The measurement of economic empowerment

An example from rural Indonesia

James C. Knowles[1]

August, 2021

[1] Economist, Chapel Hill, NC (jkinusa@aol.com)

Contents

1. Introduction

This paper focuses on issues related to the measurement of economic empowerment using a mix of analytical tools from economics, psychology and other disciplines. Specifically, the paper assesses the reliability and validity of multiple measures of economic empowerment at the household level, using a unique data set from Indonesia as a case study. The paper's ultimate purpose is to support more accurate estimates of gender differences in economic empowerment and of the effects of interventions designed to reduce them by demonstrating that giving sufficient attention to measurement issues can make a real difference. The paper is premised on the belief that insufficient attention has been given in the past to measurement issues in the collection and analysis of data in this area.

Although the paper is targeted to researchers, it uses relatively simple and widely used analytical methods (e.g., correlation and multiple linear regression analysis) so that the paper's content is accessible to a broad audience. Of course, this comes at a possible price in terms of technical soundness. Ideally, the topics discussed in this paper would be addressed through structural equations models that provide a unified framework in which to address both measurement and validation issues (Bollen 1989). However, these methods are not yet familiar to a wide audience.

The paper uses a unique household survey data set to illustrate how to identify and address measurement issues in data related to economic empowerment. These data were collected in connection with a large randomized social experiment targeting both women and men business owners in Indonesia that have previously been used in several studies (Buvinic, Johnson and others 2020, Deserrano and others 2021). Several features of this data set make it particularly suitable for this paper, including: (1) the sample is relatively large (N=4,828 established business owners, including 2,852 women business owners (WBOs) and 1,976 men business owners (MBOs), (2) the data were collected in three rounds, making it possible to compare the values of several measures at different points in time, (3) the data include a wide range of both objective and subjective measures of empowerment, including multiple alternative measures of several key outcomes (e.g., business profits, household income, personal savings); and (4) the data are of high quality.[2]

2. Conceptual framework

Economic empowerment is the capacity of women and men to participate in, contribute to and benefit from growth processes in ways that recognize the value of their contributions, respect their dignity and make it possible to negotiate a fairer distribution of the benefits of growth.[3] Economic empowerment involves both the ability to succeed and advance financially and the power to make and act on economic decisions.[4] In the case of women, economic empowerment has most often been described as a *process* involving resources, agency and achievements that cannot be adequately described by a set of final outcomes alone (Buvinic, O'Donnell and others 202):

- Resources include those of the individual woman, her household and her community;

[2] All of the data were collected by one of Indonesia's leading survey research organizations (i.e., SurveyMETER https://surveymeter.org/). The interviews were conducted using a computer-assisted personal interview (CAPI) system with pre-programmed consistency and outlier checks and with the interviewers entering information electronically (using a laptop) during the interviews.

[3] http://www.oecd.org/social/gender-development/womenseconomicempowerment.htm

[4] https://plan-international.org/eu/Youth-Economic-Empowerment-Main

- Achievements include both objective economic outcomes and subjective empowerment outcomes; and
- Women's agency plays a central role in their economic empowerment, providing the vital link between resources and achievements.

Box 1 lists the elements and dimensions of the conceptual framework that is used in this paper. It is based on a broader conceptual framework provided in Buvinic, O'Donnell and others (2020) but limited in this case to the economic empowerment outcomes analyzed in this paper. It includes all of the outcomes listed under Achievements and Agency/empowerment in the broader framework, with the exception of "Leadership roles". However, it includes less than half of the outcomes listed under Resources in the broader framework, with the omissions including the dimensions of "Health" and "Advocacy" and "Legal and Social Context." Some of the omissions reflect the limited geographical scope of the data set (i.e., rural and semi-urban villages in five regencies of northwest East Java province), while others are due to the data set's exclusive focus on established business owners.

Most of the causal linkages between the elements and dimensions in Box 1 run from Resources to Agency/empowerment and from Agency/empowerment to Achievements. However, as in Buvinic, O'Donnell and others (2020), there are also important causal linkages in the reverse direction. For example, several measures of Achievements (e.g., personal income, HH income, net assets, business capital) have a direct effect on some dimensions of Agency/empowerment (e.g., the ability to borrow and save, intra-HH and marital agency, personal work effort, mobilization of additional labor, and community participation). Also, as in Buvinic, O'Donnell and others (2020), the entire element of Agency/empowerment is included as a distinct dimension of Achievements.

Box 1. Elements and dimensions of the conceptual framework used in this paper

I. Achievements

- Personal income
- Personal assets
- Business assets
- Household income
- Quality of life (leisure time, overall well-being)
- Vulnerability to shocks (income instability)
- Agency/empowerment (as described below)

II. Agency/empowerment

- Agency (personal, intra-household and marital)
- Saving
- Borrowing
- Business investment
- Personal work effort
- Mobilization of additional labor
- Business practices
- Use of financial services
- Networking with peers
- Community participation

III. Resources

- Age/experience

- Education/skills (formal schooling, cognitive skills)
- Psychological characteristics (willingness to take risks, subjective time preference)
- Access to financial services
- Access to infrastructure (communications)
- Social capital (trust)

Source: Adapted from Buvinic, O'Donnell and others (2020)

(end of Box #1)

3. Data

The data used in this paper refer to the owners of 4,828 established non-farm business owners (2,852 WBOs and 1,976 MBOs) and were collected for an impact evaluation of a social experiment (Mobile Financial Services for Female Business Owners) that was designed to increase the utilization of recently established village agent-supported mobile phone banking services in 401 predominantly rural villages of five regencies (*kabupaten*) of East Java province, Indonesia.[5] The sample villages are rural or semi-rural villages selected by a partner bank as suitable sites for introducing branchless banking services. Village listings of business owners were prepared as the basis for randomly selecting 12 owners of established businesses in each sample village (i.e., 7 WBOs and 5 MBOs).[6] In addition to owning at least one established non-farm business, the participating business owners met the following criteria: (1) ages 18-55, (2) residents in the sample villages, and (3) having a mobile phone with an active account.

The trial interventions included business training provided to randomly selected WBOs in each village (MBOs did not receive any training).[7] The training was done in four one-day group sessions per village and focused on financial and business literacy (tracking income and expenses, setting priorities, the importance of saving, financial planning, basic bookkeeping, cash flow planning, record-keeping) and information on how to use the partner bank's mobile banking products offered by the partner bank. Apart from marketing (not included) and the instruction in the use of the branchless banking products (included), the topics covered are typical of those covered in business training trials (McKenzie and Woodruff 2014). The initial training was conducted in about 3 hours (but ranging from 1 to 4 hours) and was followed up by three group mentoring sessions of similar length that focused on addressing any questions from the trainees and on actual practices using their own individual businesses as cases.

The trial data were collected in three rounds spanning a two-year period, including a round of baseline data that were collected in two separate phases prior to the trial's interventions. Several of the questions asked in each round were identical across the three survey rounds. However, other questions were limited to one or two survey rounds. A detailed list of the questions and coded responses used to

[5] The sample regencies (formerly referred to as districts) include Bojonegoro (73 villages), Ngawi (101 villages), Tuban (72 villages), Lamongan (140 villages) and Gresik (15 villages) (Survey Meter, 2018).

[6] This was the target. However, the actual sample numbers differed in some villages that did not have a sufficient number of female and/or male business owners.

[7] In addition, supply-side trial interventions included the training of village-based banking agents in all villages, higher financial incentives for opening savings accounts to agents in randomly selected villages, and a village-level informational treatment that varied information about the level of incentives agents received in randomly selected villages. The village-level supply-side treatments are orthogonal to the demand-side training of women business owners.

obtain the data analyzed in this paper is provided in Appendix 1. The following paragraphs summarize the key features of the data collected in each survey round.

3.1 Baseline survey

The Baseline Survey (BS) of both WBOs and MBOs was conducted prior to conducting the interventions in 401 villages in two phases (November 2016-February 2017 in 107 villages and July-November 2017 in the remaining 294 villages). The BS collected data on the characteristics of all sample business owners (the "respondents"), including their age, gender, schooling (highest level completed), current marital status and relationship to the head of household. A brief test (4 questions) was also administered to all respondents to assess their cognitive skills. Household (HH) characteristics recorded in the BS include HH size, the number of the respondent's children currently residing in the HH and the HH's location. The BS also collected economic data in the following areas: (1) characteristics of the primary business and of any second business owned by the respondent (type of business, year started, whether registered with the government, number of years working in the business, number of days and hours worked by the owner, which HH members help with the business, number of paid and unpaid workers, number of customers and average monthly profits); (2) other economic activities (profits from up to two additional businesses owned and/or wages and salary earnings from up to two jobs and the number of days and hours worked in each additional business/job); (3) the current market value of all business capital by type; (4) the respondent's personal savings (amount saved in the last 12 months by saving instrument); (5) access to a bank account (whether owners have a bank account in their name, and if yes, when it was first opened, and if no, whether another HH member has a bank account), (6) use of banking services (which banking services are currently used, any existing bank loans, ever use of mobile money); (7) housing characteristics (ownership status, number of rooms, material of walls and roof, availability of electricity, water source); and (8) HH ownership of 20 durable goods.

Non-economic data collected in the BS include: (1) Decision-making in the HH (who decides: whether to buy an appliance, who and how HH members can work outside the home, whether to support family members, whether to save, or whether to sign up for a new banking product), (2) Marital agency (spouse's knowledge of the BO's earnings, share of BO's earnings used for HH expenses, share of HH expenses paid for by the spouse, who has access to the BO's earnings, who decides how the BO's earnings are spent, whether BOs have some money to spend on their own, who decides how the BO's spouse's earnings are spent); (3) Self-reported adherence to 15 recommended business practices; (1) use of mobile phones (use of a smart phone, purposes for which a smart phone is used, reliability of mobile phone signal coverage, internet access); (2) participation in voluntary activities (amount of time volunteered, charitable contributions); (3) relationships and contacts of both WBOs and MBOs with other sample WBOs (known/not known, type of relationship, frequency of contacts, whether business discussed during contacts), but not with other sample MBOs;

Subjective measures collected in the BS include: (1) Concern about the HH's access to sufficient food; (2) trust in banks, (3) willingness to take risks, and (4) subjective time preference.

3.2 Midline survey

A midline survey (MS) of 2,322 of the previously interviewed business owners in the BS was conducted in February 2018 in a non-random sub-sample of 200 villages in which the training and mentoring of agents and WBOs had been completed prior to the survey. Most of the data collected in the MS are the same as the corresponding data collected in the BS. However, there are some important differences. For example, the data on primary and second (if present) businesses include questions on average monthly business revenue and businesses expenses, making it possible to calculate an alternative estimates of

business profits for the same time period. However, the MS did not collect any data on other economic activities. The MS also collected data on current savings balances by saving instrument as well as recent savings (using the same question as in the BS, but using a reference period of the last 3 months instead of the last 12 months used in both the BS and the ES). The MS also collected detailed data on borrowing, including the amount and sources of any loans during the last 3 months and the total value of all outstanding loans by source. The MS also included questions on HH income, including total HH income during the last month and both low-end and high-end estimates of monthly HH income during the last 12 months. Although the MS did not collect any data on housing characteristics (as in the BS), it collected data not only on the simple ownership (yes-no) of 20 durable assets (as in the BS) but also on the number of each item owned.

The MS also collected similar non-economic data to that collected in the BS, including exactly the same questions on HH decision-making and marital agency. However, the MS did not collect any data on adherence to recommended business practices, participation in voluntary activities or on relationships and contacts with sample women business owners.

Although the MS did not collect any data on the same subjective outcomes for which data were collected in the BS, it collected additional subjective data on two measures of overall well-being (overall satisfaction with life and job/work satisfaction) and 15 measures of personal agency (e.g., self-confidence, assertiveness, determination/grit).

3.3 Endline survey

An endline survey (ES) of 4,644 (96.2%) of the previously interviewed business owners was conducted in all 401 sample villages in November-December 2018. The ES collected similar economic data as the BS, but with the following differences: (1) data were also collected on investment in equipment and buildings and product stock during the last 12 months; (2) data were collected on current savings balances by saving instrument (as in the MS); and (3) data were collected on the ownership of the same 20 durable assets as in the BS (but not on the number of items owned, as in the MS, or housing characteristics, as in the BS).

The ES collected similar non-economic data as the BS (including self-reported adherence to the same set of 15 recommended business practices), but with a few differences. Individual data on the age and gender of HH members age 15 and above, as well as on the number of mobile phones owned by each HH member, were collected in the ES, whereas the ES did not collect any data on participation in voluntary activities, on relationships and contacts with sample WBOs or on any subjective measures.

4. Methods

This paper assesses the reliability and validity of the economic empowerment measures in the data set described in section 3. Reliability refers to the consistency of a measure over time, while validity refers to whether the measure reflects the outcome it is intended to measure (Box 2).

Box 2. Reliability and validity in psychometrics

- Psychological researchers do not simply assume that their measures work. Instead, they conduct research to show that they work. If they cannot show that they work, they stop using them.
- There are two distinct criteria by which psychological researchers evaluate their measures: reliability and validity.

- Reliability is consistency in the measure's values across time (test-retest reliability), across items (internal consistency), and across researchers (interrater reliability). A measure's reliability is reduced by measurement error.
- Validity is the extent to which a measure actually represents the variable it is intended to measure. Validity is a judgment based on various types of evidence. The relevant evidence includes the measure's reliability, whether it covers all aspects of the construct of interest, and whether its values are correlated with variables they are expected to be correlated with and not correlated with variables that are conceptually distinct.
- The reliability and validity of a measure is not established by any single study but by the pattern of results across multiple studies, i.e., the assessment of reliability and validity is an ongoing process.

Source: Chapter 5 in Price and others (2015).

(end of Box #2)

According to classical test theory, reliability reflects the extent to which measures differ due to true differences in the variable of interest (e.g., intelligence) as distinct from measurement error. In this paper, reliability is assessed mainly by examining consistency across time (test-retest reliability) but also in a few cases more directly by comparing alternative measures of the same latent construct referring to the same period of time. When the same questions are administered to the same person under the same conditions (e.g., time of day, number of persons present) within a brief period (e.g., one day to one week), one might expect the responses to differ only due to measurement error. However, when the time periods between measurements are longer (e.g., months, as between survey rounds), test-retest measurements can be expected to vary substantially due to changes over time affecting the true values, even if the questions and respondents are the same. When outcomes are measured as part of an experiment, there is also the possibility that the experiment may have affected the responses, even when the treatments are randomly assigned.

Validity is assessed by determining whether a measure's scores are correlated with "criteria" variables (variables that one would expect them to be correlated with, based on theory and/or prior research) measured either at the same time (concurrent validity) or at a different time (predictive validity) or by comparing the values of two or more measures of the same construct (convergent validity). In this paper, criterion validity is assessed by estimating multiple regression models with the measure in question as the dependent variable and multiple "criteria" as explanatory variables. Whenever possible, the regression models are estimated with BS data to illustrate how useful information can be obtained by careful and thorough analysis of baseline data prior to finalizing survey instruments for follow-up data collection and prior to preparing a meaningful pre-analysis plan. However, the estimated relationships should not be interpreted as causal (i.e., x determines y, rather than y determines x or both x and y are determined by z). Instead, the estimated relationships are only statistical associations that indicate whether the dependent variable is measuring what it is intended to measure. However, the analysis in this paper also includes estimates of the experimental effects of the business training randomly assigned to the sample WBOs on several outcomes as additional evidence bearing on the reliability and/or validity of measured outcomes.

There are two main challenges in assessing the reliability and validity of the measures of economic empowerment in the data set analyzed in this paper. First, several of the economic variables of interest are inherently difficult to measure and are therefore likely to be subject to substantial measurement

error.[8] In addition, economic variables like business profits and revenue have highly skewed distributions, even when measured accurately, and/or have large concentrations of zero values (e.g., savings, investment). The key issue is whether and how these problematic features of key economic variables affect their reliability and validity and ultimately, the conclusions drawn on the basis of the analysis. This first main challenge is discussed more fully below (section 4.1), together with an explanation of how it is addressed in this paper.

The second main challenge is that several dimensions of economic empowerment are typically measured by asking multiple related questions, with the responses typically combined into a single summative measure of the targeted unobserved (latent) construct, such as adherence to recommended business practices or intra-HH agency. Such a measure is needed to reduce the number of relationships that need to be analyzed and interpreted and/or to obtain more reliable and stable measures that yield more meaningful results. The technical term for such summative measures is "composite variables," but they are also frequently referred to as "indexes." Two critical questions arise with respect to composite variables: First, do the multiple related items included in the composite variable all reflect the same unobserved (latent) construct such that a single number based on their individual values provides a reliable and valid measure of the construct? Second, what is the best "scale" to use to combine the individual items into a single composite variable that is as reliable and valid a measure of the construct as possible? Both of these questions are discussed below (section 4.2).

4.1 Problematic economic variables

The problematic distributions of several economic variables complicate the process of assessing their reliability and validity. When errors of measurement are large compared to the (unobserved) true value of a variable, the measured variable is unreliable by definition (Furr and Bacharach 2014). Errors of measurement can be random or non-random.[9] Large random errors of measurement can arise because a respondent's information is limited (for example, when asked how much was saved during the past 12 months). When large random measurement error is present in a variable, one can expect low correlation between successive measures of the same variable (low test-retest reliability), and particularly if the measurements are made several months apart. Although the concepts of reliability and validity are distinct (a measure can be highly reliable without being valid), the reverse is not true: unreliability reduces a measure's validity. However, it is still possible to establish that the measure is valid through multiple regression analysis with a sufficiently large sample. The simplest approach is to estimate multiple linear regression models with the measure in question as the dependent variable and several "criteria" variables as explanatory variables. Even when the true regression function is nonlinear, a linear function estimated by ordinary least-squares (OLS) may yield enough statistically significant estimated coefficients with the theoretically expected signs to provide evidence of criterion validity. In this case, the effect of random measurement error in the dependent variable is to reduce the overall explanatory power of the regression model without biasing the estimated relationship between the measure and the specified "criteria."[10] With a large enough sample, the estimated relationships can still

[8] Difficulties in the accurate measurement of many economic variables arise because the measures involve extended recall periods of up to 12 months (e.g., profits, savings) and/or their actual values vary substantially over time and because small rural businesses typically lack accurate and easily accessible records

[9] Large non-random measurement error can arise if the respondent does not want to reveal the true value (when responses are biased). Non-random measurement error may be more stable over time.

[10] However, if the values of the "criteria" variables also include large random measurement errors, the estimated relationships will generally be biased towards zero (Wooldridge 2010).

be statistically significant, even if the model's explanatory power is low (as indicated, for example, by the R^2 statistic), which is likely if the variable is mainly determined by individual (idiosyncratic) variables that are not measurable in a HH survey.

However, this is not always the case. Accordingly, this paper also uses several alternative statistical models that are designed to handle such problematic variables. For example, one common alternative is the log regression model in which the natural log of the reported variable is used as the dependent variable (Goldberger 1998).[11] Unfortunately, the log transformation is only defined for positive values.[12] Another commonly used way to reduce the influence of outliers is to "winsorize" some percentage of the most extreme values (i.e., convert them to the next highest or lowest reported values), which can be done even with a dependent variable that has multiple zero or even negative values. In addition, this paper also uses several other alternative statistical models that minimize the effect of extreme values and/or outliers (i.e. median (quantile) regression and robust regression models) or that are designed for use with dependent variables with high concentrations of zero values (i.e., Tobit and two-part regression models).[13] By comparing the results obtained with several alternative statistical models, it should be possible to assess the criterion validity of the dependent variable, even if it is highly problematic.

It is important for researchers to select an appropriate statistical model to use in analyzing a given problematic economic variable. Choosing the one that gives the "best" results is not a legitimate research strategy. It is much better to use an objective basis for selecting among the alternative models. In behavioral research, one reasonable strategy might be to identify all of the theoretically relevant right-side variables (preferably before the data are collected) and to include them all in several alternative statistical models, focusing on the model that yields the largest number of statistically significant estimated coefficients.[14] In experimental research, a similar procedure might be applied to the baseline data, using the results to prepare a pre-analysis plan that would clearly identify the statistical model that would be used to measure effects, as well as any variables (covariates) that would be used in preparing "adjusted" estimates.

4.2 Composite variables

Composite variables (often referred to as "indexes") are variables that combine the values of several variables (items) to obtain a single value that effectively reflects the values of the included variables. This paper adheres to a pre-defined protocol to assess the reliability and validity of composite variables. The first step is to use principal components analysis (PCA) to assess the dimensionality of the individual

[11] In addition to reducing the influence of outliers, a log-transformed variable is more likely to have a normal distribution (which is usually assumed with standard statistical tests) and to reduce the influence of outliers.

[12] One alternative to the natural log transformation that has been frequently used in recent years and that does not have this limitation is the inverse hyperbolic sine transformation (Ravallion 2017). However, the results are more difficult to interpret (Bellemarre and Wichman 2019) and can be misleading when applied to variables with many zero values (Knowles 2021).

[13] Median regression minimizes the sum of *absolute* deviations from the regression "line" rather than squared deviations as with OLS, while robust regression weights the observations iteratively according to their proximity to the estimated regression line. Tobit and Two-step models are nonlinear regression models that are specially designed to handle dependent variables with zero values (Wooldridge 2010).

[14] Selecting the model on the basis of the highest R-squared is not appropriate if the dependent variables are not the same or if the observations are weighted (Willett and Singer 1988). In linear regression models, OLS will always yield the lowest R^2.

"item" candidates for inclusion in the composite variable.[15] In particular, do all of the Items load mainly on the first principal component? Items that load mainly on other dimensions should not be included in the composite variable because their inclusion reduces the validity of the composite variable. In cases where multiple items load mainly on principal components other than the first, the underlying latent construct is clearly multi-dimensional, and there may be little or no value in combining the items in a single composite variable.

The second step in the pre-defined protocol applied to composite variables is to find the best "scale" to use to combine the individual items to obtain a single valued composite variable (index) with maximum reliability and validity. The simplest scale is an equally weighted mean of the items, which is equivalent to Cronbach's raw alpha (Furr and Bacharach 2014). This can work well when the items are dichotomous (0-1) variables, but it can yield unintended results when the items have varying scales (e.g., binary for some items and a 1-5 Likert scale for others). In this case, the items with more widely varying scales will be more heavily reflected in the composite variable. One way of avoiding this problem is to calculate a proportional mean (i.e., divide each item's scale by its maximum value before calculating the mean). This is one of the four alternative aggregation methods used in this paper and is referred to as the "proportional mean scale" (PM scale). Alternatively, the composite variable can be calculated as the unweighted mean of items that have been "standardized" by subtracting their means and dividing the resulting differences by their standard deviations. The resulting value is equivalent to Cronbach's *standardized* alpha (Furr and Bacharach 2014) and is referred to in this paper as the "standardized mean scale" (SM scale). This scale is also equivalent to the widely used z-score "summary indexes" (Kling and others 2007).

More complex aggregation formulas used to combine the individual items in a composite variable weight the individual items in some way. For example, asset indexes (often referred to as "wealth" indexes) that are now widely used as an indirect measure of household income (Filmer and Scott 2012) are calculated as the score[16] of the first principal component of a set of standardized indicators typically referring to both housing characteristics and the ownership of durable assets. This is referred to as the "principal component scale" (PC scale). In addition, a fourth composite variable scale is also used in this paper that is equal to the predicted latent scores obtained from 2-paramter logistic models (for dichotomous items) or graded response models for ordered categorical responses that are commonly used in Item Response Theory (IRT) and that reflect not only the responses but also the revealed "difficulty" of the questions (Furr and Bacharach 2014, Rosier 2015). This fourth alternative scale (aggregation formula) is referred to in this paper as the IRT scale.[17]

The "best" scale for combining the individual items into each composite variable is selected in this paper on the basis of the R-squared statistic (the coefficient of multiple correlation) obtained in four alternative regression models with the dependent variable (the composite variable) calculated using each of the four alternative scales described above, but standardized to make the estimated coefficients

[15] Principal components analysis also reveals whether any of the items load negatively on the first principal component. The signs of such items should be reversed so that all items included in the composite variable enter with a positive sign.

[16] The score is calculated by multiplying the standardized item variables by their corresponding eigenvector loadings.

[17] In fact, there are many alternative IRT models that can be used to develop an IRT scale. See, for example, Kim, Kwak and others (2020) and Stata Corporation (2021). The two IRT models used in this paper are among the most commonly used.

directly comparable, and the same set of explanatory (right-side) variables. Using the R^2 to select the best "scale" is consistent with the spirit of classical test theory under the assumption that the regression model with the highest explanatory power accounts for the true (signal) variation in the underlying construct and that the remaining variation is due to random measurement error (noise). When the items included in a composite variable are unidimensional, and therefore highly correlated, the alternative scales also tend to be highly correlated, and there is little or no basis for choosing among the alternative scales (i.e., the R^2s are almost the same). However, this is not always the case, as discussed in section 4.2.

5. Findings

Summary Table 1 lists all of the variables (outcomes) whose measurement properties are discussed in this paper by their conceptual framework elements and dimensions listed in Box 1, including the name of the variable (column 1), the type of variable, e.g., continuous, dichotomous, discrete, or composite and whether it is problematic or subjective (column 2), the survey rounds in which data on the variable were collected (column 3) and the Appendix 2 Tables most relevant to the variable (column 4). The 72 tables referenced in column 4 of Summary Table 1 are listed in numerical order in a separate file ("Appendix 2 Tables") and are discussed in detail in Appendix 2, which is also structured according to the conceptual framework in Box 1. With the exception of composite variables, the variables listed in Summary Table 1 are "problematic" unless indicated to be "not problematic" in column 2. The variables listed in Summary Table 1 are also "objective" unless indicated to be "subjective."[18] The main findings with respect to problematic economic variables are listed in Summary Table 2[19] and discussed in section 5.1 below. The main findings with respect to the "composite variables," are listed in Summary Table 3 and discussed in section 5.2. These two distinct groups of variables are the main focus of this paper. The measurement issues and findings with respect to the remaining variables listed in Summary Table 1 (i.e., the continuous, discrete or dichotomous variables identified in column 4 as "non-problematic") are discussed in Appendix 2 under their respective conceptual framework elements and dimensions.

5.1 Findings with respect to problematic economic variables

Column 1 of Summary Table 2 describes the characteristics of each variable that make it problematic. All of the variables in Summary Table 2 are heavily skewed to the right (the normal distribution, by comparison, has zero skewness), are most often bounded by zero on the left but often with very extreme values extending far beyond the mean or median (the number of extreme values are listed for each variable, with "extreme" defined as values that are more than three times the inter-quartile range above the 75th percentile value (or below the 25th percentile value, in the case of extreme negative values). Kurtosis (the sharpness of the peak of the distribution, with the normal distribution having kurtosis equal to three) also tends to be very high in the distributions of the problematic economic variables. In addition, Summary Table 2 also frequently reports the skewness, kurtosis and number of very extreme values in the distributions of the natural logs of the problematic variables that do not have high concentrations of zero and/or negative values.[20] Interestingly, these statistics of the log-

[18] "Subjective" variables are based on opinions or perceptions, whereas "objective" variables are based (at least in principle) on facts.

[19] The term "impact" is used instead of "experimental effect" in Summary Tables 2 and 3 to save space.

[20] An inverse hyperbolic sine (IHS) transform is sometimes used instead of the natural log for such variables in this paper, e.g., "net financial assets" (i.e., accumulated savings less outstanding loans, a few zeroes and 816 negative

transformed variables are quite close in most cases to the corresponding statistics for the normal distribution, implying that their distributions are approximately lognormal. Some of the problematic variables also have very high concentrations of zero values, particularly those referring to savings, investment, the value of outstanding loans, and the number of paid and unpaid workers in the primary business (the owner is not included in the number of unpaid workers). These concentrations of zero values can bias OLS estimates (Wooldridge 2010). Although the information in column 2 indicates clearly that the distributions of all the variables listed in Summary Table 2 are problematic (compared to a normal distribution), their reliability and validity and the effects that their problematic distributions have on estimated relationships with other key variables (for example, gender) or on estimates of treatment effects is reported in columns 2-5 of Summary Table 2.

Column 2 presents the estimates (when available) of the test-retest reliability of each variable, including the highest and lowest correlations (r) between the values in different survey rounds.[21] Unfortunately, these estimates are only available for variables that were measured in more than one survey round (Box 3 discusses a unique opportunity to assess the reliability of the investment data more directly). In the case of the problematic economic variables, the test-retest reliability estimates are generally low, compared to the test-retest reliability of 0.80 to 0.90 in well designed and carefully administered tests, such as the SAT (Furr and Bacharach 2014). At least three factors contribute to this difference. First, errors of measurement are relatively large in many economic variables, particularly when respondents are asked to recall information for up to one year in the past (e.g., profits, savings, investment). Second, the periods between successive measurements can be sufficiently long (e.g., several months to two years between survey rounds) that it is unreasonable to assume in most cases that the true (unobserved) value of the variable measured has remained unchanged. Third, the problematic economic variables (unlike test scores, which are generally bounded by zero and some upper limit) have many extreme values that can easily bias the correlations between the observed measures. For this reason, the information on test-retest reliability in column 2 includes correlations between the natural log (or IHS) transformations of the reported variables. These transformations reduce the effects of extreme values, as evidenced by their generally higher test-retest correlations.

Box 3. A rare opportunity to assess reliability directly

Data on business investment were collected separately for the primary and second (if present) businesses in the ES and were obtained using the following two questions: (1) Was any investment in "equipment, furniture and fixtures, land, buildings (including renovations)" made in the last 12 months and, if yes, how much was spent? (2) Was any investment made in the last 12 months in the form of increases in product stock and, if yes, what is the value of these increases in product stock? In addition, the level of investment during approximately the same time period can be estimated indirectly (and independently) as the change in the reported total value of business capital between the BS and the ES, creating a unique opportunity to obtain an independent assessment of the reliability of the reported investment data. Although they are based on different survey questions and do not refer to exactly the same time periods, the reported and calculated investment variables (following adjustments for their slightly different time periods) are significantly correlated (r=0.205). Assuming that both variables are measures of the same unobserved true level of investment, measurement error alone accounts for the

values) because the IHS transformation is defined for positive, negative and zero values with similar positive values to those of the natural log transformation.

[21] In this paper, we use the correlation coefficient (r) as a measure of reliability, whereas the square of r ($R=r^2$) is more often used as the measure of reliability in psychometrics (Furr and Bacharach 2014)

observed differences between these two variables. If the measurement error is random, dropping 5% of the sample with the largest absolute differences between the two investment measures should have little effect on the estimates. This turns out to be the case (cf Appendix 2 Tables 46 and 47). It is reasonable to assume that there is also a lot of random measurement error in the other problematic economic variables in Summary Table 2. However, its presence is unlikely to bias estimates in the same way as the highly skewed distributions, very extreme values and high concentrations of zero values.

(end of Box #3)

With these considerations in mind, it is surprising that some of the test-retest estimates are as high as they are. For example, the correlation between the reported BS and ES values of total value of business capital is a relatively high 0.571 (or 0.662 between the logged values). In contrast, the correlation between the reported BS and ES values of primary business profits is a relatively low 0.148 (or 0.566 between the logged values). One possible explanation for this difference is that the estimate of the total value of business capital is an estimate of current value (i.e., does not require any recall), whereas the estimate of primary business profits is an estimate of the "average value of profits in a typical month during the past 12 months." Consistent with this reasoning, the test-retest reliability of total savings during the past 12 months between the BS and ES is only 0.045 (with too many zero values to permit a meaningful log or IHS transformation). By comparison, the test-retest reliability of the total *current* savings balance, which was measured only twice (in the MS and ES) and does not involve any recall, is considerably higher (0.270). The fact that the test-retest reliability of the retrospective estimate of savings during the past 12 months is so low does not mean that it does not contain any useful information. If the large measurement error is random, it would not bias regression estimates obtained using savings during the past 12 months as the dependent variable (such as those discussed in columns 3-5 of Summary Table 2). Instead, even a large amount of random measurement error in the dependent variable would only reduce the *precision* of the estimates (i.e., their significance levels).

Column 3 summarizes the findings with respect to the criterion validity of each variable, based on multiple regression analysis using the problematic economic variable as a dependent variable in several alternative statistical models.[22] The evidence on criterion validity is classified as strong, moderate or weak depending on the maximum number of significant estimated coefficients obtained by any one of the alternative statistical models.[23] Specifically, "strong" requires 10 or more significant estimated coefficients, "moderate," 5-9 significant estimated coefficients, and "weak," fewer than 5 significant estimated coefficients. In addition, if any of the signs of significant estimated coefficients are clearly inconsistent with prior expectations, based on economic theory or previous research, or if there are sign reversals among statistically significant coefficients across the alternative statistical models, the variable's criterion validity is reduced by one level (e.g., from "strong" to "moderate" or from "moderate" to "weak"). A specific example is the "strong" criterion validity of primary business profits (row 1, column 3 of Summary Table 2) based on 11 significant estimated coefficients obtained with the log regression model, compared to only 7 significant estimated coefficients obtained with the OLS

[22] The alternative statistical models are selected for their ability to address the specific problematic features of each variable. For example, winsorized OLS, the log regression model, the quantile (median) regression model and the robust regression models are selected for their ability to reduce the effect of extreme values, whereas the Tobit and two-part models are designed for dependent variables with high concentrations of zero values.

[23] All measures of statistical significance in this paper are based on estimated standard errors adjusted for clustered sampling. The cluster-adjusted estimates are obtained using Stata's "svy" commands for most statistical models. However, the cluster-adjusted estimates for the quantile (median) regression model and the robust regression model are cluster-bootstrapped estimates.

regression model (column 1). The OLS model also fails to identify the statistically significant positive relationships with the owner's cognitive ability and willingness to take risks composite variables that are identified by several of the other models (Appendix 2 Table 1).

The information in columns 4 and 5 is provided to address the question: Does it matter which statistical model is used to analyze problematic economic variables? Column 4 answers this question with evidence from behavioral research, specifically, multivariate analysis of the relationships between the various problematic economic variables and the gender of business owners, with other relevant variables held constant. Column 5 answers the same question with examples from experimental research, specifically, estimates of the effect of the business training provided to randomly selected WBOs on the problematic economic variables.

Column 4 discusses the behavioral research findings with respect to gender. For example, the results indicate that primary business profits are significantly lower (at the 0.001 level) for WBOs with all five alternative statistical models. However, looking down column 4 to row 2, one sees that when primary profits are calculated from data on revenue and expenses, gender is significant in only three of the five alternative models (not including the OLS model), while in the case of reported business revenue (row 3) gender is significant only in the log regression model. In fact, looking further down column 4, one sees that several of the problematic economic variables are significantly related to gender, but not in the OLS model and not even in the winsorized-OLS model. For example, total savings during the last 12 months (row 13) is significantly higher among WBOs in four of six alternative statistical models in Appendix 2, Table 36 (i.e., the Tobit model, the two-part model, the quantile (median) regression model and the robust regression model), but not in either the OLS or winsorized-OLS models. The conclusion? When analyzing gender relationships with problematic economic variables, the choice of statistical model often matters.

Column 5 discusses the experimental research findings on the estimated effects of the business training provided to randomly selected WBOs on the problematic economic variables. In this paper, both unadjusted (ANOVA) and adjusted (ANCOVA) experimental effects estimates are reported in Appendix 2 tables, with the unadjusted models including only the randomized treatment variable as a right-side variable (1 if the WBO received the training, 0 if not), whereas the adjusted models include both the randomized treatment variable and the baseline value of the problematic economic variable as right-side variables (in the case of variables not measured at baseline, only unadjusted estimates are available). Returning to the example of primary business profits (row 1), column 5 reports that the estimated positive experimental effect is statistically significant in both adjusted and unadjusted models only in the log regression model (at the 0.001 level in the unadjusted model and at the 0.05 level in the adjusted model).[24] Looking down column 5, one sees that statistically significant experimental effect estimates are often obtained only in subset of the statistical models. The conclusion? When estimating the effects of randomized treatments on problematic economic variables, the choice of statistical model also often matters.

