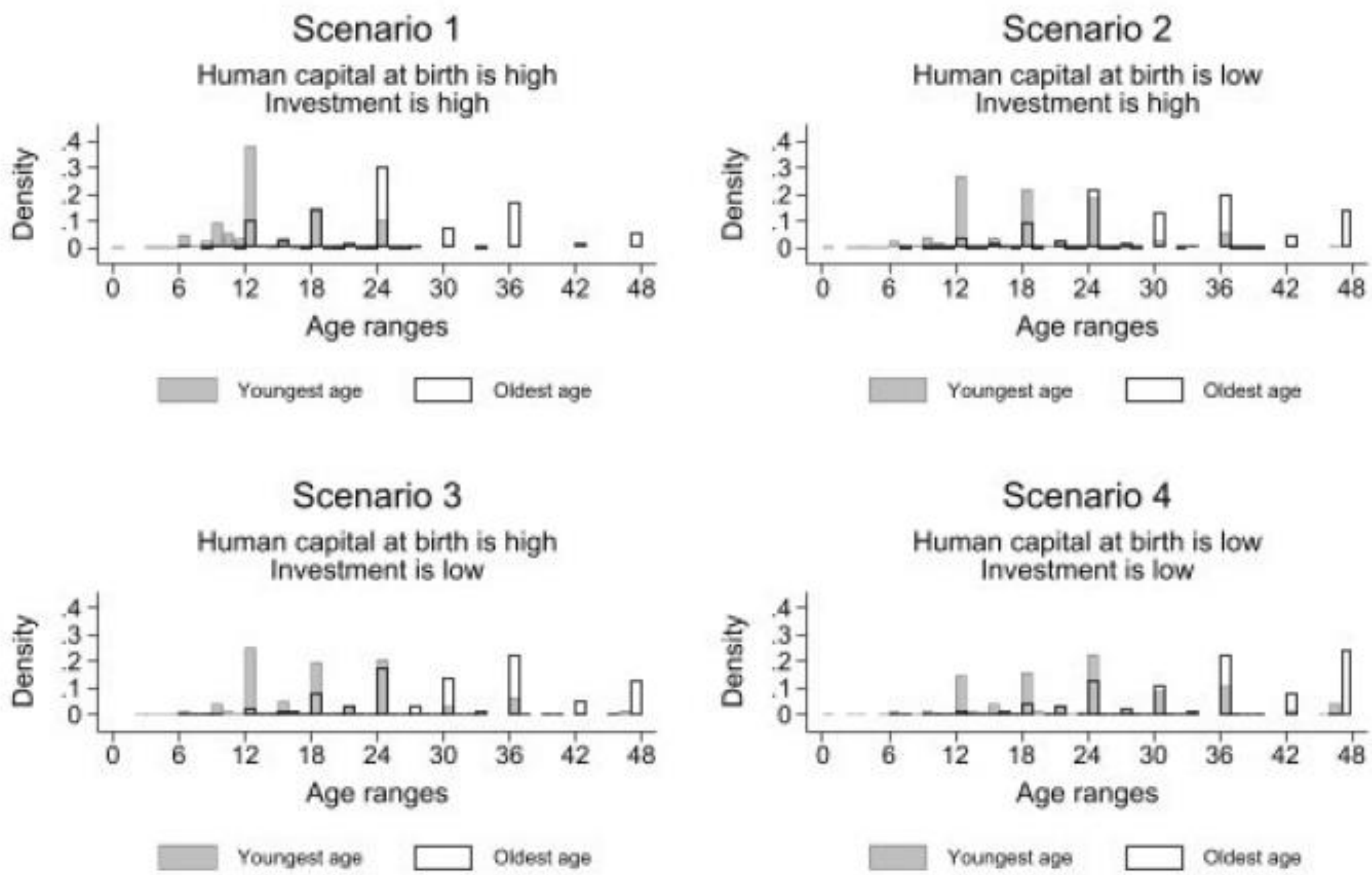
Figure 6



Notes: This figure displays the histograms of error-ridden measures of the expectation of the natural log of human capital at age two years for each scenario of human capital at birth and investments. To produce these measures, we proceed in two steps. In the first step, we transform, for each MSD item and scenario, the subjective probability data to an error-ridden, MSD-item specific, measure of the expectation of the natural log of human capital.

- Figure 7 presents the histograms of study participants' reports of the youngest and the oldest ages that children would learn how to "speak a partial sentence."
- We show the youngest age in solid gray bars, and the oldest age in white bars delimited with black lines.
- As in the subjective probability form, we observe heaping, particularly at multiples of 6 months.

Figure 7



Notes: This figure plots the histograms of maternal reports of age ranges for the MSD item "speak a partial sentence with three words or more." The solid gray bars show the youngest ages children will learn this MSD task for each scenario of human capital at birth and investments.

- In table 3, we display the mean, median, and standard error of youngest and oldest ages for each MSD item and scenario of human capital and investments.

- As in the subjective probability form, we see that mean and median tend to move in predictable patterns with the inputs.

