Norwegian University of Life Sciences (NMBU)

Gender differences in investments and risk preferences

Stein T. Holden and Mesfin Tilahun

Norwegian University of Life Sciences
Centre for Land Tenure Studies

Centre for Land Tenure Studies Working Paper 02/22
ISBN: 978-82-7490-299-2

Highlights

Gender differences in investments and risk preferences

Stein T. Holden,Mesfin Tilahun

- We study investments and risk attitudes of 822 business men and women in 111 formal business groups in Ethiopia
- Five investment variables versus three experimental tools used to obtain dis-aggregated risk preference variables
- Women invested significantly less on average but there was high inequality in investments within gender types
- Women were on average more loss averse, had higher CRRA-r and lower Prelec β than men
- Gender differences in resource endowments and risk attitudes explain some but not all of the gender difference in investments

Gender differences in investments and risk preferences[⋆,⋆⋆]

Stein T. Holden[a], Mesfin Tilahun[a,b]

[a]School of Economics and Business, Norwegian University of Life Sciences, Towerbuilding, Ås, 1432, Viken, Norway
[b]Department of Economics, Mekelle University, Mekelle, , Tigray, Ethiopia

ARTICLE INFO

Keywords:
Gender difference
Individual investment
Risk preferences
Prospect theory
Cohen's d
Business groups
Northern Ethiopia
JEL codes
C93, D90

ABSTRACT

We analyze individual investment behavior among 822 young men and women that are members of 111 formal business groups in northern Ethiopia. We collected baseline data and investment data one year later combined with incentivized field experiments to obtain dis-aggregated risk preference data. We find that business women on average invest significantly less at individual level than business men but Cohen's d values for the gender difference are moderate in size. Women are found to have higher Constant Relative Risk Aversion coefficients, to be more loss averse, but also to be more optimistic in their expectations than men. Women were also poorer in non-land assets, came from more land-poor parents and had lower incomes. The gender differences in risk attitudes and baseline endowments could explain some of but not all of the gender differences in investments.

1. Introduction

The existence of gender differences in risk preferences has received substantial attention and has been subject to many experimental studies in the behavioral and experimental literature (Eckel and Grossman, 2002, 2008a,b; Gong and Yang, 2012; Croson and Gneezy, 2009; Charness and Gneezy, 2012; Nelson, 2016; Filippin and Crosetto, 2016). While many studies have found women to be less risk tolerant and to invest less than men, some of the more recent studies have warned against stereotyping and have found the gender differences to be fairly small (Nelson, 2016) and possibly partly explained by measurement errors associated with the method used for elicitation (Filippin and Crosetto, 2016).

Charness and Gneezy (2012) assess the gender differences using the risky investment game of Gneezy and Potters (1997) (GP) and find there to be "strong evidence for gender differences in risk taking". Their study is based on the data from nine studies with six of them having university students as study subjects. Two of the studies were in developing countries and with rural villagers as study subjects. These two studies used the one-shot version of the GP game (Gneezy, Leonard and List, 2009). The study by Charness and Gneezy (2012) has been criticized by Nelson (2016) who re-analyzed the same data and concluded that there is "not-so-strong evidence for gender differences in risk taking". One reason for the criticism is that the "strong difference", which applies at aggregate level, may be misinterpreted to apply at individual level. Nelson (2016) applies additional statistical measures that allow more careful assessment of the within- versus cross-gender differences in risk taking at individual and aggregate level in the GP risky investment game based on the data from the same studies that Charness and Gneezy (2012) used.

The GP risky investment game has been considered a useful and simple tool to elicit risk tolerance in an investment setting in field experiments (Charness and Viceisza, 2016; Dave, Eckel, Johnson and Rojas, 2010; Gillen, Snowberg and Yariv, 2019) and may be associated with less cognitive problems and measurement errors than e.g. the more complicated Holt and Laury (2002) Multiple Choice List (HL-MCL) approach, especially among subjects with limited education and numeracy skills. An advantage of the GP game is that it, in the one-shot version, is easy and quick to implement and it can easily be integrated in surveys. Another potential advantage of the GP game is that it is framed as an incentivized investment choice that potentially can predict real world investments (Gillen et al., 2019). In this study we assess this for investments made the following year after the GP game was played for a subject pool of 822 young business men (68%) and women (32%) that are members of 111 rural formal business groups in northern Ethiopia.

Many different tools have been used to elicit risk preferences and there exists no consensus on what the most appropriate tool is. The choice may depend on the study subjects, their cognitive ability, especially numeracy skills, but also the context, time and budget restrictions, and the purpose of the study. Gender differences have been found to differ across risk preference experimental tools (Filippin and Crosetto, 2016) and to be higher for the GP game than for the HL-MCL approach. They suggest that the gender difference is higher in games where subjects choose between safe and risky prospects and they show that it is small in the HL-MCL.

In this study we assess the gender differences with three different experimental tools for risk preferences. In two of the tools, the GP game and a Certainty Equivalent - Multiple Choice List (CE-MCL) experiment, subjects choose between safe and risky options. In the third, an incentivized loss aversion CL experiment, subjects make choices between

[⋆]This paper is the result of two research projects funded by the Norwegian Agency for Development Cooperation (NORAD) and the Research Council of Norway.

[∗]Corresponding author
stein.holden@nmbu.no (S.T. Holden)
www.steinholden.com (S.T. Holden)
ORCID(s): 0000-0001-7502-2392 (S.T. Holden); 0000-0003-3885-2017 (M. Tilahun)

two risky prospects at each decision point. This allows us to assess the gender differences across these three tools and whether the gender differences mainly occur in the two tools that compare safe and risky prospects. The "certainty effect" (preference for certain over risky prospects) has received substantial attention in the literature (Loomes and Sugden, 1982) but few studies have assessed whether these are stronger for women that for men.

Possible measurement errors and biases have received increasing attention in the literature on elicitation of risk preferences (Andersson, Holm, Tyran and Wengström, 2016; Crosetto and Filippin, 2016; Gillen et al., 2019; Vieider, 2018). Measurement errors can be a more serious problem in field experiments for subjects with limited education and numeracy skills than in lab experiments with university students. Measurement errors may also explain so-called "certainty effects" as demonstrated by Vieider (2018), who suggest that such effects may be driven by the salience of the reference points in the tools chosen. In the GP game the initial safe amount provided is the more salient, while in each CL in our CE-MCL experiment, the risky prospect is constant and more salient while the safe amounts vary. This may trigger more risk-taking than for the GP game, based on this salience of the reference point theory (Vieider, 2018).

We assess potential measurement errors and stability of the responses in the GP game by repeating it in the second round of the data collection. Our second (CE-MCL) experiment is used to estimate a Rank Dependent Utility (RDU) model with Fechner error and contextual utility (Fechner, 1860; Wilcox, 2008). This allows separation of noise from the estimated theoretical parameters in form of a Constant Relative Risk Aversion parameter (CRRA-r) and two probability weighting parameters based on a Prelec et al. (1998) 2-parameter weighting function. Our first objective is to assess the extent of gender differences in the three elicitation tools and the estimated parameters based on a within-subject design and to assess how the design characteristics and measurement error are associated with the gender differences.

In addition to the focus on gender differences in experimental variables, we study the gender differences in investments and basic socioeconomic variables. We combine survey data collected at two points in time one year apart. The investment data for the one year period between the two data collection points were categorized into investments in durable consumer goods, livestock, productive assets and other business investments and recorded in monetary terms, in addition to the total individual investments.

Our second objective is to assess whether the gender differences in individual investments are explained by gender differences in risk preferences or by gender differences in poverty (resource endowments). Our study is in a patriarchal study area with strong traditional gender differences in division of labor and decision-power, with male dominance within families and in society in general. We may therefore expect to find large gender differences in investments but these may not be strongly associated with gender differences in risk preferences but may be driven by gender discrimination in business and cultural norms.

Skewness and censoring in investment variables were addressed through log-transformation and normalization to measure gender differences in standard deviation units (Cohen's d effect sizes (Cohen, 1992)). We demonstrate the importance of the log-transformation and handling of censoring before assessing the sizes of the gender differences. Our results demonstrate significant but moderate gender differences in the investment variables.

In our final analysis we regress the log-transformed and normalized investment variables on the normalized risk preference variables by using an IV approach to handle measurement error and attenuation bias in the GP game. We combine the RDU estimated parameters and the loss aversion rank in a separate Prospect Theory model. We find large measurement errors in the GP game while the CE-MCL experiment provided stronger estimates of RDU parameters that also contributed to explaining investment behavior.

To our knowledge this is the first study in a developing country setting that analyses gender differences in investment behavior combined with a comprehensive experimental approach to elicit risk preferences while also accounting for measurement errors in the elicitation of preferences.

2. Survey and experimental design

2.1. Survey design

Based on a census of 742 rural youth business groups in five districts in Tigray Region of Ethiopia conducted in 2016[1], we carried our a survey of 120 youth business groups with up to 12 business group members per group in July-August 2016. A new follow-up survey was carried out one year later. Both survey rounds were combined with field experiments. Some attrition and implementation errors resulted in a final sample for analysis of 822 subjects from 111 business groups.

Baseline data on the socioeconomic characteristics of the subjects were collected in the 2016 survey and individual investment data were collected in the 2017 round for investments made in the period since the first round survey. Investments were divided in four categories; consumer durables, livestock, productive assets, and other business. In addition, we include a total investment variable that summarizes the expenditure on the four investment categories.

All sampled members of a business group were interviewed simultaneously by 12 trained enumerators using tablets that were programmed using the CSPro software. Three classrooms in local schools were used as field labs with one enumerator and one study subject placed in each corner of the room, to minimize communication and disturbance during the interviews and experiments.

[1] The census and the business group activity has been documented in Holden and Tilahun (2018). In this study we focus on the individual activity as the business group activity was only a part-time activity of the members.

2.2. Experimental design and timing

2.2.1. *The GP risky investment game*

The GP game was included in the baseline survey in 2016. Each study subject was allocated 30 Ethiopian Birr (ETB)[2], approximating a daily wage rate in the study areas at the time of the survey. They were then given the chance to invest all, some or nothing of the allocated amount in a risky lottery where the experimenter triple the invested amount and with a 50-50 chance of winning the tripled amount or lose it. With the invested amount being x, the winners would get $30 + 2x$ and the losers would get $30 - x$. To identify whether they won or lost, they were asked to draw one of two paper notes from the enumerators.