The estimates in row 1 of Summary Table 2 illustrate another point of practical importance: a significant estimated experimental effect is often observed with statistical models that have the strongest criterion validity (e.g., rows 1, 6-8, 12-13). When such results are obtained from baseline data (e.g., primary

[24] The stars used in the tables reporting impact estimates are less conservative, reflecting the smaller sample size (limited to WBOs) and the typically smaller effects obtained in randomized experiments, with one star indicating significance at the 0.10 level, two stars indicating significance at the 0.05 level and three stars indicating significance at the 0.01 level.

business profits in rows 1, 12-13), they can be very useful input into pre-analysis plans. The results in column 5 illustrate another important point: low test-retest reliability in a given variable does not necessarily imply that it is difficult to obtain significant estimates of experimental effect. For example, in the case of savings during the last 12 months (row 13), the positive estimated experimental effect is significant at the 0.01 level in both the adjusted and unadjusted quantile (median) regression model and the robust regression mode (as well as at the 0.05 level in the unadjusted two-part model). The low test-retest reliability of this variable (r=0.045 from BS to ES) appears to be due mainly to the presence of large random measurement errors that do not bias the estimates of experimental effect (as in the example discussed in Box 3).

5.2 Findings with respect to composite variables

Column 1 of Summary Table 3 presents the ten available test-retest reliability estimates of the 17 composite variables listed in the table (the remaining variables are measured in only one survey round), including identifying the scales that yield the highest and lowest BS to ES correlations. There is no clear pattern.[25] The scales achieving the highest estimates are the PM scale (3 times), the PC scale (1 time), SM scale (3 times) and the IRT scale (3 times), while those achieving the lowest estimates are the PM scale (1 time), the PC scale (2 times), SM scale (3 times) and the IRT scale (4 times). Another general finding is that test-retest estimates of the composite variables in Summary Table 3 tend to be higher than those of the problematic economic variables in Summary Table 2. For example, the BS to ES test-retest reliability of the HH asset index in row 1 varies only from a low of 0.705 for the SM scale to a high of 0.749 for the PC scale. There are several reasons for this. First, the composite variables are measures based on multiple variables (20 variables, or "items," in the case of the HH asset index), and much of the random measurement error is averaged out of the composite variables. Second, most of the items included in the composite variables are either dichotomous or categorical variables with limited ranges, so that the composite variables do not have any extreme values. Despite these advantages, however, some composite variables do have relatively low test-retest reliability. The lowest is "trust in banks" (row 17) with lowest to highest BS to ES correlations ranging from 0.170 to 0.179 (similar to the corresponding test-retest estimates for primary business profits in row 1 of Summary Table 2). The most likely explanation for the low test-retest reliability of trust in banks is that it is based on three subjective ratings that appear not to be very stable over time.

Column 2 summarizes the evidence bearing on the criterion validity of the composite variables. For example, in the case of the HH asset index (row 1), the evidence supporting its criterion validity as a measure of HH income is strong, including 10 significant estimated coefficients in Appendix 2 Table 71 (Box 4 presents additional evidence bearing on the criterion validity of the HH asset index as a measure of HH income). The evidence supporting the criterion validity of the trust in banks composite variable is almost as strong (8 significant estimated coefficients in Appendix 2 Table 69), despite its relatively low test-retest reliability. In fact, column 2 indicates that most of the composite variables have either moderate or strong criterion validity. However, there are three exceptions, two of which involve subjective variables (subjective well-being in row 2 and subjective time preference in row 14). The third

[25] However, if we give one point for achieving the highest score and subtract one point for achieving the lowest, the PM scale is the winner (with 2 points), followed by the SM scale (0 points), and with the IRT and PC scales trailing with -1 and -2 points respectively.

(marital agency in row 5) is based on four objective questions.[26] However, as discussed in Appendix 2, this composite variable is quite complex, and several simplifying assumptions are made in its calculation.

Box 4. HH asset indexes are a measure of HH income, not assets

Data on the ownership of 20 household durable goods were collected in all three survey rounds (the actual goods are listed in Appendix 1). These types of data are considered relatively easy to collect and relatively reliable, and composite variables constructed from them (HH asset indexes) can be used as a cost-effective alternative to other measures of HH income based on detailed income or consumption data (Filmer and Scott 2012). However, the "face" validity of HH asset indexes as measures of HH income (as distinct from HH wealth or some other latent construct) is not obvious. One possible theoretical rationale is that the demand for durable assets is positively related to HH income, with the strength of the relationship stronger in the case of some assets (e.g., smart phones, refrigerators, cars, washing machines) with relatively high income elasticities compared to others (e.g., TVs, bicycles, simple hand phones). Table 8 shows the results of regressing either simple ownership (0-1) of the 20 durable goods (column 3) or the number of each item owned (column 4), on the natural log of the predicted measure of HH income discussed above. The results indicate that most (but not all) of the simple regressions are highly significant (i.e., 30 of the 40 t-statistics), suggesting that the estimated coefficients contain a lot of information on HH income. The question is how best to distill this information into an estimate of HH income (i.e., which scale to use). The standard method is to use the PC scale that is obtained from principal components analysis (PCA), i.e., the predicted score obtained from the estimated eigenvector loadings of each item on the first principal component of the 20 variables (shown in columns 5 and 6 of Table 8). This predicted score then becomes the asset index measure of HH income.[27] Examining the values in columns 3-6 of Table 8, one sees that the estimated eigenvector loadings in columns 5 and 6 are closely correlated with the corresponding estimated t-statistics in columns 3 and 4, establishing the criterion validity of the asset index as a measure of HH income. This is a very useful finding because it provides a cost-effective way to estimate HH income without having to measure the actual income streams that contribute to HH incomes (which in a rural setting typically include both agricultural and business income in addition to wage and salary income).

(end of Box #4)

Column 3 discusses which (if any) of the four alternative scales used in calculating the composite variables is preferred, with the preference based mainly on the estimates of test-retest reliability (if available) in column 1 and on the evidence of the variable's criterion validity summarized in column 2, but in some cases also on the experimental effect estimates discussed in column 5. The information in column 3 indicates that it is not possible to identify a clear preference for one (or even two scales) over the others for 11 of the 17 composite variables. In seven of these 11 cases, the evidence supports a slight preference for one or two of the scales over the others (rows 2, 3, 8, 11, 12, 14), as indicated in column 3, but not in the remaining five cases either because there is no basis for preferring any one or two of the scales over the others (rows 6, 10 and 17) or because the evidence is conflicting (rows 4 and 15). For the 6 composite variables where there is a clear basis for preferring one or two of the alternative scales over the others (rows 1, 5, 7, 9, 13 and 16) the proportional mean (PM) scale is

[26] The four questions are: (1) What percentage of your earnings is your spouse aware of? (1) What percentage of HH expenses are covered by your spouse? (3) Who has access to the money you make from your business? (4) Who decides how the money from your business will be spent?

[27] The estimated eigenvector loadings on the first principal component are the coefficients of the linear function that explains the highest proportion of the total variation in the 20 variables.

preferred for one variable (row 9), in combination with the IRT scale, the principal component (PC) scale is preferred for another (row 1) , the standardized mean (SM) scale is preferred for two variables (rows 5 and 7), and the IRT scale is preferred for the remaining two variables (rows 13 and 16), as well as in row 9 where it is preferred jointly with the proportional mean (PM) scale. The conclusion is that it is not possible to identify a preferred scale for composite variables analytically in many cases, but that it is possible in some cases.

The information in columns 4 and 5 is provided to address the question: Does it matter which scale is used in calculating composite variables? Column 4 answers this question with evidence from behavioral research, specifically, multivariate analysis of the relationships between the various composite variables and the gender of business owners. Column 5 answers the same question with examples from experimental research, specifically, estimates of the experimental effect of the business training provided to randomly selected WBOs on the composite variables.

Column 4 discusses the findings of behavioral research on the relationships between the composite variables and gender, with other relevant factors held constant. WBOs (and particularly currently married WBOs) have significantly higher values of the HH asset index with all four alternative scales (row 1).[28] WBOs also report significantly higher adherence to recommended business practices (row 7), particularly when this composite variable is calculated with the IRT scale, and significantly higher trust in banks (row 17), with all scales except the IRT scale.[29] In contrast, personal agency is significantly lower among WBOs with all four scales (row 3), and WBOs are significantly disadvantaged in most of the remaining composite variables, including: (1) intensity of banking services use, with all four scales (row 8) (2) mobile phone use (row 9), but consistently significant only with the IRT scale, (3) community participation, with all four scales (row 12), (4) subjective time preference, with all four scales (row 14), and (5) access to bank accounts, with all four scales (row 15). Intra-HH agency (row 4) is not significantly related to the business owner's gender, and there are no significant gender differences in cognitive ability, with all four scales (row 13). Perhaps most interestingly, the behavioral analysis indicates that the sign of the gender gap in MM use, initially favoring MBOs in the BS, is reversed in the ES (possibly due to the training provided to randomly selected WBOs, as discussed in Box 5).

Column 5 discusses the experimental research findings on the estimated effects of the business training provided to randomly selected WBOs on the composite variables. Experimental effect estimates cannot be obtained for four of the variables because only BS data were collected (rows 11-14). The available experimental effect estimates (either adjusted, unadjusted or both) for most of the remaining composite variables are statistically significant for one or more scales (the four exceptions are in rows 6 and 15-17). In most cases, the experimental effect estimates are significant with all four scales (rows 1-5, 7 and 9), whereas in two cases, the adjusted estimates are significant only with the proportional mean and principal component scales (rows 8 and 10). The finding that so many of the experimental effect estimates are significant for all four scales is consistent with the findings in column 3 that it is often difficult to identify a preferred scale. However, even when the estimates are significant with all four scales, the significance *levels* do vary in several cases (i.e., rows 1-3 and 7), which could be important in

[28] This finding is consistent with the findings for the predicted measure of HH income in Summary Table Two (row 5), which is important because (as discussed in Box 4), both variables are reliable and valid measures of HH income.

[29] Although marital agency (row 5) is also significantly higher among WBOs (at the 0.001 level), this composite variable has only weak criterion validity (column 2).

smaller samples.[30] The experimental effect estimates show only one link to the preferred scales identified in column 3, i.e., the slight preference for the proportional mean (PM) and principal component (PC) scales for the intensity of banking services use (row 8) is consistent with the experimental effect estimates, which are significant (at the 0.05 level) only for these two scales. The fact that such links are not more numerous suggests that there is only limited value in using the baseline data to identify preferred scales that could then be reflected in pre-analysis plans.

Box 5. The business training had unexpected experimental effects

The social experiment for which the survey data were collected was designed to increase the use of the newly established mobile money (MM) services made available by the partner bank in all 401 sample villages. In the BS, only 1.68% of sample business owners reported having ever used any MM services (1.58% of women business owners and 1.82% of men business owners). In the ES, however, 4.23% of business owners reported ever use of any MM services (4.73% of women business owners and 3.49% of men business owners). The business training provided exclusively to randomly assigned *women* business owners included training in the use of the partner bank's branchless banking services. Data in all three survey rounds on the use of MM services from any bank were combined into a single composite variable on "MM use" (row 9), with a clear preference for the proportional mean (PM) and IRT scales indicated in column 3. However, both adjusted and unadjusted estimates of the experimental effect of the business training on MM use are statistically significant at the 0.01 and 0.05 levels respectively with all four scales indicating that the business training encouraged WBOs to use MM services from all banks, not only those of the partner bank.

(end of Box #5)

6. Conclusions

This paper addresses issues related to the measurement of economic empowerment using a mix of analytical tools from economics, psychology and other disciplines. Specifically, the paper assesses the reliability and validity of multiple measures of economic empowerment at the household level, using a unique data set on both women and men business owners from rural Indonesia as a case study and focusing on two types of variables with special measurement issues: economic variables with problematic distributions and composite variables (indexes) representing multiple variables by a single value. In addition, the paper also addresses the question, "Does it matter?" by discussing the extent to which the findings of both behavioral and experimental research vary with the statistical model used to analyze problematic economic variables or with the scale used in calculating composite variables.

The paper analyzes 17 problematic economic variables, most having highly skewed distributions, large numbers of extreme values and/or large concentrations of zero values. The analysis finds wide variation in the test-retest reliability of the problematic economic variables, reflecting both possible changes in the true values due to relatively long intervals between successive measurements and the likely presence of large measurement errors, particularly in variables such as profits, savings and investment for which respondents are asked to provide estimates for past periods of up to the past 12 months. Using alternative statistical models that are designed to deal with the problematic features of these economic variables, the paper finds moderate to strong evidence supporting the criterion validity of most of the variables. However, the validity assessments often vary depending on the statistical model

[30] The impact estimates in this paper are based on sample sizes exceeding 2,200 WBOs, which is larger than in many experiments.

used (for example, fewer significant relationships when ordinary least-squares (OLS) estimation is applied directly to the reported variables). One important conclusion when working with problematic economic variables is that it is very important to use statistical models that are appropriate to address their specific problematic features.

The paper's findings are less conclusive regarding the best scales to use when calculating composite variables. The paper finds that the test-retest reliability of the 17 composite variables analyzed is generally higher than that of the problematic economic variables, probably because they are "averages" of multiple individual items with limited ranges. However, the paper finds that it is not always possible to identify one (or even two) preferred scales for most the composite variables. Of the 17 composite variables analyzed, a clear preference for one (or two) scales is identified for only six variables, with the IRT scale preferred for three of the six variables (including one for which the proportional mean scale is also preferred), two variables for which the standardized mean scale (i.e., a z-score) is preferred and one variable in which the principal component scale is preferred (i.e., the HH asset index, for which the principal component scale has been used from the start). Of the 13 composite variables for which impact estimates are possible (with the availability of follow-up data), all but two are either significant or insignificant across all four scales (although often with differing significance levels, which may be important in smaller samples). The overall conclusion is that it matters in some cases which scale is used in calculating composite variables, but much less than using an appropriate statistical model with problematic economic variables.

The WEE variables analyzed in this paper include seven subjective variables, including four composite variables (i.e., quality of life, personal agency, subjective time preference and trust in banks) and three non-problematic scalar variables that are discussed in Appendix 2 (i.e., concern about food insecurity and two measures of willingness to take risks). Because there is great interest in the use of more WEE subjective measures, it is useful to summarize some of the findings with respect to these measures, despite their limited representation in the data set. First, the evidence on their test-retest reliability is scant because data on all but one of the subjective measures are limited to one survey round (either the BS or the MS). The exception is trust in banks (a composite variable) for which BS to ES correlations range from 0.170 to 0.179 (depending on the scale), which is relatively low among the 17 composite variables analyzed in this paper (by comparison, the corresponding correlations in the intensity of use of banking services composite variable range between 0.425 to 0.570, depending on the scale).

Low test-retest reliability suggests the presence of substantial measurement error in this composite variable. The multivariate regression analysis used to assess the criterion validity of this variable (Appendix 2 Table 69) finds 7-8 significant estimated coefficients, depending on the scale (gender is statistically insignificant with the IRT scale). However, the comparable R^2s statistics for this composite variable are all below 0.02, which is very low and consistent with the presence of substantial measurement error (much of which is apparently random, with 7-8 of the 11 estimated regression coefficients statistically significant with theoretically plausible signs).[31] Both the adjusted and unadjusted impact estimates are also statistically insignificant with all four scales, possibly due in part to the presence of substantial random measurement error.[32] These findings point to the following conclusions: subjective variables may include substantial random measurement error resulting in low reliability.

[31] Low R^2s may also occur when a variable is largely driven by unobserved idiosyncratic variables, which is arguably more likely with subjective variables.

[32] However, significant impact estimates are obtained for the other two subjective composite variables for which impact estimates could be obtained (i.e., subjective well-being and personal agency).

This paper's ultimate purpose is to foster more accurate estimates of gender differences in economic empowerment and of the effects of interventions designed to eliminate them by encouraging program managers and researchers to give sufficient attention to the measurement properties of their data. This paper is premised on the belief that insufficient attention has been given in the past to measurement issues in the collection and analysis of data on economic empowerment, leading to possibly biased conclusions about the effects of experimental interventions on key problematic economic variables such as profits, savings or investment or possibly biased inferences about the gender differences observed in behavioral data. The sensitivity of experimental effect estimates (both adjusted and unadjusted) to the use of particular statistical models (and even particular scales for composite variables in some cases) suggests that this could be a problem even in the routine monitoring of outcomes between treated and untreated populations.

References

Behrman, Jere R. and James C. Knowles. 1999. “Household Income and Child Schooling in Vietnam.” *The World Bank Economic Review* 13(2):211-56.

Bellemarre, M. and C. Wichman. 2019. “Elasticities and the Inverse Hyperbolic Sine Transformation,” *Oxford Bulletin of Economics and Statistics*, July:1-12.

Bollen, Kenneth A. 1989. *Structural Equations with Latent Variables*. New York, NY: Wiley.

Buvinic, M., H. C. Johnson, E. Perova, and F. Witolear. 2020 “Can boosting savings and skills support female business owners in Indonesia? Evidence from a Randomized Controlled Trail,” East Asia and Pacific Gender Innovation Lab (EAPGIL), World Bank Group (March).

Buvinic, Mayra, M. O’Donnell, J. Knowles and S. Bourgault. 2020. “Measuring Women’s Economic Empowerment: A Compendium of Selected Tools.” Data2X and Center for Global Development.

Cohen, Jonathan, K.M. Ericson, D. Laibson and J.M. White. 2020. “Measuring Time Preferences.” *Journal of Economic Literature* 58(2):299-347.

Deserrano, E., G. Leon-Ciliotta and F. Witoelar. 2021. “When Transparency Fails: Financial Incentives for Local Banking Agents in Indonesia,” Unpublished manuscript (January)

Filmer, Deon and Kinnon Scott. 2008. "Assessing asset indices." Policy Research Working Paper Series 4605, The World Bank.

Furr, R. Michael and Verne R. Bacharach. 2014. *Psychometrics: An Introduction*. Second edition. Los Angeles: Sage Publishing, Inc.

Heckman, James J., Lance J. Lochner and Petra E. Todd. 2003. “Fifty Years of Mincer Earnings Regressions.” Working Paper 9732, National Bureau of Economic Research, Cambridge MA.

Kling, J., J. Liebman and L. Katz. 2007. “Experimental Analysis of Neighborhood Effects.” *Econometrica* 75(1):83-119.

Kim, Seock-Ho, Minho Kwak, and others. 2020. “Item Response Models in Psychometrika and Psychometric Textbooks.” *Frontiers in Education* 5(article 63): 1-11.

Knowles, James C. 2020. “Some caveats when using the IHS transformation.” Processed, Chapel Hill, NC (February). https://www.academia.edu/45086933/IHS_note30_08

Nering, Michael L. and Remo Ostini. editors, *Handbook of Polytomous Item Response Theory Models*. New York: Routledge, 2010.

Ravallion, Martin. 2017. “A concave log-like transformation allowing non-positive values, *Economics Letters,* 161:130-2.

Rosier, Malcolm. 2015. “A practical introduction to Item Response Theory (IRT) using Stata 14.”https://www.stata.com/meeting/australia15/abstracts/materials/oceania15_rosier.pdf

Stata Corporation. 2021. “Stata Item Response Reference Manual: Release 17.” College Station, TX: State Press. https://www.stata.com/manuals/irt.pdf

Willett, John B. and Judith D. Singer. 1988. “Another Cautionary Note About R^2: Its Use in Weighted Least-Squares Regression Analysis.” The American Statistician 42(3):236-238.

Wooldridge, Jeffrey. 2010. *Econometric Analysis for Cross-section and Panel Data*. Second edition. Cambridge, MA: MIT Press.

Summary Tables

Summary Table 1. List of variables analyzed by conceptual framework (Element and dimension)

ELEMENT	DIMENSION	VARIABLE	VARIABLE TYPE	DATA SOURCE(S)	RELEVANT TABLES
		(1)	(2)	(3)	(4)
ACHIEVEMENTS					
Achievements	**Personal income**	Primary business profits (reported)	Continuous variable	BS, MS, ES	Tables 1-2, 5, 61
Achievements	**Personal income**	Primary business profits (calculated)	Continuous variable	MS	Table 3
Achievements	**Personal income**	Primary business revenue	Continuous variable	MS	Table 4
Achievements	**Personal income**	Wage and salary earnings	Continuous variable	BS	Table 6
Achievements	**Personal assets**	Total current savings balances	Continuous variable	MS, ES	Table 11
Achievements	**Personal assets**	Net financial assets	Continuous variable	MS	Table 12
Achievements	**Business assets**	Total value of business capital	Continuous variable	BS, MS, ES	Tables 13, 14
Achievements	**Business assets**	Number of reported primary business customers	Discrete variable	BS, MS, ES	Tables 33-34
Achievements	**Business assets**	Primary business registered with the government	Dichotomous variable (not problematic)	BS, MS, ES	Table 72
Achievements	**Household income**	Household income (predicted)	Continuous variable	MS	Tables 7, 9
Achievements	**Household income**	HH asset index	Composite variable	BS, MS, ES	Tables 8, 10, 71
Achievements	**Quality of life**	Subjective well-being	Composite variable (subjective)	MS	Tables 17-18
Achievements	**Quality of life**	Number of hours worked in a typical month by business owners	Discrete variable (not problematic)	BS, MS, ES	Tables 19, 20
Achievements	**Vulnerability to shocks**	Instability in HH income	Continuous variable	MS	Table 15
Achievements	**Vulnerability to shocks**	Concern about food insecurity	Dichotomous variable (subjective, not problematic)	BS	Table 16
AGENCY/EMPOWERMENT					
Agency / empowerment	**Agency**	Personal agency	Composite variable (subjective)	BS, MS, ES	Table 21
Agency / empowerment	**Agency**	Marital agency	Composite variable	BS, MS, ES	Tables 25-27
Agency / empowerment	**Agency**	Intra-HH agency	Composite variable	BS, MS, ES	Tables 22-24

ELEMENT	DIMENSION	VARIABLE (1)	VARIABLE TYPE (2)	DATA SOURCE(S) (3)	RELEVANT TABLES (4)
Agency / empowerment	**Saving**	Any savings in the last 12 months (BS, ES) or last 3 months (MS)	Dichotomous variable (not problematic)	BS, MS, ES	Table 35
Agency / empowerment	**Saving**	Total amount saved in the last 12 months (BS, ES) or last 3 months (MS)	Continuous variable	BS, MS, ES	Tables 36, 38, 39, 65
Agency / empowerment	**Saving**	Total amount saved (as above) in relation to primary and second business profits	Continuous variable	BS, MS, ES	Tables 37, 38
Agency / empowerment	**Borrowing**	Any currently outstanding bank loans	Dichotomous variable (not problematic)	BS, MS, ES	Tables 40, 41, 43, 44
Agency / empowerment	**Borrowing**	Amount borrowed during the past 3 months	Continuous variable	MS	Table 42
Agency / empowerment	**Borrowing**	Total amount of currently outstanding loans from all sources	Continuous variable	MS	Tables 43, 44
Agency / empowerment	**Borrowing**	Ever borrowed from money lenders	Dichotomous variable (not problematic)	ES	Table 45
Agency / empowerment	**Borrowing**	Number of loans from money lenders in last 12 months	Discrete variables (not problematic)	ES	Table 45
Agency / empowerment	**Borrowing**	Ever borrowed from non-bank financial institution	Dichotomous variable (not problematic)	ES	Table 45
Agency / empowerment	**Borrowing**	Number of loans from non-bank financial institutions in last 12 months	Discrete variables (not problematic)	ES	Table 45
Agency / empowerment	**Business investment**	Reported value of business investment in the past 12 months	Continuous variable	ES	Tables 46. 49
Agency / empowerment (also listed above under Achievements/Quality of life)	**Personal work effort**	Total hours worked in a typical month in the primary and second business	Discrete variable (not problematic)	BS, MS, ES	Tables 19, 20
Agency / empowerment	**Mobilization of additional labor**	Number of unpaid workers in the primary and second business	Discrete variable	BS, MS, ES	Tables 30-31
Agency / empowerment	**Mobilization of additional labor**	Number of paid workers in the	Discrete variable	BS, MS, ES	Tables 30, 32

ELEMENT	DIMENSION	VARIABLE (1)	VARIABLE TYPE (2)	DATA SOURCE(S) (3)	RELEVANT TABLES (4)
		primary and second business			
Agency / empowerment	**Mobilization of additional labor**	Help received from other HH members	Composite variable	BS, MS, ES	Table 70
Agency / empowerment	**Business practices**	Adherence to recommended business practices	Composite variable	BS, ES	Tables 28-29
Agency / empowerment	**Use of financial services**	Access to bank accounts	Composite variable	BS, MS, ES	Table 66, 68
Agency / empowerment	**Use of financial services**	Intensity of use of banking services	Composite variable	BS, MS, ES	Tables 50-52
Agency / empowerment	**Use of financial services**	Ever use of mobile money (MM) services	Dichotomous variable	BS, MS, ES	Tables 56
Agency / empowerment	**Use of financial services**	Use of mobile money (MM) services	Composite variable	BS, MS, ES	Tables 53-56
Agency / empowerment	**Networking**	Business connections	Composite variable	BS	Table 58
Agency / empowerment	**Community participation**	Participation in voluntary activities	Composite variable	BS	Table 59
RESOURCES					
Resources	**Age/experience**	Age in years	Discrete variable (not problematic)	BS, MS, ES	see Appendix 2 (text)
Resources	**Age/experience**	Number of years worked in primary business	Discrete variable (not problematic)	BS, MS, ES	Table 62
Resources	**Education**	Highest level of schooling completed	Discrete variable (not problematic)	BS	Table 63
Resources	**Education**	Cognitive ability	Composite variable	BS	Tables 60, 62
Resources	**Psychological characteristics**	Willingness to take risks	Discrete variable (subjective, not problematic)	BS	Table 63
Resources	**Psychological characteristics**	Subjective time preference	Composite variable (subjective)	BS	Tables 64, 65
Resources	**Access to financial services**	Access to bank accounts	Composite variable	BS. MS, ES	Tables 66, 68
Resources	**Access to financial services**	Distance to banking services	Composite variable	ES	Table 67
Resources	**Access to infrastructure**	Mobile phone use	Composite variable	BS, MS, ES	Table 57
Resources	**Social capital**	Trust in banks	Composite variable (subjective)	BS	Table 69

Summary Table 2. Problematic economic variables

Variable (units)	Problematic characteristics of distributions	Test-retest reliability[33]	Criterion validity	Does the use of alternative statistical models with problematic variables make a difference? Behavioral relationships (gender)	Does the use of alternative statistical models with problematic variables make a difference? Experimental effects (business training)
	(1)	(2)	(3)	(4)	(5)
(1) Primary business profits (as reported) (Rp. millions)	Reported variable Skewness: 17.5 (BS) Kurtosis: 475.9 Extreme values: 166 (BS) Zero values: 4 (BS) Natural log transform Skewness: -0.432 (BS) Kurtosis: 4.62 Extreme values: 0	Reported variable BS to ES: 0.148 BS to MS: 0.264 MS to ES: 0.397 Natural log transform BS to ES: 0.566 BS to MS: 0.534 MS to ES: 0.596	Strong (Table 1), with 11 significant estimated coefficients with the log regression model, compared to 7 with the OLS regression model.	WBO's reported profits are significantly lower (at the 0.001 level) in all 5 models, other factors equal.	Estimated impact positive and significant in the following statistical models in Table 4: Log (unadjusted)*** / (adjusted)** OLS, (unadjusted)* Winsorized OLS (unadjjusted)* Robust (unadjusted)*
(2) Primary business profits (calculated from revenue and expenses) (Rp. millions)	Reported variable Skewness: 18.6 (MS) Kurtosis: 578.6 Extreme values: 148 Zero values: 6 Negative values: 131 Natural log transform Skewness: -2.12 (MS) Kurtosis: 3.74 Extreme values: 1	N/A (data were collected in only one survey round)	Moderate (Table 3), with 9 significant estimated coefficients with the log regression model, compared to 6 with the OLS regression model	WBO's calculated primary business profits are significantly lower (at the 0.05 level of higher) in 3 models (not including OLS).	(Not shown) Estimated impact of the business training on the calculated primary profits of WBOs is statistically insignificant in all five models
(3) Primary business revenue (Rp. millions)	Reported variable Skewness: 18.7 (MS) Kurtosis: 416.3 Extreme values: 123 Zero values: 0 Natural log transform Skewness: 0.015 (MS) Kurtosis: 3.74 Extreme values: 2	N/A (data were collected in only one survey round)	Moderate (Table 5), with 7 significant estimated coefficients with the robust regression model, compared to 3 with the OLS regression model.	WBO's reported business revenue is significantly lower (at the 0.05 level) only in the log regression model	(Not shown) Estimated impact of the business training on the primary business revenue of WBOs is statistically insignificant in all unadjusted models (adjusted models are not possible because no BS data were collected on business revenue).
(4) Wage and salary earnings (Rp. millions)	Reported variable Skewness: 13.9 (BS) Kurtosis: 348.8 Extreme values: 647 Zero values: 4178 Natural log transform Skewness: -0.305 (BS) Kurtosis: 2.77 Extreme values: 0	N/A (data were collected in only one survey round)	Moderate (Table 6), with 7 significant estimated coefficients obtained with the winsorized OLS, log and robust regression models, compared to 6 with the OLS model and 4 with the quantile (median) regression model.	WBOs' wage and salary earnings are significantly lower (at the 0.001 level) in all five statistical models (about 35% lower).	No impact estimates are possible because only BS data were collected on wage and salary earnings
(5) HH income (predicted) (Rp. millions)	Reported variable Skewness: 6.77 Kurtosis: 76.4 Extreme values: 69 Zero values: 0 Natural log transform Skewness: 0.112 Kurtosis: 3.29 Extreme values: 0 Zero values: 1	N/A (data were collected in only one survey round)	Moderate (Table 7), with 4 significant estimated coefficients, However, (not shown) the robust regression model yields 6 significant estimated coefficients, compared to 3 by the OLS regression model.	(Not shown) WBOs' reported and predicted HH incomes are significantly higher, except in the OLS model.	(Not shown) Estimated impact of the business training on the predicted HH income of WBOs is statistically insignificant in all five unadjusted models (adjusted models are not possible because there are no BS data on HH income).

[33] All estimates refer to the correlation coefficient (r, not to $R=r^2$) and are significant at the 0.05 level or higher unless otherwise indicated.

Variable (units)	Problematic characteristics of distributions	Test-retest reliability[33]	Criterion validity	Does the use of alternative statistical models with problematic variables make a difference? Behavioral relationships (gender)	Experimental effects (business training)
	(1)	(2)	(3)	(4)	(5)
(6) Total savings balances (Rp. millions)	Reported variable Skewness: 23.1 Kurtosis: 877.0 Extreme values: 287 Zero values: 409	Reported values MS to ES: 0.270 (no BS data on savings balances)	Moderate (Table 11), with 7 significant estimated coefficients with the two-part model.	No significant gender difference with the OLS or winsorized OLS models, but significantly higher savings of WBOs in 4 other models	(Not shown) Unadjusted estimates of the impact of the business training on total savings balances were positive and significant in three of the seven models for which results are reported in Table 11: OLS (0.10 level), Two-part and Robust models (0.05 level).
(7) Net financial assets (Rp. millions)	Reported variable Skewness: 25.4 (MS) Kurtosis: 871.5 Extreme values: 191 (+) and 90 (-) Zero values: 81 Negative values: 816 IHS transform Skewness: -0.122 Kurtosis: 2.35 Extreme values: 49 (+) and 29 (-)	N/A (data were collected in only one survey round)	Moderate (Table 12), with 9 significant estimated coefficients with the OLS winsorized (highest and lowest 2% of values), the IHS model, and the robust regression model, compared to only 3 with the OLS model.	No gender difference in the OLS or winsorized OLS models, but significantly positive in 3 other models (Table 12)	(Not shown): Unadjusted estimates of the impact of the business training on net financial balances are positive and significant (at the 0.10 level) only in the IHS model.
(8) Total value of business capital (Rp. millions)	Reported variable Skewness: 8.97 (BS) Kurtosis: 111.5 Extreme values: 418 Zero values: 1 Natural log transform Skewness: 0.138 Kurtosis: 3.08 Extreme values: 0	Reported variable BS to ES: 0.571 BS to MS: 0.412 MS to ES: 0.424 Natural log transform BS to ES: 0.662 BS to MS: 0.692 MS to ES: 0.705	Moderate (Table 13), with 9 significant estimated coefficients with the log regression model, compared to 6 with the OLS model.	WBOs have very significantly lower total business capital values in all five models.	Unadjusted estimates in Table 14 are consistently positive, but only significant (at the 0.05 level) in the robust regression model. However, the adjusted estimates are positive and significant at 0.05 level in the log and robust regression models and at the 0.10 level in the OLS model.
(9) Instability in HH income (ratio)	Reported variable Skewness: 11.1 Kurtosis: 171.0 Extreme values: 161 Zero values: 5 Natural log transform Skewness: -3.478 Kurtosis: 47.06 Extreme values: 0	N/A (data were collected in only one survey round)	Weak(Table 15), due to reversals in the signs of two significant estimated coefficients and to the positive sign of the significant estimated relationship with the number of earnings sources, which economic theory suggests should be negative (due to risk pooling).	Income instability is lower among women business owners and significantly in 3 of 5 models: OLS (.05 level), OLS (winsorized) and log (at 0.001 level).	(Not shown) Estimated impact of the business training on the instability of HH income is statistically insignificant in all unadjusted models (adjusted models are not possible because no BS data on HH income were collected).
(10) Number of unpaid workers in the primary business (number)	Reported variable Skewness: 13.4 (BS) Kurtosis: 438.6 Extreme values: 15 Zero values: 1449	Reported variable BS to ES: 0.293 BS to MS: 0.339 MS to ES: 0.415	Moderate (Tables 30, 31), with 10 significant estimated coefficients with the Winsorized OLS model.	WBOs employ significantly more unpaid workers in all four statistical models.	(Not shown) Both adjusted and unadjusted impact estimates are consistently positive and insignificant in all four models

Variable (units)	Problematic characteristics of distributions	Test-retest reliability[33]	Criterion validity	Does the use of alternative statistical models with problematic variables make a difference? Behavioral relationships (gender)	Experimental effects (business training)
	(1)	(2)	(3)	(4)	(5)
(11) Number of paid workers in the primary business (number)	Reported variable Skewness: 15.2 Kurtosis: 355.0 Extreme values: 782* Zero values: 4043 * Number of extreme values is equal to the number of non-zero values	Reported variable BS to ES: 0.586 BS to MS: 0.493 MS to ES: 0.664	Moderate (Tables 30, 32), with 7 significant estimated coefficients with the Winsorized OLS model.	WBOs employ significantly fewer paid workers in all four statistical models.	(Not shown) Both adjusted and unadjusted impact estimates are consistently positive and insignificant in all models
(12) Number of reported customers in the primary business (number)	Reported variable Skewness: 4.97 (BS) Kurtosis: 38.3 Extreme values: 417 Zero values: 0 Natural log transform Skewness: 0.335 Kurtosis: 2.46 Extreme values: 0	Reported variable BS to ES: 0.328 BS to MS: 0.133 MS to ES: 0.235 Natural log transform BS to ES: 0.471 BS to MS: 0.367 MS to ES: 0.519	Moderate (Table 33), with 8 significant estimated coefficients with the log regression model, compared to 5-6 with the other models	WBOs report significantly fewer customers of their primary businesses in all five statistical models.	Consistently positive and significant adjusted and unadjusted estimates in Table 34 in the log regression model (at the 0.05 and 0.10 levels respectively) and in the robust regression model (at the 0.10 level).
(13) Total savings during the last 12 months, including zeroes (MS: last 3 months) (Rp. millions)	Reported variable Skewness: 50.6 Kurtosis: 2853,2 Extreme values: 257 Zero values: 1072	Reported variable BS to ES: 0.045 BS to MS: 0.029* MS to ES: 0.186* * No adjustment for different reference peiods.	Moderate (Table 36), 8 significant estimated coefficients with the two-part model, compared to 2 with the OLS model.	WBO's have significantly higher savings in all models except OLS and winsorized OLS models.	Unadjusted impact estimates are consistently positive and statistically significant in the quantile (median) and robust regression models (at the 0.01 level) and in the two-part model (at the 0.05 level). Adjusted estimates are also consistently positive and significant at the 0.01 level in the quantile (median) and robust regression models.
(14) Total amount saved relative to annualized sum of primary and second business profits (ratio)	Reported variable Skewness: 57.1 (BS) Kurtosis: 3614.5 Extreme values: 250 Zero values: 1062	Reported variable BS to ES: 0.022* BS to MS: 0.061 MS to ES: 0121 * Not significant at the 0.05 level	Moderate (Table 37). with 9 significant estimated coefficients with the robust regression model (followed by the two-part model with 8), compared to only one with the OLS regression model.	WBO's have significantly higher savings ratios in four models, but not in the OLS or Two-part models.	(Not reported) Both adjusted and unadjusted estimates are consistently insignificant in all six statistical models.
(15) Amount of money borrowed during the past 3 months (Rp. millions)	Reported variable Skewness: 10.52 (MS) Kurtosis: 139.1 Extreme values: 406 Zero values: 1642	N/A (data were collected in only one survey round)	Weak (Table 42), with only 3 significant estimated coefficients in all four alternative statistical models.	The amount borrowed is insignificantly related to gender in all four models.	(Not shown) Unadjusted impact estimates are negative in 3 of 4 models and significant (at the 0.10 level) in the Tobit model. (Adjusted estimates are not possible due to absence of BS data on amount borrowed)

Variable (units)	Problematic characteristics of distributions	Test-retest reliability[33]	Criterion validity	Does the use of alternative statistical models with problematic variables make a difference? Behavioral relationships (gender)	Experimental effects (business training)
	(1)	(2)	(3)	(4)	(5)
(16) Total amount of currently outstanding loans **(Rp. millions)**	Reported variable Skewness: 10.82 (MS) Kurtosis: 176.0 Extreme values: 151 Zero values: 761	N/A (data were collected in only one survey round)	Moderate (Table 43), with 6 significant estimated coefficients (5 of which are the same) in all four statistical models.	WBOs owe less in currently outstanding loans in all four statistical models, but only significantly less (at the 0.10 level) in the OLS model.	(Not shown) Unadjusted impact estimates are consistently negative and statistically insignificant in all 4 models. (Adjusted estimates are not possible due to absence of BS data on amount borrowed)
(17) Reported business investment during the past 12 months **(Rp. millions)**	Reported variable Skewness: 22.2 Kurtosis: 661.9 Extreme values: 394 Zero values: 2081	N/A (data were collected in only one survey round) (r=0.205 between reported and indirectly estimated values)	Moderate (Tables 46 and 47), with 8 significant estimated coefficients with the winsorized OLS model.	WBOs have less reported investment during the last 12 months, but significantly less in only four of six statistical models (including OLS and winsorized OLS models).	Unadjusted impact estimates are positive and significant at the 0.01 level only in the quantile (median) regression model with the full sample in Table 48 (or at the 0.05 level with the sample trimmed of the 5% least reliable observations in Table 49).