Table 3: Summary Statistics of Answers to the Age-Range Elicitation Form

Panel A

Brief description of the MSD item	Scenario 1 Human capital at birth is "high" Investment is "high"						Scenario 2 Human capital at birth is "low" Investment is "high"					
	Youngest age			Oldest age			Youngest age			Oldest age		
	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
Speaks partial	13.96	12.00	5.72	25.37	24.00	9.78	17.98	18.00	7.96	30.03	30.00	10.49
sentence	(0.20)	(0.00)	(1.49)	(0.34)	(0.00)	(2.10)	(0.28)	(0.42)	(1.97)	(0.37)	(0.76)	(2.04)
Counts three	23.10	24.00	9.18	34.14	36.00	9.97	25.85	24.00	10.10	36.35	36.00	9.77
objects correctly	(0.32)	(0.25)	(1.96)	(0.35)	(0.50)	(1.88)	(0.35)	(0.25)	(2.07)	(0.34)	(0.25)	(1.89)
Knows own age	19.99	18.00	8.19	31.07	30.00	9.62	23.28	24.00	9.20	34.02	36.00	9.66
and sex	(0.29)	(0.50)	(1.80)	(0.34)	(1.01)	(1.87)	(0.32)	(0.50)	(1.94)	(0.34)	(0.50)	(1.85)
Says first and last	23.62	24.00	9.21	34.33	36.00	9.41	26.26	24.00	9.75	36.51	36.00	9.12
names together	(0.32	(0.00)	(1.94)	(0.33)	(0.50)	(1.85)	(0.34)	(0.50)	(2.01)	(0.32)	(0.25)	(1.83)

Notes: This table shows the youngest and oldest age for each MSD item and scenario.

Table 3: Summary Statistics of Answers to the Age-Range Elicitation Form, Cont'd

Brief description of the MSD item	Panel B: Scenario 3 — Human capital at birth is "high", Investment is "low": Youngest age			Oldest age			Scenario 4 — Human capital at birth is "low", Investment is "low": Youngest age			Oldest age		
	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
Speaks partial	19.07	18.00	8.33	30.79	30.00	10.14	22.78	24.00	9.65	34.32	36.00	10.41
sentence	(0.29)	(0.25)	(2.07)	(0.35)	(0.76)	(2.00)	(0.34)	(0.50)	(2.15)	(0.36)	(0.50)	(1.97)
Counts three	27.55	24.00	10.24	37.73	36.00	8.89	30.44	30.00	10.83	40.24	42.00	8.64
objects correctly	(0.36)	(0.50)	(2.04)	(0.31)	(0.50)	(1.78)	(0.38)	(0.92)	(2.13)	(0.30)	(1.01)	(1.94)
Knows own age	25.35	24.00	9.42	36.17	36.00	8.98	28.84	30.00	10.17	39.06	42.00	8.91
and sex	(0.33)	(0.00)	(1.97)	(0.31)	(0.00)	(1.79)	(0.35)	(0.50)	(2.03)	(0.31)	(0.50)	(1.89)
Says first and last	28.01	27.00	9.87	38.32	36.00	8.61	30.81	30.00	10.24	40.72	42.00	8.33
names together	(0.34)	(0.50)	(1.95)	(0.30)	(0.50)	(1.83)	(0.36)	(0.50)	(2.08)	(0.29)	(0.67)	(1.96)

Notes: This table shows the youngest and oldest age for each MSD item and scenario.

- Table 4 provides additional evidence of the influence of Δ_0 and Δ_1 on our data.
- For each MSD item and each scenario, we estimate the share of the answers in which the youngest age is above 24 months, and we assess the fraction of answers in which the oldest age is below 24 months.
- Again, we find three predictable patterns.

Table 4: The fraction of Youngest Age above 24 months or Oldest Age below 24 months

	Panel A			
	Scenario 1 Human capital at birth is "high" Investment is "high"		Scenario 2 Human capital at birth is "low" Investment is "high"	
Brief description of the MSD item	Youngest age Fraction above 24 months	Oldest age Fraction below 24 months	Youngest age Fraction above 24 months	Oldest age Fraction below 24 months
Speaks partial sentence	0.129	0.657	0.290	.0437
	(0.335)	(0.475)	(0.454)	(0.496)
Counts three objects correctly	0.573	0.280	0.642	0.186
	(0.495)	(0.449)	(0.480)	(0.389)
Knows own age and sex	0.422	0.403	0.540	0.264
	(0.494)	(0.491)	(0.499)	(0.441)
Says first and last names together	0.620	0.236	0.698	0.155
	(0.486)	(0.425)	(0.459)	(0.362)

Notes: This table shows the fraction of reported youngest (oldest) age above (below) 24 months for each MSD item and scenario. The fraction of youngest (oldest) age increases (decreases) with item difficulty and increase (decrease) with scenarios of human capital at birth and investments. Standard error in parentheses.