In the 2017 the same subjects were asked whether they remembered the game they played one year earlier, how much they invested, and whether they had won or lost in the game. After that they were asked how much they would invest in a hypothetical replay of the game. We combined the 2016 real game and the 2017 hypothetical game in combination with their cognitive memory of the game and the outcome to assess the measurement error or stability of responses in the game.

2.2.2. *Certainty Equivalent - Multiple Choice List (CE-MCL) experiment*

This experiment consisted of 12 CLs that allowed the identification of a certainty equivalent interval for a risky prospect that is constant in each CL in terms of a high (good) and a low (bad) outcome and with a fixed probability (p) of winning or losing (1-p). The probability of winning varied from 0.05 to 0.95 across CLs to allow mapping of the probability weighting function. More of the CLs had probabilities similar to the real world types of risks that the subjects faced in their natural resource management types of investment activities where low probability bad outcomes dominate. Appendix Table A1 provides an overview of the variation in the CL characteristics and Table A2 provides and example of one of the CLs. This experiment was implemented in the 2017 survey round. It is possible that the risk preferences changed from 2016 to 2017 due to exposure to shocks. We recorded subjects' self-reported shock exposure for the period. There were no serious (covariate) environmental shocks in this period in any of the study areas but some had been exposed to idiosyncratic shocks such a death in the family, sickness or other types of shocks.

To minimize the confounded starting point bias in the CLs, we randomized the starting row in each CL. We also randomized the order of the CLs for each respondent to prevent confounded order bias. Enumerator bias is another possible reason for measurement error. Each enumerator was randomly allocated one subject for each business group, making group and enumerator effects orthogonal on each other. Another type of bias leading to possible measurement error is bias towards the middle of a list and the placement of the risk-neutral row in a list (Andersson et al., 2016), The placement of the risk-neutral row is a CL-fixed characteristic but the variation across the 12 CLs allows testing for such potential bias. Our randomized starting point in each CL combined with a rapid elicitation approach that aimed to quickly narrow in on the switch point should limit the extent of bias towards the middle as the full list was not presented to the subjects. They were only presented with the risky prospect and a specific certain amount for each decision they had to make when narrowing down towards the switch point.

One of the 12 CLs, and one row in this CL were randomly chosen for payout. The preferred choice for this row, based on the location of the switch point in the CL, determined whether the subjects received the preferred certain amount or whether the risky prospect would be played, using a 20-sided die for the randomization tasks.

2.2.3. *Loss Aversion experiment*

This experiment was introduced as a single CL as the last experiment in 2017. The game implied the choice between two risky prospects with 50-50 probabilities of winning or losing in a set of binary choices. The sizes of the high and low outcomes in each of the risky prospects varied systematically in the CL such that size of the potential loss is reduced towards the bottom of the CL but so is also the expected return. The expected return is higher for the prospect (B) with highest loss, except for the first row in the CL, see Table A3 in the Appendix.

To ethically defend introducing an experiment with potential losses, the subjects had to have earned some cash in earlier experiments that they then could risk losing some of in this experiment. The design of the CL was inspired by Tanaka, Camerer and Nguyen (2010). However, we did not aim to use it to jointly estimate a loss aversion coefficient, only to get a loss aversion rank variable based on the switch point in the list.

Like in the CE-MCL experiment, the starting task row was randomized in advance and the rapid elicitation method was used to identify the switch point in the CL. The good and bad outcomes for the two prospects that are compared were demonstrated with money on the desk in front of the subjects. We expect prospect A to be chosen in task row 1 and Prospect B to be chosen in task row 9 and the switch point to occur somewhere between. The row number where subjects switch from prospect A to prospect B gives the loss aversion rank.

After the switch point has been identified, one task row is randomly chosen for real payout. The preferred risky prospect in that row is then used to play the real game and the real win/loss payment is made based on the random 50-50 draw for the preferred prospect.

3. Variable specification and estimation strategy

3.1. Selection issues

The business group program was initiated by the regional government to create youth employment and targeted

[2] They received two 10 ETB and two 5 ETB notes.

resource-poor rural youth, especially those that were landless or land-poor. These formal eligibility criteria therefore influenced the selection of business group members. There was also self-selection as the subjects who joined had to be motivated and apply to the program. They were also able to form groups by selecting group co-members from their own community. The fact that only about one third of the members were female show that males were more likely to join such a group. This may indicates a gender difference in the selection process that may also cause the gender difference among group members to be different from that in the general population. We cannot therefore generalize or findings on gender differences to the general population.

The baseline data tell us that male members are older and that female members on average come from more land-poor households and have fewer assets and less income. We cannot rule out that risk preferences affected the selection process and that more risk tolerant individuals were more likely to join such groups. This has implications for the external validity of our findings.

3.2. Assessment of gender differences and variable transformation

We calculated Cohen's $d = (\overline{Y}_m - \overline{Y}_f)/sd_Y$ to take into account the within-gender variation when assessing between-gender differences following Cohen (1992) and Nelson (2016). The investment variables need more careful treatment due to the skewness and left-censoring. To reduce the skewness, the investment variables were log-transformed. By adding a small positive value to all observations, we retain censored observations after log-transformation. For censored variables the size of the added value may influence standard deviations of the log-transformed variable. We standardized our investment variables into daily wage units and added a daily wage unit to all observations before log-transformation. We assessed the sensitivity of Cohen's d estimates to the log-transformation as well as the unit used for measuring and retaining censored observations, see Appendix C and Figure C1 there.

For the final regressions all variables were normalized by subtracting the sample mean and dividing by the standard deviation. This facilitated easy interpretation of all the regressions as all variables could be interpreted in standard deviation units. However, for the investment variables, due to their skewness and left-censoring, we had to log-transform them before we could normalize them, see Appendix C for the effect of log-transformation on the distribution of normalized total investment without and with log-transformation, see Figure C2), and the varying degree of censoring that affects also log-transformed and normalized variables (Figure C3).

3.3. Measurement errors, identification and estimation strategy

Measurement errors could cause both imprecision and bias in variables and perhaps especially in the latent risk preference variables. All experimental tools do not equally allow for the assessment of the degree of imprecision or bias. For the GP game we have observations at two points in time for each subject. The fact that the second round was hypothetical may also affect the reliability. The framing of the second game by investigating the cognitive memory of the first round game may have contributed to a more reliable estimate. We assess the mean and variance for the two rounds (Table 1) and the correlation between the two measures and how this correlation was related to the cognitive memory of the first round.

The CE-MCL experiment with 12 CLs for the estimation of 3 parameters gives enough degrees of freedom to estimate a contextualized Fechner error and how subject, CL, random starting point, and enumerator dummy variables contributed to the Fechner error[3]. The results allow us also to assess the degree of gender difference in the Fechner error (Table 1). Note that possible shock influences on the structural parameters were controlled for with three variables that captured a covariate climate shock in 2015-16 and idiosyncratic shocks over the last two years. The estimated CRRA-r and Prelec α and β parameters were also allowed to vary with the socioeconomic variables including the gender dummy, see Table 1 for the estimates. These estimated variables were also normalized before inclusion in the follow-up regression models to assess their correlations with the investment variables.

For the loss aversion rank variable we only have one switch point per subject and cannot do an independent assessment of the measurement error in this game. We randomized the starting point also in this CL and could assess whether the random starting point was correlated with the final choice. To assess the reliability of the loss aversion rank variable we assess its correlation with the investment level in the GP risky investment game as this game has been used to study and measure myopic loss aversion (Gneezy and Potters, 1997; Haigh and List, 2005). We should therefore expect a strong (negative) correlation between the loss aversion rank variable and the investment levels in the two rounds of the GP game. Finally, if the loss aversion experiment can be trusted to capture loss aversion, we may combine it with the estimated RDU parameters in the regressions with the investment variables in an expanded prospect theory assessment of investment behavior. We need to keep in mind, however, that measurement error may lead to attenuation bias for the loss aversion rank variable in the investment models.

Measurement errors may cause low correlations and attenuation bias when variables with measurement error are included as RHS variables (Gillen et al., 2019). As our identification strategy for the GP game we build on the instrumental variable approach (obviously related instrumental variables (ORIV)) proposed and demonstrated by Gillen et al. (2019). If each of the GP game rounds gives estimates of risk tolerance that are measured with error at the individual level, combining them can give estimates that

[3]The details of the estimated model are presented in Appendix B. Further details of the estimation are presented in another, yet unpublised paper by the authors.

are closer to their true latent values. We therefore used the 2017 hypothetical GP game result in combination with the cognitive memory of the 2016 real GP game one year later as instruments to predict the 2016 GP game investment level, with all variables in normalized form. The performance of this model depends on the strength of the instruments which depends on the correlation of the two GP investment measures and therefore the degree of measurement error. Low correlation implies high measurement error that implies weak instruments and low efficiency and higher attenuation bias in the coefficient on the predicted variable. Like for the parsimonious models we included two alternative specifications, one with group RE and one with group RE and community FE, for each investment variable. We use community (*tabia*) fixed effects (FE) to control for location-specific variation such as market distance, agro-ecological and other community characteristics that could be correlated with the investment levels as well as gender.

We cannot rule out some enumerator bias although careful joint training and supervision of the enumerators took place before and during the surveys and experiments. We control for such possible enumerator bias with enumerator FE. Enumerators were randomly allocated to group members and their potential influence is therefore orthogonal on groups. They may, however, contribute to measurement error in experimental variables such as the risky investment game responses.

We run the following parsimonious investment models with the un-transformed investment variables and the log-transformed and normalized investment variables (equation (1)):

$$I_{t,ij=k} = \eta_0 + \eta_1 G_F + \eta_2 E_d + (\eta_3 T_d) + g_g + \varepsilon_{gi} \tag{1}$$

where $I_{t,ij=k}$[4] represents investment type k, G_F represents the gender dummy, E_d represents enumerator fixed effects, T_d represents community FE, g_g represents business group random effects, and ε_{gi} is the error term.