Summary Table 3. Composite variables

Variable (number of items)	Test-retest reliability (see notes below)	Criterion Validity	Preferred scale	Does the use of alternative scales make a difference? Behavioral relationships (gender)	Does the use of alternative scales make a difference? Experimental effects (business training)
	(1)	(2)	(3)	(4)	(5)
(1) HH asset index **(20 items)**	Highest: PC scale BS to ES: 0.749 BS to MS: 0.794 MS to ES: 0.783 Lowest: SM scale BS to ES: 0.705 BS to MS: 0.746 MS to ES: 0.749	Strong, based on the results in Table 8 (discussed in Box 4) and on multivariate analysis (Table 71), with 10 statistically significant estimated coefficients.	There is a clear preference for the principal component scale on the basis of its relatively high R^2 in Table 71 and its more significant adjusted impact estimate in Table 10.	WBO HHs have significantly higher values of the HH asset index (0.26-0.28 standard deviations higher) with all four scales.	Unadjusted estimates are uniformly positive and statistically insignificant with all four scales (Table 10). However, the adjusted estimates are positive and statistically significant (at the 0.05 level) for 3 of the 4 scales (including the preferred principal component scale) and at the 0.10 level for the IRT scale.
(2) Subjective well-being **(2 items)**	N/A (limited to MS data)	Weak, with 7 significant estimated coefficients in Tables 17 and 18, but reduced one level due to an unexpected significant negative relationship with schooling.	Slight preference for the IRT scale, based on the results in Tables 17 and 18. However, (not shown) the estimated impact is less significant for the IRT scale.	Women business owners (including all WBOs and only currently married WBOs) report consistently higher, but not significantly higher, subjective well-being with all four scales.	(Not shown) Unadjusted estimates for all WBOs are positive and significant (at the (0.05 level) for 3 of 4 scales and significant (at the 0.10 level) for the remaining IRT scale. Unadjusted estimates for married WBOs are positive and significant (at the 0.05 level) with all four scales.
(3) Personal agency **(14 items)**	N/A (limited to MS data)	Moderate (Table 21), with 7 significant estimated coefficients for 3 of the 4 scales (versus 5 with the remaining IRT scale)	Slight preference for the proportional mean scale, based on the more significant impact for this scale (not shown).	Personal agency is significantly lower among WBOs (about 0.3 standard deviations) with all four scales	(Not shown) Unadjusted estimates are positive and significant (at the 0.01 level) for the proportional mean scale and at the 0.05 level for the other three scales.
(4) Intra-HH agency **(5 items)**	Highest: PM scale BS to ES: 0.418 BS to MS: 0.456 MS to ES: 0.444 Lowest: IRT scale BS to ES: 0.408 BS to MS: 0.466 MS to ES: 0.470	Moderate (Table 22), with 9 significant estimated coefficients with the IRT scale versus 8 with the other 3 scales.	Based on the test-retest estimates, the preferred scale is the proportional mean scale. However, based on the estimates in Table 22, the IRT scale is preferred.	Intra-HH agency is not significantly related to the business owner's gender with all four scales.	Both adjusted and unadjusted estimates obtained with the MS data (Table 23) are consistently positive and significant (at the 0.01 level) with all four scales. However, with the ES data (Table 24), both adjusted and unadjusted estimates are consistently positive and insignificant with all four scales.

Variable (number of items)	Test-retest reliability (see notes below)	Criterion Validity	Preferred scale	Does the use of alternative scales make a difference? Behavioral relationships (gender)	Experimental effects (business training)
	(1)	(2)	(3)	(4)	(5)
(5) Marital agency (4 items)	Highest: SM scale BS to ES: 0.523 BS to MS: 0.541 MS to ES: 0.555 Lowest: IRT scale BS to ES: 0.490 BS to MS: 0.496 MS to ES: 0.515	Weak (Table 25), with only 4-5 significant estimated coefficients and with some unexpected signs (i.e., the significantly higher marital agency among WBOs and the significant negative relationships with cognitive ability and total earnings)	Based on the test-retest estimates and the results in Table 25, the standardized mean scale is the preferred scale	Women business owner's marital agency is significantly higher than men's (at the 0.001 level), with all four scales.	Both adjusted and unadjusted estimates obtained with the MS data (Table 26) are consistently positive and significant with all four scales. However, adjusted and unadjusted estimates obtained with the ES data (Table 27) are consistently positive and insignificant with all four scales
(6) Help received from other HH members (5 items)	Highest: PM scale BS to ES: 0.535 BS to MS: 0.562 MS to ES: 0.524 Lowest: PC scale BS to ES: 0.496 BS to MS: 0.538 MS to ES: 0.472	Strong (Table 70), with 10 significant estimated coefficients for the proportional mean scale	No scale is clearly preferred. The proportional mean scale achieves the highest test-retest score and yields the largest number of significant estimated coefficients in Table 70, but the IRT scale has the highest R^2 among the four scales in Table 70, while the proportional mean scale has the lowest.	The mobilization of HH labor is not significantly related to the gender of business owners with all four scales.	(Not shown) Both adjusted and unadjusted estimates are consistently insignificant with all four scales.
(7) Business practices (12 items)	Highest: SM scale BS to ES: 0.504 BS to MS: N/A MS to ES: N/A Lowest: IRT scale BS to ES: 0.433 BS to MS: N/A MS to ES: N/A	Strong (Table 28), based on 10 significant estimated coefficients with 3 of the 4 scales (and 9 with the remaining IRT scale)	Based on the test-retest estimates and the R^2s in Tables 28 and 29, the standardized mean scale is preferred	WBOs report significantly higher adherence to recommended business practices in Table 28 (at the .001 level with the IRT scale, compared to the 0.05 level with the other scales)	Both adjusted and unadjusted estimates are positive and statistically significant (at the 0.01 level) for 3 scales in Table 29 and at the 0.05 level for the remaining IRT scale (the least preferred scale for business practices based on columns 1 and 2)
(8) Intensity of banking services use (11 items)	Highest: IRT scale BS to ES: 0.570 BS to MS: 0.667 MS to ES: 0.650 Lowest: SM scale BS to ES: 0.425 BS to MS: 0.468 MS to ES: 0.480	Strong (Tables 50 and 51), with 10 significant estimated coefficients	Unclear. Slight preference for the proportional mean and principal component scales, with clearly lower preference for the standardized mean scale.	Intensity of banking services use is significantly lower among woman business owners in both the BS and ES data with all four scales	Unadjusted estimates are statistically insignificant for all scales. However, the adjusted estimates are positive and significant (at the 0.05 level) in Table 52 for the proportional mean and principal component scales.
(9) Mobile money (MM) use (3 items)	Highest: PM scale BS to ES: 0.396 BS to MS: 0.404 MS to ES: 0.454 Lowest: PC scale BS to ES: 0.315 BS to MS: 0.287 MS to ES: 0.348	Moderate (Tables 53 and 54), with 7 significant estimated coefficients in Table 54 with the IRT scale	Proportional mean and IRT scales are preferred, most clearly on the basis of the BS data in Table 53 but also on the basis of the impact estimates in Table 56.	MM use is lower among WBOs, but the relationship is only consistently significant with the IRT scale.	Both the adjusted and unadjusted impact estimates are all positive and significant in Table 56 with all four scales.

				Does the use of alternative scales make a difference?	
Variable (number of items)	**Test-retest reliability (see notes below)**	**Criterion Validity**	**Preferred scale**	**Behavioral relationships (gender)**	**Experimental effects (business training)**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
(10) Intensity of mobile phone use (4 items)	Highest: IRT scale BS to ES: 0.629 BS to MS: 0.704 MS to ES: 0.734 Lowest: SM scale BS to ES: 0.608 BS to MS: 0.682 MS to ES: 0.718	Moderate(Table 57), with 8 significant estimated coefficients with all four scales	The results in Table 57 provide no basis for preferring one scale over the others.	Mobile phone use is significantly lower by about 0.31 standard deviations among women business owners with all four scales/	(Not shown) Unadjusted estimates are consistently positive but insignificant with all four scales. However, the adjusted estimates are consistently positive and significant at the 0.05 level with the proportional mean and principal component scales
(11) Business connections (13 items)	N/A (limited to BS data)	Moderate (Table 58), with 9 significant estimated coefficients for the IRT scale and 8 with the other three scales	Slight preference for the IRT scale, based on the results in Table 58.	No gender analysis (sample is limited to WBOs).	No impact analysis for this variable (only BS data were collected)
(12) Community participation (4 items)	N/A (limited to BS data)	Strong (Table 59), with 10 significant estimated coefficients (standardized mean scale)	The results in Table 59 suggest a slight preference for the standardized mean scale	Women business owner's community participation is significantly lower with all four scales.	No impact analysis for this variable (only BS data were collected)
(13) Cognitive ability (4 items)	N/A (limited to BS data)	Strong (Table 62), based on the linear relationship with highest level of completed schooling and the significant quadratic relationship with age (despite only 5 significant estimated coefficients).	Based on the results in Table 62, the IRT scale is preferred	(Not shown) Cognitive ability is not significantly related to gender with all four scales.	No impact analysis for this variable (only BS data were collected)
(14) Subjective time preference (3 items)	N/A (limited to BS data)	Weak (Table 64), with only 2-4 significant estimated coefficients. However, the significant estimated negative relationships with savings in three of six models in Table 65 support this variable's criterion validity	The principal component and IRT scales are slightly preferred, based on their higher R^2s and 4 versus only 2 significant estimated coefficients in Table 64.	Women business owners have significantly lower subjective time preference with all four scales (at the 0.001 level, but only at the 0.05 with the SM scale).	No impact analysis for this variable (only BS data were collected)
(15) Access to bank accounts (4 items)	Highest: SM scale BS to ES: 0.618 BS to MS: 0.842 MS to ES: 0.708 Lowest: IRT scale BS to ES: 0.589 BS to MS: 0.781 MS to ES: 0.689	Strong (Tables 66 and 68), with 10 significant estimated coefficients in both tables.	The IRT scale is preferred on the basis of its higher R^2 with the BS data in Table 66. However, the test-retest reliability of the IRT is slightly lower than that of the other scales.	Access to bank accounts is significantly lower among WBOs with all four scales	(Not shown) Both adjusted and unadjusted estimates are consistently positive but statistically insignificant with all four scales.

Variable (number of items)	Test-retest reliability (see notes below)	Criterion Validity	Preferred scale	Does the use of alternative scales make a difference? Behavioral relationships (gender)	Experimental effects (business training)
	(1)	(2)	(3)	(4)	(5)
(16) Distance to banking services **(3 items)**	N/A (limited to ES data)	Strong (Tables 67 and 68), including the significant negative relationship (at the 0.001 level) of this variable this variable with the "access to bank accounts" composite variable, as discussed above.	The IRT scale is preferred on the basis of the R^2s in Table 67.	N/A (The relationship between this mainly village-level distance variable and gender is misleading because more women than men business owners were included by design in all villages)	(Not shown) Unadjusted impact estimates are statistically insignificant with all four scales.
(17) Trust in banks **(3 items)**	Highest: IRT scale BS to ES: 0.179 BS to MS: 0.205 MS to ES: 0.253 Lowest: PM scale BS to ES: 0.170 BS to MS: 0.197 MS to ES: 0.205	Moderate (Table 69), with 8 significant estimated coefficients.	No basis for preferring one scale over another	Trust in banks is consistently positive with all four scales in Table 69, but significant (at the 0.05 level) with only three scales (the IRT scale is the exception).	(Not shown) Both adjusted and unadjusted impact estimates are consistently positive but statistically insignificant with all four scales

Notes: Test-retest reliability is assessed by comparing the correlation (r) between the BS and ES values, unless otherwise indicated. N/A=not available (e.g., data are limited to one survey round).

Detailed tables

Table 1. Multiple linear regression analysis of primary business profits (reported) using alternative statistical models (BS data)

	Alternative statistical models				
	OLS applied to reported dependent variable	OLS winsorized regression model	Log regression model	Quantile (median) regression	Robust regression model
	(1)	(2)	(3)	(4)	(5)
Owner's age	0.104**	0.076**	0.066***	0.065***	0.054***
	(0.038)	(0.025)	(0.016)	(0.017)	(0.015)
Owner's age squared	-0.001*	-0.001*	-0.001***	-0.001**	-0.001**
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Woman business owner	-0.708***	-0.546***	-0.401***	-0.515***	-0.436***
	(0.172)	(0.059)	(0.034)	(0.049)	(0.036)
Owner's Schooling: Lower secondary	0.117	-0.037	-0.009	0.009	0.002
	(0.144)	(0.069)	(0.041)	(0.051)	(0.039)
Owner's Schooling: Upper secondary	0.094	0.061	0.009	0.079	0.031
	(0.107)	(0.071)	(0.038)	(0.056)	(0.042)
Owner's Schooling: Tertiary	-0.344	-0.233	-0.150*	-0.121	-0.114
	(0.207)	(0.133)	(0.068)	(0.095)	(0.086)
Owner's cognitive ability (IRT scale)	0.026	0.029	0.040**	0.037	0.032*
	(0.047)	(0.026)	(0.014)	(0.019)	(0.014)
Owner's willingness to take risks (IRT scale)	0.142	0.066*	0.050**	0.058**	0.074***
	(0.077)	(0.032)	(0.017)	(0.020)	(0.019)
Days worked by owner: primary business (log)	0.572***	0.361**	0.230***	0.269**	0.325***
	(0.171)	(0.122)	(0.065)	(0.084)	(0.067)
Hours worked by owner: primary business (log)	0.088	0.137*	0.147***	0.113*	0.144***
	(0.167)	(0.065)	(0.034)	(0.044)	(0.030)
Number of paid workers: primary business	0.321***	0.179***	0.058***	0.353***	0.344***
	(0.073)	(0.041)	(0.015)	(0.054)	(0.087)
Number of unpaid workers: primary business	0.003	0.004	0.005	-0.002	0.001
	(0.037)	(0.024)	(0.012)	(0.020)	(0.018)
Total value of business capital (log)	0.424***	0.355***	0.208***	0.205***	0.159***
	(0.032)	(0.020)	(0.010)	(0.011)	(0.011)
Business practices followed (PM scale)	0.217***	0.202***	0.112***	0.136***	0.107***
	(0.057)	(0.030)	(0.015)	(0.021)	(0.018)
N	4559	4559	4556	4559	4559

* p<0.05, ** p<0.01, *** p<0.001

Notes: Models also include (results not shown) 5 dummy variables for the type of business and 4 dummy variables for regency of residence. The estimated standard errors in columns 1-3 are adjusted for clustered sampling, whereas the estimated standard errors in columns 4 and 5 are bootstrapped with 100 repetitions by cluster.

Table 2. Multiple linear regression analysis of primary business profits (reported) using alternative statistical models (MS data unless otherwise indicated)

	Alternative statistical models				
	OLS regression model	OLS winsorized regression model	Log regression model	Quantile (median) regression model	Robust regression model
	(1)	(2)	(3)	(4)	(5)
Owner's age (BS)	0.107*	0.086*	0.046*	0.016	0.014
	(0.046)	(0.033)	(0.018)	(0.022)	(0.021)
Owner's age squared (BS)	-0.001	-0.001*	-0.001*	-0.000	-0.000
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Woman business owner (BS)	-0.213	-0.350***	-0.276***	-0.369***	-0.387***
	(0.157)	(0.099)	(0.049)	(0.075)	(0.061)
Owner's Schooling: Lower secondary (BS)	0.104	0.046	0.010	0.008	0.012
	(0.139)	(0.104)	(0.050)	(0.070)	(0.051)
Owner's Schooling: Upper secondary (BS)	0.167	0.021	-0.001	0.019	-0.019
	(0.162)	(0.104)	(0.056)	(0.079)	(0.074)
Owner's Schooling: Tertiary (BS)	0.118	0.108	-0.078	0.021	-0.168
	(0.340)	(0.277)	(0.122)	(0.273)	(0.213)
Owner's cognitive ability (BS, IRT scale)	-0.003	0.028	0.024	0.023	0.034
	(0.053)	(0.039)	(0.020)	(0.030)	(0.026)
Owner's willingness to take risks (BS, IRT scale)	0.163*	0.110*	0.038	0.035	0.040
	(0.071)	(0.045)	(0.023)	(0.036)	(0.029)
Days worked by owner: primary business (log)	0.277	0.167	0.179*	0.225*	0.239***
	(0.162)	(0.141)	(0.069)	(0.093)	(0.070)
Hours worked by owner: primary business (log)	0.411***	0.300***	0.229***	0.198***	0.161***
	(0.101)	(0.078)	(0.042)	(0.049)	(0.042)
Number of paid workers: primary business	0.480*	0.313*	0.100*	0.654***	0.371***
	(0.188)	(0.120)	(0.038)	(0.106)	(0.111)
Number of unpaid workers: primary business	0.005	0.013	0.002	0.024	0.002
	(0.056)	(0.042)	(0.021)	(0.034)	(0.025)
Total value of business capital (log)	0.428***	0.385***	0.220***	0.213***	0.190***
	(0.056)	(0.038)	(0.016)	(0.020)	(0.019)
Business practices followed (BS, IRT scale)	0.126	0.103*	0.044*	0.046	0.048
	(0.069)	(0.049)	(0.020)	(0.029)	(0.028)
Personal agency (IRT scale)	0.145*	0.127**	0.086***	0.081**	0.091***
	(0.058)	(0.042)	(0.020)	(0.026)	(0.021)
N	2230	2230	2228	2230	2229

* p<0.05, ** p<0.01, *** p<0.001

Notes: Models also include (results not shown) 5 dummy variables for the type of business and 2 dummy variables for regency of residence (2 of the 5 regencies are not included in the MS sample). The estimated standard errors in columns 1-3 are adjusted for clustered sampling, whereas the estimated standard errors in columns 4 and 5 are bootstrapped with 100 repetitions by cluster.

Table 3. Multiple linear regression analysis of reported primary business profits (calculated) using alternative statistical models (MS data unless otherwise specified)

	Alternative statistical models				
	OLS regression model	OLS winsorized regression model	Log regression model	Quantile (median) regression model	Robust regression model
	(1)	(2)	(3)	(4)	(5)
Owner's age (BS)	0.154	0.170	0.036	0.063	0.068
	(0.279)	(0.125)	(0.025)	(0.047)	(0.044)
Owner's age squared (BS)	-0.002	-0.002	-0.000	-0.001	-0.001
	(0.004)	(0.002)	(0.000)	(0.001)	(0.001)
Woman business owner	-0.288	0.180	-0.117*	-0.341**	-0.330***
	(0.683)	(0.411)	(0.059)	(0.122)	(0.098)
Owner's Schooling: Lower secondary (BS)	1.076	0.218	0.024	0.051	0.079
	(1.066)	(0.496)	(0.065)	(0.120)	(0.095)
Owner's Schooling: Upper secondary (BS)	-0.169	0.085	0.002	0.009	0.000
	(1.055)	(0.660)	(0.076)	(0.137)	(0.111)
Owner's Schooling: Tertiary (BS)	-1.610	-0.056	-0.206	-0.104	-0.056
	(1.333)	(0.870)	(0.158)	(0.338)	(0.269)
Owner's cognitive ability (BS, IRT scale)	-0.059	-0.350*	-0.025	0.049	0.115*
	(0.376)	(0.140)	(0.024)	(0.059)	(0.046)
Owner's willingness to take risks (BS, IRT scale)	0.928	0.254	0.056	0.098	0.113**
	(0.740)	(0.196)	(0.030)	(0.055)	(0.043)
Days worked by owner: primary business (log)	0.991	-0.212	0.085	0.369*	0.318**
	(0.949)	(0.555)	(0.069)	(0.154)	(0.122)
Hours worked by owner: primary business (log)	0.364	0.657*	0.134**	0.211*	0.271**
	(0.682)	(0.272)	(0.047)	(0.096)	(0.088)
Number of paid workers: primary business	2.875	-0.191	0.061	0.715***	0.340*
	(1.917)	(0.744)	(0.034)	(0.154)	(0.161)
Number of unpaid workers: primary business	0.677**	0.480**	0.099***	0.106	0.085
	(0.247)	(0.155)	(0.028)	(0.065)	(0.047)
Total value of business capital (log)	1.276*	0.927***	0.200***	0.379***	0.283***
	(0.557)	(0.180)	(0.019)	(0.049)	(0.034)
Business practices followed (IRT scale)	-1.173	-0.073	-0.004	0.023	0.074
	(0.758)	(0.220)	(0.032)	(0.066)	(0.046)
Individual agency (IRT scale)	0.662*	0.575**	0.073**	0.089	0.071
	(0.307)	(0.206)	(0.026)	(0.055)	(0.042)
N	2212	2212	2212	2212	2211

* p<0.05, ** p<0.01, *** p<0.001

Notes: Models also include (results not shown) 5 dummy variables for the type of business and 2 dummy variables for regency of residence (2 of the 5 regencies are not included in the MS sample). The estimated standard errors in columns 1-3 are adjusted for clustered sampling, whereas the estimated standard errors in columns 4 and 5 are bootstrapped with 100 repetitions by cluster.

Table 4. Estimates of the Impact of business training on the primary profits of women business owners using alternative statistical models (ES and BS data)

	Alternative statistical models				
	OLS regression model	OLS winsorized regression model	Log regression model	Quantile (median) regression model	Robust regression model
	(1)	(2)	(3)	(4)	(5)
UNADJUSTED					
Received business training (BS)	0.136*	0.110*	0.110***	0.100	0.061*
	(0.081)	(0.059)	(0.039)	(0.072)	(0.035)
N	2657	2657	2657	2657	2657
ADJUSTED					
Received business training (BS)	0.088	0.069	0.078**	0.042	0.032
	(0.078)	(0.056)	(0.035)	(0.032)	(0.030)
Primary profits (BS: log)	0.681***	0.574***	0.436***	0.373***	0.304***
	(0.071)	(0.036)	(0.021)	(0.021)	(0.023)
N	2644	2644	2644	2644	2644

* p<0.10, ** p<0.05, *** p<0.01 (impact estimates only)

Notes: Models also include (results not shown) 5 dummy variables for the type of business and 4 dummy variables for regency of residence. The estimated standard errors in columns 1-3 are adjusted for clustered sampling, whereas the estimated standard errors in columns 4 and 5 are bootstrapped with 100 repetitions by cluster.

Table 5. Multiple linear regression analysis of primary business revenue using alternative statistical models (MS data unless otherwise indicated)

	Alternative statistical models				
	OLS regression model	OLS winsorized regression model	Log regression model	Quantile (median) regression model	Robust regression model
Owner's age	2.122	0.762*	0.093***	0.521**	0.471***
	(1.123)	(0.349)	(0.023)	(0.169)	(0.143)
Owner's age squared	-0.028	-0.008	-0.001***	-0.006**	-0.006**
	(0.015)	(0.005)	(0.000)	(0.002)	(0.002)
Woman business owner	0.686	-1.087	-0.110*	-0.659	-0.365
	(3.820)	(0.858)	(0.055)	(0.484)	(0.350)
Owner's Schooling: Lower secondary	2.311	-0.221	-0.021	-0.226	-0.123
	(4.525)	(0.901)	(0.059)	(0.431)	(0.370)
Owner's Schooling: Upper secondary	-6.210	-1.568	-0.176**	-1.060	-1.078**
	(3.373)	(0.921)	(0.059)	(0.550)	(0.341)
Owner's Schooling: Tertiary	-6.849	-2.301	-0.347*	-1.747	-2.056*
	(5.905)	(2.424)	(0.136)	(1.327)	(0.829)
Owner's cognitive ability (BS, IRT scale)	-0.634	0.077	0.057*	0.218	0.240
	(1.361)	(0.340)	(0.023)	(0.156)	(0.170)
Owner's willingness to take risks (BS, IRT scale)	1.141	0.303	0.030	0.220	0.117
	(2.489)	(0.445)	(0.027)	(0.196)	(0.158)
Days worked by owner: primary business (log)	9.624*	3.133***	0.319***	1.340	1.491***
	(4.165)	(0.827)	(0.076)	(0.684)	(0.381)
Hours worked by owner: primary business (log)	2.797	1.568*	0.266***	1.539***	1.360***
	(3.166)	(0.689)	(0.049)	(0.337)	(0.276)
Number of paid workers: primary business	22.838**	2.952***	0.170***	3.166***	2.136***
	(8.105)	(0.546)	(0.027)	(0.713)	(0.334)
Number of unpaid workers: primary business	1.629	0.961*	0.083***	0.728***	0.467**
	(1.558)	(0.381)	(0.024)	(0.219)	(0.163)
Total value of business capital (log)	3.272	3.596***	0.289***	1.365***	1.012***
	(1.977)	(0.301)	(0.017)	(0.171)	(0.121)
Business practices followed (IRT scale)	-2.700	0.133	0.043	0.159	0.200
	(2.405)	(0.451)	(0.025)	(0.268)	(0.178)
Individual agency (IRT scale)	2.010	1.045**	0.072**	0.321	0.305*
	(1.109)	(0.381)	(0.024)	(0.188)	(0.153)
N	2233	2233	2233	2233	2232

* p<0.05, ** p<0.01, *** p<0.001

Notes: Models also include (results not shown) 5 dummy variables for the type of business and 2 dummy variables for regency of residence (2 of the 5 regencies are not included in the MS sample). The estimated standard errors in columns 1-3 are adjusted for clustered sampling, whereas the estimated standard errors in columns 4 and 5 are bootstrapped with 100 repetitions by cluster.

Table 6. Multiple linear regression analysis of monthly wage and salary earnings of business owners (log) using alternative statistical models and by gender (BS data)

	Full sample					WBOs	MBOs
	Alternative statistical models					Statistical models	
	OLS regression model	OLS winsorized regression model	Log regression model	Quantile (median) regression	Robust regression	Log regression model	Log regression model
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Number of days worked (log)	0.328***	0.286***	0.418***	0.261***	0.226***	0.345***	0.480***
	(0.058)	(0.040)	(0.054)	(0.036)	(0.032)	(0.092)	(0.063)
Number of hours worked (log)	0.481***	0.447***	0.635***	0.312***	0.312***	0.669***	0.593***
	(0.071)	(0.055)	(0.068)	(0.030)	(0.043)	(0.094)	(0.091)
Owner's age	0.104	0.114**	0.112***	0.066*	0.069**	0.052	0.153***
	(0.060)	(0.035)	(0.034)	(0.029)	(0.027)	(0.052)	(0.043)
Owner's age squared	-0.001	-0.001**	-0.001**	-0.001	-0.001*	-0.000	-0.002**
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Woman business owner	-0.568***	-0.530***	-0.651***	-0.544***	-0.441***		
	(0.112)	(0.077)	(0.080)	(0.057)	(0.053)		
Owner's Schooling: Lower secondary	0.146	0.177	0.207	0.031	0.079	0.128	0.340*
	(0.161)	(0.099)	(0.111)	(0.072)	(0.073)	(0.166)	(0.151)
Owner's Schooling: Upper secondary	0.432*	0.354***	0.309**	0.134	0.152*	0.177	0.460**
	(0.210)	(0.100)	(0.111)	(0.080)	(0.073)	(0.171)	(0.140)
Owner's Schooling: Tertiary	0.626**	0.563***	0.501***	0.229	0.221*	0.461**	0.553**
	(0.208)	(0.133)	(0.126)	(0.121)	(0.111)	(0.174)	(0.179)
Owner's cognitive ability (IRT scale)	0.082*	0.063	0.020	0.002	0.013	-0.001	0.036
	(0.041)	(0.037)	(0.037)	(0.035)	(0.024)	(0.053)	(0.050)
N	646	646	646	646	646	295	351

* p<0.05, ** p<0.01, *** p<0.001

Notes: Models also include (results not shown) 4 dummy variables for regency of residence. The estimated standard errors in columns 1-3 and 6-7 are adjusted for clustered sampling, whereas those in columns 4-5 are bootstrapped with 100 repetitions by cluster.

Table 7. Multiple linear regression analysis of HH income (MS data unless otherwise indicated)

Estimation samples→	Total sample		Women business owners	Men business owners
Dependent variables→	Reported HH income (log)	Predicted HH income (log)	Predicted HH income (log)	Predicted HH income (log)
	(1)	(2)	(3)	(4)
Owner's age (BS)	0.010	0.020	0.024	0.018
	(0.018)	(0.016)	(0.019)	(0.024)
Owner's age squared (BS)	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Woman business owner	0.169***	0.113***		
	(0.034)	(0.031)		
Owner's Schooling: Lower secondary (BS)	0.037	0.067	0.042	0.101
	(0.049)	(0.039)	(0.045)	(0.064)
Owner's Schooling: Upper secondary (BS)	0.165**	0.190***	0.159**	0.222***
	(0.052)	(0.042)	(0.057)	(0.058)
Owner's Schooling: Tertiary (BS)	0.543***	0.525***	0.450***	0.600***
	(0.079)	(0.073)	(0.109)	(0.105)
Owner's cognitive ability (BS, IRT scale)	0.023	0.025	0.051**	-0.011
	(0.018)	(0.015)	(0.020)	(0.021)
Household size (BS)	0.022	0.014	0.006	0.027
	(0.012)	(0.010)	(0.010)	(0.014)
Primary and second profits (log)	0.504***	0.461***	0.414***	0.531***
Constant	(0.021)	(0.018)	(0.023)	(0.025)
R-squared	0.35	0.40	0.34	0.48
N	2192	2208	1280	928

* p<0.05, ** p<0.01, *** p<0.001

Notes: Variables are from the MS if not otherwise indicated. All models also include (results not shown) 2 dummy variables for regency of residence (2 regencies are not included in the MS). The estimated standard errors in columns 1-4 are adjusted for clustered sampling

Table 8. Analysis of the ownership of 20 durable assets by item (MS data)

	Sample means (standard deviation)		Estimated coefficients from linear regressions of ownership on the log of HH income (t-statistics)		Principal components analysis: Loadings of items on first principal component	
Item	Item owned 0-1)	Number of items owned (N)	Item owned (0-1)	Number of items owned (N)	Item owned (0-1)	Number of items owned (N)
	(1)	(2)	(3)	(4)	(5)	(6)
TV	0.984	1.338	0.009	0.159	0.1427	0.3258
	(0.127)	(0.625)	(3.04)**	(10.15)***		
DVD/VCD	0.612	0.838	0.072	0.166	0.2163	0.1964
	(0.487)	(0.919)	(7.09)***	(8.12)***		
Satellite dish	0.196	0.205	0.075	0.078	0.2516	0.1744
	(0.397)	(0.436)	(8.15)***	(7.60)***		
Microwave	0.035	0.037	0.023	0.024	0.1986	0.1572
	(0.185)	(0.198)	(5.09)***	(4.46)***		
Refrigerator	0.696	0.748	0.080	0.112	0.2917	0.2595
	(0.460)	(0.554)	(7.70)***	(8.31)***		
Gas cylinder (3 Kg +)	0.958	1.329	0.008	0.132	0.1160	0.2300
	(0.201)	(0.786)	(1.90)	(6.03)***		
Washing machine	0.341	0.349	0.142	0.158	0.3656	0.2983
	(0.474)	(0.498)	(12.86)***	(11.77)***		
Air conditioner	0.017	0.022	0.017	0.024	0.2081	0.1626
	(0.130)	(0.183)	(4.16)***	(3.90)***		
Telephone	0.003	0.003	0.003	0.003	0.0909	0.0980
	(0.051)	(0.051)	(1.81)	(1.81)		
Simple hand phone	0.843	1.335	-0.010	0.122	-0.0768	0.1011
	(0.364)	(0.998)	(-1.21)	(4.60)***		
Smart hand phone	0.838	1.457	0.084	0.392	0.2758	0.3511
	(0.369)	(1.084)	(10.18)***	(15.17)***		
Computer / laptop	0.255	0.302	0.112	0.153	0.3559	0.3244
	(0.436)	(0.562)	(11.65)***	(11.56)***		
Tablet	0.203	0.214	0.073	0.078	0.1881	0.1694
	(0.403)	(0.435)	(7.79)***	(7.58)***		
Camcorder / camera	0.056	0.061	0.033	0.037	0.2643	0.2009
	(0.231)	(0.265)	(6.18)***	(5.52)***		
Water heater	0.011	0.011	0.007	0.007	0.1176	0.0906
	(0.105)	(0.105)	(2.01)*	(2.01)*		
Electric pump / jet pump	0.694	0.722	0.052	0.072	0.1492	0.1484
	(0.461)	(0.510)	(4.11)***	(5.14)***		
Generator	0.061	0.062	0.030	0.031	0.2164	0.1767
	(0.239)	(0.247)	(4.45)***	(4.52)***		
Car	0.141	0.166	0.106	0.128	0.3512	0.2961
	(0.348)	(0.452)	(11.90)***	(10.63)***		
Boat / motor boat	0.004	0.005	0.005	0.006	0.0493	0.029
	(0.066)	(0.083)	(2.39)*	(2.20)*		
Motorbike / motorcycle	0.924	1.711	0.019	0.261	0.1542	0.3117
	(0.265)	(1.008)	(2.76)**	(10.19)***		
N	2261	2261	2261	2261	2261	2261

* p<0.05, ** p<0.01, *** p<0.001

Notes: All data are from the MS. Estimated standard errors in the regressions reported in columns 3-4 are adjusted for clustered sampling.