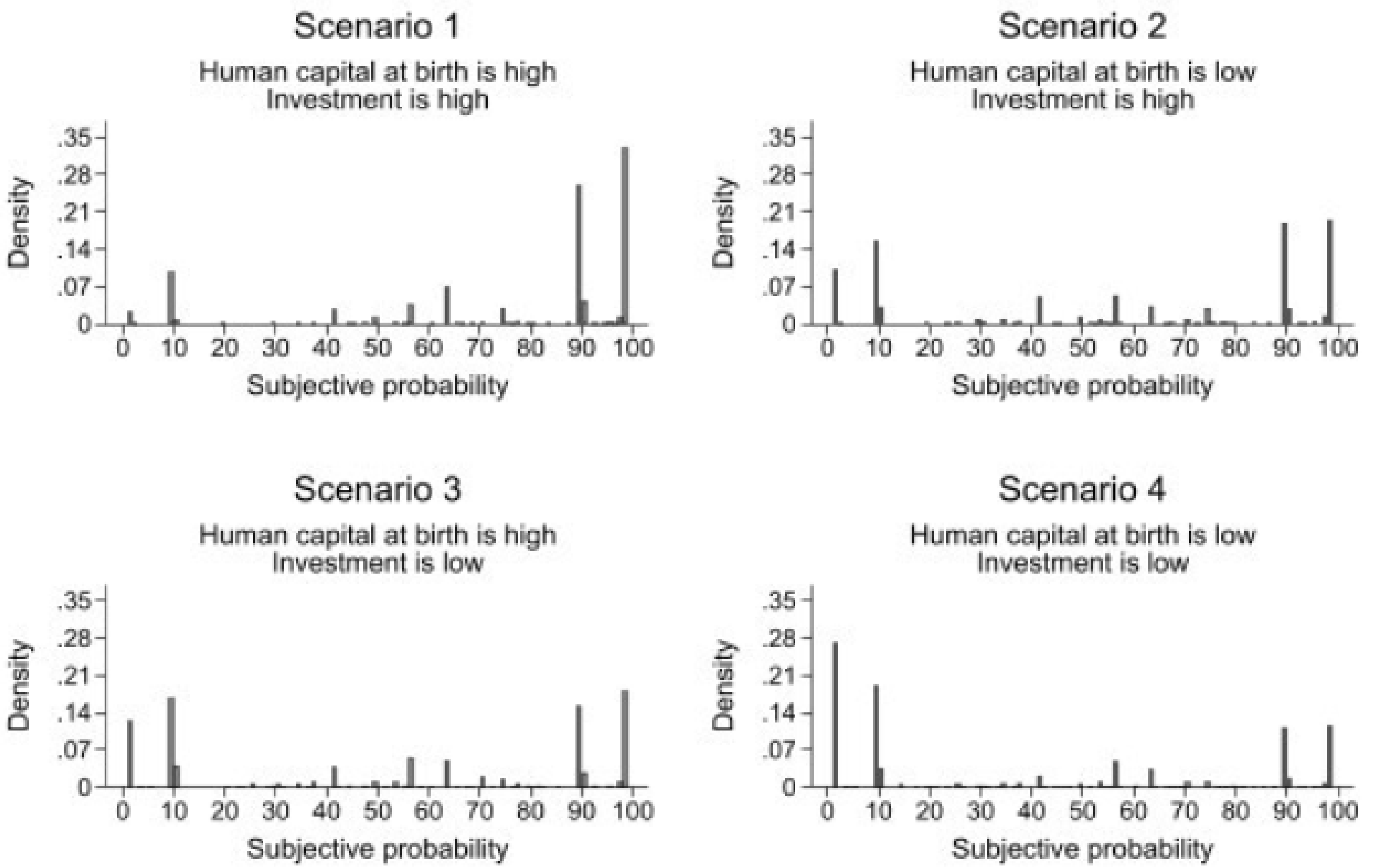
Table 4: The fraction of Youngest Age above 24 months or Oldest Age below 24 months, Cont'd

	Panel B			
	Scenario 3 Human capital at birth is "high" Investment is "low"		Scenario 4 Human capital at birth is "low" Investment is "low"	
Brief description of the MSD item	Youngest age Fraction above 24 months	Oldest age Fraction below 24 months	Youngest age Fraction above 24 months	Oldest age Fraction below 24 months
Speaks partial sentence	0.335	0.378	0.505	0.263
	(0.472)	(0.485)	(0.500)	(0.440)
Counts three objects correctly	0.706	0.125	0.792	0.078
	(0.456)	(0.331)	(0.406)	(0.268)
Knows own age and sex	0.635	0.161	0.740	0.105
	(0.482)	(0.367)	(0.439)	(0.306)
Says first and last names together	0.740	0.111	0.811	0.072
	(0.439)	(0.314)	(0.391)	(0.258)

Notes: This table shows the fraction of reported youngest (oldest) age above (below) 24 months for each MSD item and scenario. The fraction of youngest (oldest) age increases (decreases) with item difficulty and increase (decrease) with scenarios of human capital at birth and investments. Standard error in parentheses.

- Figure 8 provides the histograms of the estimates of the subjective probability based on age ranges.
- The histograms show not only heaping but also that respondents concentrate answers at low or high values of the probability range.
- When human capital at birth and investments are both "high," the heaping is pronounced at high probabilities.
- When both are "low," we observe the opposite.
- When one is "low" and the other is "high," then the heaping is more or less equal at both extremes of the probability range.