To assess the impact of the GP game investment on the real investments we run the following IV model where the 2016 investment level ($r_{t-1,gi}$) in the game is instrumented for with the 2017 investment level ($r_{t,gi}$) in the game and the cognitive memory index ($cm_{t,gi}$) for the game played one year earlier.

$$\begin{aligned} I_{t,ij=k} &= \beta_0 + \beta_1 G_F + \beta_2 \hat{r}_{t-1,gi} + \beta_3 E_d + (\beta_4 T_d) + g_g + \epsilon_{gi} \\ r_{t-1,gi} &= \alpha_0 + \alpha_1 r_{t,gi} + \alpha_2 cm_{t,gi} + \omega_{gi} \end{aligned} \tag{2}$$

Next, we estimate Prospect Theory models with the three predicted RDU variables ($\hat{\Theta}_{t,gi}$) and the loss aversion rank variable ($\lambda_{t,gi}$). The three predicted RDU variables are the CRRA-r, Prelec α and Prelec β. The details of the estimated models and instruments are presented in Appendix B.

$$\begin{aligned} I_{t,ij=k} &= \gamma_0 + \gamma_1 G_F + \gamma_2 \hat{\Theta}_{t,gi} + \gamma_3 \lambda_{t,gi} + \gamma_4 E_d \\ &\quad + (\gamma_5 T_d) + g_g + \upsilon_{gi} \\ \Theta_{t,gi} &= RDU model^{5} \end{aligned} \tag{3}$$

Next we run models to investigate whether the baseline individual resource endowment variables ($z_{t-1,gi}$) can explain the gender differences in investment levels:

$$I_{t,ij=k} = \delta_0 + \delta_1 G_F + \delta_2 z_{t-1,gi} + \delta_3 E_d + g_g + \nu_{gi} \tag{4}$$

In the final regression we combine the PT variables and the baseline variables to assess their joint contribution to explain the gender differences in investment levels:

$$\begin{aligned} I_{t,ij=k} &= \tau_0 + \tau_1 G_F + \tau_2 \hat{\Theta}_{t,gi} + \tau_3 \lambda_{t,gi} + \tau_4 z_{t-1,gi} \\ &\quad + \tau_5 E_d + (\tau_6 T_d) + g_g + \varsigma_{gi} \\ \Theta_{t,gi} &= RDU model \end{aligned} \tag{5}$$

In the investment regressions with the RDU and loss aversion rank variable we can have less attenuation bias for the RDU parameters given that the structural model is well specified and the Fechner error has removed random error and reduced estimation bias. A critical assessment of the instrumentation is in order here. The variables included in the Fechner error specification were the random starting point in each CL, the random order of the CL, and the placement of the risk-neutral row in the CL. In addition, enumerator dummies and subject characteristics were allowed to influence the Fechner error. The CL-related variables can be claimed to have no influence on the investment variables of the subjects and can therefore serve as valid instruments for the identification of the RDU parameters that we use in the second stage investment regressions. The fact that two of these CL-related variables also were significant at the 1% level is an indication of the strength of these instruments although we are unable to perform standard F-tests of their strength by combining our first stage structural RDU model with the second stage investment models. The inclusion of covariate and idiosyncratic shock variables should also control for preference change between the two points in time for the surveys and experiments. The use of four investment types and total investments should contribute furthermore to the assessment of how gender differences in risk preferences are related to gender differences in investments.

All variables are normalized in the second stage investment models. To correct the standard errors of the estimated and predicted variables, the standard errors in the models with predicted variables are corrected using bootstrapping, re-sampling business groups. Several of the baseline variables such as the livestock and the income variables may be endogenous and closely related to investments in the following period. It is therefore possible that these endogenous baseline and investment variables jointly are influenced

[4] We denote 2016 as $t-1$ to emphasize the timing of the game and the variables used. The real investment variables were for the period 2016-2017 and are therefore flow variables denoted as t variables based on the time when they were collected. We suppress the timing subscript for the enumerator, business group variables, and the error term to keep the notation simple.

by the risk preference variables. However, we do not have instruments to deal with this and can only run models without and with the baseline variables when assessing the potential influence of the risk preference variables. Our final set of investment models allow us to assess whether there still exist substantial gender differences that are unexplained by the included variables. By jointly comparing all models we assess the extent to which investments are driven by risk preferences directly, indirectly through endowments, and the relative importance of the different PT variables, and the robustness of these effects across models.

4. Results

4.1. Descriptive analysis

The sample size is 822 subjects with 558 men (67.9%) and 264 women (32.1%). Our sample is therefore gender biased compared to the general population in the study areas. Table 1 provides an overview of the key investment variables by gender, the experimentally elicited and estimated latent risk attitude variables, as well as the observable and self-reported socioeconomic baseline variables. Four out of the five investment variables collected in 2017 demonstrate significant gender differences with women investing less than men. The women in the sample were on average 4.4 years younger, but were not significantly different in number of years of education and birth rank. However, women came from parents with significantly smaller farm sizes, indicating that their parents' poverty may be a driver of their business group participation as farming is the main source of livelihood in the area and very land-poor parents may be less able to take care of their daughters. The business women also had significantly less livestock, durable assets and income at the time of the baseline survey compared to the male counterparts in these groups. We therefore see a confounding of poverty and gender in the sample. We also assessed the within-gender combined inequality in total investments in our sample using gini-coefficients. We found the gini=0.622 for men, 0.621 for women, and 0.627 for the full sample. This signals substantial inequality in investments among men as well as among women.

The first two of the risk preference variables are the investment shares in the GP risky investment game that was played twice with the respondents in 2016 and 2017, with the first being real and the second being hypothetical. This game has in many contexts revealed significant gender differences with women being characterized as less risk tolerant. Table 1 demonstrates the same type of pattern in both game rounds and the average investment level is similar in the two rounds although slightly lower in the hypothetical game than in the real game played one year earlier. The next variable is the loss aversion rank variable for which the women are found to be significantly more loss averse than men.

Last, the table presents a number of estimated parameters based on the RDU structural model. Table 1 shows that women have on average significantly more concave utility (CRRA-r), have similar average Prelec α, and have significantly smaller Prelec β parameters than men. The latter indicator implies that women are more optimistic in form of a more elevated probability weighting function. This should stimulate investment but the effect is countered by their more concave utility function and stronger loss aversion as found with the loss aversion experiment. Finally, when inspecting whether there is a gender difference in the noise in the RDU model, we see from Table 1 that the noise was higher for men than for women and with a Cohen's d value of 0.22. Based on these findings we cannot rule out that the gender differences in risk preferences are an important driver of the gender differences in investment levels.

The Cohen's d (effect size) values are also included in Table 1. Based on their sizes for the investment variables the gender differences may be classified as low. However, the skewness of the investment variables may cause a downward bias in the Cohen's d values as outlier observations can inflate standard deviations. We illustrate this in Fig.1, where the Cohen's d values for the un-transformed investment variables are shown to the left and the log-transformed values are shown to the right. We found that Cohen's d values were sensitive to how this log-transformation was done and the value added to retain censored observations. We illustrate this for the total investment variable in Fig. C1 in Appendix C. We see from Fig.1 that the Cohen's d values increase from 0.22 for the un-transformed total investment variable to 0.38 after log-transformation into daily wage units. We see a similar tendency for the other investment variables. Furthermore, Fig. 2 shows the cumulative probability distributions of the log-transformed total investment variable by gender. This demonstrates stochastic dominance in investments by men but also that there is substantial overlap and that a fairly small share of the men and women in the sample had zero individual investment levels, despite their poverty.

Fig. 1 shows that, after log-transformation, for four out of five investment variables men invest significantly more than women on average but the Cohen's d "effect" sizes are small to medium also after log-transformation[6]. The Cohen's d effect sizes for the baseline socioeconomic variables (Table 1 and Fig. 3) indicate medium gender differences in age and income and moderate to low or no gender differences for the other variables. The gender differences in investment levels can therefore potentially be explained by the gender differences in the baseline variables. We assess this in the analysis.

For the risk attitude variables the gender differences in terms of the Cohen's d effect sizes are medium to low for four out of six variables. The real and hypothetical GP risk tolerance measures are similar with Cohen's d values around 0.3, demonstrating that men invest more in this game, similar to what has been found in earlier studies, see Charness and Gneezy (2012) for a review. Nelson (2016) reviewed the same studies and came to the opposite conclusion that there were not so strong evidences for gender differences in risk taking, giving more emphasis on individual variation rather than aggregate average differences. Most of the reviewed

[6]Cohen suggested that a d=0.2 is small, d=0.5 is medium, and d=0.8 is large in psychological studies (Cohen 1992).

Table 1
Gender differences (means) in investment, risk tolerance parameters and other individual characteristics

	Males	Females	Difference	S.E.	Cohen's d	Obs.
Investments 2016-2017, ETB						
Consumer goods ETB	950	892	57	(144)	0.028	822
Livestock ETB	3639	2219	1419***	(343)	0.293	822
Productive assets ETB	1213	856	357**	(139)	0.206	822
Other Business ETB	4795	2911	1884**	(899)	0.133	822
Total Investment ETB	10597	6880	3717***	(1054)	0.223	822
Risk attitude variables						
Riskshare 2016, real	0.466	0.388	0.078***	(0.017)	0.317	822
Riskshare 2017, hypothetical	0.470	0.373	0.097***	(0.020)	0.345	810
Loss aversion rank	5.149	5.576	-0.427***	(0.160)	-0.199	822
CRRA-r 2017, estimated	0.585	0.839	-0.254***	(0.018)	-1.061	822
Prelec α 2017, estimated	0.606	0.612	-0.007**	(0.003)	-0.145	822
Prelec β 2017, estimated	0.958	0.864	0.095***	(0.008)	0.883	822
Fechner error, residual	0.144	0.133	0.011***	(0.003)	0.255	822
Individual characteristics 2016						
Age	29.952	25.576	4.376***	(0.591)	0.500	822
Education, years	5.369	5.792	-0.422	(0.304)	-0.107	822
Birth rank	3.066	3.102	-0.036	(0.150)	-0.018	822
Married, dummy	0.608	0.587	0.020	(0.037)		822
Farm size of parents, *tsimdi*	2.442	1.964	0.477***	(0.151)	0.221	822
Livestock (Tropical livestock units)	1.372	0.862	0.510***	(0.106)	0.309	822
Durable assets, number	1.392	1.061	0.332***	(0.099)	0.229	822
Log(Income), 1000 ETB	1.789	1.315	0.474***	(0.080)	0.442	822

The CRRA-r, Prelec parameters and Fechner errors are jointly estimated, see Appendix A2 for details.
dummy and other covariates. T-tests: *** p<0.01, ** p<0.05, * p<0.1.