Table 9. Simple linear regression analysis of alternative asset indexes (composite variables) on the log of predicted HH income by alternative scales (MS data)

Dependent variables: Asset indexes based on ownership alone (0-1) using alternative scales				
	Alternative scales			
	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale
	(1)	(2)	(3)	(4)
Predicted HH income (log)	0.457***	0.463***	0.441***	0.465***
	(0.021)	(0.023)	(0.024)	(0.021)
R-squared	0.18	0.18	0.17	0.18
N	2261	2261	2261	2261
Dependent variables: Asset indexes based on the number of items owned using alternative scales				
	Alternative scales			
	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale
	(1)	(2)	(3)	(4)
Predicted HH income (log)	0.475***	0.486***	0.476***	0.494***
	(0.026)	(0.027)	(0.027)	(0.022)
R-squared	0.19	0.20	0.20	0.21
N	2261	2261	2261	2261

* p<0.05, ** p<0.01, *** p<0.001

Notes: Estimated standard errors are adjusted for clustered sampling

Table 10. Impact of business training on the asset indexes (composite variable) of women business owners by alternative scales (ES and MS data)

	Alternative scales			
Dependent variables→	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale
	(1)	(2)	(3)	(4)
UNADJUSTED				
Received business training (BS)	0.045	0.030	0.042	0.033
	(0.035)	(0.036)	(0.035)	(0.036)
R-squared	0.00	0.00	0.00	0.00
N	2724	2724	2724	2724
ADJUSTED				
	Alternative scales			
Dependent variables→	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale
	(1)	(2)	(3)	(4)
Received business training (BS)	0.059**	0.051**	0.063**	0.047*
	(0.026)	(0.025)	(0.026)	(0.026)
Baseline HH asset index (BS)	0.720***	0.736***	0.691***	0.734***
	(0.016)	(0.018)	(0.019)	(0.015)
R-squared	0.52	0.54	0.49	0.54
N	2724	2724	2724	2724

* p<0.10, ** p<0.05, *** p<0.01 (impact estimates only)

Notes: Estimated standard errors are adjusted for clustered sampling

Table 11. Multiple regression analysis of total current savings balances by alternative statistical models (ES data unless otherwise indicated)

	Alternative statistical models					
	OLS regression model	OLS winsorized regression model	Tobit model	Two-part model	Quantile (median) regression model	Robust regression model
	(1)	(2)	(3)	(4)	(5)	(6)
Owner's age (BS)	0.343	0.178	0.550	-0.012	-0.008	-0.002
	(0.268)	(0.144)	(0.301)	(0.241)	(0.058)	(0.050)
Owner's age squared (BS)	-0.005	-0.002	-0.008	-0.000	0.000	-0.000
	(0.004)	(0.002)	(0.004)	(0.003)	(0.001)	(0.001)
Woman business owner	0.592	0.449	2.240**	3.370***	0.628***	0.794***
	(0.574)	(0.364)	(0.788)	(0.572)	(0.139)	(0.106)
Owner's Schooling: Lower secondary (BS)	0.434	0.172	1.172	1.715**	0.148	0.195
	(0.958)	(0.389)	(1.172)	(0.638)	(0.176)	(0.134)
Owner's Schooling: Upper secondary (BS)	0.212	0.550	1.349	2.887***	0.462**	0.517***
	(0.862)	(0.414)	(1.029)	(0.656)	(0.177)	(0.135)
Owner's Schooling: Tertiary (BS)	3.533*	3.438**	4.849**	4.391***	1.422	0.598*
	(1.782)	(1.045)	(1.880)	(1.273)	(0.823)	(0.247)
Owner's cognitive ability (BS, IRT scale)	-0.174	0.156	-0.089	0.591*	0.086	0.131**
	(0.412)	(0.165)	(0.422)	(0.252)	(0.068)	(0.050)
Primary/second profits (log)	3.810***	2.752***	4.376***	4.876***	1.125***	0.777***
	(0.765)	(0.184)	(0.754)	(0.341)	(0.074)	(0.050)
HH asset index (PC scale)	4.750***	3.188***	5.210***	4.678***	1.569***	0.807***
	(0.558)	(0.237)	(0.610)	(0.366)	(0.103)	(0.053)
N	4525	4525	4525	4525	4525	4525

* p<0.05, ** p<0.01, *** p<0.001

Notes: Data are from the Endline Survey (ES) unless otherwise indicated. Estimated standard errors in columns 1-4 are adjusted for clustered sampling, whereas those in columns 5-6 are bootstrapped with 100 repetitions by cluster.

Table 12. Multiple regression analysis of net financial asset balances by alternative statistical models (MS data unless otherwise indicated)

	Alternative statistical models				
	OLS regression model	OLS winsorized regression model	Inverse hyperbolic sine (IHS)-transformed dependent variable	Quantile (median) regression model	Robust regression model
	(1)	(2)	(3)	(4)	(5)
Owner's age (BS)	0.771	0.154	-0.064	0.177	-0.210
	(1.921)	(0.566)	(0.053)	(0.454)	(0.195)
Owner's age squared (BS)	0.001	0.000	0.001	-0.001	0.002
	(0.026)	(0.008)	(0.001)	(0.006)	(0.003)
Woman business owner	7.311	2.646	0.487***	2.318*	1.662***
	(6.276)	(1.452)	(0.127)	(1.147)	(0.402)
Owner's Schooling: Lower secondary (BS)	7.143	1.671	0.270	0.456	0.721
	(6.236)	(1.561)	(0.149)	(1.165)	(0.459)
Owner's Schooling: Upper secondary (BS)	1.001	1.473	0.202	-0.185	0.218
	(6.523)	(1.836)	(0.168)	(1.344)	(0.519)
Owner's Schooling: Tertiary (BS)	40.422	1.475	0.318	7.426	1.294
	(32.483)	(5.261)	(0.436)	(9.895)	(1.820)
Owner's cognitive ability (BS: IRT scale)	0.131	0.205	0.019	0.431	0.085
	(2.506)	(0.715)	(0.066)	(0.445)	(0.236)
Number of banks with accounts	1.677	-3.562*	-0.869***	0.124	-2.127***
	(8.605)	(1.598)	(0.128)	(1.163)	(0.418)
Primary and second profits (log)	2.211	2.118*	0.068	2.273***	0.497*
	(3.029)	(0.826)	(0.068)	(0.507)	(0.209)
Total value of business capital (log)	2.312	1.626***	0.116**	1.953***	0.375*
	(1.804)	(0.481)	(0.043)	(0.388)	(0.148)
HH asset index (PC scale)	8.962	3.082***	0.209**	4.313***	0.883*
	(5.736)	(0.885)	(0.080)	(0.753)	(0.359)
Uses bank account to save for emergencies	7.750	5.660**	0.660***	5.662	2.184**
	(7.443)	(2.143)	(0.169)	(2.922)	(0.800)
Saves for emergencies by purchasing non-financial assets	19.443*	7.059***	0.707***	6.277*	1.258*
	(8.096)	(1.996)	(0.164)	(2.476)	(0.506)
Borrows for emergencies	-10.552*	-5.613***	-1.149***	-4.049***	-2.963***
	(4.148)	(1.446)	(0.148)	(1.196)	(0.513)
Saves to pay off debt	-12.871**	-5.237*	-0.910**	-4.789*	-3.026**
	(4.302)	(2.461)	(0.276)	(2.211)	(1.107)
HH income instability[a]	7.715*	2.847**	0.213**	1.332	0.210
	(3.435)	(0.958)	(0.080)	(1.080)	(0.239)
Has wage and salary employment (BS)	6.342	2.547	0.288	0.722	0.598
	(11.289)	(2.287)	(0.190)	(1.870)	(0.588)
N	2135	2135	2135	2135	2135

* p<0.05, ** p<0.01, *** p<0.001

Notes: Estimated standard errors are adjusted for clustered sampling in columns 1-3, whereas they are bootstrapped with 100 repetitions by cluster in columns 4-5.

[a] This variable is defined as the difference between the high-end and low-end estimates of HH income as a proportion of HH income in the last month.

Table 13. Multiple regression analysis of the total value of business capital using alternative statistical models and by gender (BS data)

	Full sample					WBOs	MBOs
	Alternative statistical models					Statistical models	
	OLS regression model	OLS winsorized regression model	Log regression model	Quantile (median) regression model	Robust regression model	Log regression model	Log regression model
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Owner currently married	-7.633	-5.211	-0.232*	-1.788*	-0.875	-0.318*	-0.346*
	(6.914)	(3.489)	(0.093)	(0.802)	(0.461)	(0.148)	(0.148)
Owner's age	2.498	0.489	-0.008	-0.122	-0.196	-0.007	0.029
	(1.624)	(0.849)	(0.025)	(0.202)	(0.124)	(0.034)	(0.039)
Owner's age squared	-0.028	-0.004	0.000	0.002	0.003	0.000	-0.000
	(0.022)	(0.011)	(0.000)	(0.003)	(0.002)	(0.000)	(0.000)
Woman business owner	-15.658***	-12.270***	-0.566***	-3.777***	-2.279***		
	(4.660)	(2.801)	(0.094)	(0.800)	(0.470)		
Owner's Schooling: Lower secondary	1.240	1.692	0.181**	0.917*	0.757*	0.107	0.285**
	(3.799)	(2.144)	(0.065)	(0.453)	(0.308)	(0.079)	(0.108)
Owner's Schooling: Upper secondary	4.165	4.281	0.301***	1.719***	1.707***	0.221*	0.394***
	(4.542)	(2.465)	(0.066)	(0.419)	(0.312)	(0.086)	(0.100)
Owner's Schooling: Tertiary	38.729*	18.481**	0.746***	8.896***	4.424***	0.881***	0.599**
	(16.126)	(5.815)	(0.119)	(1.681)	(0.577)	(0.153)	(0.187)
Owner's cognitive ability (IRT scale)	3.132*	1.725*	0.072**	0.316*	0.214	0.027	0.131***
	(1.467)	(0.747)	(0.024)	(0.160)	(0.115)	(0.029)	(0.037)
Household size	0.139	-0.585	-0.011	-0.192	-0.133	-0.024	0.020
	(1.376)	(0.591)	(0.019)	(0.147)	(0.081)	(0.024)	(0.031)
Number of children in HH	-3.599	-1.362	-0.078*	-0.463	-0.340*	-0.055	-0.134**
	(1.920)	(1.101)	(0.031)	(0.276)	(0.155)	(0.040)	(0.051)
Spouse of HH head	-7.560	-1.703	0.114	0.171	0.665	0.225	
	(5.490)	(3.244)	(0.101)	(0.800)	(0.503)	(0.156)	
Child of HH head	-23.080***	-10.929***	-0.164	-1.940*	-0.019	-0.067	-0.186
	(5.471)	(3.281)	(0.098)	(0.916)	(0.479)	(0.162)	(0.131)
Total profits (log)	16.437***	11.367***	0.447***	2.510***	1.511***	0.414***	0.502***
	(2.063)	(0.972)	(0.023)	(0.207)	(0.111)	(0.029)	(0.036)
Wage and salary earnings	4.322	2.379	0.015	1.027	0.036	-0.055	0.032
	(4.910)	(1.532)	(0.025)	(0.700)	(0.154)	(0.075)	(0.025)
HH asset index (PC scale)	29.888***	19.105***	0.547***	4.622***	1.960***	0.507***	0.590***
	(3.526)	(1.458)	(0.026)	(0.413)	(0.124)	(0.032)	(0.039)
N	4552	4552	4551	4552	4552	2674	1877

* p<0.05, ** p<0.01, *** p<0.001

Notes: Data are from the Baseline Survey (BS) unless otherwise indicated. Estimated standard errors in columns 1-3 and 6-7 are adjusted for clustered sampling, whereas those in columns 4-5 are bootstrapped with 100 repetitions by cluster.

Table 14. Impact of business training on the total value of business capital of women business owners using alternative statistical models (ES and BS data)

	Alternative statistical models				
Statistical models→	OLS regression model	OLS winsorized regression model	Log regression model	Quantile (median) regression model	Robust regression model
	(1)	(2)	(3)	(4)	(5)
Unadjusted					
Received business training (BS)	4.180	1.396	0.081	0.750	0.720**
	(3.590)	(2.394)	(0.063)	(0.537)	(0.333)
N	2597	2597	2597	2597	2597
Adjusted					
Received business training (BS)	6.145*	2.654	0.120**	0.556	0.789**
	(3.452)	(2.236)	(0.052)	(0.498)	(0.347)
Baseline value of business capital (log, BS)	19.430***	15.893***	0.575***	4.432***	2.474***
	(2.610)	(1.264)	(0.018)	(0.273)	(0.156)
N	2457	2457	2457	2457	2456

* p<0.10, ** p<0.05, *** p<0.01 (impact estimates only)

Notes: The estimated standard errors in columns 1-3 are adjusted for clustered sampling, whereas the estimated standard errors in columns 4 and 5 are bootstrapped with 100 repetitions by cluster.

Table 15. Multiple regression analysis of income instability using alternative statistical models (MS data unless otherwise indicated)

	Alternative statistical models				
	OLS regression model	OLS regression model (winsorized)	Log regression model	Quantile (median) regression model	Robust regression model
	(1)	(2)	(3)	(4)	(5)
Owner's age (BS)	0.119	0.028	-0.003	-0.002	-0.009
	(0.083)	(0.039)	(0.018)	(0.015)	(0.012)
Owner's age squared (BS)	-0.001	-0.000	0.000	0.000	0.000
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
Woman business owner	-0.542*	-0.306***	-0.136***	-0.028	-0.018
	(0.215)	(0.089)	(0.039)	(0.030)	(0.021)
Owner's Schooling: Lower secondary (BS)	0.160	0.074	-0.000	-0.009	-0.039
	(0.311)	(0.140)	(0.054)	(0.035)	(0.025)
Owner's Schooling: Upper secondary (BS)	-0.160	-0.103	-0.051	-0.013	-0.038
	(0.256)	(0.124)	(0.050)	(0.038)	(0.029)
Owner's Schooling: Tertiary (BS)	-0.658*	-0.293	-0.049	0.076	-0.006
	(0.331)	(0.227)	(0.096)	(0.082)	(0.056)
Owner's cognitive ability (BS, IRT scale)	-0.130	-0.018	0.003	0.007	0.011
	(0.107)	(0.047)	(0.021)	(0.015)	(0.013)
Owner currently married (BS)	0.337	0.096	0.066	0.035	0.038
	(0.189)	(0.145)	(0.065)	(0.054)	(0.038)
Household size (BS)	-0.006	-0.013	-0.006	-0.001	-0.002
	(0.054)	(0.025)	(0.011)	(0.008)	(0.007)
Number of children in HH (BS)	-0.186	-0.057	-0.010	0.009	0.013
	(0.152)	(0.056)	(0.025)	(0.020)	(0.017)
Owner's willingness to take risks (BS, IRT scale)	0.086	-0.025	0.000	0.017	0.012
	(0.164)	(0.059)	(0.024)	(0.017)	(0.011)
Number of earnings sources (BS)	0.514*	0.295***	0.138***	0.106***	0.067**
	(0.225)	(0.088)	(0.036)	(0.026)	(0.023)
Owner has wage employment (BS)	-0.858*	-0.422*	-0.211**	-0.161**	-0.067
	(0.396)	(0.199)	(0.077)	(0.057)	(0.044)
Primary and second profits (log)	-0.904***	-0.358***	-0.062*	0.016	0.061***
	(0.219)	(0.079)	(0.028)	(0.019)	(0.013)
HH income (predicted, log)	1.132***	0.485***	0.057	-0.059*	-0.114***
	(0.263)	(0.096)	(0.035)	(0.027)	(0.018)
N	2171	2171	2167	2171	2171

* p<0.05, ** p<0.01, *** p<0.001

Notes: Estimated standard errors in columns 1-3 are adjusted for clustered sampling, whereas those in columns 4-5 are bootstrapped with 100 repetitions by cluster.

Table 16. Multiple regression analysis (OLS models) of concern about food insecurity by gender (BS data)

	Full sample	Women business owners	Men business owners
	(1)	(2)	(3)
Owner's age	-0.004	-0.002	-0.006
	(0.004)	(0.005)	(0.007)
Owner's age squared	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
Woman business owner	-0.011		
	(0.008)		
Owner's Schooling: Lower secondary	-0.032**	-0.040**	-0.021
	(0.012)	(0.015)	(0.020)
Owner's Schooling: Upper secondary	-0.037**	-0.041**	-0.031
	(0.012)	(0.016)	(0.017)
Owner's Schooling: Tertiary	-0.051**	-0.046*	-0.060**
	(0.016)	(0.022)	(0.020)
Owner's cognitive ability (IRT scale)	-0.009*	-0.004	-0.016*
	(0.004)	(0.005)	(0.006)
Household size	0.003	0.001	0.010*
	(0.004)	(0.005)	(0.005)
Number of children in HH	0.017**	0.016*	0.016
	(0.006)	(0.008)	(0.009)
Total earnings from all sources (log)	-0.014***	-0.013**	-0.018*
	(0.004)	(0.005)	(0.007)
Number of earnings sources	-0.001	-0.003	0.002
	(0.006)	(0.009)	(0.009)
Owner has wage employment	0.020	0.014	0.022
	(0.013)	(0.019)	(0.018)
HH asset index (PC scale)	-0.032***	-0.034***	-0.029***
	(0.004)	(0.005)	(0.006)
R-squared	0.027	0.031	0.027
N	4780	2827	1953

* p<0.05, ** p<0.01, *** p<0.001

Notes: Estimated standard errors are adjusted for clustered sampling.

Table 17. Multiple regression analysis of subjective well-being (composite variable) among all business owners by alternative scales and by gender (MS data unless otherwise indicated)

	Alternative scales				IRT scale	
	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale	Women business owners	Men business owners
	(1)	(2)	(3)	(4)	(5)	(6)
Owner's age (BS)	-0.020	-0.020	-0.020	-0.012	-0.003	-0.008
	(0.022)	(0.022)	(0.022)	(0.021)	(0.029)	(0.035)
Owner's age squared (BS)	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Woman business owner	0.054	0.054	0.054	0.047		
	(0.048)	(0.048)	(0.048)	(0.045)		
Owner's Schooling: Lower secondary (BS)	-0.081	-0.079	-0.079	-0.109	-0.100	-0.148
	(0.057)	(0.057)	(0.057)	(0.058)	(0.073)	(0.088)
Owner's Schooling: Upper secondary (BS)	-0.234***	-0.230***	-0.230***	-0.288***	-0.291***	-0.297***
	(0.056)	(0.056)	(0.056)	(0.058)	(0.073)	(0.083)
Owner's Schooling: Tertiary (BS)	-0.375**	-0.368**	-0.368**	-0.411***	-0.365*	-0.456**
	(0.113)	(0.112)	(0.112)	(0.111)	(0.153)	(0.173)
Owner's cognitive ability (BS, IRT scale)	-0.030	-0.030	-0.030	-0.042	-0.018	-0.069*
	(0.021)	(0.021)	(0.021)	(0.022)	(0.030)	(0.032)
Owner currently married (BS)	0.207*	0.210*	0.210*	0.168*	0.178	0.172
	(0.082)	(0.082)	(0.082)	(0.080)	(0.107)	(0.117)
Number of children in HH	-0.001	0.001	0.001	-0.002	0.010	-0.025
	(0.030)	(0.030)	(0.030)	(0.029)	(0.035)	(0.047)
Household size (BS)	-0.001	-0.001	-0.001	-0.001	-0.007	0.008
	(0.014)	(0.014)	(0.014)	(0.012)	(0.013)	(0.025)
Primary and second profits (log)	0.077**	0.074**	0.074**	0.100***	0.117***	0.077*
	(0.025)	(0.025)	(0.025)	(0.024)	(0.031)	(0.038)
HH income (log)	0.068*	0.069*	0.069*	0.072*	0.075*	0.066
	(0.029)	(0.029)	(0.029)	(0.030)	(0.038)	(0.045)
Net assets (Rp. millions)	-0.000	-0.000	-0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Days worked in primary business	0.010*	0.010*	0.010*	0.010*	0.012*	0.007
	(0.004)	(0.004)	(0.004)	(0.004)	(0.006)	(0.006)
Hours worked in primary business	-0.009	-0.009	-0.009	-0.010	-0.011	-0.007
	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)	(0.011)
HH agency (PM scale)	-0.038	-0.038	-0.038	-0.031	-0.007	-0.065
	(0.021)	(0.021)	(0.021)	(0.022)	(0.027)	(0.037)
Personal agency (PM scale)	0.142***	0.143***	0.143***	0.151***	0.132***	0.181***
	(0.020)	(0.020)	(0.020)	(0.020)	(0.027)	(0.033)
Connectedness (PC scale)	0.009	0.009	0.009	0.005	0.008	0.005
	(0.020)	(0.020)	(0.020)	(0.020)	(0.026)	(0.034)
R-squared	0.06	0.06	0.06	0.07	0.07	0.08
N	2226	2226	2226	2226	1288	938

* p<0.05, ** p<0.01, *** p<0.001

Notes: Estimated standard errors are adjusted for clustered sampling.

Table 18. Multiple regression analysis of subjective well-being (composite variable) among currently married business owners by alternative scales and by gender (MS data unless otherwise indicated)

	Full sample				WBOs	MBOs
	Alternative scales					
	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale	IRT scale	IRT scale
	(1)	(2)	(3)	(4)	(5)	(6)
Owner's age (BS)	-0.008	-0.008	-0.008	0.003	0.003	0.055
	(0.024)	(0.024)	(0.024)	(0.025)	(0.031)	(0.052)
Owner's age squared (BS)	0.000	0.000	0.000	-0.000	-0.000	-0.001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Woman business owner	0.096	0.097	0.097	0.070		
	(0.060)	(0.060)	(0.060)	(0.059)		
Owner's Schooling: Lower secondary (BS)	-0.139*	-0.137*	-0.137*	-0.136*	-0.149	-0.161
	(0.063)	(0.063)	(0.063)	(0.065)	(0.078)	(0.121)
Owner's Schooling: Upper secondary (BS)	-0.306***	-0.302***	-0.302***	-0.349***	-0.301***	-0.454***
	(0.065)	(0.065)	(0.065)	(0.068)	(0.079)	(0.116)
Owner's Schooling: Tertiary (BS)	-0.333**	-0.328**	-0.328**	-0.405**	-0.437*	-0.343
	(0.121)	(0.121)	(0.121)	(0.125)	(0.170)	(0.188)
Owner's cognitive ability (BS: IRT scale)	-0.014	-0.014	-0.014	-0.031	-0.014	-0.065
	(0.026)	(0.026)	(0.026)	(0.028)	(0.033)	(0.044)
Number of children in HH	-0.011	-0.010	-0.010	-0.011	-0.004	-0.071
	(0.033)	(0.033)	(0.033)	(0.033)	(0.038)	(0.063)
Household size (BS)	0.002	0.001	0.001	0.000	-0.009	0.041
	(0.017)	(0.017)	(0.017)	(0.015)	(0.013)	(0.035)
Primary and second profits (log)	0.079**	0.076**	0.076**	0.098***	0.128***	0.036
	(0.028)	(0.028)	(0.028)	(0.028)	(0.034)	(0.048)
HH income (predicted, log)	0.058	0.060	0.060	0.064	0.052	0.076
	(0.032)	(0.032)	(0.032)	(0.033)	(0.040)	(0.058)
Net assets	-0.000	-0.000	-0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Days worked in primary business	0.011*	0.011*	0.011*	0.010*	0.014*	0.004
	(0.005)	(0.005)	(0.005)	(0.004)	(0.006)	(0.007)
Hours worked in primary business	-0.010	-0.010	-0.010	-0.013	-0.016	-0.006
	(0.007)	(0.007)	(0.007)	(0.008)	(0.009)	(0.013)
HH agency (PM scale)	-0.013	-0.013	-0.013	-0.011	0.017	-0.082
	(0.024)	(0.024)	(0.024)	(0.026)	(0.029)	(0.053)
Marital agency (IRT scale)	-0.062*	-0.062*	-0.062*	-0.065*	-0.048	-0.096*
	(0.026)	(0.026)	(0.026)	(0.026)	(0.031)	(0.047)
Personal agency (PM scale)	0.133***	0.133***	0.133***	0.144***	0.115***	0.205***
	(0.022)	(0.022)	(0.022)	(0.022)	(0.029)	(0.038)
Connectedness (PC scale)	-0.012	-0.013	-0.013	-0.010	-0.007	-0.012
	(0.021)	(0.021)	(0.021)	(0.023)	(0.027)	(0.045)
R-squared	0.06	0.06	0.06	0.07	0.06	0.12
N	1635	1635	1635	1635	1108	527

* p<0.05, ** p<0.01, *** p<0.001

Notes: Estimated standard errors are adjusted for clustered sampling.

Table 19. Total number of hours worked by business owners in a typical month using alternative statistical models and by gender (BS data)

	Full sample					WBOs	MBOs
	Alternative statistical models						
	OLS regression model	OLS regression model (winsorized)	Log regression model	Quantile (median) regression model	Robust regression model	OLS regression model	OLS regression model
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Owner's age	-0.787	-0.787	0.002	-0.199	-1.023	-0.791	-2.935
	(1.758)	(1.758)	(0.008)	(2.352)	(2.029)	(2.390)	(2.642)
Owner's age squared	0.000	0.000	-0.000	-0.007	0.004	0.006	0.021
	(0.023)	(0.023)	(0.000)	(0.031)	(0.027)	(0.032)	(0.033)
Woman business owner	21.123***	21.123***	0.081***	34.308***	23.220***		
	(3.385)	(3.385)	(0.015)	(5.115)	(3.502)		
Owner's Schooling: Lower secondary	-9.433*	-9.433*	-0.049**	-13.761	-9.492*	-12.416*	-2.615
	(4.026)	(4.026)	(0.017)	(7.042)	(3.990)	(5.236)	(6.403)
Owner's Schooling: Upper secondary	-8.799	-8.799	-0.046*	-11.469	-8.194	-7.354	-6.464
	(4.606)	(4.606)	(0.020)	(7.071)	(4.409)	(6.108)	(6.310)
Owner's Schooling: Tertiary	2.169	2.169	0.012	1.265	1.730	-12.097	23.125*
	(7.260)	(7.260)	(0.028)	(9.527)	(6.787)	(9.515)	(10.945)
Owner's cognitive ability (IRT scale)	-2.065	-2.065	-0.008	-1.296	-1.972	-2.790	-0.797
	(1.523)	(1.523)	(0.007)	(2.335)	(1.768)	(2.039)	(2.463)
Owner currently married	-10.843	-10.843	-0.045	-16.724*	-11.417*	-10.142	-11.595
	(5.556)	(5.556)	(0.027)	(7.594)	(5.722)	(6.842)	(9.559)
Number of children in HH	-1.420	-1.420	-0.004	-3.882	-1.588	-5.102*	4.745
	(2.024)	(2.024)	(0.009)	(2.848)	(2.019)	(2.578)	(3.240)
Household size	-2.232*	-2.232*	-0.012*	-1.830	-2.278*	-1.852	-3.334
	(0.975)	(0.975)	(0.006)	(1.494)	(1.044)	(1.050)	(1.940)
Number of hours per month volunteered	-0.465**	-0.465**	-0.002*	-0.530	-0.463**	-0.709*	-0.384*
	(0.168)	(0.168)	(0.001)	(0.302)	(0.174)	(0.295)	(0.185)
HH wealth index (PC scale)	-9.386***	-9.386***	-0.044***	-10.159***	-9.762***	-11.799***	-5.745
	(1.827)	(1.827)	(0.009)	(3.007)	(1.837)	(2.382)	(3.004)
Spouse helps with primary business	19.035***	19.035***	0.078***	23.720***	19.425***	18.082***	26.301**
	(3.343)	(3.343)	(0.015)	(5.376)	(3.932)	(4.270)	(8.612)
Son helps with primary business	18.519***	18.519***	0.075***	21.857*	19.535***	24.319***	12.730
	(5.120)	(5.120)	(0.021)	(8.548)	(5.137)	(5.849)	(12.113)
Daughter helps with primary business	18.743***	18.743***	0.079***	15.675*	18.595**	19.775***	24.151
	(5.218)	(5.218)	(0.022)	(7.995)	(6.185)	(5.626)	(13.301)
Other male HH member helps with primary business	14.016*	14.016*	0.051*	14.222	14.750*	6.840	28.662**
	(5.892)	(5.892)	(0.024)	(8.380)	(6.469)	(8.516)	(10.354)
Other female HH member helps with primary business	19.686***	19.686***	0.099***	23.336***	20.111***	26.197***	17.788
	(4.403)	(4.403)	(0.019)	(6.963)	(5.302)	(5.434)	(11.448)
Total earnings (log)	7.085***	7.085***	0.036***	8.327**	7.214***	7.699***	5.480
	(1.683)	(1.683)	(0.008)	(2.718)	(1.884)	(2.106)	(2.900)
Has second business	13.784**	13.784**	0.033	18.608**	15.601**	4.747	26.233***
	(4.985)	(4.985)	(0.017)	(7.040)	(5.039)	(6.632)	(6.982)
Number of different earnings sources	75.299***	75.299***	0.279***	79.770***	77.031***	73.547***	77.041***
	(2.602)	(2.602)	(0.010)	(3.504)	(2.663)	(3.862)	(3.657)
Number of paid workers in primary business	-2.053	-2.053	-0.012	-1.930	-2.369	-4.585	-0.810
	(1.256)	(1.256)	(0.006)	(2.067)	(1.704)	(3.345)	(1.206)
Number of unpaid workers in primary business	0.149	0.149	0.003	0.153	0.176	0.636	-7.735
	(0.797)	(0.797)	(0.004)	(2.467)	(2.208)	(0.872)	(7.023)
Total value of business capital (log)	7.120***	7.120***	0.034***	8.630***	7.658***	11.833***	1.216
	(1.101)	(1.101)	(0.005)	(1.784)	(1.224)	(1.359)	(1.657)
N	4550	4550	4550	4550	4550	2673	1877

* p<0.05, ** p<0.01, *** p<0.001

Notes: Notes: Estimated standard errors in columns 1-3 and 6-7 are adjusted for clustered sampling, whereas those in columns 4-5 are bootstrapped with 100 repetitions by cluster.

Table 20. Estimates of the Impact of business training on the total number of hours worked in a typical month by women business owners using alternative statistical models (ES and BS data)

Statistical models→	OLS regression model	OLS winsorized regression model (highest 5% of values)	Log regression model	Quantile (median) regression model	Robust regression model
	(1)	(2)	(3)	(4)	(5)
UNADJUSTED					
Received business training (BS)	6.780	7.411*	0.034	No estimates obtained	8.195**
	(4.189)	(4.048)	(0.025)		(4.089)
N	2679	2679	2678		2679
ADJUSTED					
Received business training (BS)	5.710	6.380*	0.028	8.182	7.188*
	(3.950)	(3.815)	(0.024)	(5.534)	(4.109)
Total number of hours worked (BS)	0.338***	0.326***	0.002***	0.424***	0.375***
	(0.020)	(0.019)	(0.000)	(0.036)	(0.022)
N	2679	2679	2678	2679	2679

* p<0.10, ** p<0.05 *** p<0.01 (impact estimates only)

Notes: No estimates could be obtained for the unadjusted quantile (median) regression model in column 4. The estimated standard errors in columns 1-3 are adjusted for clustered sampling, whereas the estimated standard errors in columns 4 and 5 are bootstrapped with 100 repetitions by cluster.

Table 21. Multiple regression analysis of personal agency (composite variable) by alternative scales (MS data unless otherwise indicated)

	Alternative scales			
	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale
	(1)	(2)	(3)	(4)
Owner's age (BS)	-0.003	0.002	0.007	0.003
	(0.022)	(0.022)	(0.022)	(0.022)
Owner's age squared (BS)	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Woman business owner	-0.309***	-0.320***	-0.310***	-0.335***
	(0.073)	(0.073)	(0.074)	(0.075)
Owner's Schooling: Lower secondary (BS)	-0.023	-0.010	-0.020	-0.025
	(0.055)	(0.054)	(0.055)	(0.055)
Owner's Schooling: Upper secondary (BS)	0.136*	0.139*	0.122*	0.104
	(0.058)	(0.058)	(0.059)	(0.061)
Owner's Schooling: Tertiary (BS)	0.308**	0.290*	0.266*	0.231
	(0.116)	(0.117)	(0.117)	(0.121)
Owner's cognitive ability (BS, IRT scale)	0.026	0.034	0.028	0.027
	(0.021)	(0.021)	(0.022)	(0.021)
Owner currently married (BS)	-0.169*	-0.167*	-0.166*	-0.186*
	(0.078)	(0.078)	(0.079)	(0.078)
Number of children in HH (BS)	0.008	0.013	0.015	0.004
	(0.032)	(0.032)	(0.032)	(0.031)
Head of household (BS)	0.025	0.002	-0.006	-0.023
	(0.072)	(0.071)	(0.073)	(0.073)
Household size (BS)	-0.001	0.002	-0.005	0.009
	(0.015)	(0.014)	(0.015)	(0.014)
Primary and second profits (log)	0.077**	0.081**	0.074**	0.091***
	(0.024)	(0.025)	(0.025)	(0.025)
HH income (log)	0.079**	0.072**	0.081**	0.063*
	(0.028)	(0.028)	(0.028)	(0.028)
Net financial assets	-0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Any savings in last 3 months	0.082	0.068	0.086	0.040
	(0.051)	(0.049)	(0.051)	(0.048)
Any current bank loans	0.038	0.033	0.034	0.037
	(0.045)	(0.045)	(0.044)	(0.045)
Primary business registered	-0.034	-0.049	-0.039	-0.049
	(0.050)	(0.052)	(0.051)	(0.055)
Business practices followed (PC scale)	0.091***	0.098***	0.098***	0.090***
	(0.021)	(0.021)	(0.021)	(0.021)
R-squared	0.10	0.10	0.09	0.09
N	2226	2226	2226	2226

* p<0.05, ** p<0.01, *** p<0.001

Notes: Estimated standard errors are adjusted for clustered sampling.