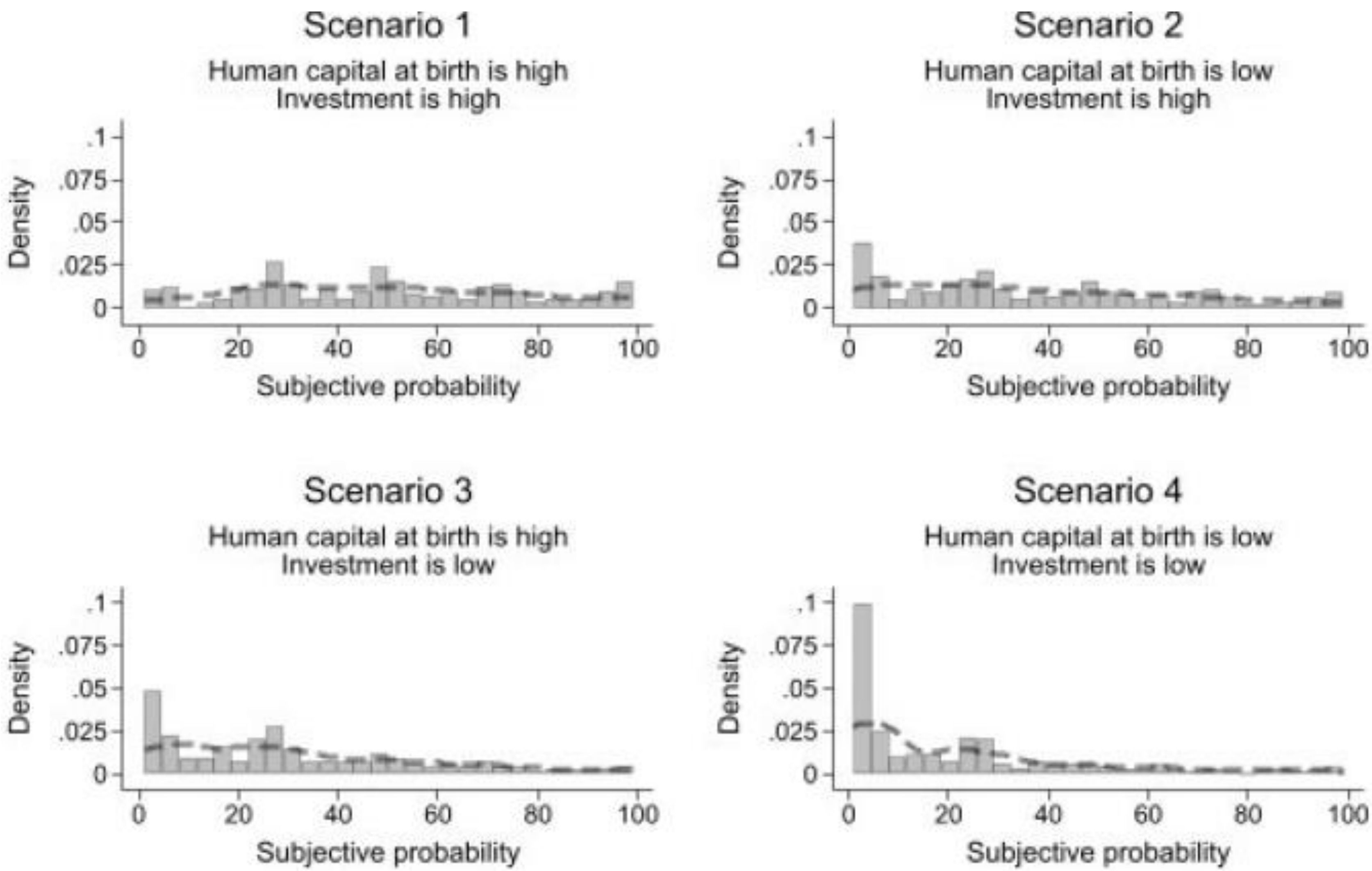
Figure 8



Notes: This figure shows the subjective probability data after transformation from age ranges. In this figure, we show the data for the MSD item "speak a partial sentence with three words or more."

- Figure 9 shows the histograms of the averaged (or 26 aggregated) probabilities.
- For all of the scenarios, we can eliminate heaping at high levels of probabilities when we average across MSD items.
- For Scenarios 3 and 4, there still is some heaping at low levels of probabilities.