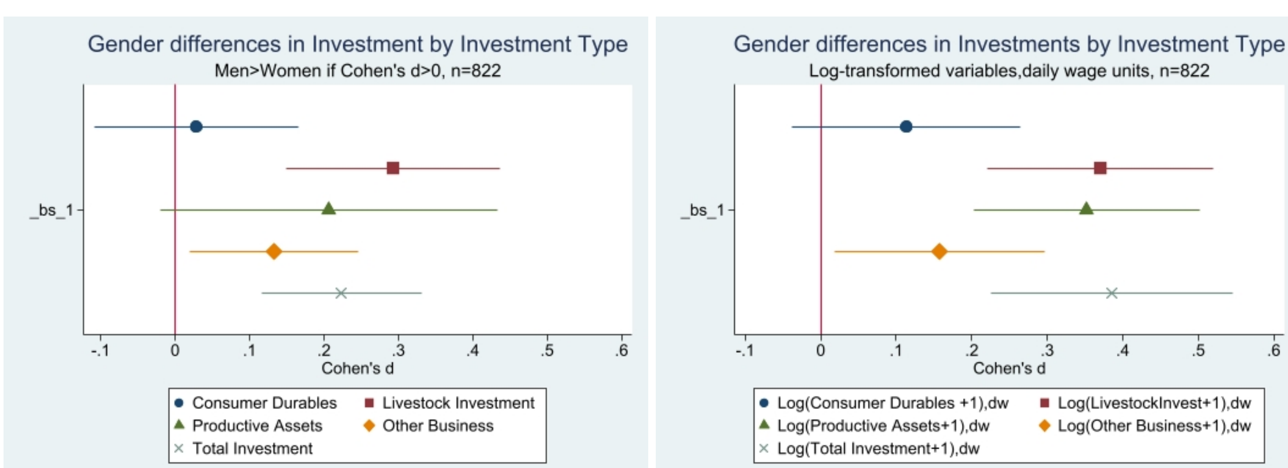

Figure 1: Gender differences in Investments: Skewness, log-transformation and Cohen's ds

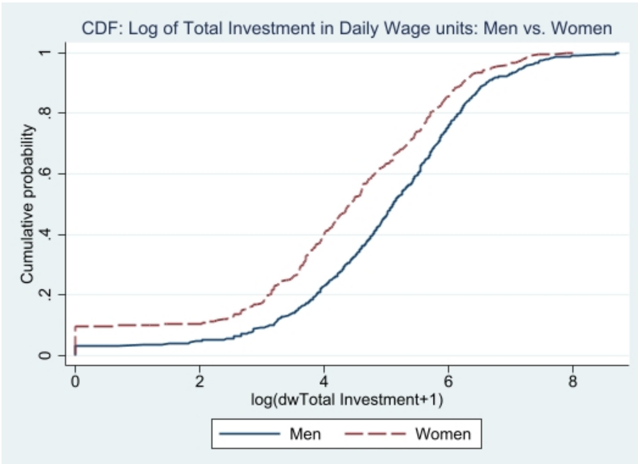

Figure 2: CDF of Gender differences in log-transformed Total Investment

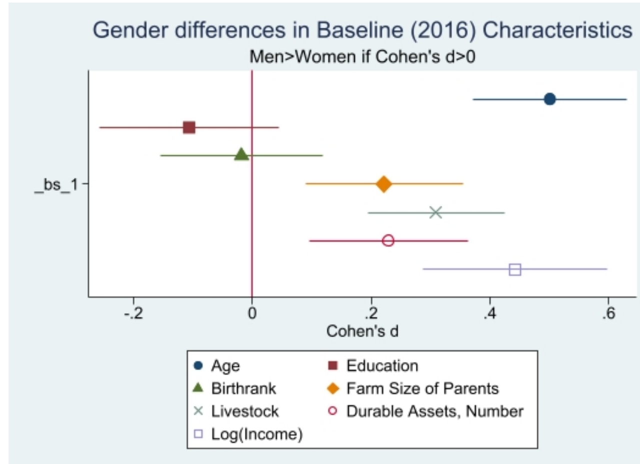

Figure 3: Gender differences in Baseline variables

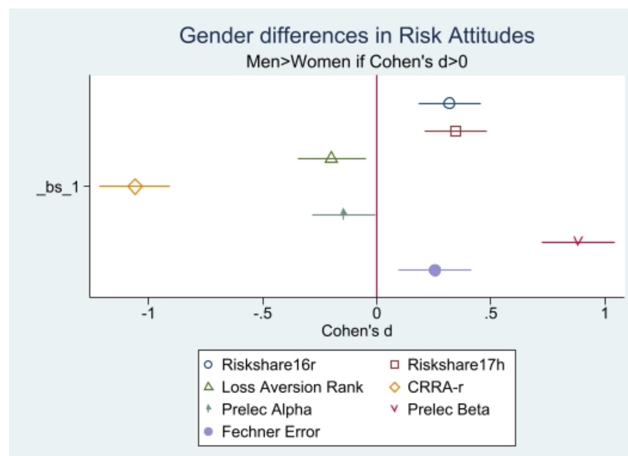

Figure 4: Gender differences in risk attitude variables

studies had small samples and did not give significant gender differences measured in form of Cohen's d effect sizes. However, also Nelson (2016) found a significant gender difference in the pooled sample of studies with a mean Cohen's d of 0.49, with a confidence interval of [0.40,0.58]. Our study estimate of about 0.3 is outside and below this confidence interval for these previous pooled studies. This may be because the business groups we study attract women that are more willing to take risk as we see a smaller and more select sample than for men that traditionally are more responsible for household investment decisions than women as household heads in this patriarchal society.

Let us then look at the other risk tolerance measures. The loss aversion rank variable indicates that women are significantly more loss averse than men, with a Cohen's d effect size of -0.20. This indicates an even smaller gender difference than for the risky investment game. This is consistent with the finding of Crosetto and Filippin (2016) that the risky investment game gives larger gender differences than games where two risky choices are compared like in the Holt and Laury (2002) and in our loss aversion experiment.

Table 1 and Fig. 4 demonstrate significant and larger gender differences for two of the three estimated RDU variables. With an average CRRA-r of 0.82 for women and 0.57 for men we get a Cohen's d effect size of -1.02 which may be considered large and points towards women being substantially more risk averse than men. When it comes to the Prelec α estimates, they were on the other hand very similar for men (0.60) and women (0.61) and with a slightly negative Cohen's d of -0.16. For the Prelec β, the mean estimate is 0.97 for men and 0.87 for women and the Cohen's d effect size=0.85. The larger Cohen's d values could be due to the removal of noise from these estimates and that may have reduced the standard deviations in the estimated variables.

4.2. Investment models

4.2.1. Parsimonious models

Table 2 presents parsimonious models with the five untransformed investment variables with the gender dummy, enumerator FE, group RE and without and with community FE. These models give a feel for the real money investment levels, especially the model without community FE. Table 2 shows that the average total investment level is about 10,100 ETB for men and about 5,600 ETB for women. The constant terms for the models with community FE are much larger and indicate that the baseline community on average has more than three times as high average investment level for men, indicating substantial local heterogeneity.

We are primarily interested in the gender differences and use log-transformed and normalized investment variables to assess the gender difference in parsimonious models that are presented in Table 3. Table 3 shows that females invest on average about 0.4 standard deviations less than men. The coefficients on the gender variable in Table 3 are directly comparable to the Cohen's d effect sizes as the normalized variables are also measured in standard deviation units

Table 2
Parsimonious panel investment models with un-transformed investment and gender dummy variables

VARIABLES	(1) Consumer Durables (CInv)	(2) Livestock (AInv)	(3) Productive Assets (PInv)	(4) Other Business (OBInv)	(5) Total Investment (TInv)
Group RE					
Sex (Female=1 dummy)	-150.140	-1,616.877***	-345.270**	-2,273.395***	-4,514.374***
	(153.698)	(403.394)	(152.778)	(790.958)	(1,049.678)
Constant	1,050.415***	3,469.514***	1,000.854***	4,636.402***	10,108.219***
	(219.411)	(799.259)	(125.595)	(1,653.325)	(2,162.321)
Group RE and community FE					
Sex (Female=1 dummy)	-78.853	-1,765.437***	-259.851	-2,605.372**	-4,925.296***
	(160.255)	(397.803)	(181.575)	(1,038.869)	(1,322.651)
Constant	2,590.994*	3,646.206***	1,627.249*	25,952.672	33,674.233*
	(1,388.102)	(1,224.222)	(913.030)	(18,411.813)	(18,146.828)
Observations	822	822	822	822	822
Number of groups	111	111	111	111	111

Un-transformed investment variables in ETB. Alternative group RE without and with community FE models, all with enumerator FE. Cluster-robust standard errors in parentheses, clustering on business groups, *** p<0.01, ** p<0.05, * p<0.1.

Table 3
Gender difference with log-transformed and normalized (z-score) investment variables

Z-SCORE VARIABLES	(1) z_logdwCInv Consumer Durables	(2) z_logdwAInv Livestock	(3) z_logdwPInv Productive Assets	(4) z_logdwOBInv Other Business	(5) z_logdwTInv Total Investment
Group RE					
Sex (Female=1 dummy)	-0.128	-0.373***	-0.292***	-0.168**	-0.409***
	(0.085)	(0.084)	(0.069)	(0.074)	(0.089)
Constant	0.204*	-0.143	-0.159	-0.084	-0.229
	(0.118)	(0.127)	(0.124)	(0.123)	(0.145)
Group RE and community FE					
Sex (Female=1 dummy)	-0.097	-0.372***	-0.223***	-0.166**	-0.379***
	(0.092)	(0.078)	(0.068)	(0.080)	(0.089)
Constant	0.496	-0.376*	-0.470	0.378	0.466*
	(0.393)	(0.219)	(0.345)	(0.628)	(0.252)
Observations	822	822	822	822	822
Number of groups	111	111	111	111	111

Z-score of log-transformed investment variables from daily wage units. Alternative group RE without and with community FE models, all with enumerator FE. z_logdwCInv=Z-score(log((Consumer Durables Investment 2016-17/Daily wage rate)+1), etc. Cluster-robust standard errors in parentheses, clustering on business groups, *** p<0.01, ** p<0.05, * p<0.1.

after having been log-transformed to reduce skewness of the variables. These models show higher stability when we compare the gender difference for models without and with community FE than what we found for the un-transformed investment variables in Table 2. It is noteworthy that the gender difference is larger for the total investments than for each of sub-categories of investments. All the differences are medium to small according to Cohen's d effect size classification.