Table 22. Multiple regression analysis of intra-HH agency (composite variable) by alternative scales (BS data)

	Alternative scales			
	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale
	(1)	(2)	(3)	(4)
Spouse of HH head	-0.523***	-0.502***	-0.508***	-0.521***
	(0.078)	(0.078)	(0.078)	(0.082)
Child of HH head	-0.848***	-0.810***	-0.821***	-0.661***
	(0.070)	(0.070)	(0.070)	(0.074)
Owner's age	0.091***	0.090***	0.091***	0.080***
	(0.017)	(0.017)	(0.017)	(0.017)
Owner's age squared	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Woman business owner	-0.016	-0.014	-0.011	0.037
	(0.077)	(0.077)	(0.077)	(0.080)
Owner's Schooling: Lower secondary	-0.050	-0.050	-0.049	-0.053
	(0.035)	(0.035)	(0.035)	(0.036)
Owner's Schooling: Upper secondary	-0.064	-0.060	-0.060	-0.093**
	(0.034)	(0.034)	(0.034)	(0.035)
Owner's Schooling: Tertiary	0.014	0.030	0.027	0.048
	(0.066)	(0.065)	(0.065)	(0.067)
Owner's cognitive ability (IRT scale)	0.048***	0.050***	0.049***	0.033*
	(0.013)	(0.013)	(0.013)	(0.013)
Owner currently married	-0.082	-0.116	-0.106	-0.439***
	(0.077)	(0.077)	(0.077)	(0.079)
Number of children in HH	0.032	0.029	0.030	0.019
	(0.019)	(0.019)	(0.019)	(0.018)
Household size	-0.028*	-0.026*	-0.027*	-0.018
	(0.012)	(0.012)	(0.012)	(0.010)
Total earnings (log)	0.047***	0.046**	0.046**	0.049***
	(0.014)	(0.014)	(0.014)	(0.014)
R-squared	0.17	0.15	0.16	0.13
N	4781	4781	4781	4781

* p<0.05, ** p<0.01, *** p<0.001

Notes: Estimated standard errors are adjusted for clustered sampling.

Table 23. Impact of business training on the intra-HH agency (composite variable) of women business owners by alternative scales (MS and BS data)

	Alternative scales			
Dependent variable→	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale
	(1)	(2)	(3)	(4)
UNADJUSTED				
Received business training (BS)	0.206***	0.207***	0.206***	0.206***
	(0.055)	(0.056)	(0.056)	(0.053)
R-squared	0.010	0.010	0.010	0.011
N	1345	1345	1345	1345
ADJUSTED				
	Alternative scales			
Dependent variable→	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale
	(1)	(2)	(3)	(4)
Received business training (BS)	0.445***	0.446***	0.447***	0.491***
	(0.031)	(0.032)	(0.032)	(0.028)
Intra-HH agency (BS, PM scale)	0.194***	0.193***	0.192***	0.181***
	(0.050)	(0.050)	(0.050)	(0.046)
R-squared	0.207	0.209	0.210	0.253
N	1345	1345	1345	1345

* p<0.10, ** p<0.05, *** p<0.01 (impact estimates only)

Notes: Estimated standard errors are adjusted for clustered sampling.

Table 24. Impact of business training on the intra-HH agency (composite variable) of women business owners by alternative scales (ES and BS data)

	Alternative scales			
Dependent variable→	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale
	(1)	(2)	(3)	(4)
UNADJUSTED				
Received business training (BS)	0.027	0.030	0.028	0.027
	(0.039)	(0.039)	(0.039)	(0.038)
R-squared	0.000	0.000	0.000	0.000
N	2724	2724	2724	2724
ADJUSTED				
	Alternative scales			
Dependent variable→	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale
	(1)	(2)	(3)	(4)
Received business training (BS)	0.017	0.020	0.019	0.015
	(0.038)	(0.039)	(0.039)	(0.036)
Intra-HH agency (BS, PM scale)	0.236***	0.242***	0.241***	0.282***
	(0.027)	(0.026)	(0.026)	(0.026)
R-squared	0.055	0.058	0.058	0.078
N	2724	2724	2724	2724

* p<0.10, ** p<0.05, *** p<0.01 (impact estimates only)

Notes: Estimated standard errors are adjusted for clustered sampling.

Table 25. Multiple regression analysis of marital agency (composite variable) among currently married business owners by alternative scales (BS data)

	Alternative scales			
	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale
	(1)	(2)	(3)	(4)
Spouse of HH head	-0.122	-0.123	-0.117	-0.118
	(0.086)	(0.086)	(0.083)	(0.090)
Child of HH head	0.024	0.041	0.037	0.035
	(0.066)	(0.063)	(0.061)	(0.071)
Owner's age	0.011	0.014	0.012	0.012
	(0.015)	(0.015)	(0.015)	(0.015)
Owner's age squared	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Woman business owner	1.029***	1.022***	1.070***	0.941***
	(0.085)	(0.086)	(0.083)	(0.089)
Owner's Schooling: Lower secondary	-0.027	-0.028	-0.015	-0.042
	(0.037)	(0.037)	(0.037)	(0.038)
Owner's Schooling: Upper secondary	-0.005	-0.009	0.007	-0.024
	(0.037)	(0.037)	(0.037)	(0.038)
Owner's Schooling: Tertiary	0.041	0.027	0.041	0.026
	(0.069)	(0.069)	(0.068)	(0.071)
Owner's cognitive ability (BS, IRT)	-0.035**	-0.043**	-0.042**	-0.033*
	(0.013)	(0.013)	(0.013)	(0.014)
Number of children in HH	0.032*	0.033*	0.036*	0.028
	(0.016)	(0.016)	(0.017)	(0.016)
Household size	0.001	-0.005	-0.006	0.003
	(0.007)	(0.007)	(0.007)	(0.007)
Total earnings (log)	-0.047***	-0.053***	-0.057***	-0.040**
	(0.014)	(0.014)	(0.014)	(0.014)
Constant	-0.832**	-0.870**	-0.851**	-0.783**
	(0.284)	(0.282)	(0.281)	(0.286)
R-squared	0.22	0.22	0.25	0.18
N	4337	4337	4337	4337

* p<0.05, ** p<0.01, *** p<0.001

Note: Sample is limited to married business owners with some earned income. Estimated standard errors are adjusted for clustered sampling.

Table 26. Impact of business training on the marital agency (composite variable) of women business owners by alternative scales (MS and BS data)

	Alternative scales			
Dependent variable→	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale
	(1)	(2)	(3)	(4)
UNADJUSTED				
Received business training (BS)	0.166**	0.191**	0.177**	0.168**
	(0.057)	(0.059)	(0.057)	(0.057)
R-squared	0.007	0.009	0.008	0.007
N	1210	1210	1210	1210
ADJUSTED				
	Alternative scales			
Dependent variable→	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale
	(1)	(2)	(3)	(4)
Received business training (BS)	0.137*	0.160**	0.145**	0.146**
	(0.053)	(0.054)	(0.053)	(0.053)
Marital agency (BS, PM scale)	0.383***	0.397***	0.390***	0.375***
	(0.030)	(0.030)	(0.030)	(0.030)
R-squared	0.149	0.161	0.156	0.143
N	1197	1197	1197	1197

* p<0.10, ** p<0.05, *** p<0.01 (impact estimates only)

Notes: Estimated standard errors are adjusted for clustered sampling.

Table 27. Impact of business training on the marital agency (composite variable) of women business owners by alternative scales (ES and BS data)

	Alternative scales			
Dependent variable→	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale
	(1)	(2)	(3)	(4)
UNADJUSTED				
Received business training (BS)	0.052	0.056	0.053	0.059
	(0.041)	(0.041)	(0.041)	(0.041)
R-squared	0.001	0.001	0.001	0.001
N	2440	2440	2440	2440
ADJUSTED				
	Alternative scales			
Dependent variable→	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale
	(1)	(2)	(3)	(4)
Received business training (BS)	0.060	0.066	0.061	0.070
	(0.038)	(0.039)	(0.038)	(0.038)
Marital agency (BS, PM scale)	0.404***	0.401***	0.390***	0.410***
	(0.019)	(0.020)	(0.020)	(0.019)
R-squared	0.158	0.155	0.149	0.161
N	2399	2399	2399	2399

* p<0.10, ** p<0.05, *** p<0.01 (impact estimates only)

Notes: Estimated standard errors are adjusted for clustered sampling.

Table 28. Multiple regression analysis of business practices (composite variable) by alternative scales (BS)

	Alternative scales			
	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale
	(1)	(2)	(3)	(4)
Owner's age	-0.051**	-0.052**	-0.052**	-0.044**
	(0.017)	(0.017)	(0.017)	(0.016)
Owner's age squared	0.000*	0.001*	0.000*	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)
Woman business owner	0.078*	0.078*	0.074*	0.141***
	(0.031)	(0.032)	(0.031)	(0.033)
Owner's Schooling: Lower secondary	0.005	0.005	0.001	0.025
	(0.035)	(0.035)	(0.034)	(0.038)
Owner's Schooling: Upper secondary	0.168***	0.161***	0.173***	0.156***
	(0.037)	(0.038)	(0.037)	(0.039)
Owner's Schooling: Tertiary	0.518***	0.512***	0.539***	0.407***
	(0.085)	(0.087)	(0.087)	(0.080)
Owner's cognitive ability (BS, IRT)	-0.004	-0.007	0.000	-0.009
	(0.015)	(0.015)	(0.015)	(0.015)
Willingness to take risks (proportional mean)	0.287***	0.243***	0.283***	0.191***
	(0.054)	(0.053)	(0.054)	(0.055)
Number of paid workers: primary business	0.032***	0.032**	0.032***	0.024**
	(0.009)	(0.010)	(0.010)	(0.009)
Number of unpaid workers: primary business	-0.008	-0.014	-0.011	0.010
	(0.010)	(0.009)	(0.011)	(0.011)
Total business profits (log)	0.120***	0.105***	0.116***	0.097***
	(0.013)	(0.013)	(0.013)	(0.014)
Total value of business capital (log)	0.113***	0.116***	0.112***	0.124***
	(0.009)	(0.009)	(0.009)	(0.009)
R-squared	0.145	0.132	0.145	0.119
N	4553	4553	4553	4553

* p<0.05, ** p<0.01, *** p<0.001

Notes: Estimated standard errors are adjusted for clustered sampling.

Table 29. Estimates of the impact of business training on business practices (composite variable) of women business owners by alternative scales (ES and BS data)

	Alternative scales			
Dependent variable→	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale
	(1)	(2)	(3)	(4)
UNADJUSTED				
Received business training (BS)	0.100***	0.100***	0.101***	0.077**
	(0.037)	(0.036)	(0.037)	(0.035)
R-squared	0.003	0.003	0.003	0.002
N	2679	2679	2679	2679
ADJUSTED				
	Alternative scales			
Dependent variable→	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale
	(1)	(2)	(3)	(4)
Received business training (BS)	0.094***	0.094***	0.094***	0.070**
	(0.033)	(0.033)	(0.033)	(0.033)
Business practices (BS, count)	0.454***	0.431***	0.462***	0.371***
	(0.022)	(0.022)	(0.022)	(0.019)
R-squared	0.201	0.181	0.209	0.139
N	2679	2679	2679	2679

* p<0.10, ** p<0.05, *** p<0.01 (impact estimates only)

Notes: Estimated standard errors are adjusted for clustered sampling.

Table 30. Multiple regression analysis (OLS model) of the total number of employed workers (paid and unpaid) by gender (BS)

	Unpaid workers			Paid workers		
	All business owner	Women business owners	Men business owners	All business owner	Women business owners	Men business owners
	(1)	(2)	(3)	(4)	(5)	(6)
Owner's age	-0.025	-0.029	-0.025	0.016	0.008	0.035
	(0.017)	(0.026)	(0.022)	(0.019)	(0.016)	(0.036)
Owner's age squared	0.000	0.000	0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Woman business owner	0.214***			-0.207***		
	(0.060)			(0.057)		
Owner's Schooling: Lower secondary	-0.040	0.007	-0.096	-0.061	-0.026	-0.080
	(0.038)	(0.051)	(0.058)	(0.058)	(0.034)	(0.142)
Owner's Schooling: Upper secondary	0.022	0.095	-0.065	-0.044	0.113	-0.257*
	(0.053)	(0.086)	(0.055)	(0.069)	(0.076)	(0.126)
Owner's Schooling: Tertiary	0.127	0.100	0.182	-0.133	0.089	-0.442
	(0.075)	(0.104)	(0.116)	(0.119)	(0.092)	(0.241)
Owner's cognitive ability (IRT scale)	0.013	-0.006	0.036	-0.001	-0.027	0.029
	(0.015)	(0.022)	(0.019)	(0.025)	(0.023)	(0.048)
Owner currently married	0.232***	0.325*	0.131	0.068	-0.021	0.097
	(0.057)	(0.128)	(0.083)	(0.067)	(0.050)	(0.137)
Household size	0.057***	0.051**	0.082***	-0.004	0.002	-0.037
	(0.017)	(0.019)	(0.022)	(0.011)	(0.007)	(0.041)
Female working-age HH members	0.124***	0.161***	0.082*	-0.022	-0.006	-0.060
	(0.032)	(0.048)	(0.038)	(0.037)	(0.034)	(0.072)
Male working-age HH members	0.054	0.061	0.025	0.034	-0.014	0.126
	(0.029)	(0.040)	(0.035)	(0.039)	(0.026)	(0.104)
Number of children in HH	0.019	0.038	-0.016	-0.022	0.029	-0.083
	(0.027)	(0.037)	(0.033)	(0.034)	(0.033)	(0.076)
Spouse of HH head	0.068	0.032		-0.137	0.032	
	(0.072)	(0.126)		(0.076)	(0.058)	
Child of HH head	0.162*	0.174	0.060	-0.213***	-0.057	-0.195
	(0.065)	(0.121)	(0.092)	(0.063)	(0.056)	(0.103)
Total value of business capital (log)	0.014	0.026	0.001	0.125***	0.043***	0.212***
	(0.012)	(0.018)	(0.013)	(0.018)	(0.011)	(0.039)
Primary business profits (log)	-0.023	-0.024	-0.018	0.183***	0.133***	0.275***
	(0.016)	(0.022)	(0.022)	(0.031)	(0.027)	(0.068)
HH asset index (PC scale)	0.025	0.041	0.004	0.195***	0.108***	0.291***
	(0.023)	(0.037)	(0.023)	(0.039)	(0.024)	(0.084)
Constant	0.437	0.508	0.679	-0.015	-0.153	-0.483
	(0.346)	(0.494)	(0.444)	(0.383)	(0.359)	(0.708)
R-squared	0.05	0.03	0.05	0.09	0.04	0.10
N	4430	2587	1843	4430	2587	1843

* p<0.05, ** p<0.01, *** p<0.001

Notes: Dependent variable includes both paid and unpaid workers (not including the business owner). Estimated standard errors are adjusted for clustered sampling. There are no observations of WBOs who are spouses of the head of HH.

Table 31. Multiple regression analysis of the number of unpaid workers using alternative statistical models (BS unless otherwise indicated)

	Alternative statistical models			
	OLS regression model	OLS (winsorized) regression model	Two-part model	Poisson regression model
	(1)	(2)	(3)	(4)
Owner's age	-0.025	-0.029*	-0.044*	-0.028
	(0.017)	(0.014)	(0.022)	(0.017)
Owner's age squared	0.000	0.000*	0.001*	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)
Woman business owner	0.214***	0.197***	0.279***	0.196**
	(0.060)	(0.055)	(0.084)	(0.061)
Owner's Schooling: Lower secondary	-0.040	-0.039	-0.053	-0.038
	(0.038)	(0.035)	(0.054)	(0.039)
Owner's Schooling: Upper secondary	0.022	-0.018	-0.021	0.023
	(0.053)	(0.035)	(0.053)	(0.054)
Owner's Schooling: Tertiary	0.127	0.116	0.185	0.120
	(0.075)	(0.070)	(0.101)	(0.067)
Owner's cognitive ability (IRT scale)	0.013	0.004	0.007	0.014
	(0.015)	(0.012)	(0.019)	(0.015)
Owner currently married	0.232***	0.217***	0.332***	0.276***
	(0.057)	(0.055)	(0.084)	(0.068)
Household size	0.057***	0.050***	0.099***	0.032***
	(0.017)	(0.014)	(0.020)	(0.009)
Female working-age HH members (ES)	0.124***	0.112***	0.158***	0.131***
	(0.032)	(0.023)	(0.035)	(0.029)
Male working-age HH members (ES)	0.054	0.050*	0.070*	0.055*
	(0.029)	(0.022)	(0.032)	(0.025)
Number of children in HH	0.019	0.029	0.025	0.039
	(0.027)	(0.019)	(0.029)	(0.022)
Spouse of HH head	0.068	0.054	0.110	0.099
	(0.072)	(0.060)	(0.090)	(0.072)
Child of HH head	0.162*	0.148*	0.208*	0.200**
	(0.065)	(0.061)	(0.092)	(0.066)
Total value of business capital (log)	0.014	0.021*	0.030*	0.014
	(0.012)	(0.009)	(0.014)	(0.012)
Primary business profits (log)	-0.023	-0.011	-0.021	-0.022
	(0.016)	(0.014)	(0.022)	(0.015)
HH asset index (PC scale)	0.025	-0.000	-0.002	0.029
	(0.023)	(0.014)	(0.022)	(0.022)
N	4430	4430	4430	4430

* p<0.05, ** p<0.01, *** p<0.001

Notes: The number of unpaid workers does not include the business owner. Estimated standard errors are adjusted for clustered sampling in columns 1-4.

Table 32. Multiple regression analysis of the number of paid workers using alternative statistical models (BS data)

	Alternative statistical models			
	OLS	OLS (winsorized)	Two-part model	Poisson regression
	(1)	(2)	(3)	(4)
Owner's age	0.016	0.003	0.006	0.027
	(0.019)	(0.011)	(0.026)	(0.022)
Owner's age squared	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Woman business owner	-0.207***	-0.186***	-0.433***	-0.386***
	(0.057)	(0.039)	(0.108)	(0.107)
Owner's Schooling: Lower secondary	-0.061	-0.025	-0.034	-0.013
	(0.058)	(0.028)	(0.069)	(0.067)
Owner's Schooling: Upper secondary	-0.044	-0.017	-0.023	0.008
	(0.069)	(0.034)	(0.071)	(0.074)
Owner's Schooling: Tertiary	-0.133	-0.011	-0.001	-0.052
	(0.119)	(0.067)	(0.110)	(0.087)
Owner's cognitive ability (IRT scale)	-0.001	0.008	0.006	0.006
	(0.025)	(0.012)	(0.026)	(0.026)
Owner currently married	0.068	0.015	-0.018	0.034
	(0.067)	(0.044)	(0.091)	(0.081)
Household size	-0.004	0.013*	0.018	-0.006
	(0.011)	(0.006)	(0.011)	(0.018)
Female working-age HH members	-0.022	-0.025	-0.035	-0.022
	(0.037)	(0.021)	(0.041)	(0.038)
Male working-age HH members	0.034	0.009	0.014	0.034
	(0.039)	(0.017)	(0.038)	(0.039)
Number of children in HH	-0.022	-0.033*	-0.050	-0.030
	(0.034)	(0.016)	(0.032)	(0.036)
Spouse of HH head	-0.137	-0.062	0.028	0.056
	(0.076)	(0.042)	(0.114)	(0.124)
Child of HH head	-0.213***	-0.130**	-0.193*	-0.134
	(0.063)	(0.043)	(0.093)	(0.069)
Total value of business capital (log)	0.125***	0.085***	0.141***	0.119***
	(0.018)	(0.009)	(0.018)	(0.019)
Primary business profits (log)	0.183***	0.135***	0.275***	0.150***
	(0.031)	(0.015)	(0.031)	(0.032)
HH asset index (PC scale)	0.195***	0.120***	0.171***	0.076**
	(0.039)	(0.016)	(0.025)	(0.024)
N	4430	4430	4430	4430

* p<0.05, ** p<0.01, *** p<0.001

Notes: Estimated standard errors are adjusted for clustered sampling in columns 1-4.

Table 33. Multiple regression analysis of the number of primary business customers using alternative statistical models (BS data)

	OLS applied to reported variable	OLS applied to winsorized reported variable (highest 5% of values)	Log regression model	Quantile (median) regression model	Robust regression model
	(1)	(2)	(3)	(4)	(5)
Owner's age	-5.468	-6.688	-0.023	-0.611	0.923
	(8.474)	(6.299)	(0.026)	(1.383)	(0.658)
Owner's age squared	0.057	0.072	0.000	0.005	-0.015
	(0.115)	(0.083)	(0.000)	(0.018)	(0.009)
Woman business owner	-151.920***	-110.421***	-0.378***	-14.944***	-10.036***
	(22.467)	(15.744)	(0.051)	(3.059)	(1.690)
Owner's Schooling: Lower secondary	-48.982*	-41.131*	-0.157*	-3.888	-0.590
	(22.527)	(16.943)	(0.067)	(3.592)	(1.924)
Owner's Schooling: Upper secondary	-28.432	-22.036	-0.058	0.966	-0.863
	(27.348)	(18.939)	(0.074)	(4.273)	(2.108)
Owner's Schooling: Tertiary	-5.514	-14.205	0.029	0.625	0.885
	(49.014)	(31.731)	(0.124)	(5.691)	(2.567)
Owner's cognitive ability (IRT scale)	4.870	2.421	0.014	-1.648	-0.227
	(8.241)	(6.260)	(0.027)	(1.236)	(0.650)
Number of hours worked in typical month	0.020	0.051	0.001***	0.012	0.016**
	(0.073)	(0.054)	(0.000)	(0.011)	(0.006)
Number of paid workers in the primary business	-9.136**	-9.374**	-0.107***	-3.987**	-1.508**
	(3.371)	(2.847)	(0.021)	(1.316)	(0.531)
Number of unpaid workers in the primary business	8.486	9.862**	0.064**	4.008*	0.122
	(4.915)	(3.775)	(0.020)	(1.911)	(0.442)
Total value of business capital (log)	16.986**	17.168***	0.120***	4.940***	2.026***
	(5.568)	(4.124)	(0.017)	(0.946)	(0.438)
Business practices (PM scale)	7.133	1.961	0.005	0.148	-0.304
	(8.143)	(5.765)	(0.024)	(1.189)	(0.727)
Mobile phone use (PM scale)	-25.076**	-22.389***	-0.098***	-3.272*	-0.855
	(9.359)	(6.657)	(0.028)	(1.648)	(0.711)
Connectedness with peers (PM scale)	12.469	10.276	0.024	-0.065	-0.441
	(10.180)	(6.778)	(0.026)	(1.231)	(0.601)
Number of hours volunteered in a typical month	-0.975	-1.087	-0.009***	-0.318	-0.152**
	(0.916)	(0.709)	(0.003)	(0.167)	(0.054)
R-squared	0.12	0.15	0.20		0.06
N	4442	4442	4442	4442	4442

* p<0.05, ** p<0.01, *** p<0.001

Notes: Models also include (results not shown) 5 dummy variables for the type of business and 2 dummy variables for regency of residence (2 of the 5 regencies are not included in the MS sample). The estimated standard errors in columns 1-3 are adjusted for clustered sampling, while the estimated standard errors in columns 4-5 are bootstrapped with 100 repetitions by cluster.

Table 34. Estimates of the impact of business training on the number of customers in the primary businesses of women business owners using alternative statistical models (ES and BS data)

	Alternative statistical models				
Dependent variable→	OLS applied to reported variable	OLS applied to winsorized variable (highest 5% of values)	Log regression model	Quantile (median) regression	Robust regression model
	(1)	(2)	(3)	(4)	(5)
UNADJUSTED					
Received business training (BS)	27.994	31.291	0.125**	No estimates obtained	29.181*
	(24.850)	(20.700)	(0.059)		(16.163)
N	2675	2675	2675		2675
ADJUSTED					
	Alternative statistical models				
Dependent variable→	OLS applied to reported variable	OLS applied to winsorized variable (highest 5% of values)	Log regression model	Quantile (median) regression	Robust regression model
	(1)	(2)	(3)	(4)	
Received business training (BS)	18.577	22.318	0.105*	10.078	25.487*
	(23.898)	(19.617)	(0.057)	(11.948)	(13.672)
Business practices (BS, count)	0.418***	0.398***	0.001***	0.388***	0.358***
	(0.045)	(0.038)	(0.000)	(0.029)	(0.035)
Constant	424.425***	412.196***	5.263***	282.171***	308.174***
	(20.999)	(17.367)	(0.054)	(12.274)	(15.294)
N	2674	2674	2674	2674	2674

* p<0.10, ** p<0.05, *** p<0.01 (impact estimates only)

Notes: The estimated standard errors in columns 1-3 are adjusted for clustered sampling, whereas the estimated standard errors in columns 4 and 5 are bootstrapped with 100 repetitions by cluster.

Table 35. Multiple regression analysis (OLS model) of whether any money was saved during the past 12 months by gender (BS data)

	Full sample	Women business owners	Men business owners
	(1)	(2)	(3)
Owner's age	0.007	0.012	-0.011
	(0.007)	(0.008)	(0.013)
Owner's age squared	-0.000*	-0.000*	0.000
	(0.000)	(0.000)	(0.000)
Woman business owner	0.139***		
	(0.037)		
Owner's Schooling: Lower secondary	0.030	0.033	0.018
	(0.019)	(0.021)	(0.032)
Owner's Schooling: Upper secondary	0.066***	0.050*	0.083**
	(0.019)	(0.022)	(0.030)
Owner's Schooling: Tertiary	0.093***	0.059	0.144**
	(0.027)	(0.031)	(0.048)
Owner's cognitive ability (IRT scale)	0.018**	0.021**	0.014
	(0.007)	(0.008)	(0.011)
Subjective time preference (IRT scale)	0.007	0.006	0.007
	(0.006)	(0.007)	(0.011)
Household size	0.003	0.001	0.010
	(0.003)	(0.003)	(0.010)
Number of children in HH	-0.005	0.002	-0.019
	(0.008)	(0.009)	(0.017)
Spouse of HH head	0.038	0.067	
	(0.037)	(0.062)	
Child of HH head	0.002	0.050	-0.035
	(0.037)	(0.066)	(0.046)
Total earnings (log)	0.040***	0.036***	0.049***
	(0.007)	(0.008)	(0.013)
HH asset index (PC scale)	0.015*	0.020*	0.006
	(0.007)	(0.008)	(0.011)
Intra-HH agency (PM scale)	0.016*	0.014	0.023
	(0.007)	(0.008)	(0.017)
Marital agency (IRT scale)	0.004	0.010	-0.014
	(0.007)	(0.008)	(0.015)
R-squared	0.08	0.05	0.06
N	4337	2579	1758

* p<0.05, ** p<0.01, *** p<0.001

Notes: Estimated standard errors are adjusted for clustered sampling.

Table 36. Multiple regression analysis of total savings during the last 12 months using alternative statistical models (BS data)

	Alternative statistical models					
	OLS model	OLS (winsorized) model	Tobit model	Two-part model	Quantile (median) regression model	Robust regression model
	(1)	(2)	(3)	(4)	(5)	(6)
Owner's age	-0.826	0.236	0.592	0.163	-0.019	-0.012
	(1.299)	(0.160)	(1.309)	(0.235)	(0.065)	(0.051)
Owner's age squared	0.009	-0.004	-0.020	-0.004	-0.000	-0.000
	(0.017)	(0.002)	(0.016)	(0.003)	(0.001)	(0.001)
Woman business owner	-3.538	0.445	15.314**	1.227**	0.694***	0.714***
	(2.500)	(0.343)	(5.927)	(0.464)	(0.149)	(0.105)
Owner's Schooling: Lower secondary	1.215	0.066	5.454	0.607	0.163	0.125
	(1.551)	(0.373)	(3.455)	(0.621)	(0.196)	(0.109)
Owner's Schooling: Upper secondary	3.161	0.845*	10.373	1.727**	0.423*	0.345*
	(3.310)	(0.406)	(6.521)	(0.645)	(0.196)	(0.135)
Owner's Schooling: Tertiary	4.510	2.767**	12.918*	3.573***	1.946**	0.853*
	(4.351)	(0.981)	(5.851)	(1.014)	(0.707)	(0.361)
Owner's cognitive ability (IRT scale)	-0.248	0.167	1.793	0.370	0.128	0.153**
	(0.545)	(0.160)	(1.055)	(0.216)	(0.079)	(0.050)
Owner's subjective time preference (PM scale)	1.774	-0.365*	1.346	-0.486*	-0.127	-0.102*
	(1.934)	(0.145)	(2.075)	(0.223)	(0.078)	(0.050)
Owner currently married	-1.515	0.215	-0.830	0.307	0.041	0.030
	(1.729)	(0.476)	(3.065)	(0.734)	(0.225)	(0.170)
Household size	-0.829	0.004	-0.756	-0.096	-0.021	-0.046
	(0.889)	(0.088)	(1.010)	(0.138)	(0.056)	(0.030)
Number of children in HH	5.471	-0.370	5.679	-0.583*	-0.056	-0.099
	(4.859)	(0.196)	(5.404)	(0.291)	(0.089)	(0.070)
Intensity of bank use (IRT scale)	2.192**	1.201***	6.157**	1.805***	0.510***	0.310***
	(0.708)	(0.161)	(2.029)	(0.236)	(0.091)	(0.064)
Total earnings (log)	5.675***	2.704***	10.591**	4.365***	0.855***	0.563***
	(1.537)	(0.179)	(3.408)	(0.310)	(0.089)	(0.074)
HH asset index (PC scale)	6.410**	2.357***	7.925**	3.037***	0.950***	0.345***
	(2.137)	(0.221)	(2.840)	(0.297)	(0.120)	(0.090)
N	4754	4754	4754	4754	4754	4753

* p<0.05, ** p<0.01, *** p<0.001

Notes: The estimated standard errors in columns 1-4 are adjusted for clustered sampling, whereas the estimated standard errors in columns 5 and 6 are bootstrapped with 100 repetitions by cluster.

Table 37. Multiple regression analysis of total savings during the last 12 months as a proportion of primary and second profits using alternative statistical models (BS data)

	Alternative statistical models					
	OLS model	OLS (winsorized) model	Tobit model	Two-part model	Quantile (median) regression model	Robust regression model
	(1)	(2)	(3)	(4)	(5)	(6)
Owner's age	-0.184	0.174	0.668	0.105	-0.059	0.041
	(0.397)	(0.121)	(0.551)	(0.166)	(0.070)	(0.037)
Owner's age squared	0.000	-0.003	-0.017	-0.003	0.000	-0.001*
	(0.004)	(0.002)	(0.009)	(0.002)	(0.001)	(0.000)
Woman business owner	-2.816	0.676**	7.820**	0.551	0.640***	0.480***
	(2.296)	(0.232)	(2.733)	(0.335)	(0.105)	(0.076)
Owner's Schooling: Lower secondary	-0.389	0.313	1.690	0.600	0.155	0.213*
	(1.460)	(0.286)	(1.642)	(0.439)	(0.133)	(0.094)
Owner's Schooling: Upper secondary	1.519	1.040***	5.510*	1.370**	0.401*	0.332*
	(0.860)	(0.285)	(2.688)	(0.450)	(0.165)	(0.130)
Owner's Schooling: Tertiary	2.224	3.729***	6.911**	4.457***	1.756***	0.925***
	(2.727)	(0.737)	(2.464)	(0.784)	(0.518)	(0.233)
Owner's cognitive ability (IRT scale)	-0.244	0.041	0.920	0.228	0.107	0.147***
	(0.475)	(0.114)	(0.509)	(0.157)	(0.059)	(0.040)
Owner's subjective time preference (PM scale)	-0.671	-0.158	-1.131	-0.377*	-0.067	-0.048
	(0.654)	(0.108)	(0.989)	(0.158)	(0.054)	(0.040)
Owner currently married	0.630	0.692	0.775	0.298	0.238	-0.006
	(0.778)	(0.371)	(1.733)	(0.511)	(0.216)	(0.151)
Household size	-0.145	0.009	-0.108	-0.069	-0.026	-0.047
	(0.222)	(0.110)	(0.306)	(0.101)	(0.045)	(0.039)
Number of children in HH	2.030	-0.229	2.086	-0.438*	-0.085	-0.052
	(1.851)	(0.155)	(2.101)	(0.212)	(0.068)	(0.055)
Intensity of bank use (IRT scale)	2.563*	1.001***	4.995*	1.268***	0.332***	0.230***
	(1.140)	(0.126)	(2.269)	(0.176)	(0.068)	(0.041)
Total earnings (log)	-7.669	-1.877***	-6.068	-2.822***	-0.542***	-0.191***
	(4.461)	(0.157)	(3.916)	(0.226)	(0.079)	(0.050)
HH asset index (PC scale)	4.838	1.050***	5.819	1.924***	0.482***	0.259***
	(2.878)	(0.132)	(3.417)	(0.195)	(0.068)	(0.049)
N	4753	4753	4753	4753	4753	4752

* p<0.05, ** p<0.01, *** p<0.001

Notes: The estimated standard errors in columns 1-4 are adjusted for clustered sampling, whereas the estimated standard errors in columns 5 and 6 are bootstrapped with 100 repetitions by cluster.

Table 38. Multiple regression analysis (using the robust regression model) of total savings during the past 12 months and of the ratio of total savings to total earnings (using robust regression model) by gender (BS data)

Dependent variable→	Total savings during the past 12 months			Total savings during the past 12 months as a ratio to annualized total earnings		
	Full sample	WBOs	MBOs	Full sample	WBOs	MBOs
	(1)	(2)	(3)	(4)	(5)	(6)
Owner's age	-0.012	0.019	-0.081	0.041	0.050	-0.018
	(0.047)	(0.070)	(0.102)	(0.037)	(0.074)	(0.057)
Owner's age squared	-0.000	-0.001	0.000	-0.001*	-0.001	-0.000
	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
Woman business owner	0.714***			0.480***		
	(0.116)			(0.076)		
Owner's Schooling: Lower secondary	0.125	0.058	0.271	0.213*	0.149	0.158
	(0.127)	(0.154)	(0.223)	(0.094)	(0.150)	(0.111)
Owner's Schooling: Upper secondary	0.345*	0.173	0.615**	0.332*	0.238	0.301**
	(0.154)	(0.177)	(0.215)	(0.130)	(0.188)	(0.109)
Owner's Schooling: Tertiary	0.853**	0.618	2.432	0.925***	0.837*	0.703*
	(0.327)	(0.335)	(1.275)	(0.233)	(0.347)	(0.346)
Owner's cognitive ability (IRT scale)	0.153**	0.132*	0.227**	0.147***	0.175**	0.095*
	(0.053)	(0.056)	(0.079)	(0.040)	(0.061)	(0.040)
Owner's subjective time preference (PM scale)	-0.102*	-0.094	-0.145	-0.048	-0.072	-0.027
	(0.048)	(0.059)	(0.090)	(0.040)	(0.071)	(0.042)
Owner currently married	0.030	0.096	-0.176	-0.006	0.052	0.008
	(0.186)	(0.187)	(0.447)	(0.151)	(0.204)	(0.257)
Household size	-0.046	-0.018	-0.194*	-0.047	-0.066	-0.012
	(0.030)	(0.031)	(0.083)	(0.039)	(0.074)	(0.046)
Number of children in HH	-0.099	-0.066	-0.022	-0.052	-0.053	-0.100
	(0.060)	(0.074)	(0.143)	(0.055)	(0.095)	(0.064)
Intensity of bank use (IRT scale)	0.310***	0.320***	0.330***	0.230***	0.291***	0.153***
	(0.054)	(0.071)	(0.100)	(0.041)	(0.077)	(0.045)
Total earnings (log)	0.563***	0.499***	0.836***	-0.191***	-0.407***	0.033
	(0.068)	(0.062)	(0.177)	(0.050)	(0.077)	(0.052)
HH asset index (PC scale)	0.345***	0.395***	0.257*	0.259***	0.386***	0.131*
	(0.076)	(0.109)	(0.131)	(0.049)	(0.092)	(0.056)
N	4753	2808	1945	4752	2807	1945

* p<0.05, ** p<0.01, *** p<0.001

Notes: The estimated standard errors are bootstrapped with 100 repetitions by cluster.