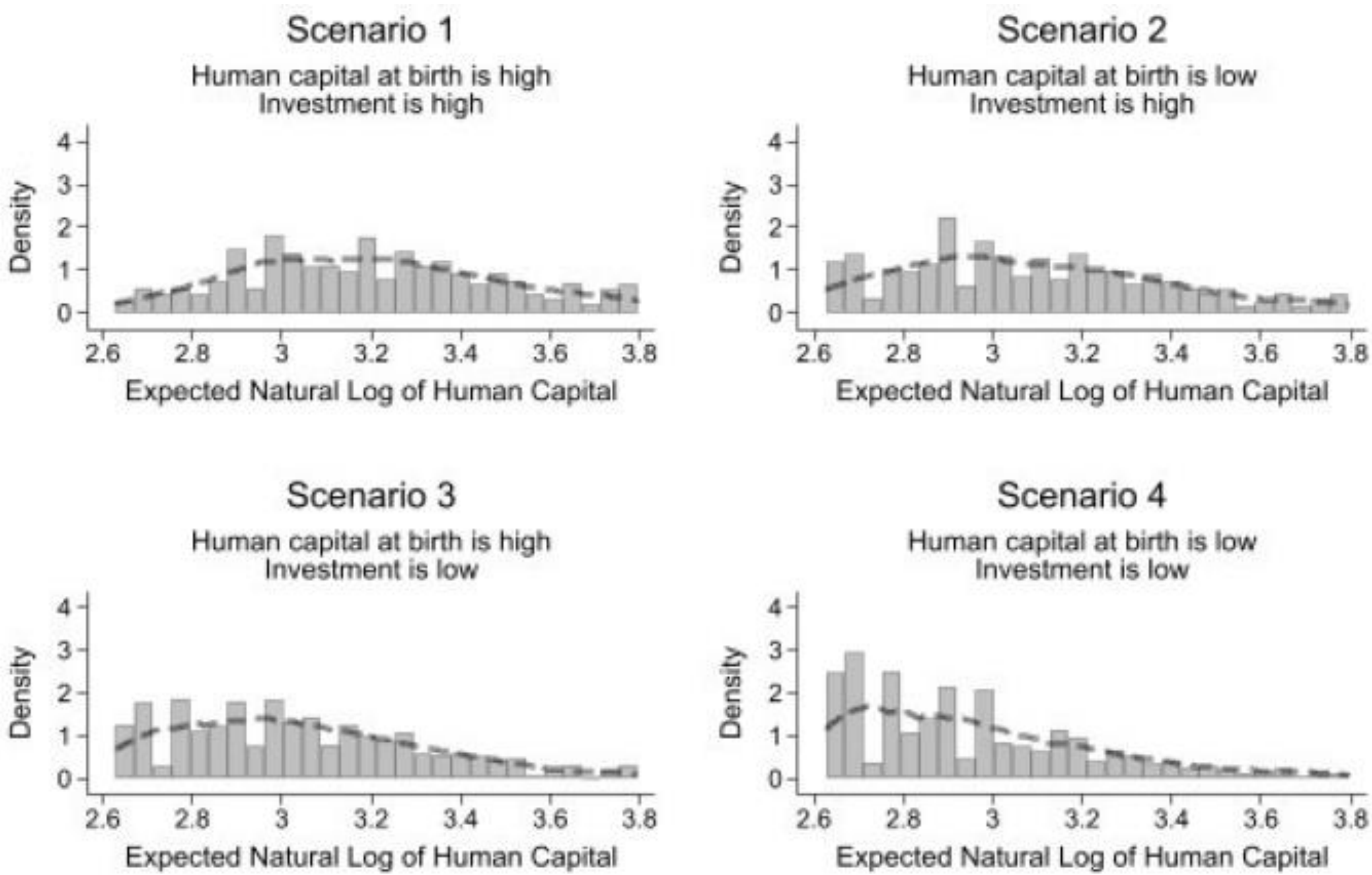
Figure 9



Notes: This figure shows the histograms of the subjective probability average across MSD items for each scenario.

- Figure 10 shows the results.
- It is interesting to contrast Figure 10 with the corresponding one from the elicitation of subjective probability (Figure 6).
- In Figure 6, the distribution of answers for Scenario 1 is left-skewed.

Figure 10



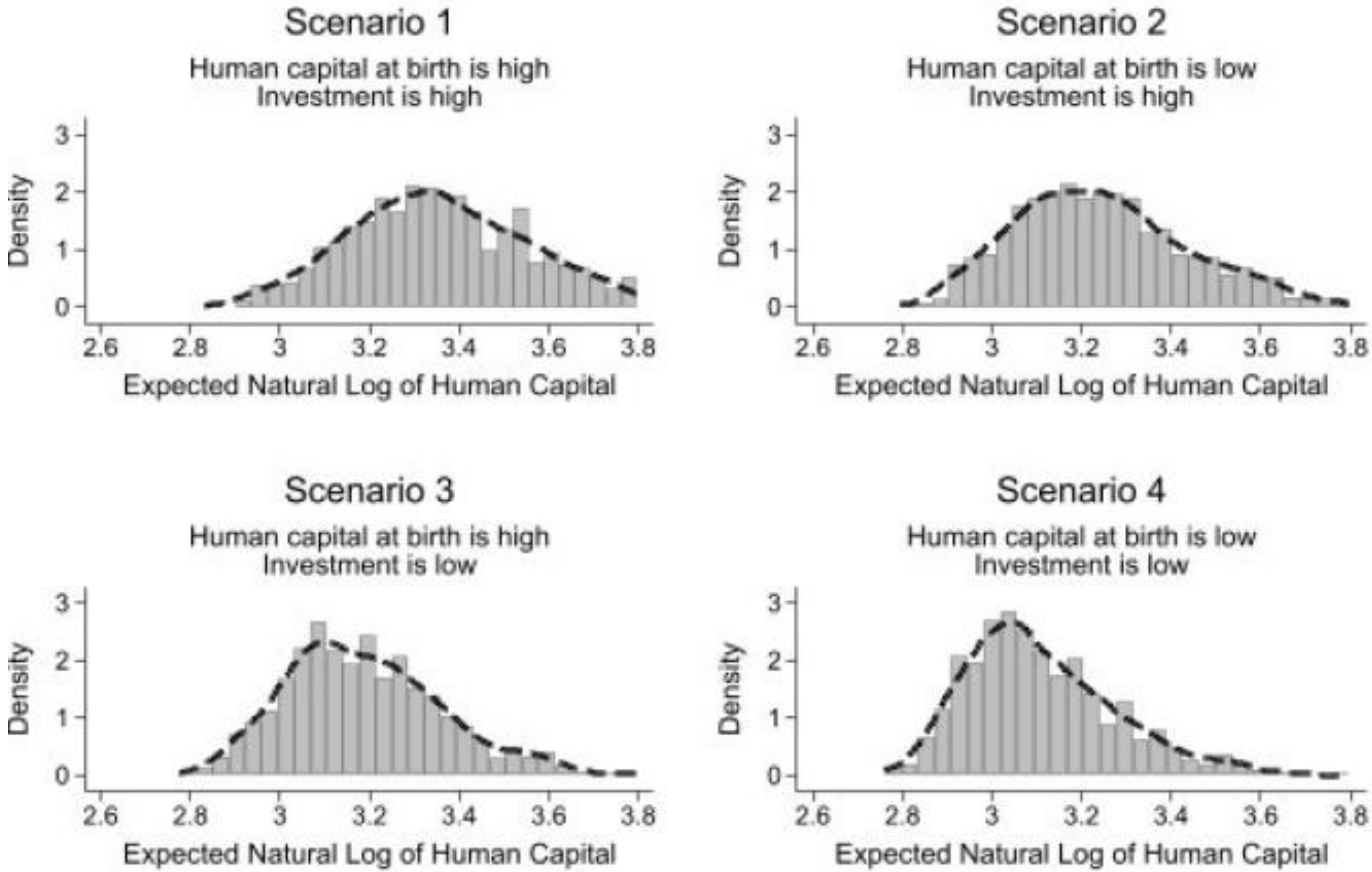
Notes: This figure shows the histograms of the averaged error-ridden measures of the expectation of the natural log of human capital for each scenario when we average across MSD items for the age ranges elicitation form.