4.2.2. IV-models: Can risk preferences explain gender differences in investments?

An instrumental variable approach can control for endogeneity in RHS variables as well as attenuation bias due to measurement errors in RHS variables given that sufficiently strong and valid instruments can be identified. Potential measurement errors in the GP risky investment game have largely been overlooked in the previous literature, with the notable exception of Gillen et al. (2019). The fact that we repeated the risky investment game for the same subjects in 2016 and 2017 allows us to assess the degree of measurement error in this game. The within subject correlation coefficient between the two rounds is as low as

0.135. This makes us ask whether this low correlation and therefore high measurement error[7] can contribute to higher standard deviations and thereby negative bias in the Cohen's d effect sizes? Certainly it implies that the individual level precision of the estimate of the latent risk tolerance based on the game is very low. This can also lead to attenuation bias when the variable is included as a RHS variable in regressions such as in our planned investment models (Gillen et al., 2019). We build on the approach of Gillen et al. and use obviously related instrumental variables (ORIV) to instrument for the pre-investment GP risk tolerance. We combine the investment in the game one year later with the individual cognitive memory index for the game that took place one year earlier. Like for the previous models, we ran these ORIV models without and with community FE and group RE. However, the results were discouraging and point towards too high measurement errors for the ORIV approach to work well. Rather these models give results that are no better than models without IV. We have included the results in Appendix D for interested readers for inspection.

The IV PT investment panel data models build on the first stage RDU models and the three predicted CRRA-r, Prelec α and β variables and the experimental instruments that we rely on for their validity and identification. The CL characteristics have no direct effect on the investment decisions in the previous year so the theoretical validity cannot be questioned. Our main hypothesis is that the inclusion of the PT risk attitude variables reduces the coefficients on the gender dummy variable in each investment model (hypothesis H1). We included the loss aversion rank variable to expand from an RDU model into a PT model. We have not been able to correct for measurement error in this variable so there is a risk that it suffers from attenuation bias. We assess whether it can provide additional explanatory power as loss aversion may potentially have a negative effect on real world investments (hypothesis H2). We also hypothesize that a higher CRRA-r is associated with lower investment (hypothesis H3). Furthermore, we hypothesize (H4) that a more elevated probability weighting function is associated with higher investment (more optimistic behavior), implying a negative effect of Prelec β on investment. A more inverted S-shape of the probability weighting function makes subjects less risk averse for low probability gains and more risk averse for high probability gains and its effect on investment is therefore ambiguous and depends on the subjective probabilities of gains and losses associated with the actual investments made by the subjects. We leave it to the data to speak on this. The results are presented in Table 4 without and with community FE to allow for a robustness check of these hypotheses tests for each type of investment.

We first inspect the models without community FE. The gender dummy remains significant in four of five models but the sizes of the coefficients have declined in four of the models lending some support to our H1 hypothesis. Next, we see that the loss aversion variable has a negative sign in all five models and is significant at 5% level in two of the models. The coefficient is less than 0.1 standard deviation in absolute value but we need to keep in mind that it may have been negatively affected by attenuation bias due to measurement error. We cannot reject hypothesis H2 that loss aversion contributes to lower investment in total or in livestock.

Next, we assess the CRRA-r variable. We hypothesized that a higher CRRA-r is associated with lower investment. Table 4 shows that the coefficient is negative in all five models but significant in only two of the models (at 5 and 1% levels). A higher CRRA-r is associated with significantly lower total investments and investments in productive assets, lending support to our hypothesis H3.

Our last hypothesis was that subjects with a more elevated probability weighting function (lower Prelec β) invest more. Table 4 shows a negative sign in all five investment models for Prelec β and the coefficient is significant (at 5 and 10% levels) in three of the models. This is evidence in support of our hypothesis H4. Finally, we assess the results for the Prelec α variable. We see that the coefficient is negative in all the investment models but significant (at 5% level) only in one, the one for Productive Assets. The results points in direction of more investments by those that overweigh low probabilities more.

The second panel in Table 4 with community FE is a robustness check of the results above. The findings for hypothesis H1 are not robust based on the inspection of the gender coefficients in Table 3 versus Table 4 for the models with community FE. The absolute value of the coefficient even increases in three of the models after introducing the PT variables, pointing towards stronger gender differences after controlling for risk preferences. For the loss aversion rank variable the sign of the coefficients remain negative in four of five models but is significant in only one model, the model for total investments. Hypothesis H2 cannot therefore be rejected. For the CRRA-r variable the results are not robust and the signs have changed to become positive in four of the models and are even significant (at 5 and 10% levels) in two of the models, while the sign remains negative and highly significant (at 1% level) in the model for productive assets. Likewise, the robustness check for our hypothesis H4 reveal a change in the sign for this variable in four of five models and the previous result was only robust in the model with productive assets. For the Prelec α variable the coefficients remained negative in all models and became highly significant (at 1% level) in three of the models, pointing in direction of higher investments by those who overvalue low probabilities more.

We need to make one more cautionary comment to the last results. Covariate risk due to a drought shock in the 2015-16 year before our baseline survey was found to affect the risk preferences making subjects more risk tolerant[8].

[7] This is based on the the finding that the investment level in the game was not significantly affected by recent idiosyncratic and covariate shocks and we assuming that the true latent risk tolerance has not changed much for other reasons over the one year period.

[8] We have shown this in a separate yet unpublished paper that is available from the authors.

Table 4
Z-score PT panel data models with normalized log-transformed investment variables

Z-SCORE VARIABLES	(1) z_logdwCInv Consumer Durables	(2) z_logdwAInv Livestock	(3) z_logdwPInv Productive Assets	(4) z_logdwOBInv Other Business	(5) z_logdwTInv Total Investment
Group RE					
Sex (Female=1 dummy)	-0.182*	-0.280***	-0.113	-0.153*	-0.291***
	(0.100)	(0.089)	(0.082)	(0.089)	(0.098)
z_Loss aversion rank	-0.036	-0.075**	-0.019	-0.011	-0.094**
	(0.038)	(0.034)	(0.033)	(0.044)	(0.037)
z_CRRA-r, predicted	-0.208	-0.103	-0.458***	-0.064	-0.370**
	(0.155)	(0.195)	(0.170)	(0.171)	(0.170)
z_Prelec α, predicted	-0.063	-0.020	-0.099**	-0.054	-0.071
	(0.041)	(0.050)	(0.046)	(0.052)	(0.045)
z_Prelec β, predicted	-0.331**	-0.023	-0.302*	-0.067	-0.302*
	(0.158)	(0.181)	(0.164)	(0.169)	(0.159)
Constant	0.212*	-0.141	-0.202*	-0.086	-0.239
	(0.111)	(0.132)	(0.119)	(0.120)	(0.145)
Group RE and community FE					
Sex (Female=1 dummy)	-0.329***	-0.379***	0.013	-0.297***	-0.441***
	(0.101)	(0.098)	(0.084)	(0.103)	(0.104)
z_Loss aversion rank	-0.028	-0.058	-0.021	0.002	-0.074**
	(0.040)	(0.041)	(0.032)	(0.038)	(0.036)
z_CRRA-r, predicted	0.603*	0.258	-0.804***	0.673**	0.484
	(0.333)	(0.311)	(0.267)	(0.327)	(0.316)
z_Prelec α, predicted	-0.192***	-0.063	-0.032	-0.217***	-0.214***
	(0.065)	(0.066)	(0.053)	(0.071)	(0.059)
z_Prelec β, predicted	0.324	0.254	-0.600**	0.540*	0.397
	(0.299)	(0.283)	(0.242)	(0.298)	(0.290)
Constant	0.687*	-0.321	-0.624*	0.582	0.615*
	(0.379)	(0.364)	(0.376)	(0.402)	(0.331)
Observations	822	822	822	822	822
Number of groups	111	111	111	111	111

All models with group RE and enumerator FE. Bootstrapped standard errors in parentheses, 500 replications, re-sampling business groups, *** p<0.01, ** p<0.05, * p<0.1.

As the covariate shock typically affected all subjects within a community, they are correlated with the community FE in the models used in the robustness analysis and may contribute to spurious correlations between the risk preference and community FE. We suspect this contributes to the strange changes in the results. We therefore are tempted to trust the results from the models without community FE more than the results from the models with community FE. We leave it to the readers also to assess this critically.

4.3. Are gender differences in investments rather explained by poverty?

Table 1 revealed that the business women came from more land-poor parents, and they had less livestock, durable assets and lower income than the business men. These differences could therefore also potentially explain the gender differences in investments. This is the fifth hypothesis (H5) that we want to test. We do this in Table 5 for the five investment variables with all variables, except the two dummy variables for gender and being married, normalized for easy interpretation. These are models without community FE. For the testing of the hypothesis we do like for the risk preference variables and assess how the inclusion of these baseline variables affects the sizes of the coefficients on the gender dummy varible in Table 5 compared to Table 3. In addition we assess the sign and significance for each of the endowment variables that are indicators of relative poverty differences.