Table 39. Estimates of the impact of business training on savings during the last 12 months of women business owners using alternative statistical models (ES and BS data)

Dependent variable→	Any savings	Total savings (including zeroes)					
Statistical model→	OLS regression model	OLS regression model	OLS (winsorized) regression model	Tobit model	Two-part model	Quantile (median) regression model	Robust regression model
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
UNADJUSTED							
Randomly assigned to receive training (BS)	0.029***	0.439	0.423	0.935	0.887**	0.500***	0.394***
	(0.011)	(0.526)	(0.405)	(0.609)	(0.412)	(0.193)	(0.139)
N	2724	2819	2819	2819	2819	2819	2819
ADJUSTED							
Randomly assigned to receive training (BS)	0.027**	0.343	0.346	0.816	1.795	0.372***	0.332**
	(0.011)	(0.516)	(0.397)	(0.593)	(1.222)	(0.133)	(0.147)
Any nonzero savings (BS)	0.131***						
	(0.019)						
Total savings (BS)		0.327***	0.260***	0.344***	0.552	0.296***	0.117**
		(0.068)	(0.038)	(0.072)	(0.400)	(0.038)	(0.054)
N	2724	2797	2797	2797	2797	2797	2796

* p<0.10, ** p<0.05, *** p<0.01 (impact estimates only)

Notes: The estimated standard errors in columns 1 are adjusted for clustered sampling, while those in columns 6 and 7 are bootstrapped with 100 repetitions by cluster.

Table 40. Multiple regression analysis of any currently outstanding bank loans by gender (BS data)

	Full sample	Women business owners	Men business owners
	(1)	(2)	(3)
Owner's age	0.011*	0.010	0.011
	(0.006)	(0.007)	(0.010)
Owner's age squared	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)
Woman business owner	-0.073***		
	(0.011)		
Owner's Schooling: Lower secondary	-0.010	-0.013	-0.006
	(0.014)	(0.017)	(0.022)
Owner's Schooling: Upper secondary	-0.021	-0.024	-0.020
	(0.014)	(0.017)	(0.022)
Owner's Schooling: Tertiary	-0.140***	-0.143***	-0.138**
	(0.031)	(0.037)	(0.049)
Owner's cognitive ability (IRT scale)	-0.011*	-0.013*	-0.007
	(0.005)	(0.006)	(0.008)
Owner currently married	0.073***	0.039	0.119***
	(0.018)	(0.024)	(0.032)
Household size	0.005*	0.006*	0.002
	(0.002)	(0.003)	(0.006)
Number of children in HH	0.012	0.016	0.005
	(0.006)	(0.008)	(0.011)
Number of banks in which accounts are held	0.120***	0.120***	0.117***
	(0.019)	(0.030)	(0.025)
Intensity of bank use (IRT scale)	0.226***	0.194***	0.271***
	(0.011)	(0.017)	(0.015)
Total earnings (log)	0.023***	0.023***	0.021*
	(0.005)	(0.006)	(0.009)
HH asset index (PC scale)	-0.016**	-0.011	-0.022*
	(0.006)	(0.007)	(0.009)
N	4781	2827	1954

* p<0.05, ** p<0.01, *** p<0.001

Notes: The estimated standard errors are adjusted for clustered sampling.

Table 41. Multiple regression analysis (OLS model) of borrowing during the last three months by gender and by sources of loans (MS data unless otherwise indicated)

	Any borrowing			Source of loan			
	Total	WBOs	MBOs	Bank	Friends/family	ROSCA	Other
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Owner's age (BS)	0.008	0.009	-0.003	0.004	0.001	0.002	0.008
	(0.011)	(0.014)	(0.017)	(0.005)	(0.008)	(0.006)	(0.007)
Owner's age squared (BS)	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Woman business owner	0.071**			-0.012	0.017	0.052***	0.024
	(0.026)			(0.014)	(0.018)	(0.014)	(0.016)
Owner's Schooling: Lower secondary (BS)	0.020	0.017	0.010	-0.011	0.017	-0.011	0.008
	(0.028)	(0.037)	(0.040)	(0.013)	(0.018)	(0.014)	(0.018)
Owner's Schooling: Upper secondary (BS)	-0.019	-0.023	-0.023	-0.028	-0.011	-0.003	-0.007
	(0.027)	(0.036)	(0.042)	(0.015)	(0.019)	(0.015)	(0.017)
Owner's Schooling: Tertiary (BS)	-0.146**	-0.160*	-0.154*	-0.033	-0.068*	-0.005	-0.029
	(0.045)	(0.063)	(0.067)	(0.031)	(0.028)	(0.030)	(0.030)
Owner's cognitive ability (BS, iRT scale)	0.004	0.007	0.001	0.001	0.009	-0.001	-0.018*
	(0.010)	(0.013)	(0.015)	(0.006)	(0.008)	(0.006)	(0.007)
Borrows for emergencies	0.150***	0.192***	0.099**	0.018	0.141***	0.007	0.029
	(0.027)	(0.039)	(0.036)	(0.014)	(0.020)	(0.015)	(0.017)
Number of banks in which accounts are held	0.065***	0.049*	0.079**	0.066***	0.007	0.006	-0.000
	(0.018)	(0.023)	(0.026)	(0.011)	(0.013)	(0.011)	(0.010)
Primary and second profits (log)	-0.008	0.002	-0.026	-0.003	0.000	0.003	-0.002
	(0.011)	(0.014)	(0.018)	(0.006)	(0.007)	(0.006)	(0.008)
HH income (predicted, log)	0.066***	0.056**	0.080***	0.030***	0.005	0.015	0.027**
	(0.014)	(0.018)	(0.023)	(0.008)	(0.010)	(0.009)	(0.009)
HH asset index (PC scale)	-0.019	-0.017	-0.022	-0.003	-0.006	0.002	-0.007
	(0.012)	(0.017)	(0.018)	(0.008)	(0.007)	(0.007)	(0.008)
Current total savings balance	-0.000	-0.000**	0.000	0.000	0.000	-0.000*	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
HH agency (PM scale)	0.007	0.015	-0.016	0.003	0.005	-0.003	0.015*
	(0.011)	(0.014)	(0.019)	(0.006)	(0.008)	(0.006)	(0.007)
Marital agency (SM scale)	-0.008	-0.019	0.024	-0.011*	-0.001	-0.001	0.001
	(0.011)	(0.013)	(0.021)	(0.006)	(0.007)	(0.006)	(0.007)
Individual agency (IRT scale)	0.008	-0.001	0.019	0.004	-0.001	0.003	0.010
	(0.010)	(0.013)	(0.014)	(0.006)	(0.007)	(0.006)	(0.007)
R-squared	0.04	0.05	0.05	0.05	0.04	0.01	0.02
N	2002	1172	830	2002	2002	2002	2002

* p<0.05, ** p<0.01, *** p<0.001

Notes: The estimated standard errors are adjusted for clustered sampling.

Table 42. Multiple regression analysis of the amount of money borrowed during the past 3 months using alternative statistical models and by gender (MS data unless indicated otherwise)

Estimation sample→	All business owners				WBOs	MBOs
Statistical model→	OLS model	OLS (winsorized) model	Tobit model	Two-part model	Tobit model	Tobit model
	(1)	(2)	(3)	(4)	(5)	(6)
Owner's age (BS)	0.480	0.077	1.179	0.271	0.536	1.682
	(0.346)	(0.133)	(1.011)	(0.344)	(0.974)	(1.960)
Owner's age squared (BS)	-0.007	-0.001	-0.019	-0.003	-0.010	-0.029
	(0.004)	(0.002)	(0.013)	(0.004)	(0.013)	(0.024)
Woman business owner	-0.806	-0.161	3.448	-0.195		
	(1.004)	(0.332)	(2.480)	(0.709)		
Owner's Schooling: Lower secondary (BS)	-0.080	-0.253	1.539	0.095	-0.545	3.170
	(0.780)	(0.314)	(2.486)	(0.733)	(1.963)	(5.645)
Owner's Schooling: Upper secondary (BS)	-0.919	-0.296	-2.224	0.388	-1.497	-4.257
	(0.755)	(0.347)	(2.324)	(0.793)	(2.163)	(5.202)
Owner's Schooling: Tertiary (BS)	5.446	-0.663	0.154	0.929	-4.088	1.668
	(4.528)	(0.848)	(8.114)	(2.067)	(7.297)	(14.024)
Owner's cognitive ability (BS, iRT scale)	-0.335	-0.032	-0.213	0.053	0.243	-0.547
	(0.393)	(0.136)	(0.973)	(0.284)	(0.798)	(2.169)
Borrows for emergencies	1.534*	0.860**	11.066***	2.401*	8.236***	13.209**
	(0.640)	(0.319)	(2.431)	(0.933)	(2.136)	(4.814)
Number of banks in which accounts are held	2.852**	1.296***	8.105***	2.191**	5.006**	11.972**
	(0.930)	(0.263)	(2.034)	(0.743)	(1.598)	(3.869)
Primary and second profits (log)	0.673	0.135	0.615	0.507	1.258	-0.289
	(0.405)	(0.141)	(1.006)	(0.359)	(0.876)	(2.121)
HH income (log)	1.864***	0.956***	7.114***	2.209**	3.956**	10.912***
	(0.479)	(0.181)	(1.508)	(0.760)	(1.324)	(3.144)
HH asset index (PC scale)	0.623	0.192	-0.088	0.418	0.747	-1.408
	(0.554)	(0.178)	(1.294)	(0.367)	(1.220)	(2.394)
Current total savings balance	0.014	0.001	0.017	0.004	-0.005	0.052
	(0.016)	(0.002)	(0.022)	(0.003)	(0.008)	(0.036)
Intra-HH agency PM scale)	-0.263	0.086	0.085	0.331	1.507	-4.051
	(0.369)	(0.144)	(1.045)	(0.371)	(0.890)	(2.678)
Marital agency (SM scale)	0.014	-0.121	-0.736	-0.662	-1.472	3.062
	(0.362)	(0.153)	(1.060)	(0.392)	(0.952)	(2.773)
Personal agency (IRT scale)	0.521	0.134	1.253	0.601	-0.102	3.165
	(0.378)	(0.128)	(0.952)	(0.325)	(0.710)	(1.862)
N	2002	2002	2002	2002	1172	830

* p<0.05, ** p<0.01, *** p<0.001

Notes: The estimated standard errors are adjusted for clustered sampling.

Table 43. Multiple regression analysis of any currently outstanding loans and of the amount still owed using alternative statistical models and by gender (MS data unless otherwise indicated)

Estimation sample→	Full sample					WBOs	MBOs
Dependent variable→	Any outstanding loans	Amount still owed					
		Alternative statistical models				Statistical models	
Statistical model→	OLS	OLS	OLS (winsorized)	Tobit model	Two-part model	Two-part model	Two-part model
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Owner's age (BS)	0.016	0.930	0.593	2.184*	0.884	0.470	0.773
	(0.009)	(0.700)	(0.332)	(1.092)	(0.613)	(0.486)	(1.491)
Owner's age squared (BS)	-0.000	-0.010	-0.007	-0.027	-0.009	-0.004	-0.009
	(0.000)	(0.010)	(0.004)	(0.015)	(0.008)	(0.006)	(0.019)
Woman business owner	-0.063**	-3.874*	-1.875	-3.125	-2.778		
	(0.022)	(1.597)	(0.960)	(2.310)	(1.462)		
Owner's Schooling: Lower secondary (BS)	-0.051*	-1.771	-1.143	-3.607	-1.725	-1.418	-1.638
	(0.023)	(1.475)	(0.821)	(2.275)	(1.432)	(1.079)	(3.731)
Owner's Schooling: Upper secondary (BS)	-0.074**	-0.974	-1.194	-4.811	-2.183	-1.842	-2.277
	(0.027)	(1.982)	(0.973)	(2.598)	(1.498)	(1.275)	(3.683)
Owner's Schooling: Tertiary (BS)	-0.131*	-2.819	-2.487	-11.459	-5.344	-2.256	-12.567*
	(0.054)	(5.312)	(2.636)	(7.003)	(3.066)	(2.887)	(5.996)
Owner's cognitive ability (BS: iRT scale)	0.026**	-0.842	-0.094	0.058	0.152	0.321	-0.383
	(0.009)	(0.672)	(0.370)	(0.939)	(0.599)	(0.460)	(1.434)
Borrows for emergencies	0.026	5.008**	3.100***	10.262***	6.308***	4.335***	7.969*
	(0.020)	(1.513)	(0.857)	(2.445)	(1.430)	(1.201)	(3.380)
Number of banks in which accounts are held	0.343***	12.387***	8.217***	21.126***	12.673***	6.418***	22.316***
	(0.016)	(2.288)	(0.646)	(3.573)	(1.288)	(1.137)	(3.482)
Primary and second profits (log)	0.036***	3.444***	1.665***	4.825***	2.832***	2.092***	2.238
	(0.011)	(0.972)	(0.431)	(1.354)	(0.744)	(0.611)	(1.676)
HH income (log)	0.023	4.634***	3.038***	6.375***	4.621***	2.071**	9.662***
	(0.012)	(0.943)	(0.529)	(1.392)	(1.059)	(0.785)	(2.613)
HH asset index (PC scale)	-0.014	4.244***	2.260***	3.974***	2.819***	1.244*	5.643**
	(0.011)	(1.050)	(0.432)	(1.173)	(0.670)	(0.585)	(1.820)
Current total savings balance	-0.000	0.005	-0.001	0.002	0.003	0.002	0.004
	(0.000)	(0.015)	(0.006)	(0.018)	(0.004)	(0.005)	(0.009)
HH agency (PM scale)	0.004	-1.013	-0.216	-0.926	0.752	0.661	0.331
	(0.011)	(0.754)	(0.351)	(1.091)	(0.658)	(0.465)	(1.912)
Marital agency (SM scale)	-0.014	-0.986	-0.941**	-0.815	-2.355***	-1.266**	-4.031
	(0.009)	(0.574)	(0.332)	(0.918)	(0.632)	(0.449)	(2.181)
Individual agency (IRT scale)	-0.004	-0.418	0.008	0.037	0.736	0.551	1.051
	(0.009)	(0.608)	(0.346)	(0.855)	(0.588)	(0.483)	(1.425)
N	2000	1998	1998	1998	1998	1170	828

* p<0.05, ** p<0.01, *** p<0.001

Notes: The estimated standard errors are adjusted for clustered sampling.

Table 44. Multiple regression analysis (OLS model) of any currently unpaid loans by source (MS data unless otherwise indicated)

	Loan source			
Source→	Bank	Friends or family	ROSCA	Other
	(1)	(2)	(3)	(4)
Owner's age (BS)	0.016	0.014	0.016	0.002
	(0.009)	(0.010)	(0.009)	(0.007)
Owner's age squared (BS)	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Woman business owner	-0.063**	-0.025	0.108***	0.008
	(0.022)	(0.022)	(0.024)	(0.015)
Owner's Schooling: Lower secondary (BS)	-0.051*	0.018	-0.040	-0.027
	(0.023)	(0.023)	(0.023)	(0.019)
Owner's Schooling: Upper secondary (BS)	-0.074**	-0.025	-0.029	-0.025
	(0.027)	(0.025)	(0.025)	(0.018)
Owner's Schooling: Tertiary (BS)	-0.131*	-0.014	-0.044	-0.045
	(0.054)	(0.046)	(0.044)	(0.032)
Owner's cognitive ability (BS, IRT scale)	0.026**	0.002	0.010	0.008
	(0.009)	(0.010)	(0.009)	(0.006)
Borrows for emergencies	0.026	0.165***	-0.025	0.029
	(0.020)	(0.023)	(0.023)	(0.017)
Number of banks in which accounts are held	0.343***	0.001	0.004	-0.009
	(0.016)	(0.014)	(0.014)	(0.011)
Primary and second profits (log)	0.036***	0.004	-0.010	-0.001
	(0.011)	(0.009)	(0.010)	(0.007)
HH income (MS, log)	0.023	-0.022	0.019	0.014
	(0.012)	(0.013)	(0.012)	(0.010)
HH asset index (PC scale)	-0.014	-0.011	-0.001	-0.000
	(0.011)	(0.010)	(0.009)	(0.007)
Current total savings balance	-0.000	-0.000	-0.000**	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
HH agency (PM scale)	0.004	0.006	-0.001	0.006
	(0.011)	(0.010)	(0.011)	(0.006)
Marital agency (SM scale)	-0.014	-0.001	-0.008	0.015*
	(0.009)	(0.010)	(0.011)	(0.008)
Personal agency (IRT scale)	-0.004	-0.001	0.008	-0.004
	(0.009)	(0.008)	(0.009)	(0.006)
R-squared	0.28	0.05	0.02	0.01
N	2000	2002	2001	2001

* p<0.05, ** p<0.01, *** p<0.001

Notes: The estimated standard errors are adjusted for clustered sampling.

Table 45. Multiple regression analysis (OLS model) of borrowing from money lenders and non-bank financial institutions (ES data unless otherwise indicated)

Source→	Money lenders				Non-bank financial institutions	
Dependent variable→	Ever borrowed from this source	Number of loans in last 12 months			Ever borrowed from this source	Number of loans in last 12 months
		Alternative statistical models				
Statistical model→	OLS regression model	OLS regression model	OLS (winsorized) model	Poisson regression model	OLS regression model	OLS regression model
	(1)	(2)	(3)	(4)	(5)	(6)
Owner's age (BS)	0.011**	0.001	0.011	0.085	0.024**	0.022*
	(0.004)	(0.012)	(0.008)	(0.110)	(0.008)	(0.010)
Owner's age squared (BS)	-0.000*	0.000	-0.000	-0.001	-0.000**	-0.000
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Woman business owner	-0.004	-0.019	0.001	-0.218	-0.040**	0.018
	(0.009)	(0.028)	(0.019)	(0.281)	(0.015)	(0.022)
Owner's Schooling: Lower secondary (BS)	-0.029*	-0.132	-0.063	-0.783**	0.017	0.050
	(0.013)	(0.075)	(0.033)	(0.296)	(0.019)	(0.029)
Owner's Schooling: Upper secondary (BS)	-0.014	-0.134	-0.066*	-0.859**	0.019	0.017
	(0.013)	(0.073)	(0.032)	(0.290)	(0.021)	(0.028)
Owner's Schooling: Tertiary (BS)	-0.070***	-0.115*	-0.080	-0.778	-0.019	-0.059
	(0.017)	(0.055)	(0.044)	(0.559)	(0.039)	(0.049)
Owner's cognitive ability (BS, IRT scale)	0.001	-0.024	-0.010	-0.177	0.019**	0.011
	(0.004)	(0.023)	(0.009)	(0.141)	(0.007)	(0.010)
Primary and second business profits (log)	0.007	0.070	0.017	0.579	0.023**	0.018
	(0.004)	(0.057)	(0.011)	(0.333)	(0.007)	(0.011)
HH asset index (PC scale)	-0.018***	-0.038**	-0.029**	-0.357***	-0.003	-0.033**
	(0.005)	(0.014)	(0.010)	(0.103)	(0.008)	(0.012)
Borrows for emergencies	-0.018	-0.003	0.005	0.007	0.248***	0.160
	(0.038)	(0.072)	(0.074)	(0.728)	(0.060)	(0.131)
Number of banks in which accounts are held	0.011	-0.010	0.019	-0.064	0.046***	0.045*
	(0.007)	(0.032)	(0.017)	(0.237)	(0.011)	(0.021)
Intensity of banking services use (IRT scale)	0.001	-0.017	-0.004	-0.141	0.003	-0.023
	(0.005)	(0.026)	(0.012)	(0.216)	(0.009)	(0.012)
Bad experience with a bank (BS)	-0.018**	-0.018	-0.028	-0.139	-0.047***	-0.042**
	(0.007)	(0.021)	(0.019)	(0.132)	(0.010)	(0.014)
N	4525	4525	4525	4525	4525	4525

* p<0.05, ** p<0.01, *** p<0.001

Notes: The estimated standard errors are adjusted for clustered sampling.

Table 46. Multivariate analysis of reported investment during the last 12 months in the primary and second business using alternative statistical models (ES data unless otherwise indicated)

	Alternative statistical models					
	OLS model	OLS (winsorized) model	OLS applied to IHS transformed variable	Tobit model	Two-part model	Quantile (median) regression model
	(1)	(2)	(3)	(4)	(5)	(6)
Owner's age (BS)	-0.190	-0.226*	-0.047*	-0.933	-0.339*	-0.047
	(0.281)	(0.100)	(0.019)	(0.583)	(0.163)	(0.024)
Owner's age squared (BS)	0.002	0.002	0.000	0.009	0.003	0.000
	(0.004)	(0.001)	(0.000)	(0.008)	(0.002)	(0.000)
Woman business owner	-1.475*	-0.663**	-0.088*	-2.011	-0.360	-0.093*
	(0.635)	(0.205)	(0.038)	(1.191)	(0.305)	(0.043)
Owner's Schooling: Lower secondary (BS)	-1.532	-0.019	0.060	1.095	0.602	0.042
	(1.469)	(0.230)	(0.045)	(2.014)	(0.426)	(0.039)
Owner's Schooling: Upper secondary (BS)	-2.733	0.125	0.058	-1.832	0.822*	0.039
	(1.954)	(0.251)	(0.048)	(2.643)	(0.418)	(0.033)
Owner's Schooling: Tertiary (BS)	3.988	1.885**	0.414***	8.463	2.978***	0.518
	(5.414)	(0.613)	(0.105)	(6.861)	(0.731)	(0.356)
Owner's cognitive ability (BS, IRT scale)	-0.148	0.015	0.014	0.024	0.190	0.006
	(0.365)	(0.094)	(0.017)	(0.643)	(0.150)	(0.016)
Owner currently married (BS)	-0.293	0.049	0.019	-1.147	0.451	-0.035
	(0.645)	(0.305)	(0.058)	(1.718)	(0.513)	(0.044)
Number of children in HH (BS)	1.301*	0.294*	0.043*	2.043*	0.254	0.012
	(0.545)	(0.117)	(0.021)	(0.852)	(0.194)	(0.021)
Household size (BS)	0.256	-0.039	-0.006	0.547	-0.095	-0.000
	(0.385)	(0.048)	(0.011)	(0.557)	(0.129)	(0.013)
HH asset index (score)	2.962*	0.661***	0.135***	4.166*	1.091***	0.128***
	(1.398)	(0.142)	(0.024)	(1.802)	(0.185)	(0.036)
Primary and second profits (log)	4.375***	1.324***	0.267***	8.639***	2.274***	0.173***
	(0.994)	(0.118)	(0.020)	(1.848)	(0.214)	(0.029)
Total current savings balance (IHS)	1.164**	0.405***	0.091***	2.593***	0.640***	0.121***
	(0.380)	(0.088)	(0.015)	(0.647)	(0.128)	(0.024)
Any current bank loans	1.810**	0.633**	0.133***	4.648***	0.892**	0.130**
	(0.587)	(0.199)	(0.037)	(1.331)	(0.296)	(0.050)
N	4522	4522	4522	4522	4522	4522

* p<0.05, ** p<0.01, *** p<0.001

Notes: The estimated standard errors in columns 1-5 are adjusted for clustered sampling, whereas the estimated standard errors in column 6 were bootstrapped with 100 repetitions by cluster. IHS refers to the inverse hyperbolic transformation.

Table 47.Multivariate analysis of reported investment during the last 12 months in the primary and second business during the past 12 months using alternative statistical models (ES data: least reliable 5% of observations dropped, as discussed in Appendix 2)

	Alternative statistical models					
	OLS model	OLS (winsorized) model	OLS applied to IHS transformed dependent variable	Tobit model	Two-part model	Quantile (median) regression model
	(1)	(2)	(3)	(4)	(5)	(6)
Owner's age (BS)	-0.220	-0.184*	-0.043*	-0.594	-0.257*	-0.062*
	(0.244)	(0.092)	(0.019)	(0.396)	(0.126)	(0.026)
Owner's age squared (BS)	0.002	0.002	0.000	0.005	0.002	0.001*
	(0.003)	(0.001)	(0.000)	(0.005)	(0.002)	(0.000)
Woman business owner	-1.148**	-0.645***	-0.076*	-1.459*	-0.237	-0.081*
	(0.375)	(0.192)	(0.038)	(0.673)	(0.247)	(0.032)
Owner's Schooling: Lower secondary (BS)	-0.257	0.010	0.059	1.135	0.492	0.023
	(0.363)	(0.186)	(0.041)	(0.796)	(0.327)	(0.038)
Owner's Schooling: Upper secondary (BS)	0.364	0.312	0.081	1.268	0.777*	0.035
	(0.471)	(0.208)	(0.045)	(0.919)	(0.326)	(0.045)
Owner's Schooling: Tertiary (BS)	2.795	2.134***	0.439***	5.268**	2.560***	0.512
	(1.493)	(0.610)	(0.104)	(1.953)	(0.592)	(0.294)
Owner's cognitive ability (BS, IRT scale)	-0.106	0.008	0.012	-0.064	0.150	0.005
	(0.229)	(0.083)	(0.017)	(0.375)	(0.114)	(0.016)
Owner currently married (BS)	-0.075	0.074	0.025	-0.539	0.377	-0.037
	(0.440)	(0.298)	(0.058)	(0.998)	(0.416)	(0.049)
Number of children in HH (BS)	0.705*	0.254*	0.034	1.040*	0.157	0.011
	(0.350)	(0.117)	(0.021)	(0.485)	(0.153)	(0.024)
Household size (BS)	0.056	-0.029	-0.004	0.208	-0.074	-0.001
	(0.105)	(0.045)	(0.010)	(0.187)	(0.101)	(0.014)
HH asset index (PC scale)	1.157*	0.523***	0.123***	1.792**	0.793***	0.099***
	(0.477)	(0.122)	(0.022)	(0.647)	(0.130)	(0.024)
Primary and second profits (log)	2.082***	1.034***	0.233***	4.255***	1.617***	0.135***
	(0.382)	(0.103)	(0.019)	(0.733)	(0.161)	(0.023)
Total current savings balance (IHS transform)	0.838**	0.341***	0.084***	1.585***	0.487***	0.115***
	(0.308)	(0.079)	(0.015)	(0.468)	(0.100)	(0.024)
Any current bank loans	1.247**	0.601**	0.119**	2.664***	0.699**	0.114*
	(0.379)	(0.194)	(0.037)	(0.707)	(0.236)	(0.047)
N	4311	4311	4311	4311	4311	4311

* p<0.05, ** p<0.01, *** p<0.001

Notes: The estimates reported in this table were obtained using a sample in which the 5% of observations with the largest absolute difference between reported total investment and calculated investment. The estimated standard errors in columns 1-5 are adjusted for clustered sampling, whereas the estimated standard errors in column 6 are bootstrapped with 100 repetitions by cluster. IHS refers to the inverse hyperbolic transformation.

Table 48. Estimates (unadjusted) of the impact of business training on reported investment by women business owners in their primary and second businesses using alternative statistical models (ES data only)

	Alternative statistical models					
	OLS model	OLS (winsorized) model	OLS applied to IHS transformed dependent variable	Tobit model	Two-part model	Quantile (median) regression model
	(1)	(2)	(3)	(4)	(5)	(6)
Randomly assigned to receive training (BS)	-0.051	0.049	0.049	0.700	0.290	0.190***
	(0.417)	(0.181)	(0.041)	(0.711)	(0.267)	(0.054)
Estimated effect in standard deviations	-0.002	0.007	0.038	0.024	0.010	0.007
N	2677	2677	2677	2677	2677	2677

* p<0.10, ** p<0.05, *** p<0.01 (impact estimates only)

Notes: The estimated standard errors in columns 1-5 are adjusted for clustered sampling, whereas the estimated standard errors in column 6 are bootstrapped with 100 repetitions by cluster. IHS refers to the inverse hyperbolic transformation.

Table 49. Estimates (unadjusted) of the impact of business training on reported investment by women business owners in their primary and second businesses using alternative statistical models (ES data only with least reliable 5% of observations dropped, as discussed in Appendix 2)

	OLS model	OLS (winsorized) model	OLS applied to IHS transformed dependent variable	Tobit model	Two-part model	Quantile (median) regression model
	(1)	(2)	(3)	(4)	(5)	(6)
Randomly assigned to receive training (BS)	-0.378	-0.090	0.034	0.043	0.169	0.150**
	(0.310)	(0.179)	(0.042)	(0.480)	(0.224)	(0.067)
Estimated effect in standard deviations	-0.026	-0.015	0.028	0.003	0.012	0.010
N	2603	2603	2603	2603	2603	2603

* p<0.10, ** p<0.05, *** p<0.01 (impact estimates only)

Notes: The estimated standard errors in columns 1-5 are adjusted for clustered sampling, whereas the estimated standard errors in column 6 are bootstrapped with 100 repetitions by cluster. IHS refers to the inverse hyperbolic transformation.

Table 50. Multiple regression analysis of intensity of bank use (composite variable) by alternative scales (BS data)

	Alternative scales			
Dependent variable→	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale
	(1)	(2)	(3)	(4)
Owner's age	0.048***	0.043**	0.045***	0.039**
	(0.013)	(0.014)	(0.013)	(0.014)
Owner's age squared	-0.001***	-0.001**	-0.001**	-0.001**
	(0.000)	(0.000)	(0.000)	(0.000)
Woman business owner	-0.242***	-0.227***	-0.220***	-0.200***
	(0.032)	(0.032)	(0.033)	(0.032)
Owner's Schooling: Lower secondary	0.074	0.082*	0.048	0.087*
	(0.038)	(0.039)	(0.039)	(0.039)
Owner's Schooling: Upper secondary	0.184***	0.195***	0.125**	0.209***
	(0.041)	(0.041)	(0.040)	(0.042)
Owner's Schooling: Tertiary	0.410***	0.446***	0.285***	0.510***
	(0.066)	(0.065)	(0.072)	(0.064)
Owner's cognitive ability (IRT scale)	0.038*	0.041**	0.014	0.048**
	(0.015)	(0.015)	(0.014)	(0.015)
Owner's trust in banks (IRT scale)	0.054***	0.054***	0.044**	0.050***
	(0.014)	(0.015)	(0.014)	(0.015)
Total earnings (log)	0.109***	0.104***	0.111***	0.094***
	(0.014)	(0.014)	(0.014)	(0.014)
HH wealth index (PC scale)	0.182***	0.185***	0.174***	0.182***
	(0.015)	(0.015)	(0.016)	(0.015)
Constant	-0.880***	-0.803**	-0.828***	-0.716**
	(0.251)	(0.257)	(0.250)	(0.267)
R-squared	0.117	0.117	0.097	0.112
N	4780	4780	4780	4780

* p<0.05, ** p<0.01, *** p<0.001

Notes: The estimated standard errors are adjusted for clustered sampling.

Table 51. Multiple regression analysis of intensity of bank use (composite variable) by alternative scales (ES data unless otherwise indicated)

	Alternative scales			
	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale
	(1)	(2)	(3)	(4)
Owner's age (BS)	0.046***	0.043**	0.032*	0.037**
	(0.014)	(0.014)	(0.014)	(0.014)
Owner's age squared (BS)	-0.001***	-0.001***	-0.000*	-0.001**
	(0.000)	(0.000)	(0.000)	(0.000)
Woman business owner	-0.134***	-0.121***	-0.102***	-0.100**
	(0.031)	(0.030)	(0.030)	(0.031)
Owner's Schooling: Lower secondary (BS)	0.074	0.074	0.060	0.074
	(0.039)	(0.039)	(0.040)	(0.039)
Owner's Schooling: Upper secondary (BS)	0.159***	0.165***	0.090*	0.183***
	(0.042)	(0.041)	(0.044)	(0.041)
Owner's Schooling: Tertiary (BS)	0.427***	0.459***	0.245**	0.536***
	(0.066)	(0.065)	(0.075)	(0.063)
Owner's cognitive ability (BS, IRT scale)	0.060***	0.059***	0.052***	0.057***
	(0.015)	(0.015)	(0.015)	(0.015)
Owner's trust of banks (IRT scale)	0.044**	0.045**	0.033*	0.043**
	(0.014)	(0.014)	(0.015)	(0.014)
Distance to banking services (IRT scale)	-0.059***	-0.060***	-0.036*	-0.067***
	(0.015)	(0.015)	(0.015)	(0.015)
Primary / second profits (log)	0.129***	0.125***	0.124***	0.109***
	(0.015)	(0.014)	(0.016)	(0.014)
HH asset index (PC scale)	0.156***	0.160***	0.181***	0.146***
	(0.017)	(0.018)	(0.025)	(0.016)
R-squared	0.114	0.115	0.097	0.106
N	4523	4523	4523	4523

* p<0.05, ** p<0.01, *** p<0.001

Notes: The estimated standard errors are adjusted for clustered sampling.

Table 52. Impact analysis: Impact of the business training on the Intensity of bank use (composite variable) by women business owners by alternative scales (ES)

	Alternative scales			
Dependent variable→	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale
	(1)	(2)	(3)	(4)
UNADJUSTED				
Received business training (BS)	0.022	0.022	0.011	0.007
	(0.038)	(0.038)	(0.038)	(0.039)
R-squared	0.000	0.000	0.000	0.000
N	2728	2728	2728	2728
ADJUSTED				
	Alternative scales			
Dependent variable→	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale
	(1)	(2)	(3)	(4)
Received business training (BS)	0.067**	0.066**	0.048	0.051
	(0.030)	(0.031)	(0.032)	(0.032)
Intensity of bank use (BS, PM scale)	0.563***	0.565***	0.481***	0.566***
	(0.016)	(0.016)	(0.022)	(0.016)
R-squared	0.311	0.310	0.227	0.300
N	2728	2728	2728	2728

* p<0.10, ** p<0.05, *** p<0.01 (impact estimates only)

Notes: The estimated standard errors are adjusted for clustered sampling.

Table 53.Multiple regression analysis of mobile money (MM) use (composite variable) by alternative scales (BS data)

	Alternative scales			
	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale
	(1)	(2)	(3)	(4)
Owner's age	0.015	0.017	0.016	0.016
	(0.015)	(0.015)	(0.015)	(0.015)
Owner's age squared	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Woman business owner	-0.070*	-0.034	-0.046	-0.064*
	(0.032)	(0.031)	(0.031)	(0.032)
Owner's Schooling: Lower secondary	-0.031	-0.032	-0.032	-0.029
	(0.029)	(0.034)	(0.032)	(0.031)
Owner's Schooling: Upper secondary	0.033	0.018	0.023	0.027
	(0.032)	(0.038)	(0.036)	(0.035)
Owner's Schooling: Tertiary	0.363***	0.266*	0.302**	0.345**
	(0.107)	(0.109)	(0.109)	(0.105)
Owner's cognitive ability (IRT scale)	0.005	-0.005	-0.002	-0.002
	(0.013)	(0.014)	(0.014)	(0.014)
Owner's trust of banks (IRT scale)	0.011	0.008	0.009	0.014
	(0.014)	(0.015)	(0.015)	(0.015)
Intensity of bank use (IRT scale)	0.137***	0.099***	0.113***	0.131***
	(0.014)	(0.013)	(0.013)	(0.014)
Total earnings (log)	0.034*	0.023	0.027	0.034*
	(0.015)	(0.015)	(0.014)	(0.014)
HH wealth index (PC scale)	0.103***	0.078***	0.088***	0.094***
	(0.017)	(0.018)	(0.018)	(0.017)
R-squared	0.069	0.037	0.047	0.062
N	4780	4780	4780	4780

* p<0.05, ** p<0.01, *** p<0.001

Notes: The estimated standard errors are adjusted for clustered sampling.