4.3 Measurement Error and Testable Restrictions

- These “super-aggregate” measures still produce the same predictable patterns, but now heaping is no longer a significant feature of the data.

- It is the variation across scenarios in Figure 11 that the RCM estimator will explore to estimate the individual-level MSE parameters.

Figure 11



Notes: This figure shows the histograms of the averaged error-ridden measures of the expectation of the natural log of human capital for each scenario when we average both across MSD items and elicitation forms.

- Table 5 shows the results of our analysis.
- For the subjective probability elicitation form, we find that six factors summarize the informational content of the data.
- The six factors have patterns precisely as predicted by the model presented in Section 2.7.
- Factors 2–5 capture information due to the term $\lambda_{i,j}^{f}$.

Table 5: Factor Loadings of the Measurements of Expectation of Human Capital at Age 2

Subjective Probability-Elicitation Form

			Factor Variances					
			1	2	3	4	5	6
			2.951	2.608	2.538	2.362	2.320	0.733
			Factor Loadings					
Equation Number	Latent Variable	MSD Item	1	2	3	4	5	6
1	$\mu_{\psi,3}$	1	0.008	-0.034	-0.041	**-0.886**	-0.044	**0.170**
2	$\mu_{\psi,3}$	2	-0.043	-0.053	**-0.842**	-0.062	-0.048	**0.257**
3	$\mu_{\psi,3}$	3	-0.048	-0.070	-0.033	-0.031	**-0.861**	**0.277**
4	$\mu_{\psi,3}$	4	-0.056	**-0.831**	-0.081	-0.030	-0.056	**0.249**
5	$\mu_{\psi,2}$	1	**0.693**	-0.047	-0.040	**0.489**	-0.100	-0.033
6	$\mu_{\psi,2}$	2	**0.777**	-0.054	**0.369**	-0.061	-0.047	0.008
7	$\mu_{\psi,2}$	3	**0.827**	-0.059	-0.068	-0.057	**0.381**	-0.002
8	$\mu_{\psi,2}$	4	**0.767**	**0.401**	-0.043	-0.051	-0.073	0.015
9	$\mu_{\psi,1}$	1	-0.024	0.054	0.045	**0.841**	0.056	**0.300**
10	$\mu_{\psi,1}$	2	0.029	0.076	**0.826**	0.036	0.035	**0.341**
11	$\mu_{\psi,1}$	3	0.008	0.056	0.046	0.045	**0.841**	**0.347**
12	$\mu_{\psi,1}$	4	0.011	**0.801**	0.052	0.068	0.103	**0.336**

Notes: This table displays the factor loadings when we estimate the latent variable model (18) in Section 2.7. The data come from the subjective elicitation form.

- The results in Table 6 show that the factor analysis of the age-range form generates a factor structure that satisfies the predictions of our model.

- As expected, the analysis generates four factors (Factors 1–4) that summarize the correlation within MSD item measures.

Table 6: Factor Loadings of the Measurements of Expectation of Human Capital at Age 2

Age-Range Elicitation Form

			Factor Variances					
			1	2	3	4	5	6
			2.951	2.608	2.538	2.362	2.320	0.733
			Factor Loadings					
Equation Number	Latent Variable	MSD Item	1	2	3	4	5	6
1	$\mu_{\psi,3}$	1	**-0.930**	-0.020	-0.026	-0.024	0.046	**0.056**
2	$\mu_{\psi,3}$	2	0.024	-0.004	**-0.896**	-0.001	0.021	**0.129**
3	$\mu_{\psi,3}$	3	-0.001	**-0.908**	-0.026	-0.027	-0.014	**0.147**
4	$\mu_{\psi,3}$	4	-0.001	-0.025	0.011	**-0.888**	0.003	**0.180**
5	$\mu_{\psi,2}$	1	**0.755**	-0.010	-0.019	-0.020	**0.188**	-0.054
6	$\mu_{\psi,2}$	2	0.010	-0.013	**0.638**	0.004	**0.411**	-0.024
7	$\mu_{\psi,2}$	3	0.016	**0.605**	-0.051	-0.049	**0.447**	0.005
8	$\mu_{\psi,2}$	4	-0.026	-0.005	-0.026	**0.617**	**0.422**	0.020
9	$\mu_{\psi,1}$	1	**0.858**	0.007	-0.014	0.002	-0.023	**0.203**
10	$\mu_{\psi,1}$	2	0.017	0.036	**0.886**	-0.002	0.011	**0.183**
11	$\mu_{\psi,1}$	3	0.019	**0.865**	0.034	0.021	0.010	**0.188**
12	$\mu_{\psi,1}$	4	0.032	0.001	0.027	**0.863**	-0.003	**0.164**

Notes: This table displays the factor loadings when we estimate the latent variable model (18) in Section 2.7 for the data from the age ranges form.

4.4 Estimates of MSE about the Technology of Skill Formation

- Among other things, Table 7 presents the percentage of respondents for whom we can reject the null hypothesis.