Comparing the gender coefficients in Tables 3 and 5 we see that the inclusion of the baseline variables resulted in a reduction of the absolute values of parameters on the gender dummy variable in all models. This is an indication that gender differences in baseline variables contribute to explain the gender differences in the investment variables. However, the gender differences remain significant (at 1% level) in three of the models (livestock, productive assets and total investment). This indicates that relative resource poverty can explain only a part of the gender difference in investments. Table 5 furthermore shows that the farm size of parents (significant in four of fime models), the subjects' own endowments of durable assets (significant in four of five

Table 5
Z-score Investment Models with baseline socioeconomic variables

Z-SCORE VARIABLES	(1) z_logdwCInv Consumer Durables	(2) z_logdwAInv Livestock	(3) z_logdwPInv Productive Assets	(4) z_logdwOBInv Other Business	(5) z_logdwTInv Total Investment
Sex (Female dummy)	-0.057	-0.282***	-0.212***	-0.132	-0.295***
	(0.087)	(0.091)	(0.074)	(0.081)	(0.094)
z_Age, years	-0.073	-0.083	0.079*	-0.028	-0.018
	(0.049)	(0.053)	(0.042)	(0.065)	(0.051)
z_Education, years	0.175***	0.001	0.018	0.053	0.052
	(0.041)	(0.043)	(0.039)	(0.046)	(0.041)
Married, dummy	-0.127	0.204*	0.170*	0.204**	0.130
	(0.091)	(0.107)	(0.095)	(0.102)	(0.108)
z_Birth rank	-0.016	0.051	-0.013	-0.021	-0.001
	(0.035)	(0.035)	(0.030)	(0.035)	(0.033)
z_Farm size of parents	0.035	0.083**	0.106***	0.085**	0.099***
	(0.036)	(0.041)	(0.034)	(0.037)	(0.032)
z_Livestock endowment	-0.012	0.002	0.018	-0.003	0.012
	(0.043)	(0.054)	(0.051)	(0.051)	(0.049)
z_Durable assets, number	0.116**	0.135***	0.187***	0.072	0.167***
	(0.047)	(0.051)	(0.053)	(0.054)	(0.049)
z_Log(Income)	0.195***	0.164***	0.029	0.039	0.148***
	(0.051)	(0.043)	(0.039)	(0.050)	(0.044)
Constant	0.202	-0.315**	-0.253**	-0.230*	-0.355**
	(0.130)	(0.136)	(0.129)	(0.124)	(0.146)
Observations	822	822	822	822	822
Number of groups	111	111	111	111	111

All models with group RE and enumerator FE. Cluster-robust standard errors in parentheses, clustering on business groups, *** p<0.01, ** p<0.05, * p<0.1.

models) and incomes (significant in three of five models) stimulated investments in the following year. This is strong evidence that variation in the baseline resource endowments partly explain the gender differences in investments. The unexplained gender difference in total investments is reduced from 0.4 standard deviation to 0.3 standard deviation after the introduction of the baseline variables, indicating that most of the gender gap in investments is still unexplained.

Finally, we assess the effect of combining the PT risk attitude variables and the baseline socioeconomic variables to assess the extent to which they jointly may explain the gender gap in investment levels in Table 6. This further contributes to the robustness analysis for all the hypotheses. If both risk attitudes and resource endowment variables contribute to explain the gender gap, the gender gap should be further reduced in the models that combine the risk attitude and resource endowment (baseline) variables.

Comparing the gender differences in the models in Table 6 with those in Tables 3, 4 and 5 for the models without community FE, we see a further reduction in the absolute values of the coefficients. However, the reductions are not incremental, showing that the effects of the risk attitude variables and the baseline resource endowment variables are not additive but confounded. The gender difference parameters in Table 6 remain negative and are significant (at 5 and 1% levels) in three of the five models. The remaining difference for the total investments is still 0.28 of a standard deviation and is larger than the part that is explained by the included risk attitude and resource endowment/baseline variables. Our study area is a patriarchal society with clear gender division of labor and with possible gender discrimination in market participation and business. Our best guess of additional gender difference explanations are therefore such additional cultural and discriminatory explanations. Our study and findings will be discussed in relation to the wider literature in the next section.

5. Discussion and Conclusions

The overall theme of this study is to assess the extent of gender differences in risk attitudes and investments among young business men and women is our study area in northern Ethiopia. Our study relates to the general literature on gender differences in risk taking in different contexts. It contributes to the limited number of comprehensive studies that assess gender differences in investment behavior and risk attitudes in a developing country context outside the usual university context where most such studies have been implemented before. The fact that our study is in a patriarchal society with traditional strong gender differences in division of labor and decision-power with men typically serving as heads of households and with women being more responsible for household shores should point towards stronger gender

Table 6
Caption

Z-SCORE VARIABLES	(1) z_logdwCInv Consumer Durables	(2) z_logdwAInv Livestock	(3) z_logdwPInv Productive Assets	(4) z_logdwOBInv Other Business	(5) z_logdwTInv Total Investment
Sex (Female dummy)	-0.069	-0.257***	-0.213**	-0.136	-0.281**
	(0.112)	(0.099)	(0.100)	(0.102)	(0.117)
z_Loss aversion rank	-0.031	-0.067**	-0.013	-0.009	-0.088**
	(0.038)	(0.033)	(0.032)	(0.044)	(0.037)
z_CRRA-r, predicted	-1.122**	0.787	-0.699	0.247	-0.158
	(0.536)	(0.502)	(0.487)	(0.510)	(0.570)
z_Prelec α, predicted	-0.478*	0.308	-0.381	0.045	-0.084
	(0.262)	(0.270)	(0.237)	(0.254)	(0.284)
z_Prelec β, predicted	-1.107**	0.785	-0.643	0.250	-0.147
	(0.555)	(0.511)	(0.504)	(0.521)	(0.585)
z_Age, years	-0.674**	0.311	-0.400	0.041	-0.121
	(0.313)	(0.313)	(0.285)	(0.320)	(0.340)
z_Education, years	-0.112	0.196	-0.172	0.095	0.011
	(0.178)	(0.165)	(0.159)	(0.167)	(0.188)
Married, dummy	-0.124	0.211*	0.178*	0.208**	0.140
	(0.094)	(0.110)	(0.091)	(0.102)	(0.106)
z_Birth rank	-0.014	0.048	-0.015	-0.023	-0.002
	(0.036)	(0.033)	(0.028)	(0.035)	(0.032)
z_Farm size of parents	0.044	0.079**	0.120***	0.088**	0.100***
	(0.039)	(0.039)	(0.036)	(0.036)	(0.031)
z_Livestock endowment	-0.012	0.002	0.017	-0.002	0.011
	(0.047)	(0.056)	(0.060)	(0.054)	(0.050)
z_Durable assets, number	0.119**	0.131**	0.195***	0.072	0.164***
	(0.047)	(0.055)	(0.055)	(0.055)	(0.050)
z_Log(Income)	0.195***	0.160***	0.030	0.039	0.144***
	(0.053)	(0.042)	(0.038)	(0.053)	(0.044)
Constant	0.182	-0.285**	-0.272**	-0.221*	-0.343**
	(0.125)	(0.139)	(0.128)	(0.119)	(0.146)
Observations	822	822	822	822	822
Number of groups	111	111	111	111	111

All models with group RE and enumerator FE. Bootstrapped standard errors in parentheses, 500 replications, re-sampling business groups, *** p<0.01, ** p<0.05, * p<0.1.

differences than may be found in typical Western cultures where the gender roles are less specialized. However, a number of reforms including family law and land tenure reforms have also strengthened women's rights in Ethiopia and this may have affected their positions also in business (Holden and Tilahun, 2020). However, there have been few studies investigating this. One exception is Holden and Tilahun (2021b) who studied the role of mobile phones and gender differences in leadership and board membership in business groups in our study area. The study showed that women group members were much less likely to become group leaders and board members than men. Mobile phone ownership to a large extent increased the likelihood that male members became group leaders and board members but the same was not the case for women, who were much less likely to own a mobile phone and indicating that gender discrimination is still an important issue in business in the study area.

With the above finding in mind we were surprised to find only medium to low gender differences in individual investment levels measured in form of Cohen's d effect sizes, or after imposing other controls in regression models after log-transformation and normalization of the variables. When it came to risk taking in the GP investment game we found a significantly lower gender difference than in the pooled sample analyzed by Charness and Gneezy (2012) and Nelson (2016) which may be considered surprising given our cultural context. This may be because the business groups we study attract women that are more willing to take risk as we see a smaller and more select sample than for men that as household heads traditionally are more responsible for household investment decisions than women in this patriarchal society.

By repeating the game one year later for the same subjects we found evidence of large measurement error in this game, an issue that has been overlooked in most earlier studies. It is still possible that the average investment level in the game is unbiased but we cannot rule out that even a random error in individual investment contributes to inflating the standard deviation in our sample and thereby causing a downward bias in the Cohen's d effect sizes for gender differences. In our study we find that the two rounds of the GP game give a poor proxy for individual risk tolerance due to their low correlation. We caution that other studies need also to critically assess the measurement error issue when applying this game to measure individual risk tolerance. The use of the IV approach of Gillen et al. (2019) did not remedy the problem in our investment models as the measurement error was too large.

The within-gender inequalities in investment levels were high for both genders with intra-gender and overall gini-coefficients above 0.6. The skewness and censoring of the investment variables gave additional challenges in making gender comparisons that took the intra-gender inequality into account. Log-transformation and the treatment of censored observations affect the Cohen's d and standard deviation effect sizes. Our variable transformation approach could only partly remove the skewness problem in the investment variables but we show in Appendix C, Fig. C1, that it makes a big difference to log-transform the total investment variable before normalization. However, the larger the share of censored observations, the less is log-transformation able to create an unskewed distribution. This is demonstrated with the kdensity distributions of the log-transformed and normalized investment variables in Appendix C, Fig. C3. The extent of the censoring problem is lowest for the total investment and productive asset investments. These were also the investment categories for which the investment models seemed to perform better in terms consistency od the results across models.

Our study also speaks to the literature that is comparing different experimental tools for the elicitation of risk preferences in field settings. While the GP game has been recommended for its simplicity, the ease of comprehension and implementation, we think more studies are needed to assess the extent of within-subject randomness and thereby measurement error if the game if it is to be utilized to obtain a measure of individual risk tolerance. While the game may say something about general risk tolerance in a population group, our findings indicate that individual responses in the game are poor predictors of the same individuals' latent risk tolerance. While the GP game has been found to be less cognitively demanding to respond to than the more complex Holt and Laury (2002) game, especially for subjects with limited numeracy skills (Charness and Viceisza, 2016), we think the CE-MCL experiment we have used generates both more reliable and more comprehensive estimates in form of dis-aggregated risk attitude variables and that have stronger predictive power. The advantage of this approach is that there is only one risky prospect with fixed outcomes and probabilities to relate to varying certain amounts. This is much less cognitively demanding than comparing two risky prospects such as the Holt and Laury MCL approach, where the probabilities are changing for every row in the CL.