Table 54. Multiple regression analysis of mobile money (MM) use (composite variable) by alternative scales (ES data unless otherwise indicated)

	Alternative scales			
	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale
	(1)	(2)	(3)	(4)
Owner's age (BS)	-0.011	-0.018	-0.017	-0.021
	(0.016)	(0.017)	(0.017)	(0.018)
Owner's age squared (BS)	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Woman business owner	-0.047	-0.067	-0.052	-0.076*
	(0.034)	(0.035)	(0.034)	(0.035)
Owner's Schooling: Lower secondary (BS)	0.072*	0.067*	0.069*	0.072*
	(0.031)	(0.032)	(0.031)	(0.032)
Owner's Schooling: Upper secondary (BS)	0.120**	0.106**	0.109**	0.109**
	(0.039)	(0.040)	(0.040)	(0.039)
Owner's Schooling: Tertiary (BS)	0.503***	0.495***	0.491***	0.534***
	(0.104)	(0.106)	(0.106)	(0.104)
Owner's cognitive ability (BS, IRT scale)	0.006	0.003	0.005	0.006
	(0.015)	(0.015)	(0.015)	(0.015)
Owner's trust of banks (IRT scale)	0.029*	0.034*	0.035*	0.027
	(0.014)	(0.014)	(0.014)	(0.014)
Intensity of bank use (IRT scale)	0.182***	0.174***	0.173***	0.184***
	(0.015)	(0.015)	(0.015)	(0.015)
Distance to banking services (PC scale)	-0.012	-0.008	-0.011	-0.006
	(0.016)	(0.016)	(0.016)	(0.016)
Primary/second profits (log)	0.030	0.029	0.028	0.035*
	(0.017)	(0.017)	(0.017)	(0.017)
HH asset index (PC scale)	0.160***	0.158***	0.156***	0.157***
	(0.022)	(0.023)	(0.022)	(0.021)
R-squared	0.110	0.104	0.102	0.113
N	4163	4163	4163	4163

* p<0.05, ** p<0.01, *** p<0.001

Notes: Estimated standard errors are adjusted for clustered sampling.

Table 55. Multiple regression analysis of mobile money (MM) use (composite variable) by gender (BS, ES, proportional mean scale)

	Baseline survey (BS)			Endline survey (ES)		
Estimation sample→	All business owners	Women business owners	Men business owners	All business owners	Women business owners	Men business owners
	(1)	(2)	(3)	(4)	(5)	(6)
Owner's age (BS)	0.012	-0.001	0.022	-0.015	-0.038	-0.001
	(0.015)	(0.019)	(0.024)	(0.016)	(0.021)	(0.024)
Owner's age squared (BS)	-0.000	-0.000	-0.000	0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Woman business owner	-0.070*			-0.043		
	(0.031)			(0.032)		
Owner's Schooling: Lower secondary (BS)	-0.030	-0.024	-0.041	0.049	0.080*	0.012
	(0.029)	(0.037)	(0.050)	(0.031)	(0.033)	(0.061)
Owner's Schooling: Upper secondary (BS)	0.035	0.040	0.025	0.113**	0.152**	0.064
	(0.032)	(0.046)	(0.048)	(0.038)	(0.046)	(0.060)
Owner's Schooling: Tertiary (BS)	0.364***	0.302*	0.448*	0.509***	0.477***	0.552**
	(0.108)	(0.127)	(0.178)	(0.103)	(0.130)	(0.168)
Owner's cognitive ability (BS, IRT scale)	0.010	0.019	0.000	0.010	0.018	0.003
	(0.013)	(0.016)	(0.022)	(0.014)	(0.018)	(0.023)
Owner's trust of banks (IRT scale)	0.010	-0.001	0.027	0.025	-0.002	0.066**
	(0.015)	(0.020)	(0.021)	(0.013)	(0.016)	(0.023)
Intensity of bank use (PM scale)	0.129***	0.100***	0.162***	0.172***	0.147***	0.202***
	(0.014)	(0.016)	(0.021)	(0.015)	(0.020)	(0.023)
Primary/second profits (log)	0.020	0.010	0.035	0.030	0.028	0.025
	(0.014)	(0.016)	(0.025)	(0.016)	(0.018)	(0.029)
HH asset index (PC scale)	0.129***	0.102***	0.159***	0.153***	0.132***	0.175***
	(0.020)	(0.024)	(0.034)	(0.020)	(0.026)	(0.033)
R-squared	0.072	0.046	0.101	0.105	0.086	0.127
N	4788	2827	1961	4524	2653	1871

* p<0.05, ** p<0.01, *** p<0.001

Notes: Estimated standard errors are adjusted for clustered sampling.

Table 56. Impact analysis: Impact of the business training on mobile money (MM) use (composite variable) of women business owners by alternative scales (ES and BS data)

		Alternative scales			
Dependent variable→	Ever use of MM	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale
	(1)	(2)	(3)	(4)	(5)
UNADJUSTED					
Received business training (BS)	0.014*	0.116***	0.110***	0.111***	0.110***
	(0.008)	(0.034)	(0.033)	(0.034)	(0.034)
R-squared	0.001	0.004	0.004	0.004	0.004
N	2728	2728	2728	2728	2728
ADJUSTED					
		Alternative scales			
Dependent variable→	Ever use of MM	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale
	(1)	(2)	(3)	(4)	
Received business training (BS)	0.011	0.082**	0.079**	0.079**	0.077**
	(0.008)	(0.032)	(0.031)	(0.032)	(0.031)
Ever used MM (BS)	0.468***				
	(0.073)				
Intensity of bank use (BS, PM scale)		0.345***	0.320***	0.327***	0.336***
		(0.030)	(0.033)	(0.031)	(0.035)
R-squared	0.073	0.128	0.119	0.119	0.131
N	2728	2728	2728	2728	2728

* p<0.10, ** p<0.05, *** p<0.01 (impact estimates only)

Notes: Estimated standard errors are adjusted for clustered sampling.

Table 57. Multiple regression analysis of mobile phone use (composite variable) by alternative scales (BS data)

	Alternative scales			
	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale
	(1)	(2)	(3)	(4)
Owner's age	-0.129***	-0.127***	-0.134***	-0.118***
	(0.013)	(0.013)	(0.014)	(0.013)
Owner's age squared	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Woman business owner	-0.313***	-0.314***	-0.313***	-0.317***
	(0.025)	(0.025)	(0.025)	(0.025)
Owner's Schooling: Lower secondary	0.063*	0.063*	0.061*	0.073**
	(0.027)	(0.027)	(0.027)	(0.028)
Owner's Schooling: Upper secondary	0.350***	0.353***	0.339***	0.373***
	(0.032)	(0.032)	(0.031)	(0.032)
Owner's Schooling: Tertiary	0.920***	0.925***	0.899***	0.945***
	(0.057)	(0.057)	(0.058)	(0.056)
Owner's cognitive ability (IRT scale)	0.033**	0.033**	0.032**	0.034**
	(0.011)	(0.011)	(0.011)	(0.011)
HH asset index (PC scale)	0.205***	0.206***	0.202***	0.209***
	(0.013)	(0.013)	(0.013)	(0.013)
R-squared	0.327	0.328	0.325	0.329
N	4812	4812	4812	4812

* p<0.05, ** p<0.01, *** p<0.001

Notes: Estimated standard errors are adjusted for clustered sampling.

Table 58. Multiple regression analysis of peer connections between sample women business owners (composite variable) by alternative scales (BS data)

	Alternative scales			
	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale
	(1)	(2)	(3)	(4)
Owner's age	0.059**	0.058**	0.058**	0.066***
	(0.018)	(0.018)	(0.019)	(0.019)
Owner's age squared	-0.001**	-0.001*	-0.001*	-0.001**
	(0.000)	(0.000)	(0.000)	(0.000)
Owner's Schooling: Lower secondary	0.085	0.091	0.080	0.098
	(0.053)	(0.053)	(0.053)	(0.054)
Owner's Schooling: Upper secondary	-0.045	-0.042	-0.049	-0.028
	(0.055)	(0.055)	(0.055)	(0.057)
Owner's Schooling: Tertiary	-0.264**	-0.261**	-0.269**	-0.218*
	(0.093)	(0.094)	(0.093)	(0.096)
Owner's cognitive ability (IRT scale)	0.049*	0.050*	0.050*	0.067***
	(0.020)	(0.020)	(0.020)	(0.020)
Owner currently married	-0.124	-0.113	-0.134	-0.083
	(0.070)	(0.070)	(0.071)	(0.070)
Household size	0.027	0.028	0.026	0.028*
	(0.015)	(0.015)	(0.015)	(0.013)
Number of children in HH	0.004	0.004	0.003	0.001
	(0.026)	(0.027)	(0.026)	(0.026)
Business practices followed (PM scale)	0.143***	0.140***	0.147***	0.137***
	(0.022)	(0.022)	(0.023)	(0.021)
Intensity of mobile phone use (PC scale)	0.019	0.021	0.019	0.018
	(0.024)	(0.024)	(0.024)	(0.023)
Number of customers of primary business	0.000*	0.000*	0.000*	0.000**
	(0.000)	(0.000)	(0.000)	(0.000)
Total earnings (log)	0.063***	0.062***	0.065***	0.053**
	(0.018)	(0.018)	(0.018)	(0.019)
HH wealth index (PC scale)	0.053*	0.053*	0.052*	0.061**
	(0.024)	(0.024)	(0.024)	(0.024)
R-squared	0.054	0.053	0.055	0.056
N	2814	2814	2814	2814

* p<0.05, ** p<0.01, *** p<0.001

Notes: Estimated standard errors are adjusted for clustered sampling.

Table 59. Multiple regression analysis of participation in voluntary activities (composite variable) by alternative scales and by gender (BS data)

Estimation sample→	All business owners				Women business owners	Men business owners
Scale→	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale	Proportional mean scale	Proportional mean scale
	(1)	(2)	(3)	(4)	(5)	(6)
Owner's age	0.042**	0.045**	0.043**	0.051***	0.028	0.057*
	(0.016)	(0.016)	(0.015)	(0.015)	(0.018)	(0.026)
Owner's age squared	-0.000*	-0.000*	-0.000*	-0.001**	-0.000	-0.001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Woman business owner	-0.167***	-0.172***	-0.171***	-0.125***		
	(0.032)	(0.032)	(0.032)	(0.031)		
Owner's Schooling: Lower secondary	0.146***	0.155***	0.148***	0.100**	0.182***	0.086
	(0.035)	(0.035)	(0.034)	(0.036)	(0.040)	(0.058)
Owner's Schooling: Upper secondary	0.254***	0.259***	0.253***	0.189***	0.274***	0.223***
	(0.037)	(0.037)	(0.037)	(0.038)	(0.047)	(0.059)
Owner's Schooling: Tertiary	0.561***	0.572***	0.569***	0.497***	0.510***	0.646***
	(0.089)	(0.090)	(0.090)	(0.085)	(0.110)	(0.152)
Owner's cognitive ability (IRT scale)	0.052***	0.047**	0.050***	0.043**	0.045**	0.061*
	(0.015)	(0.015)	(0.014)	(0.014)	(0.017)	(0.024)
Owner currently married	-0.012	-0.024	-0.013	-0.009	0.082	-0.110
	(0.051)	(0.051)	(0.051)	(0.050)	(0.060)	(0.099)
Household size	-0.012	-0.009	-0.012	-0.014	-0.018*	0.005
	(0.007)	(0.007)	(0.007)	(0.008)	(0.007)	(0.021)
Number of children in HH	-0.045*	-0.048*	-0.047*	-0.038	-0.019	-0.075*
	(0.019)	(0.019)	(0.019)	(0.020)	(0.024)	(0.035)
Total earnings (log)	0.029	0.029	0.032*	0.041**	0.012	0.063*
	(0.015)	(0.015)	(0.015)	(0.015)	(0.016)	(0.029)
HH wealth index (PM score)	0.156***	0.158***	0.160***	0.166***	0.117***	0.191***
	(0.017)	(0.017)	(0.017)	(0.018)	(0.020)	(0.029)
R-squared	0.08	0.08	0.08	0.07	0.06	0.09
N	4780	4741	4780	4780	2826	1954

* p<0.05, ** p<0.01, *** p<0.001

Notes: Estimated standard errors are adjusted for clustered sampling.

Table 60. Multiple regression analysis (OLS model) of cognitive ability (composite variable) and wage and salary earnings with and without observations with obvious age misreporting Examples of the effects of age misreporting (BS)

	Cognitive ability score (IRT scale)		Wage and salary earnings (log)	
	Full sample	Without age misreporting	Full sample	Without age misreporting
	(1)	(2)	(3)	(4)
Owner's age (BS)	0.048**	0.049**	0.120***	0.096**
	(0.015)	(0.016)	(0.034)	(0.036)
Owner's age squared (BS)	-0.001**	-0.001**	-0.001**	-0.001*
	(0.000)	(0.000)	(0.000)	(0.000)
Woman business owner	-0.040	-0.020	-0.661***	-0.704***
	(0.029)	(0.032)	(0.080)	(0.086)
Owner's schooling: Lower secondary	0.257***	0.241***	0.185	0.117
	(0.041)	(0.044)	(0.109)	(0.119)
Owner's schooling: Upper secondary	0.475***	0.459***	0.271*	0.253*
	(0.037)	(0.040)	(0.109)	(0.120)
Owner's schooling: Tertiary secondary	0.628***	0.622***	0.472***	0.472***
	(0.068)	(0.073)	(0.126)	(0.141)
Days worked in wage and salary jobs (log)			0.422***	0.445***
			(0.054)	(0.059)
Hours worked in wage and salary jobs (log)			0.651***	0.675***
			(0.068)	(0.071)
Owner's cognitive ability (IRT scale)			-0.004	0.007
			(0.037)	(0.039)
	0.04	0.04	0.37	0.40
N	4812	4186	646	566

* p<0.05, ** p<0.01, *** p<0.001

Notes: Estimated standard errors are adjusted for clustered sampling.

Table 61. Multiple regression analysis of primary business profits (reported), including both age and experience, using alternative statistical models (BS data)

	Alternative statistical models				
Statistical models→	OLS regression model	OLS winsorized regression model	Log regression model	Quantile (median) regression model	Robust regression model
	(1)	(2)	(3)	(4)	(5)
Owner's age	**0.087***	**0.057***	**0.053****	**0.051****	**0.049****
	(0.042)	**(0.027)**	**(0.016)**	**(0.016)**	**(0.017)**
Owner's age squared	**-0.001***	**-0.001***	**-0.001****	**-0.001****	**-0.001****
	(0.001)	**(0.000)**	**(0.000)**	**(0.000)**	**(0.000)**
Years worked in primary business	**0.058***	**0.054*****	**0.032*****	**0.025****	**0.021****
	(0.023)	**(0.011)**	**(0.006)**	**(0.009)**	**(0.008)**
Years worked in primary business (squared)	**-0.001**	**-0.001**	**-0.001***	**-0.000**	**-0.000**
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Woman business owner	-0.744***	-0.573***	-0.414***	-0.515***	-0.447***
	(0.169)	(0.059)	(0.033)	(0.043)	(0.036)
Owner's Schooling: Lower secondary	0.136	-0.023	-0.002	0.005	0.003
	(0.143)	(0.069)	(0.041)	(0.045)	(0.038)
Owner's Schooling: Upper secondary	0.120	0.081	0.019	0.084	0.040
	(0.107)	(0.071)	(0.037)	(0.048)	(0.042)
Owner's Schooling: Tertiary	-0.322	-0.215	-0.141*	-0.144	-0.111
	(0.209)	(0.134)	(0.068)	(0.089)	(0.086)
Owner's cognitive ability (theta)	0.028	0.030	0.040**	0.032	0.031*
	(0.047)	(0.026)	(0.015)	(0.019)	(0.013)
Owner's willingness to take risks (IRT scale)	0.144	0.068*	0.051**	0.060**	0.076***
	(0.077)	(0.032)	(0.017)	(0.020)	(0.019)
Days worked by owner: primary business (log)	0.556**	0.348**	0.223***	0.286***	0.322***
	(0.174)	(0.124)	(0.066)	(0.081)	(0.068)
Hours worked by owner: primary business (log)	0.087	0.137*	0.147***	0.097*	0.141***
	(0.166)	(0.064)	(0.033)	(0.040)	(0.029)
Number of paid workers: primary business	0.321***	0.179***	0.058***	0.353***	0.343***
	(0.072)	(0.041)	(0.014)	(0.055)	(0.086)
Number of unpaid workers: primary business	-0.001	-0.000	0.003	-0.016	-0.002
	(0.035)	(0.022)	(0.011)	(0.018)	(0.017)
Total value of business capital (log)	0.412***	0.346***	0.203***	0.209***	0.158***
	(0.031)	(0.019)	(0.010)	(0.010)	(0.011)
Business practices followed (PM scale)	0.230***	0.213***	0.118***	0.135***	0.111***
	(0.058)	(0.030)	(0.015)	(0.021)	(0.018)
N	4559	4559	4556	4559	4559

* p<0.05, ** p<0.01, *** p<0.001

Notes: This model is the same as in Table 1 except for the addition of linear and quadratic measures of the number of years of experience working in the primary business (in boldface). Models also include (results not shown) 5 dummy variables for the type of business and 4 dummy variables for regency of residence, as in Table 1. Estimated standard errors in columns 1-3 are adjusted for clustered sampling, whereas the estimated standard errors in columns 4 and 5 are bootstrapped with 100 repetitions by cluster.

Table 62. Multiple regression analysis of cognitive ability (composite variable) by alternative scales (BS data)

	Alternative scales			
	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale
	(1)	(2)	(3)	(4)
Owner's age	0.046**	0.044**	0.043**	0.048**
	(0.015)	(0.016)	(0.016)	(0.015)
Owner's age squared	-0.001**	-0.001*	-0.001*	-0.001**
	(0.000)	(0.000)	(0.000)	(0.000)
Woman business owner	-0.018	-0.013	0.006	-0.040
	(0.029)	(0.030)	(0.030)	(0.029)
Owner's Schooling: Lower secondary	0.230***	0.224***	0.202***	0.257***
	(0.041)	(0.041)	(0.040)	(0.041)
Owner's Schooling: Upper secondary	0.424***	0.413***	0.371***	0.475***
	(0.038)	(0.038)	(0.038)	(0.037)
Owner's Schooling: Tertiary	0.556***	0.550***	0.497***	0.628***
	(0.068)	(0.067)	(0.066)	(0.068)
R-squared	0.031	0.030	0.024	0.040
N	4812	4808	4812	4812

* p<0.05, ** p<0.01, *** p<0.001

Notes: Estimated standard errors are adjusted for clustered sampling.

Table 63. Multiple regression analysis of willingness to take risks (BS and MS data)

Data source→	Baseline survey (BS data)		Midline survey (MS data unless otherwise indicated)	
Dependent variable→	Willingness to take risks (1-10)		Ability to get rich by taking risks (1-5)	
	PM scale	IRT scale		
	(1)	(2)	(3)	(4)
Owner's age (BS)	-0.028	-0.024	0.003	-0.001
	(0.015)	(0.015)	(0.021)	(0.021)
Owner's age squared (BS)	0.000	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Woman business owner	-0.270***	-0.261***	-0.283***	-0.286***
	(0.032)	(0.031)	(0.040)	(0.040)
Owner's Schooling: Lower secondary (BS)	-0.020	-0.022	0.157**	0.165**
	(0.042)	(0.042)	(0.051)	(0.051)
Owner's Schooling: Upper secondary (BS)	0.056	0.066	0.415***	0.419***
	(0.042)	(0.042)	(0.054)	(0.054)
Owner's Schooling: Tertiary (BS)	0.261***	0.273***	0.470***	0.486***
	(0.067)	(0.067)	(0.092)	(0.090)
Owner's cognitive ability (BS, IRT scale)	0.036*	0.034*	0.083***	0.083***
	(0.014)	(0.015)	(0.020)	(0.019)
Primary / second profits (BS/MS, log)	0.076***	0.068***	0.076***	0.074***
	(0.014)	(0.014)	(0.019)	(0.020)
HH asset index (BS/MS, PC scale)	0.014	0.019	0.022	
	(0.016)	(0.016)	(0.021)	
Household income (MS)				0.004
				(0.004)
R-squared	0.051	0.049	0.139	0.138
N	4787	4787	2261	2230

* p<0.05, ** p<0.01, *** p<0.001

Notes: Dependent variables in columns 1-4 are standardized to make the estimated coefficients comparable. Estimated standard errors are adjusted for clustered sampling. BS/MS indicates that BS data are used in columns 1-2 and MS data are used for the same variable in columns 3-4. HH income was only measured directly in the MS. Estimated standard errors are adjusted for clustered sampling.

Table 64. Multiple regression analysis of subjective time preference (composite variable) by alternative scales (BS data)

	Alternative scales			
	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale
	(1)	(2)	(3)	(4)
Owner's age	0.013	0.013	0.012	0.013
	(0.015)	(0.014)	(0.015)	(0.014)
Owner's age squared	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Woman business owner	-0.121***	-0.157***	-0.079*	-0.155***
	(0.032)	(0.032)	(0.032)	(0.032)
Owner's Schooling: Lower secondary	-0.007	-0.036	0.020	-0.029
	(0.041)	(0.041)	(0.041)	(0.040)
Owner's Schooling: Upper secondary	-0.084*	-0.115**	-0.051	-0.101*
	(0.041)	(0.040)	(0.042)	(0.040)
Owner's Schooling: Tertiary	-0.127	-0.142*	-0.100	-0.162*
	(0.074)	(0.072)	(0.075)	(0.074)
Owner's cognitive ability (IRT scale)	-0.073***	-0.090***	-0.053***	-0.087***
	(0.015)	(0.015)	(0.015)	(0.015)
Intensity of bank use (PM scale)	-0.026	-0.029	-0.023	-0.022
	(0.016)	(0.015)	(0.016)	(0.015)
Total earnings (log)	-0.009	-0.008	-0.009	-0.003
	(0.014)	(0.014)	(0.014)	(0.014)
HH wealth index (PC scale)	-0.014	-0.020	-0.008	-0.019
	(0.017)	(0.017)	(0.016)	(0.017)
R-squared	0.017	0.024	0.011	0.022
N	4780	4780	4780	4780

* p<0.05, ** p<0.01, *** p<0.001

Notes: Estimated standard errors are adjusted for clustered sampling.

Table 65. Multiple regression analysis of total savings during the last 12 months using alternative statistical models (BS data)

	Alternative statistical models					
	OLS regression model	OLS winsorized regression model	Tobit model	Two-part model	Quantile (median) regression model	Robust regression model
	(1)	(2)	(3)	(4)	(5)	(6)
Owner's age	-0.826	0.236	0.592	0.163	-0.019	-0.012
	(1.299)	(0.160)	(1.309)	(0.237)	(0.065)	(0.051)
Owner's age squared	0.009	-0.004	-0.020	-0.004	-0.000	-0.000
	(0.017)	(0.002)	(0.016)	(0.003)	(0.001)	(0.001)
Woman business owner	-3.538	0.445	15.314**	1.227**	0.694***	0.714***
	(2.500)	(0.343)	(5.927)	(0.471)	(0.149)	(0.105)
Owner's Schooling: Lower secondary	1.215	0.066	5.454	0.607	0.163	0.125
	(1.551)	(0.373)	(3.455)	(0.616)	(0.196)	(0.109)
Owner's Schooling: Upper secondary	3.161	0.845*	10.373	1.727**	0.423*	0.345*
	(3.310)	(0.406)	(6.521)	(0.640)	(0.196)	(0.135)
Owner's Schooling: Tertiary	4.510	2.767**	12.918*	3.573***	1.946**	0.853*
	(4.351)	(0.981)	(5.851)	(1.002)	(0.707)	(0.361)
Owner's cognitive ability (IRT scale)	-0.248	0.167	1.793	0.370	0.128	0.153**
	(0.545)	(0.160)	(1.055)	(0.219)	(0.079)	(0.050)
Subjective time preference (proportional mean scale)	**1.774**	**-0.365***	**1.346**	**-0.486***	**-0.127**	**-0.102***
	(1.934)	**(0.145)**	**(2.075)**	**(0.225)**	**(0.078)**	**(0.050)**
Owner currently married	-1.515	0.215	-0.830	0.307	0.041	0.030
	(1.729)	(0.476)	(3.065)	(0.742)	(0.225)	(0.170)
Household size	-0.829	0.004	-0.756	-0.096	-0.021	-0.046
	(0.889)	(0.088)	(1.010)	(0.132)	(0.056)	(0.030)
Number of children in HH	5.471	-0.370	5.679	-0.583*	-0.056	-0.099
	(4.859)	(0.196)	(5.404)	(0.288)	(0.089)	(0.070)
Intensity of bank use (IRT scale)	2.192**	1.201***	6.157**	1.805***	0.510***	0.310***
	(0.708)	(0.161)	(2.029)	(0.237)	(0.091)	(0.064)
Total earnings (log)	5.675***	2.704***	10.591**	4.365***	0.855***	0.563***
	(1.537)	(0.179)	(3.408)	(0.307)	(0.089)	(0.074)
HH asset index (PC scale)	6.410**	2.357***	7.925**	3.037***	0.950***	0.345***
	(2.137)	(0.221)	(2.840)	(0.296)	(0.120)	(0.090)
N	4754	4754	4754	4754	4754	4753

* p<0.05, ** p<0.01, *** p<0.001

Notes: This is the same savings model as in Table 36, but with the addition of subjective time preference (in boldface) as an additional explanatory variable. The estimated standard errors in columns 1-4 are adjusted for clustered sampling, while those in columns 5 and 6 are bootstrapped with 100 repetitions by cluster.

Table 66. Multiple regression analysis of access to bank accounts (composite variable) by alternative scales and by gender (BS data)

	Alternative scales				Gender	
	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale	Women business owners	Men business owners
	(1)	(2)	(3)	(4)	(5)	(6)
Owner's age	0.041**	0.046***	0.047***	0.043**	0.044*	0.039
	(0.014)	(0.014)	(0.014)	(0.014)	(0.018)	(0.022)
Owner's age squared	-0.001**	-0.001**	-0.001**	-0.001**	-0.001*	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Woman business owner	-0.131***	-0.152***	-0.140***	-0.126***		
	(0.031)	(0.030)	(0.030)	(0.030)		
Owner's Schooling: Lower secondary	0.089*	0.086*	0.092*	0.088*	0.044	0.128*
	(0.039)	(0.038)	(0.038)	(0.038)	(0.047)	(0.060)
Owner's Schooling: Upper secondary	0.205***	0.203***	0.216***	0.212***	0.110*	0.329***
	(0.043)	(0.042)	(0.042)	(0.041)	(0.055)	(0.061)
Owner's Schooling: Tertiary	0.545***	0.634***	0.613***	0.659***	0.569***	0.774***
	(0.061)	(0.063)	(0.064)	(0.065)	(0.085)	(0.099)
Owner's cognitive ability (IRT scale)	0.048***	0.046**	0.045**	0.048***	0.061***	0.033
	(0.014)	(0.014)	(0.014)	(0.014)	(0.018)	(0.021)
Household size	-0.010	-0.009	-0.011	-0.009	-0.012	-0.004
	(0.008)	(0.007)	(0.007)	(0.007)	(0.008)	(0.015)
Owner currently married	-0.103*	-0.129**	-0.109*	-0.100*	-0.164*	-0.027
	(0.050)	(0.048)	(0.049)	(0.049)	(0.066)	(0.079)
Total earnings (log)	0.102***	0.106***	0.108***	0.115***	0.094***	0.153***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.017)	(0.024)
HH wealth index (PC scale)	0.245***	0.244***	0.248***	0.260***	0.259***	0.254***
	(0.015)	(0.015)	(0.015)	(0.015)	(0.019)	(0.025)
R-squared	0.146	0.157	0.159	0.171	0.148	0.184
N	4780	4740	4780	4780	2826	1954

* p<0.05, ** p<0.01, *** p<0.001

Notes: Estimated standard errors are adjusted for clustered sampling.

Table 67. Multiple regression analysis of distance to banking services (composite variable) by alternative scale (ES data)

	Alternative scales			
	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale
	(1)	(2)	(3)	(4)
Urban (as distinct from semi-urban) village	-0.611***	-0.638***	-0.626***	-0.713***
	(0.079)	(0.086)	(0.081)	(0.090)
R-squared	0.069	0.072	0.068	0.089
N	4633	4261	4633	4633

* p<0.05, ** p<0.01, *** p<0.001

Notes: These models also include 28 dummy variables representing the 29 sample sub-districts (not shown in the table). The 28 dummy variables are jointly significant at the 0.001 level in columns 1-4.

Table 68. Multiple regression analysis of access to bank accounts (composite variable) by alternative scale (ES data)

	Alternative scales			
	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale
	(1)	(2)	(3)	(4)
Owner's age (BS)	0.039**	0.046***	0.047***	0.045***
	(0.013)	(0.014)	(0.014)	(0.014)
Owner's age squared (BS)	-0.001**	-0.001**	-0.001**	-0.001**
	(0.000)	(0.000)	(0.000)	(0.000)
Woman business owner	-0.064*	-0.077*	-0.073*	-0.063*
	(0.030)	(0.031)	(0.031)	(0.030)
Owner's Schooling: Lower secondary (BS)	0.072	0.079*	0.081*	0.081*
	(0.038)	(0.038)	(0.038)	(0.038)
Owner's Schooling: Upper secondary (BS)	0.179***	0.196***	0.201***	0.200***
	(0.040)	(0.041)	(0.040)	(0.040)
Owner's Schooling: Tertiary (BS)	0.609***	0.700***	0.695***	0.695***
	(0.061)	(0.068)	(0.068)	(0.066)
Owner's cognitive ability (IRT scale)	0.054***	0.052***	0.051***	0.052***
	(0.014)	(0.015)	(0.015)	(0.015)
Distance to banking services (IRT scale)	**-0.064*** **	**-0.063*** **	**-0.063*** **	**-0.064*** **
	(0.014)	**(0.014)**	**(0.014)**	**(0.014)**
Primary/second profits (log)	0.101***	0.110***	0.109***	0.110***
	(0.014)	(0.014)	(0.014)	(0.014)
HH asset index (PC scale)	0.179***	0.188***	0.193***	0.200***
	(0.014)	(0.015)	(0.015)	(0.015)
R-squared	0.118	0.131	0.134	0.138
N	4523	4521	4523	4523

* p<0.05, ** p<0.01, *** p<0.001

Notes: This model is similar to the model in Table 68 except that it includes the bank distance composite variable (in boldface) and is estimated with ES data. Estimated standard errors are adjusted for clustered sampling.

Table 69. Multiple regression analysis of trust in banks (composite variable) by alternative scales (BS data)

	Alternative scales			
	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale
	(1)	(2)	(3)	(4)
Owner's age	0.036*	0.038*	0.040*	0.042**
	(0.015)	(0.015)	(0.016)	(0.015)
Owner's age squared	-0.000*	-0.000*	-0.000*	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)
Woman business owner	0.101**	0.101**	0.099**	0.056
	(0.032)	(0.032)	(0.032)	(0.031)
Owner's Schooling: Lower secondary	-0.064	-0.066	-0.056	-0.124**
	(0.041)	(0.041)	(0.041)	(0.040)
Owner's Schooling: Upper secondary	-0.028	-0.029	-0.026	-0.084*
	(0.041)	(0.041)	(0.041)	(0.040)
Owner's Schooling: Tertiary	0.044	0.039	0.042	-0.094
	(0.073)	(0.073)	(0.072)	(0.073)
Owner's cognitive ability (IRT scale)	0.035*	0.036*	0.047**	0.028
	(0.015)	(0.016)	(0.016)	(0.015)
Number of banks with accounts	0.077**	0.082***	0.102***	0.088***
	(0.024)	(0.024)	(0.024)	(0.025)
Prior bad experience with banks	-0.292***	-0.295***	-0.316***	-0.228***
	(0.063)	(0.063)	(0.065)	(0.063)
Total earnings (log)	0.045**	0.046**	0.046**	0.049**
	(0.016)	(0.016)	(0.016)	(0.015)
HH wealth index (PC scale)	-0.038*	-0.038*	-0.026	-0.046**
	(0.017)	(0.017)	(0.017)	(0.017)
R-squared	0.015	0.015	0.019	0.017
N	4780	4780	4780	4780

* p<0.05, ** p<0.01, *** p<0.001

Notes: The HH wealth index is the HH asset index with the addition of housing characteristics. Estimated standard errors are adjusted for clustered sampling.

Table 70. Multiple regression analysis of help received by business owners from other HH members (composite variable) by alternative scales and by gender (ES data unless otherwise indicated)

	Alternative scales				Gender	
	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale	Women business owners	Men business owners
	(1)	(2)	(3)	(4)	(5)	(6)
Owner's age (BS)	-0.081***	-0.060***	-0.065***	-0.048**	-0.056*	-0.019
	(0.019)	(0.018)	(0.017)	(0.017)	(0.022)	(0.025)
Owner's age squared (BS)	0.001**	0.000	0.000*	0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Woman business owner (BS)	-0.002	-0.009	-0.051	0.033		
	(0.031)	(0.030)	(0.031)	(0.030)		
Owner's Schooling: Lower secondary (BS)	-0.001	0.011	0.000	0.019	0.063	-0.035
	(0.037)	(0.036)	(0.037)	(0.035)	(0.048)	(0.046)
Owner's Schooling: Upper secondary (BS)	0.101**	0.106**	0.096*	0.115**	0.158**	0.065
	(0.038)	(0.037)	(0.038)	(0.037)	(0.052)	(0.044)
Owner's Schooling: Tertiary (BS)	0.267***	0.337***	0.303***	0.330***	0.313**	0.316**
	(0.079)	(0.079)	(0.076)	(0.075)	(0.100)	(0.110)
Owner's cognitive ability (BS, IRT scale)	-0.001	-0.006	-0.003	-0.007	-0.005	-0.015
	(0.013)	(0.013)	(0.013)	(0.012)	(0.018)	(0.019)
Number of hours worked by owner in a typical month (log)	-0.027	-0.002	-0.031	0.009	-0.022	0.039
	(0.022)	(0.020)	(0.022)	(0.020)	(0.028)	(0.029)
Number of paid workers: primary business	0.014	0.011	0.014*	0.008	0.001	0.013
	(0.007)	(0.006)	(0.007)	(0.006)	(0.012)	(0.007)
Number of unpaid workers: primary business	-0.159***	0.101***	-0.176***	0.164***	0.189***	0.137***
	(0.036)	(0.019)	(0.041)	(0.028)	(0.026)	(0.039)
Household size (BS)	0.052*	0.063**	0.058*	0.066**	0.058*	0.097***
	(0.021)	(0.023)	(0.023)	(0.022)	(0.024)	(0.020)
Number of children in HH	-0.200***	-0.222***	-0.225***	-0.219***	-0.236***	-0.220***
	(0.022)	(0.023)	(0.023)	(0.023)	(0.026)	(0.026)
Number of working-age women in HH	0.275***	0.315***	0.234***	0.370***	0.491***	0.178***
	(0.031)	(0.030)	(0.031)	(0.031)	(0.041)	(0.036)
(Number of working-age males in HH	0.056*	0.050*	0.033	0.013	-0.010	0.058
	(0.025)	(0.025)	(0.026)	(0.024)	(0.031)	(0.036)
Primary/second profits (log)	0.001	-0.005	0.005	-0.006	0.024	-0.061**
	(0.014)	(0.014)	(0.014)	(0.014)	(0.018)	(0.022)
Total value of business capital (log)	-0.027**	-0.014	-0.017	-0.014	-0.006	-0.021
	(0.010)	(0.009)	(0.010)	(0.009)	(0.013)	(0.012)
HH asset index (PC scale)	0.010	-0.000	-0.001	-0.004	-0.029	0.043*
	(0.015)	(0.016)	(0.015)	(0.015)	(0.021)	(0.021)
R-squared	0.184	0.211	0.201	0.248	0.270	0.232
N	4422	4422	4422	4422	2584	1838

* p<0.05, ** p<0.01, *** p<0.001

Notes: Estimated standard errors are adjusted for clustered sampling

Table 71. Multiple regression analysis of the HH asset index (composite variable) by alternative scales and by gender (BS data)

					Gender	
	Proportional mean scale	Principal component scale	Standardized mean scale	IRT scale	Women business owners	Men business owners
	(1)	(2)	(3)	(4)	(5)	(6)
Owner's age	0.004	-0.001	0.000	-0.001	0.010	-0.025
	(0.016)	(0.016)	(0.016)	(0.015)	(0.020)	(0.024)
Owner's age squared	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Woman business owner	0.262***	0.268***	0.246***	0.276***		
	(0.032)	(0.032)	(0.032)	(0.031)		
Owner's Schooling: Lower secondary	0.185***	0.156***	0.145***	0.194***	0.191***	0.075
	(0.040)	(0.040)	(0.042)	(0.040)	(0.048)	(0.072)
Owner's Schooling: Upper secondary	0.370***	0.351***	0.317***	0.403***	0.411***	0.241***
	(0.040)	(0.040)	(0.042)	(0.041)	(0.052)	(0.068)
Owner's Schooling: Tertiary	0.683***	0.726***	0.607***	0.786***	0.762***	0.668***
	(0.078)	(0.080)	(0.082)	(0.072)	(0.090)	(0.136)
Owner's cognitive ability (IRT scale)	0.009	-0.003	-0.005	0.005	-0.012	0.009
	(0.015)	(0.015)	(0.015)	(0.015)	(0.017)	(0.025)
Owner currently married	0.191***	0.168**	0.193***	0.174**	0.240***	0.083
	(0.053)	(0.052)	(0.055)	(0.053)	(0.069)	(0.086)
Household size	0.062***	0.056***	0.058***	0.061***	0.050**	0.070***
	(0.016)	(0.017)	(0.016)	(0.017)	(0.019)	(0.019)
Number of children in HH	0.029	0.039	0.028	0.044	0.027	0.068
	(0.022)	(0.023)	(0.023)	(0.022)	(0.028)	(0.035)
Total value of business capital (log)	0.151***	0.160***	0.149***	0.157***	0.144***	0.181***
	(0.010)	(0.010)	(0.010)	(0.009)	(0.012)	(0.016)
Any current bank loans	0.089**	0.084**	0.099**	0.085**	0.082*	0.078
	(0.030)	(0.031)	(0.032)	(0.030)	(0.041)	(0.047)
Total personal savings in past 12 months (log)	0.108***	0.110***	0.107***	0.108***	0.114***	0.099***
	(0.010)	(0.010)	(0.011)	(0.010)	(0.013)	(0.017)
Total earnings (log)	0.091***	0.095***	0.098***	0.082***	0.072***	0.154***
	(0.017)	(0.017)	(0.017)	(0.016)	(0.020)	(0.030)
R-squared	0.258	0.276	0.244	0.277	0.244	0.332
N	3544	3544	3544	3544	2241	1303

* p<0.05, ** p<0.01, *** p<0.001

Notes: Estimated standard errors are adjusted for clustered sampling.