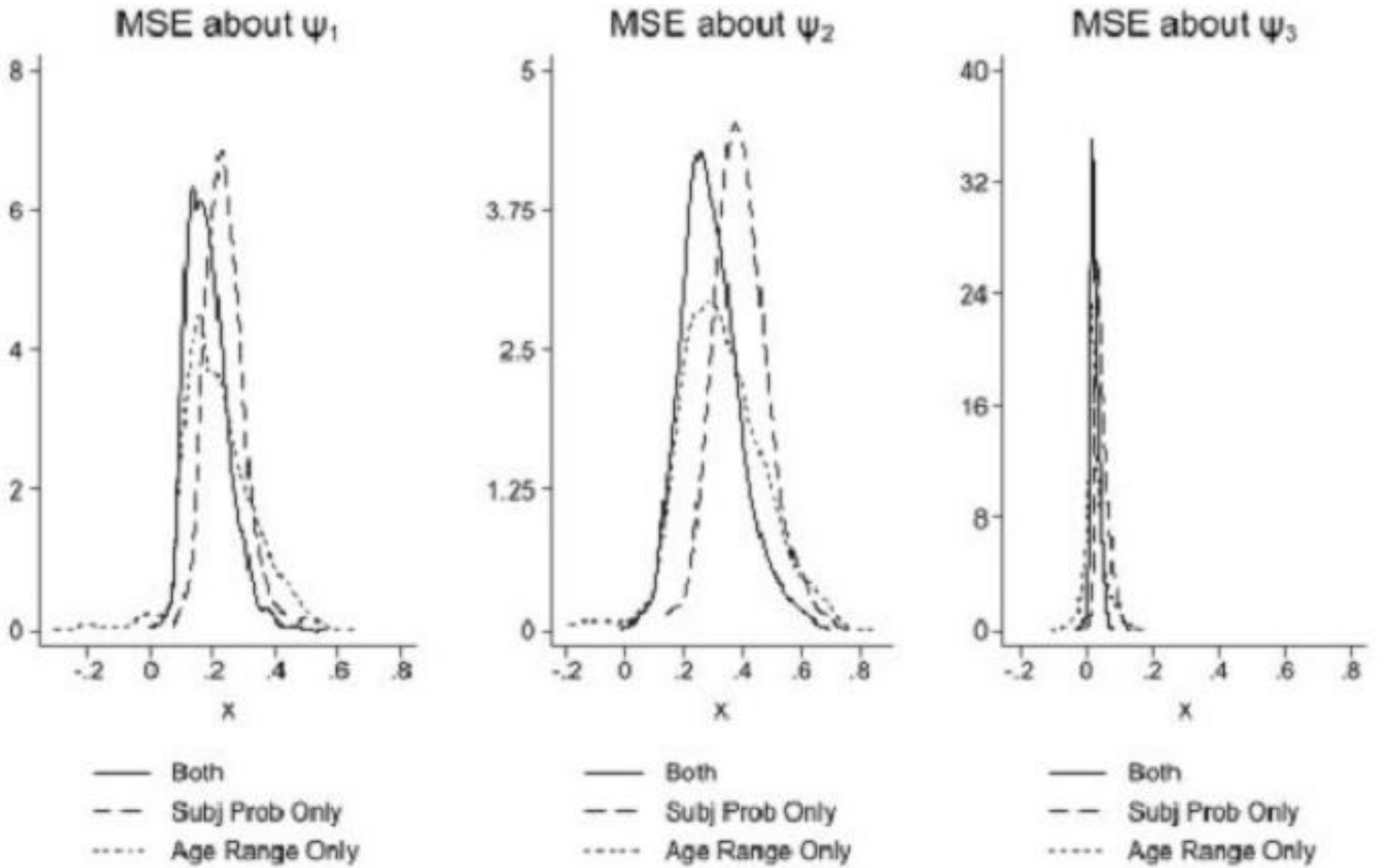
Table 7: Means of Maternal Beliefs about the Technology of Skill Formation

	Both Forms	Subjective-Probability Form	Age-Range Form
$\hat{\mu}_{\psi,1}$	0.182	0.238	0.216
	(0.008)	(0.010)	(0.011)
$\hat{\mu}_{\psi,2}$	0.293	0.396	0.330
	(0.009)	(0.011)	(0.012)
$\hat{\mu}_{\psi,3}$	0.028	0.050	0.029
	(0.006)	(0.009)	(0.009)
	Test of Parameter Constancy		
$\text{H}_0: \mu_{i,\psi,j} = \mu_{\psi,j}\ \forall\ i, j$ $\text{H}_1: \mu_{i,\psi,j} \neq \mu_{\psi,j}$ *for at least one* i, j	Reject H_0	Reject H_0	Reject H_0
	Hypotheses Tests (% Reject H_0)		
$\text{H}_0: \mu_{i,\psi,1} = 0;\ H_1: \mu_{i,\psi,1} \neq 0$	68.0%	92.0%	44.3%
$\text{H}_0: \mu_{i,\psi,2} = 0;\ H_1: \mu_{i,\psi,2} \neq 0$	88.2%	98.7%	71.3%
$\text{H}_0: \mu_{i,\psi,3} = 0;\ H_1: \mu_{i,\psi,3} \neq 0$	5.7%	14.5%	2.6%

Notes: Generalized least squares standard error in parentheses

- Figure 12 displays the marginal densities of MSE about ψ_j for j = 1,2,3.
- We fix the horizontal axis so that they take common values across all j.
- It is, therefore, easy to see that the heterogeneity in MSE about ψ_3 is of little importance, regardless of whether we combine data from different forms or use data from each one of the forms separately.
- Figure 12 also shows that the marginal distributions of MSE implied by the age range form have lower means and higher variances for both ψ_1 and ψ_2.

Figure 12



Notes: This figure shows the kernel density estimates of the individual-level MSE parameters for each elicitation form separately and jointly. We fix the horizontal axis to be able to compare the densities for the different parameters.