Testing four tools, including the GP game and the HL-MCL approach, Crosetto and Filippin (2016) found evidence of less measurement error, proxied by a Fechner error specification in the estimation, in the GP game than in the HL-MCL method which generated more inconsistent choices in a sample of undergraduate students in Germany. However, they also found that the tools that included a safe and a risky option, where the GP game is one of them, exhibited stronger gender differences than tools that used only risky alternatives, such as the HL-MCL method. Filippin and Crosetto (2016) used 54 replications of the HL-MCL method and found that gender differences appeared in less than 10% of the studies. They suggested that the availability of a safe option enhances the salience of potential loss in the risky prospect and this may cause a "certainty bias" that may cause violation of Expected Utility Theory (EUT) (Loomes and Sugden, 1982).

However, it is not entirely clear that the availability of a safe option is the main reason for potential bias as shown by Vieider (2018). Reference dependence and the salience of safe versus risky prospects based on Prospect Theory (PT) may influence behavior. The standard GP game makes the safe amount more salient through the initial endowment allocation. This may cause subjects to appear less risk tolerant. Holden and Tilahun (2021a) show that when subjects alternatively are allocated an initial risky prospect that they are allowed to trade with a safe option, they take substantially higher risk in the game that is otherwise identical to the standard one-shot GP game.

In our study we apply three different risk eliciting tools, of which two of them, the GP game and the CE-MCL experiments, involve a safe option, while the incentivized loss aversion CL includes two risky prospects. The fact that we found a lower Cohen's d effect size for the loss aversion experiment than in the two rounds of the GP game may still be due to a certainty effect for the GP game where the starting point is a certain amount (reference point). Both the GP game and the loss aversion experiment are supposed to be driven by loss aversion and give small and only slightly different gender differences.

When it comes to the relative salience of risky versus safe prospects in the different experiments, we may consider the safe amount in the GP game to be more salient and this should give a bias towards preferring the safe option or risking less. In the CE-MCL game it is the risky prospect that is constant in each CL and this risky prospect may therefore be considered more salient and this could pull in opposite direction of subjects being willing to take more risk in this game. Finally, the loss aversion CL contains two risky prospects and should therefore not lead to a bias towards safe options. Rather it may be the size of the loss that attracts attention and that drives behavior in the game, depending on how loss averse subjects are.

Our final major contribution in the paper is to assess whether the gender differences in investments can be explained by gender differences in risk attitude variables or gender differences in resource endowments or a combination of these. We used IV methods to predict risk tolerance based on the GP game and three RDU parameters based on the CE-MCL experiment. The ORIV models based on the GP game did not perform well due to large measurement errors and are only included in the Appendix for documentation. The models based on the RDU CE-MCL experiment and loss aversion experiment performed better. There were also some challenges with the identification of the risk preference models with community fixed effects as a covariate drought shock in 2015 affected community level risk preference parameters. These models are also included for completeness and documentation. But we think the models without community FE are more reliable. The findings with these models are in line with our hypotheses. The loss aversion, CRRA-r and Prelec β variables respectively appeared to have the hypothesized effects on investment levels in terms of signs although they were only significant in some of the models. There were also indications that a lower Prelec α was associated with higher investment levels, indicating that subjects that were more likely to overweigh low probabilities invested more. Overall, we also found that higher initial baseline resource endowments contributed to higher investment levels and that the inclusion of these variables as well partly reduced the gender differences. However, a 0.28 standard deviation in gender difference in total investments remains after controlling for both risk attitudes and initial resource endowments. We suspect that gender discrimination due to traditional norms is responsible for at least some of this remaining unexplained gender differences in investments.

A. Experimental designs

Table A1 (7) presents an overview of the characteristics of the 12 CLs included in the CE-MCL experiment. Table A2 (8) presents an example of one of the CLs in the CE-MCL experiment. Table A3 (9) presents the loss aversion CL.

B. Estimation of the RDU model based on the CE-MCL experiment

This is a summary of the estimation we build on and that is presented in more detail in a separate, yet unpublished, paper by the authors. Each row in each CL represents a choice between a risky prospect and a certain amount. The risky prospect gives a good outcome (x) with probability p and a bad outcome (y) with probability $1-p$. The choice between the risky and safe prospect is framed as a Rank Dependent Utility (RDU) model (Quiggin, 1982). The net utility return for any row can be specified as:

$$\Delta RDU = w(p)u(x) + [1-w(p)]u(y) - u(s) \tag{6}$$

where $w(p)$ is the probability weighting function. The ΔRDU switches from being negative to becoming positive at the switch point between the risky and certain amounts in the CL. The experiment captures only non-negative risky and safe amounts and the probability weighting function is therefore modeled in the gains domain only with a Prelec et al. (1998) 2-parameter weighting function:

$$w(p) = e^{-\beta(-\ln p)^{\alpha}}, \alpha > 0, \beta > 0 \tag{7}$$

where α captures the degree of (inverse) S-shape of the weighting function[9], and the β captures the elevation of the function, with $\beta < 1$ giving more elevated (optimistic) and $\beta > 1$ giving less elevated (pessimistic) weighting of prospects. The function is strictly increasing and continuous within the interval [0, 1] with $w(0) = 0$ and $w(1) = 1$.

Utility is captured with a Constant Relative Risk Aversion (CRRA) function:

$$u(x) = (1-r)^{-1}((b+x)^{1-r} - 1) \tag{8}$$

where r is the CRRA coefficient and b is the base consumption or asset integration level[10].

Noise or measurement error in the data are captured with a heteroscedastic Fechner (1860) type error (ξ) and the prospects are standardized with Wilcox (2008) with a CL-level contextual utility framing.

The probability of the respondent choosing the risky lottery is formulated with a probit function:

$$Pr(Risky) = \phi\left(\frac{\Delta RDU_{gimk}}{\xi_{gim}[u(x_m) - u(y_m)]}\right) \tag{9}$$

Subscripts i, m and k represent subjects, CLs, and row numbers in the CLs. The model flexibility allows respondent errors in the identification of switch points within CLs. The latent Fechner error (ξ_{gim}) can be assessed at the within-subject CL level as a measure of subject response inconsistency across CLs or to assess e.g. gender differences in such inconsistency that we are interested in.

The log-likelihood function for the risk experiment is obtained by summing the natural logs over the cumulative density functions resulting from equation (5) and summing them over CLs (subscript m) and subjects:

$$\begin{aligned} &\ln L(\Omega_{gi}(IS_{gi,t-n}, CS_{g,t-2}, z_i), \xi_{gim}(c_m, z_i, E_d)) = \\ &\sum_{imk} (\ln \Theta(\Delta RDU) \mid_{Choice_{imk}=1}) + \\ &(\ln \Theta(1-\Delta RDU) \mid_{Choice_{imk}=0}) \end{aligned} \tag{10}$$

Ω_{gi} is a vector of subject-specific risk preference parameters (r_i, α_i, β_i) that are modeled linearly on the lagged idiosyncratic and covariate shock variables (IS_{t-n}, CS_{t-2})

[9] $\alpha = 1$ implies $w(p) = p$, for $\alpha < 1$ the inverted S-shape becomes stronger as α declines

[10] We set the base consumption equal to 30 ETB which was equivalent to a daily wage. This is similar to what Andersen, Harrison, Lau and Rutström (2008) did in their field experiment in Denmark for the elicitation of risk preferences.

Table 7
CE-Multiple Choice List Treatment Overview

Choice List	Prob (good)	Bad outcome (ETB)	Good outcome (ETB)	CE-range min, max (ETB)
1	0.95	0	100	50,100
2	0.90	0	100	50,100
3	0.80	0	100	50,100
4	0.70	0	100	30,80
5	0.50	0	100	10,60
6	0.95	20	100	50,100
7	0.90	20	100	50,100
8	0.80	20	100	50,100
9	0.70	20	100	30,80
10	0.50	20	100	40,100
11	0.25	20	300	20,90
12	0.05	20	1500	20,90

Table 8
Example of CE-MCL Choice List

CL no.	Start point	Task no.	Prob. good outcome	Low outcome	High outcome	Choice	Certain amount	Choice
8		1	0.80	20	100		100	
8		2	0.80	20	100		95	
8		3	0.80	20	100		90	
8		4	0.80	20	100		85	
8		5	0.80	20	100		80	
8		6	0.80	20	100		75	
8		7	0.80	20	100		70	
8		8	0.80	20	100		65	
8		9	0.80	20	100		60	
8		10	0.80	20	100		50	

and the observable respondent variables (z_i) such as sex, age, and education.

$$\Omega_{gi} = \eta_0 + \eta_1 IS_{gi,t-n} + \eta_2 CS_{g,t-2} + \eta_3 z_{gi} + \epsilon_{gi} \quad (11)$$

The Fechner error (ξ_{im}) is modeled linearly on the CL characteristics (CL_m); including the random starting point in each CL, the random order of the CL, and the position

Table 9
Table A3. Choice List in Loss aversion experiment

CL no.	Start point	Task no.	Prob. Win	Prospect Win	A (ETB) Loss	Choice	Prospect Win	B (ETB) Loss
13		1	0.5	50	-10		60	-40
13		2	0.5	30	-10		60	-40
13		3	0.5	20	-10		60	-40
13		4	0.5	10	-10		60	-40
13		5	0.5	5	-10		60	-40
13		6	0.5	5	-10		60	-30
13		7	0.5	5	-15		60	-30
13		8	0.5	5	-15		60	-25
13		9	0.5	5	-15		60	-20

of the risk-neutral row in the CL; enumerator dummies (E_d) and subject characteristics (z_i):

$$\xi_{gim} = \rho_1 + \rho_2 CL_m + \rho_3 z_i + \rho_4 E_d + u_{gim} \tag{12}$$

The likelihood function was estimated with the Broyden-Fletcher-Goldfarb-Shanno optimization algorithm in Stata while clustering standard errors at the subject level to derive the predicted RDU variables for the second stage investment models and for direct inspection of gender differences.