- Due to space constraints, Table 8 shows only the estimated coefficients on dummies for household income.

- We single out the income dummies because they are the only ones that systematically correlate with all of the MSE parameters in whatever way we estimate them, that is, when we combine data from both forms, when we use data only from one of the forms separately, or when we apply different interpolation methods for the age-range forms.

Table 8: Correlation between MSE and Household Income of PHD Study Participants

	Both Forms			Subjective-Probability Form			Age-Range Form		
Dummies for household income per year (y)	$\mu_{i,\psi,1}$	$\mu_{i,\psi,2}$	$\mu_{i,\psi,3}$	$\mu_{i,\psi,1}$	$\mu_{i,\psi,2}$	$\mu_{i,\psi,3}$	$\mu_{i,\psi,1}$	$\mu_{i,\psi,2}$	$\mu_{i,\psi,3}$
$\mathbf{1}(\$25{,}000 \leq y < \$55{,}0000)$	0.22**	0.35***	0.19*	-0.02	0.32***	0.02	0.29***	0.28***	0.24**
	(0.10)	(0.09)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.09)	(0.11)
$\mathbf{1}(\$55{,}000 \leq y < \$105{,}0000)$	-0.17	0.37***	-0.25**	-0.35***	0.45***	-0.22*	-0.09	0.25**	-0.14
	(0.13)	(0.12)	(0.11)	(0.16)	(0.12)	(0.13)	(0.12)	(0.12)	(0.10)
$\mathbf{1}(y \geq \$105{,}000)$	-0.51***	0.47***	-0.53***	-0.67***	0.60***	-0.40***	-0.45***	0.26*	-0.37***
	(0.13)	(0.14)	(0.12)	(0.12)	(0.14)	(0.13)	(0.13)	(0.14)	(0.11)
Observations	822	822	822	822	822	822	822	822	822
R^2	0.071	0.064	0.090	0.080	0.062	0.080	0.058	0.071	0.046

Notes: This table shows the association of MSE parameters with household income. We standardized MSE parameters so that they have mean = 0 and variance = 1.

- To create Table 9, we residualized the "structural" factors to make sure that we purge any correlation with "measurement error" factors.

- We then standardized the residualized factors to aid in the interpretation of our findings.

- We analyze the relationship between "structural" factors and MSE as estimated from subjective probability and age-range forms separately.

Table 9: Relationship between MSE Parameters and Structural Factors

Variable	Subjective-Probability Form		
	$\mu_{i,\psi,1}$	$\mu_{i,\psi,2}$	$\mu_{i,\psi,3}$
Structural Factor 1	-0.294***	0.966***	-0.052
	(0.010)	(0.023)	(0.033)
Structural Factor 6	0.892***	0.031*	0.677***
	(0.017)	(0.017)	(0.043)
Observations	822	822	822
R^2	0.935	0.871	0.468

Variable	Age-Range Form		
	$\mu_{i,\psi,1}$	$\mu_{i,\psi,2}$	$\mu_{i,\psi,3}$
Structural Factor 5	-0.038***	0.855***	-0.015
	(0.014)	(0.022)	(0.035)
Structural Factor 6	0.923***	0.147***	0.506***
	(0.019)	(0.022)	(0.042)
Observations	822	822	822
R^2	0.849	0.727	0.255

Notes: This table shows the relationships between the Structural Factors we introduced in 2.7 with the MSE parameters in equation (2).

4.5 MSE and Correlation with the HOME Score

- Table 10 shows that $\mu_{i,\psi,2}$ consistently predicts higher levels of investments as measured by the scores on the HOME scale.

- One standard deviation in $\mu_{i,\psi,2}$ is associated with 11% of a standard deviation in investments even after we control for the mother's race, ethnicity, marital status, educational attainment, and household income.

Table 10: Correlation between MSE Latent Variables and HOME Score

Variable	Both		Subjective Probability		Age Range	
Standardized $\mu_{i,v,1}$	-0.024	-0.002	-0.095	-0.058	-0.014	0.031
	(0.081)	(0.074)	(0.080)	(0.074)	(0.059)	(0.053)
Standardized $\mu_{i,v,2}$	0.167***	0.114***	0.119***	0.098**	0.170***	0.083**
	(0.045)	(0.039)	(0.043)	(0.040)	(0.045)	(0.038)
Standardized $\mu_{i,v,3}$	-0.086	0.010	-0.040	0.034	-0.058	-0.014
	(0.067)	(0.061)	(0.066)	(0.062)	(0.048)	(0.042)
Demographic variables included	No	Yes	No	Yes	No	Yes
Observations	687	687	687	687	687	687
R^2	0.037	0.270	0.034	0.271	0.031	0.265

Notes: This table shows the correlation between the measure of investment (the Home Observation for the Measurement of the Environment – HOME scores) and the MSE parameters.

- Table 11 presents the results.
- We mark with an “x” the variables the LASSO regression included in the prediction model.

Table 11: LASSO Regression

Dependent Variable: Standardized HOME Scores

	Both Forms		Subjective Probability		Age Range	
Variable	CV	Adaptive	CV	Adaptive	CV	Adaptive
$\mu_{iv,1}$			x		x	
$\mu_{iv,2}$	x	x	x	x	x	x
$\mu_{iv,3}$						

Notes: This table shows the MSE parameters that are included in the final prediction model, as estimated by LASSO regressions. The dependent variable is the measure of investment (the Home Observation for the Measurement of the Environment – HOME scores).

4.6 Robustness

5. Conclusion