C. Skewness and censoring of investment variables and distributional effects of log-transformation and normalization

Fig. 5 demonstrates the sensitiveness of the Cohen's d effect sizes to log-transformation and the unit used to handle variable censoring in the case of the total investment variable. Fig. 6 demonstrates the effect of normalization of the skewed and censored total investment variable without and with log-transformation. The figure shows that the combined log-transformation and normalization works quite well for this variable. Fig. 7 shows the kdensity distributions for all the five log-transformed and normalized investment variables. The figure shows that the transformations work less well for the investment variables where a larger share of the observations are left-censored.

D. ORIV Investment models with the predicted GP riskshare16r variable

Table 10 presents the instrumented models using the 2017 hypothetical GP game, the cognitive memory index for the 2016 game one year later and the random outcome of the 2016 real GP game as instruments. Models without and with community FE were tested. The results are presented for completeness and are not regarded as reliable due to the large measurement error that was detected with the low correlation between the two game responses per subject.

CRediT authorship contribution statement

Stein T. Holden: Conceptualization of this study, Methodology, Training of field staff, Data checking and organization, Analysis, Write-up. **Mesfin Tilahun:** Training of field staff, Field testing, Fieldwork organization and supervision, Data checking and cleaning, Commenting on drafts.

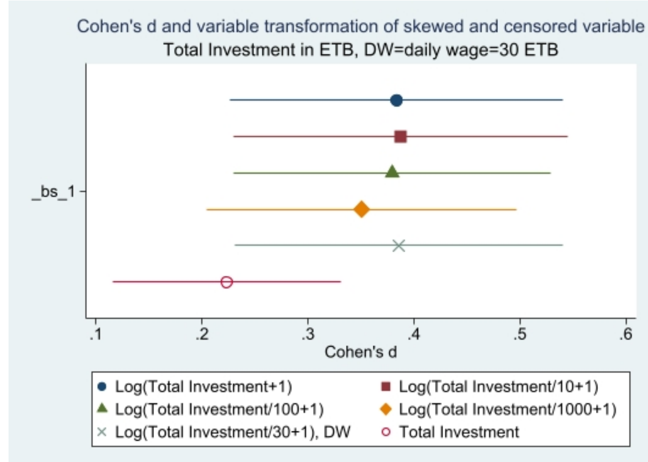

Figure 5: Figure C1. Cohen's d sensitivity to alternative transformations of skewed and censored investment variable

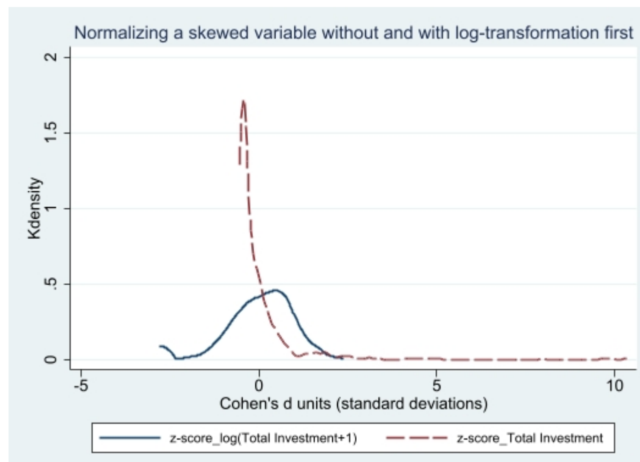

Figure 6: Figure C2. Normalization without and with log-transformation for the total investment variable

References

Andersen, S., Harrison, G.W., Lau, M.I., Rutström, E.E., 2008. Eliciting risk and time preferences. Econometrica 76, 583–618.

Andersson, O., Holm, H.J., Tyran, J.R., Wengström, E., 2016. Risk aversion relates to cognitive ability: Preferences or noise? Journal of the European Economic Association 14, 1129–1154.

Charness, G., Gneezy, U., 2012. Strong evidence for gender differences in risk taking. Journal of Economic Behavior & Organization 83, 50–58.

Charness, G., Viceisza, A., 2016. Three risk-elicitation methods in the field-evidence from rural senegal. Review of Behavioral Economics 3, 145–171.

Cohen, J., 1992. Statistical power analysis. Current directions in psychological science 1, 98–101.

Crosetto, P., Filippin, A., 2016. A theoretical and experimental appraisal of four risk elicitation methods. Experimental Economics 19, 613–641.

Croson, R., Gneezy, U., 2009. Gender differences in preferences. Journal of Economic literature 47, 448–74.

Dave, C., Eckel, C.C., Johnson, C.A., Rojas, C., 2010. Eliciting risk preferences: When is simple better? Journal of Risk and Uncertainty 41, 219–243.

Eckel, C.C., Grossman, P.J., 2002. Sex differences and statistical stereotyping in attitudes toward financial risk. Evolution and Human Behavior 23, 281–295.

Eckel, C.C., Grossman, P.J., 2008a. Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. Journal of Economic Behavior & Organization 68, 1–17.

Eckel, C.C., Grossman, P.J., 2008b. Men, women and risk aversion: Experimental evidence. Handbook of experimental economics results 1, 1061–1073.

Fechner, G.T., 1860. Elemente der psychophysik. volume 2. Breitkopf u. Härtel.

Table 10
Table A4. IV normalized investment models with predicted normalized riskshare16r

Z-SCORE VARIABLES	(1) z_logdwCInv Consumer Durables	(2) z_logdwAInv Livestock	(3) z_logdwPInv Productive Assets	(4) z_logdwOBInv Other Business	(5) z_logdwTInv Total Investment
Group RE					
Sex (Female dummy)	-0.017	-0.411***	-0.225**	-0.235**	-0.399***
	(0.116)	(0.109)	(0.105)	(0.120)	(0.114)
z_riskshare16r, predicted	0.329	-0.118	0.130	-0.143	0.065
	(0.222)	(0.248)	(0.227)	(0.258)	(0.235)
Constant	0.213*	-0.238*	-0.161	-0.126	-0.290**
	(0.122)	(0.135)	(0.135)	(0.115)	(0.146)
Group RE and community FE					
Sex (Female dummy)	-0.000	-0.423***	-0.126	-0.291**	-0.396***
	(0.121)	(0.113)	(0.108)	(0.138)	(0.122)
z_riskshare16r, predicted	0.310	-0.151	0.222	-0.331	0.045
	(0.248)	(0.251)	(0.222)	(0.309)	(0.264)
Constant	0.562	-0.609**	-0.396	0.185	0.299
	(0.387)	(0.252)	(0.307)	(0.570)	(0.219)
Observations	809	809	809	809	809
Number of business groups	110	110	110	110	110

Panel IV-2SLS models with z_riskshare17h, cognitive memory index for the 2016 GP game and outcome dummy for risky investment game 2016 as instruments. All models with group RE and enumerator FE. Cluster-robust standard errors in parentheses, clustering on business groups, *** p<0.01, ** p<0.05, * p<0.1.

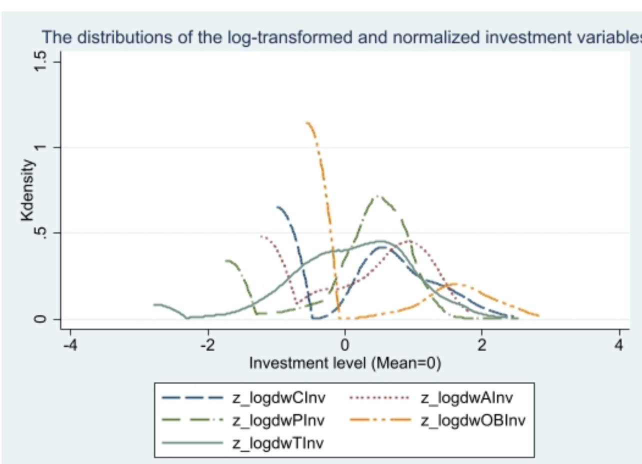

Figure 7: Figure C3. The kdensity distributions of the log-transformed and normalized investment variables

Filippin, A., Crosetto, P., 2016. A reconsideration of gender differences in risk attitudes. Management Science 62, 3138–3160.

Gillen, B., Snowberg, E., Yariv, L., 2019. Experimenting with measurement error: Techniques with applications to the caltech cohort study. Journal of Political Economy 127, 1826–1863.

Gneezy, U., Leonard, K.L., List, J.A., 2009. Gender differences in competition: Evidence from a matrilineal and a patriarchal society. Econometrica 77, 1637–1664.

Gneezy, U., Potters, J., 1997. An experiment on risk taking and evaluation periods. The Quarterly Journal of Economics 112, 631–645.

Gong, B., Yang, C.L., 2012. Gender differences in risk attitudes: Field experiments on the matrilineal mosuo and the patriarchal yi. Journal of Economic Behavior & Organization 83, 59–65.

Haigh, M.S., List, J.A., 2005. Do professional traders exhibit myopic loss aversion? an experimental analysis. The Journal of Finance 60, 523–534.

Holden, S.T., Tilahun, M., 2018. The importance of ostrom's design principles: Youth group performance in northern ethiopia. World Development 104, 10–30.

Holden, S.T., Tilahun, M., 2020. Farm size and gender distribution of land: Evidence from ethiopian land registry data. World Development 130, 104926.

Holden, S.T., Tilahun, M., 2021a. Endowment effects in the risky investment game? Theory and Decision , 1–16.

Holden, S.T., Tilahun, M., 2021b. Mobile phones, leadership and gender in rural business groups. World Development Perspectives 24, 100370.

Holt, C.A., Laury, S.K., 2002. Risk aversion and incentive effects. American Economic Review 92, 1644–1655.

Loomes, G., Sugden, R., 1982. Regret theory: An alternative theory of rational choice under uncertainty. The economic journal 92, 805–824.

Nelson, J.A., 2016. Not-so-strong evidence for gender differences in risk taking. Feminist Economics 22, 114–142.

Prelec, D., et al., 1998. The probability weighting function. Econometrica 66, 497–528.

Quiggin, J., 1982. A theory of anticipated utility. Journal of Economic Behavior & Organization 3, 323–343.

Tanaka, T., Camerer, C.F., Nguyen, Q., 2010. Risk and time preferences: Linking experimental and household survey data from vietnam. American Economic Review 100, 557–71.

Vieider, F.M., 2018. Violence and risk preference: experimental evidence from afghanistan: comment. American Economic Review 108, 2366–82.

Wilcox, N.T., 2008. Stochastic models for binary discrete choice under risk: A critical primer and econometric comparison. Risk Aversion in Experiments 12, 197–292.