Table 72. Multiple regression analysis of primary business registered with the government by gender (BS data)

	(1)	(2)	(3)
	Total	Females	Males
	b/se	b/se	b/se
Owner's age	-0.006	-0.009	-0.004
	(0.005)	(0.006)	(0.009)
Owner's age squared	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
Years worked in primary business	0.006***	0.007***	0.004
	(0.002)	(0.003)	(0.004)
Years worked squared	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)
Woman business owner	-0.028***		
	(0.011)		
Owner's Schooling: Lower secondary	0.001	-0.001	0.001
	(0.013)	(0.015)	(0.022)
Owner's Schooling: Upper secondary	0.006	-0.004	0.019
	(0.013)	(0.017)	(0.022)
Owner's Schooling: Tertiary	-0.013	-0.059**	0.052
	(0.025)	(0.026)	(0.046)
Owner's cognitive ability (IRT scale)	-0.008*	-0.003	-0.015*
	(0.005)	(0.006)	(0.008)
Adherence to recommended business practices (SM scale))	0.042***	0.039***	0.041***
	(0.006)	(0.008)	(0.010)
Total business profits (log)	0.026***	0.023***	0.031***
	(0.005)	(0.006)	(0.009)
HH asset index (PC scale)	0.046***	0.028***	0.068***
	(0.006)	(0.007)	(0.010)
N	4784	2827	1957

* p<0.05, ** p<0.01, *** p<0.001

Notes: Estimated standard errors are adjusted for clustered sampling.

Appendix 1. Detailed description of the data set

			Questionnaires		
Element / Dimension	**Question**	**Coding**	**BS**	**MS**	**ES**
ACHIEVEMENTS					
Personal income	Average monthly profit in primary business in past year	Rupiah	H16_H1A	N22_H1A	N22_H1A
Personal income	Average monthly profit in second business in past year	Rupiah	H16_H1B	N22_H1B	N22_H1B
Personal income	Average monthly profit from other work activities	Rupiah	I07_Act1, I07_Act2		
Personal income	Average monthly wage/salaries from other work activities	Rupiah	I08_Act1, I08_Act2		
Personal income	Average monthly business revenue (including production for own consumption)	Rupiah		N16_H1A	
Personal income	BE1: Average monthly expenses to pay employees	Rupiah		N17_H1A	
Personal income	BE2: Average monthly expenses on goods, inventory, stock)	Rupiah		N18_H1A	
Personal income	BE3: Average monthly operational expenses	Rupiah		N19_H1A	
Personal income	BE4: Average monthly expenses for electricity	Rupiah		N20_H1A	
Personal income	BE5: Average monthly expenses for internet	Rupiah		N21_H1A	
Personal income	At least one other work activity in past year	1 Yes, 2 No	I01_Act1		
Personal income	At least two other work activities in past year	1 Yes, 2 No	I01_Act2		
Personal income	Type of first other work activity	1 Agriculture, 2 Non-agricultural	I02_Act1		
Personal income	Type of second other work activity	1 Agriculture, 2 Non-agricultural	I02_Act1		
Personal income	Type of income from first other work activity	1 Profit, 6 Wage/salary	I07x_Act1		
Personal income	Type of income from second other work activity	1 Profit, 6 Wage/salary	I07x_Act2		
Personal assets	Current savings balance	A Formal bank account, B Electronic savings account, C Hiding place at home, D With friends or family, E Cooperative, F Informal saving network, G BMT, H ROSCA, I Other HH member's saving, J Physical assets (e.g., jewelry), K Simakmur (Bank Mandiri), L Other e-savings, M LAKUPANDAI (other than Bank Mandiri), V Other X Refuse to answer (+ Endline Survey I LAKU PANDAi/ SIMAKMUR J LKD/E-CASH no codes K L M)		G03a	G03a (some changes from MS)
Business assets (K=business capital)	K1: Value of own shop premises	Rupiah (if have)	N01	N01	N01

			Questionnaires		
Element / Dimension	**Question**	**Coding**	**BS**	**MS**	**ES**
Business assets (K=business capital)	K2: Value of advances paid for rented shop premises	Rupiah (if have)	N02	N02	N02
Business assets (K=business capital)	K3:Value of furniture and fixtures	Rupiah (if have)	N03	N03	N03
Business assets (K=business capital)	K4: Value of equipment	Rupiah (if have)	N04	N04	N04
Business assets (K=business capital)	K5: Value of product stock	Rupiah (if have)	N05	N05	N05
Business assets (K=business capital)	K6: Value of other business assets	Rupiah (if have)	N06	N06	N06
Business assets	Type of primary business	1 Grocery, 2 Restaurant, 3 Retail shop, 4 Services, 5 Processing, 6 Other	H02_H1A	H02_H1A	H02_H1A
Business assets	Has second business	1 Yes, 3 No	H01_H1B	H01_H1b	H01_H1b
Business assets	When did primary business start?	1 < 1 year ago, 2 1-5 years ago, 3 5-10 years ago, 4 > 10 years ago	H09_H1A	H09_H1A	H09_H1A
Business assets	Is primary business registered with government?	1 Yes, 3 No	H08_H1A	H08_H1A	H08_H1A
Household income	Household income in last month	Rupiah		I0	
Household income	Low end estimate in the last year	Rupiah		I01	
Household income	High end estimate in the last year	Rupiah		I02	
Household income	Housing characteristics (HC)	Composite indicator (see items below)	L01-L09		
Household income	HC1: Dwelling status	1 Self-owned, 2 Occupied, 3 Rented/contracted, 5 Other	L01		
Household income	HC2: Number of rooms	Number of rooms	L02		
Household income	HC3: Number of bedrooms	Number of bedrooms	L03		
Household income	HC4: Material of walls	01 Brick, 02 Wall (?), 03 Prefab brick, 04 Wood, 05 Zinc, 06 Clay, 07 Bamboo, 08 Canvas, cloth, 09 Concrete block, 95 Other	L04		
Household income	HC5: Material of roof	01 Brick, 02 Concrete block, 03 Prefab brick, 04 Wood, 05 Zinc sheets, 06 Clay, 07 Bamboo, 08 Canvas, cloth, 09 Concrete, 10 Roof tile, 11 Shingle, 12 Zinc, 13 Asbestos, 14 Palm fibers, 95 Other	L05		
Household income	HC6: Dwelling utilizes electricity	1 Yes, 3 No	L06		
Household income	HC7: Main source of drinking water	01 Pipe water, 02 Mineral water, 03 Well/pump (electric, hand), 04 Well water, 05 Spring water, 06 Rain water, 07 River/creek water, 08 Pond/fishpond, 09 Collection basin, 95 Other	L07		
Household income	HC8: Is water used for non-drinking purposes drawn from the same source?	1 Yes →L10, 3 No	L08		
Household income	HC9: Water source for non-drinking purposes	01 Pipe water, 02 Mineral water, 03 Well/pump (electric, hand), 04 Well	L09		

			Questionnaires		
Element / Dimension	**Question**	**Coding**	**BS**	**MS**	**ES**
		water, 05 Spring water, 06 Rain water, 07 River/creek water, 08 Pond/fishpond, 09 Collection basin, 95 Other			
Household income	DA1: TV	1 Yes, 3 No (+ number in Midline Survey)	Ma	Ma	Ma
Household income	DA2: DVD/VCD	Same coding as Ma	Mb	Mb	Mb
Household income	DA3: Satellite dish	Same coding as Ma	Mc	Mc	Mc
Household income	DA4: Microwave	Same coding as Ma	Md	Md	Md
Household income	DA5: Refrigerator	Same coding as Ma	Me	Me	Me
Household income	DA6: Gas cylinder (3 Kg +)	Same coding as Ma	Mf	Mf	Mf
Household income	DA7: Washing machine	Same coding as Ma	Mg	Mg	Mg
Household income	DA8: Air conditioner	Same coding as Ma	Mh	Mh	Mh
Household income	DA9: Telephone	Same coding as Ma	Mi	Mi	Mi
Household income	DA10: Simple hand phone	Same coding as Ma	Mj	Mj	Mj
Household income	DA11: Smart phone	Same coding as Ma	Mk	Mk	Mk
Household income	DA12: Computer/laptop	Same coding as Ma	Ml	Ml	Ml
Household income	DA13: Tablet	Same coding as Ma	Mm	Mm	Mm
Household income	DA14: Handycam/ camera	Same coding as Ma	Mn	Mn	Mn
Household income	DA15: Water heater	Same coding as Ma	Mo	Mo	Mo
Household income	DA16: Electric pump/ jet pump	Same coding as Ma	Mp	Mp	Mp
Household income	DA17: Generator	Same coding as Ma	Mq	Mq	Mq
Household income	DA18: Car / truck	Same coding as Ma	Mr	Mr	Mr
Household income	DA19: Boat/ motor boat	Same coding as Ma	Ms	Ms	Ms
Household income	DA20: Motor cycle / motorbike	Same coding as Ma	Mt	Mt	Mt
Quality of life (leisure time, overall well-being)	Subjective well-being (overall)	1 Very unhappy, 2 Somewhat unhappy, 3 Neither happy nor unhappy, 4 Somewhat happy, 5 Very happy		SW01	
Quality of life (leisure time, overall well-being)	Work/job satisfaction	1 Very unsatisfied, 2 Somewhat unsatisfied, 3 Neither satisfied nor unsatisfied, 4 Somewhat satisfied, 5 Very satisfied		SW02	
Vulnerability to shocks (income instability)	Concern about food shortage in past 7 days (also Current savings balance above)	1 Yes, 3 No	L10		
AGENCY/EMPOWERMENT					
Agency (Personal: AS=assertiveness)	AS1: Degree of comfort speaking out at a meeting of other (women/men)	1 Not at all comfortable, 2 Great difficulty, 3 Little difficulty, 4 Fairly comfortable, 5 Very comfortable		AS01	
Agency (Personal: AS=assertiveness)	AS2: Degree of comfort talking to people who work	Same coding as AS01		AS03	

			Questionnaires		
Element / Dimension	**Question**	**Coding**	**BS**	**MS**	**ES**
	for you about a disagreement				
Agency (Personal: AS=assertiveness)	AS3: Degree of comfort refusing someone who has asked to pay less than a fair price	Same coding as AS01		AS04	
Agency (Personal: AS=assertiveness)	AS4: Degree of comfort bargaining with a supplier over price	Same coding as AS01		AS05	
Agency (Personal: AS=assertiveness)	AS5: Degree of comfort speaking out about a money issue with your spouse	Same coding as AS01		AS06	
Agency (Personal: IA= individual attributes)	IA1: I plan tasks carefully	1 Strongly disagree, 2 Disagree, 3 Neither agree nor disagree, 4 Agree, 5 Strongly agree		AT14	
Agency (Personal: IA= individual attributes)	IA 2; I save regularly	Same coding as AT14		AT15	
Agency (Personal: IA= individual attributes)	IA 3: I can think of many times when I persisted with work when others quit	Same coding as AT1		AT17	
Agency (Personal: IA= individual attributes)	IA 4: A person can get rich by taking risks	Same coding as AT1		AT21	
Agency (Personal: IA= individual attributes)	IA 5: I would rather direct an activity rather than just help out	Same coding as AT1		AT24	
Agency (Personal: IA= individual attributes)	IA 6: I try harder when I'm in competition with others	Same coding as AT1		AT25	
Agency (Personal: IA= individual attributes)	IA 7: I enjoy planning things and deciding what others should do	Same coding as AT1		AT27	
Agency (Personal: IA= individual attributes)	IA 8: I like to have a lot of control over the events around me	Same coding as AT1		AT29	
Agency (Personal: IA= individual attributes)	IA 9: My family and friends would say I am a very organized person	Same coding as AT1		AT31	
Agency (Personal: IA= individual attributes)	IA 10: Overall I expect more good things to happen to me than bad	Same coding as AT1		AT34	
Agency (Personal: AS=assertiveness)	AS1: Degree of comfort speaking out at a meeting of other (women/men)	1 Not at all comfortable, 2 Great difficulty, 3 Little difficulty, 4 Fairly comfortable, 5 Very comfortable		AS01	
Agency (Personal: AS=assertiveness)	AS2: Degree of comfort talking to people who work for you about a disagreement	Same coding as AS01		AS03	
Agency (Personal: AS=assertiveness)	AS3: Degree of comfort refusing someone who has asked to pay less than a fair price	Same coding as AS01		AS04	
Agency (Personal: AS=assertiveness)	AS4: Degree of comfort bargaining with a supplier over price	Same coding as AS01		AS05	
Agency (Personal: AS=assertiveness)	AS5: Degree of comfort speaking out about a money issue with your spouse	Same coding as AS01		AS06	

			Questionnaires		
Element / Dimension	**Question**	**Coding**	**BS**	**MS**	**ES**
Agency (Marital)	SR1: Out of every Rupiah 100,000 of your business earnings, how much does spouse know about?	Rupiah (<100,000)	O05	O05	O05
Agency (Marital)	SR2: Out of every Rupiah 100,000 of your business earnings, how much go to HH expenses?	Rupiah (<100,000)	O06	O06	O06
Agency (Marital)	SR3: Out of every Rupiah 100,000 of HH expenses, how much comes from your spouse?	Rupiah (<100,000)	O07	O07	O07
Agency (Marital)	SR4: In the past 12 months, did your spouse ever ask you for money that you did not want to give?	1 Yes, 3 No	O08	O08	O08
Agency (Marital)	SR5: Who has access to the money you earn from your business?	A Only myself, B Myself and spouse jointly, C Spouse only, D Myself and other persons (not my spouse)	O09	O09	O09
Agency (Marital)	SR6: Who decides how money from your business will be spent?	A Myself, B Spouse C Other HH member, D Other persons (not a HH member) (multiple responses possible)	O10	O10	O10
Agency (Marital)	SR7: Who has the most influence in this decision?	1 Myself, 2 Spouse, 3 Other HH member, 4 Other persons (not a HH member)			EO10a
Agency (Marital)	SR8: Is there some money that you have sole control over?	1 Yes, 3 No	O12	O12	O12
Agency (Marital)	SR9: Your involvement in decisions about how your spouse's earnings are spent	1 Spouse alone, 2 Spouse primarily after consulting me, 3 Spouse and I have equal say, 4 Me primarily after consulting spouse, 5 I alone, 6 Spouse has no income	O13	O13	O13
Agency (Intra-HH: HD=HH decision-making)	HD1: Who decides whether to buy an appliance	1 Business owner, 2 Business owner and spouse, 3 Spouse, 4 Business owner and other HH member (not spouse), 5 Only other HH member (+ in ES: 6 HH head, 7 HH head and spouse, 8 Spouse of HH head)	O01_a	O01_a	O01_a
Agency (Intra-HH: HD=HH decision-making)	HD2: Who decides who can work outside the home	same coding as O01_a	O01_b	O01_b	O01_b
Agency (Intra-HH: HD=HH decision-making)	HD3: Who decides whether to support family members	same coding as O01_a	O01_c	O01_c	O01_c
Agency (Intra-HH: HD=HH decision-making)	HD4: Who decides whether to save for the future	same coding as O01_a	O01_d	O01_d	O01_d
Agency (Intra-HH: HD=HH decision-making)	HD5: Who decides whether to sign up for a new banking product	same coding as O01_a	O01_e	O01_e	O01_e
Saving	Any savings in last 12 months?	1 Yes, 3 No	G01	G01 (last 3 months)	G01
Saving	Amount saved in last 12 months	A Formal bank account, B Electronic savings account, C Hiding place at	G03	G03 (some changes from	G03 (some changes from

			Questionnaires		
Element / Dimension	**Question**	**Coding**	**BS**	**MS**	**ES**
		home, D With friends or family, E Cooperative, F Informal saving network, G BMT, H ROSCA, V Other, X Refuse to answer (+Midline Survey: I Other HH member's saving, J Physical assets (e.g., jewelry), K Simakmur (Bank Mandiri), L Other e-savings, M LAKUPANDAI (other than Bank Mandiri)) (+ Endline Survey: I LAKU PANDAi/SIMAKMUR, J LKD/E-CASH)		Baseline Survey)	Baseline and Midline Surveys)
Saving	(If no) Reasons why no savings?	1 No money, 2 Don't know where to save, 3 Spouse is one saving, 4 People will ask to borrow money, 5 Other reason	G02	G02	G02
Saving	Type of savings instrument for emergencies	A Formal bank account, B Electronic savings account, C Hiding place at home, D With friends/family, E Cooperative, F Informal saving network, G BMT, H ROSCA, I Other HH member saving, J Sold the assets, V Other, X Refused to answer (+Midline Survey: K SIMAKMUR savings account at Bank Mandiri, L e-savings account at other bank, M LAKUPANDAI other than Bank Mandiri) (+Endline Survey: K E-Cash, L Simakmur)	G06	G06 (some changes)	G06 (some changes)

			Questionnaires		
Element / Dimension	**Question**	**Coding**	**BS**	**MS**	**ES**
Saving	Reasons for saving	A School fees/education, B Home improvements, C Expand primary business, D Expand second business, E Start a new business, F Health emergencies, G Other emergencies, H Retirement, I Pay off debt, J Buy vehicle, K Holiday, L Wedding, M Hajj pilgrimage, V Other reason	G09	G09	G09
Saving	Would you like to save more?	1 More, 2 Less, 3 About the same	G13	G13	EG07
Saving	(If More) Reasons for not saving more	A Not enough money, B Pressure/ obstacles from family members, C Other	G14	G14	
Saving	(If More) Reasons for not saving more (Endline Survey responses)	A Not enough money, B Pressure/ obstacles from family members, C Cash flow management, D Lots of expenses, E Like to keep cash on hand, F Put off until later, G No financial goal, V Other			G14
Borrowing	Business owner has loan(s) with a bank	Yes/no. If yes (name of bank: a Bank Mandiri, b Bank BNI, c Bank BRI, d Bank BTPN, v Other bank, W None)	G12		G12
Borrowing	Source of loan(s) in last 3 months (multiple responses)	A Formal bank account, D Friends/family, E Cooperative, G BMT, V Other, W No loans		G12ax	
Borrowing	Total value of loans in last 3 months by saving instrument (Rupiah)	A Formal bank account, D Friends/family, E Cooperative, G BMT, V Other, W No loans		G12aa-G12av	G12ba-B12Bv
Borrowing	Total value of all outstanding loan(s) by source (Rupiah)	A Formal bank account, D Friends/family, E Cooperative, G BMT, V Other, W No loans 1 Yes, 3 No Number of loans			
Borrowing	Business owner ever had a loan with a money lender (individual)	1 Yes, 3 No			EG01a
Borrowing	Number of loans from a money lender in last 12 months	Number of loans			EG02a
Borrowing	Business owner ever had a loan with a non-bank financial institution	1 Yes, 3 No			EG01b
Borrowing	Number of loans from a nonbank financial institution in last 12 months	Number of loans			EG02b
Business investment	Amount of business investment in past 12 months	Rupiah (millions)			EH01, EH02
Personal work effort	Days worked in primary business in typical month	Number of days	H11_H1A	H11_H1A	H11_H1A

			Questionnaires		
Element / Dimension	**Question**	**Coding**	**BS**	**MS**	**ES**
Personal work effort	Hours worked in primary business in typical day	Number hours	H12_H1A	H12_H1A	H12_H1A
Personal work effort	Days worked in second business (if any) in typical month	Number of days (Note: The data on days worked in the second business appear to overlap with days worked in the primary business)	H11_H1B	H11_H1B	H11_H1B
Personal work effort	Hours worked in second business (if any) in typical day	Number hours (Note: The data on hours worked in the second business appear to overlap with hours worked in the primary business)	H12_H1B	H12_H1B	H12_H1B
Personal work effort	Days worked in first other activity (if any) in typical month	Number of days (Note: The days worked in other activities may also overlap)	I03_Act1		
Personal work effort	Hours worked in first other activity (if any) in typical day	Number hours (Note: The hours worked in other activities may also overlap)	I04_Act1		
Personal work effort	Days worked in second other activity (if any) in typical month	Number of days (Note: The days worked in other activities may also overlap)	I03_Act2		
Personal work effort	Hours worked in second other activity (if any) in typical day	Number hours (Note: The hours worked in other activities may also overlap)	I04_Act2		
Personal work effort	Number of paid workers in primary business in typical month	Number of workers	H14_H1A	H14_H1A	H14_H1A
Personal work effort	Number of paid workers in second business (if any) in typical month	Number of workers	H14_H1B	H14_H1B	H14_H1B
Personal work effort	Number of customers in primary business in typical month	Number of customers	H15_H1A	H15_H1A	H15_H1A
Personal work effort	Number of customers in second business (if any) in typical month	Number of customers	H15_H1B	H15_H1B	H15_H1B
Mobilization of additional labor	Which HH members help with business?	A Spouse, B Son(s), C Daughter(s), D Other male HH member, E Other female HH member, W No other HH member	H07a_H1A	H07a_H1A	H07a_H1A
Mobilization of additional labor	Number of HH/unpaid workers in primary business in typical month	Number of workers	H13_H1A	H13_H1A	H13_H1A
Mobilization of additional labor	Number of HH/unpaid workers in the second business (if any) in typical month	Number of workers (Note: There may be overlap with the number of unpaid workers in the primary business)	H13_H1B	H13_H1B	H13_H1B
Business practices (BP)	BP1: Ask a supplier which products are selling well	1 Yes, 3 No, 6 Not applicable	P10		P10
Business practices (BP)	BP2: Used a special offer to attract customers in last 3 months	1 Yes, 3 No	P11		P11
Business practices (BP)	BP3: Done any advertising in last 6 months	1 Yes, 3 No (skip next question)	P12		P12
Business practices (BP)	BP4: Done anything to measure the effect of the advertising	1 Yes, 3 No	P13		P13
Business practices (BP)	BP5: Tried to get a lower price from supplier in last 3 months	1 Yes, 3 No	P14		P14

			Questionnaires		
Element / Dimension	**Question**	**Coding**	**BS**	**MS**	**ES**
Business practices (BP)	BP6: Has record system that informs about stocks of goods or raw materials	1 Yes, 3 No	P15		P15
Business practices (BP)	BP7: Keeps written business records	1 Yes, 3 No (skip next 3 questions)	P17		P17
Business practices (BP)	BP8: Records every business purchase/sale	1 Yes, 3 No	P18		P18
Business practices (BP)	BP9: Knows cash on hand at any point	1 Yes, 3 No	P19		P19
Business practices (BP)	BP10: Knows whether products are selling month to month	1 Yes, 3 No	P20		P20
Business practices (BP)	BP11: Knows cost of each main product	1 Yes, 3 No	P21		P21
Business practices (BP)	BP12: Has a written budget for business	1 Yes, 3 No	P23		P23
Business practices (BP)	BP14: Sells any goods on credit	1 Yes, 3 No (skip next question)	P24		P24
Business practices (BP)	BP15: Has record of how much is owed by customers	1 Yes, 3 No	P25		P25
Business practices (BP)	BP16: Has records needed to apply for a bank loan	1 Yes, 3 No	P26		P26
Business practices (BP)	BP17: Keeps business money separate from HH money	1 Yes, 3 No		EP27	
Business practices (BP)	BP18: Has bank account that is used for the business	1 Yes, 3 No		EP28	
Use of financial services	Which banking products used?	A savings account, B Check account balance, C Home mortgage, D Certificate of deposit, E Letter of credit, F business loan, G Vehicle loan, H Personal loan, I Health/life insurance, J Micro credit, V Other, W None	F17	F17	F17
Networking with peers	Belongs to any business-related organization/group	1 Yes, 3 No	K01		
Networking with peers	Proportion of female trainees known	Excluding K02==6	K02		
Networking with peers	Proportion of female trainees with whom talks business	Excluding K02==6	K05		
Community participation	Any voluntary activities in past year?	1 Yes, 3 No	J01		
Community participation	(If yes) How many hours per month volunteered?	Number of hours	J03		
Community participation	Amount contributed to charitable purposes in past year	Rupiah, 0 if J03=3	J04		
RESOURCES					
Gender	Gender of respondent	1=MBO, 3=WBO	brt_cov3		
Age/experience	Age of respondent	Years	A03	A03	A03
Age/experience	Number of years working in primary business	Number of years	H10_H1A	H10_H1A	H10_H1A
Education	Highest level of schooling completed by business owner	1 Primary or less, 2 Lower secondary, 3 Upper secondary, 4 Tertiary	edlevel (recoded)		
Education (CA=cognitive ability)	CA1: Ability to add and multiply	Number	X01		

			Questionnaires		
Element / Dimension	**Question**	**Coding**	**BS**	**MS**	**ES**
Education (CA=cognitive ability)	CA2: Ability to divide	Rp.	X02		
Education (CA=cognitive ability)	CA3: Ability to calculate annual interest	01 <90,000, 02 90,000 to 100,000, 03 Exactly 100,000, 04 105,000 to 115,000, 05 >115,000 98 Don't know	X03		
Education (CA=cognitive ability)	CA4: Ability to calculate monthly interest (ambiguous question)	01 <90,000, 02 90,000 to 100,000, 03 Exactly 100,000, 04 105,000 to 115,000, 05 >115,000 98 Don't know	X04		
Subjective characteristics	Willingness to take risks	Self-rating (1-10), 1=never, 10=always	Q01		
Subjective characteristics (TP=subjective time preference)	TP1: No time preference	1 Prefer payment in 6 months, 2 Prefer same payment in 7 months	Q03		
Subjective characteristics (TP=subjective time preference)	TP2: Moderate time preference	1 Prefer payment in 6 months, 2 Prefer 50% higher payment in 7 months	Q04		
Subjective characteristics (TP=subjective time preference)	TP3: Strong time preference	1 Prefer payment in 6 months, 2 Prefer 100% higher payment in 7 months	Q05		
Demographic characteristics	Business owner's current marital status	1 Not married, 2 Married, 3 Divorced, 4 Widowed	A04		
Demographic characteristics	Relationship of HH members to HH head	1 HH head, 2 Spouse, 3 Child, 4 Son/daughter-in-law, 5 Parents, 6 Sibling			AR02
Demographic characteristics	Number of children in HH	0 +	A09, A10		
Demographic characteristics	Relationship of business owner to HH head	1 HH head, 2 Spouse of HH head, 3 Other HH member (+ in Endline Survey: 3 Child, 4 Son/daughter-in-law, 5 Parents, 6 Sibling, 7 Brother/sister-in-law, 8 Grandchild, 9 Grandparent, 10 Uncle/aunt, 11 Nephew/niece, 12 Cousin, 13 Other HH member 95 Other	brt_cov2	brt_cov2	brt_cov2
Demographic characteristics	Business owner is head of household	1 Respondent, 2 Spouse, 3 Mother/father, 4 Other relative, 5 Other	A07		
Demographic characteristics	Household size (number of HH members)	1+	A08		
Access to financial services	Has registered bank account in own name?	1 Yes, 3 No	F18	F18	F18A-F18V
Access to financial services	(If yes) What year was account opened?	Year (obtained for each bank in Endline Survey)	F20		F20A-F20V

			Questionnaires		
Element / Dimension	**Question**	**Coding**	**BS**	**MS**	**ES**
Access to financial services	Number of banks in which business owner has an account	length of string (F19): string contains the following codes indicating each bank in which the business owner has an account: A Bank Mandiri, B Bank BNI, C Bank BRI, D Bank BTPN, V Other bank + (in Endline Survey: E Bank BCA, F Bank Jatim)	F19	F19	F18A-F18V
Access to financial services (DF=Distance to financial services)	DF1: Distance to nearest bank branch	1. <=0.5 km, 2. >0.5 to 1 km, 3. >1 to 5 km, 4. >5 to 10 km, 5. >10 to 15 km, 6. > 15 km			EF47
Access to financial services (DF=Distance to financial services)	DF2: Distance to nearest branchless banking agent	1. <=0.5 km, 2. >0.5 to 1 km, 3. >1 to 5 km, 4. >5 to 10 km, 5. >10 to 15 km, 6. > 15 km			EF49
Access to financial services (DF=Distance to financial services)	DF3: Distance to nearest bank ATM	1. <=0.5 km, 2. >0.5 to 1 km, 3. >1 to 5 km, 4. >5 to 10 km, 5. >10 to 15 km, 6. > 15 km			EF51
Access to infrastructure (communications)	Business owner has a mobile phone	1 Yes, 3 No	B01	B01	B05a, AR07 (by HH member)
Access to infrastructure (communications)	Type of mobile phone	A simple phone, B smart phone, C. Both types	B06	B06	B06 (type used)
Access to infrastructure (communications)	Uses of mobile phone (multiple choices permitted)	A Make phone calls, B Send message, C Use *Whatsapp*, D Browse on the internet, E Play games, F Use social media, G Mobile money, H Banking transactions	B07	B07	B07
Access to infrastructure (communications)	Expenses for mobile phone by business owner in last 3 months	Rupiah			B07a
Access to infrastructure (communications)	Spouse owns mobile phone?	1 Yes, 3 No			AR07
Access to infrastructure (communications)	Number of mobile phones owned by spouse	Number			AR07
Access to infrastructure (communications)	Number of mobile phones owned by HH members	Number			AR07
Access to infrastructure (communications)	Does any HH member own a smart phone	1 Yes, 3 No	B06a	B06a	B06a
Access to infrastructure (communications)	Signal coverage problems with mobile network?	If yes, Number of days per week	B10a	B10a	B10a
Access to infrastructure (communications)	Internet access problems with mobile network?	If yes, Number of days per week	B10e	B10e	B10e
Access to markets	Semi-urban village	1 Semi-urban, 2 Rural	LK06	LK06	LK06
Access to markets	District of residence	Official locality code	LK02	LK02	LK02
Access to markets	Sub-district of residence	Official locality code	LK03	LK03	LK03
Social capital (TR=trust in banks)	TR1: How much confidence in enforcement of contracts between state-owned banks	1 Not confident at all, 2 Somewhat not confident, 3 Neither confident nor not confident, 4 Somewhat confident, 5 Extremely confident	E02	E02	E02

			Questionnaires		
Element / Dimension	**Question**	**Coding**	**BS**	**MS**	**ES**
Social capital (TR=trust in banks)	TR2: How much confidence in enforcement of contracts between non-state-owned banks	1 Not confident at all, 2 Somewhat not confident, 3 Neither confident nor not confident, 4 Somewhat confident, 5 Extremely confident	E03	E03	E03
Social capital (TR=trust in banks)	TR3: How confident get back savings deposit in bank?	1 Not confident at all, 2 Somewhat not confident, 3 Neither confident nor not confident, 4 Somewhat confident, 5 Extremely confident	E06		

Appendix 2. Detailed analysis

Achievements

Personal income

This dimension includes both wage and salary earnings and business profits (i.e., all earned income). In principle, it should also include unearned incomes (e.g., rent, interest, and net transfers). However, the data set does not include any data on unearned incomes. The personal income data are also limited to the incomes of the responding WBO or MBO (i.e., the "respondents"). In particular, no data on the personal income of spouses or other HH members are available in the data set.

Primary business profits (reported) (Tables 1, 2, 5)

Data on average monthly business profits during the last 12 months were collected in all three survey rounds for both the primary and, if present (17% of the sample), the second business.[34] It is possible to calculate a meaningful combined average monthly profits measure for both the primary and second business. However, there is clearly overlap in some of the reported labor inputs (average number of days and hours worked in a typical month) between the primary and second businesses, as discussed below.[35] A second problem is that the value of capital inputs is reported only for all businesses combined. Under these circumstances, the analysis of business profits focuses on the primary business, acknowledging that the capital input variable is measured with error when more than one business is owned.

The data on primary business profits indicate low test-retest reliability for the reported values (i.e., r=0.17 between ES and BS values), but substantially higher reliability for the natural log transformations of the reported values (r=0.57). These differences reflect the highly skewed distribution of the reported primary profits variable in both the BS and ES (skewness=17.5 and 43.7 respectively, with the number of extreme outliers equal to 166 and 140 respectively).[36] In contrast, the corresponding statistics for the natural log transformation of primary business profits are -0.43 and -0.10 respectively, with only 0 and 1 extreme outliers respectively.

Table 1 reports the results of multiple regression analysis of the BS data with reported primary business profits as the dependent variable, using five alternative statistical models. Most of the significant estimated coefficients have the expected signs (including the positive linear and negative quadratic terms in age, as a proxy for experience), supporting the criterion validity of primary profits. Of note are the highly significant estimates indicating that women business owners earn 49-67% less than men business owners (depending on the statistical model). The results are reasonably robust with respect to the alternative statistical models in columns 1-5 (although there are only 7 significant estimated coefficients in the OLS model, compared to 11 in the log regression model). However, the OLS estimates

[34] In addition, data were collected on profits from up to two additional businesses in the BS.

[35] For example, if labor inputs are summed across primary and second businesses, the total number of days worked per typical months exceeds 760 of the 821 cases reporting a second business, while the total number of hours worked per typical day exceeds 16 in 139 of the 821 cases.

[36] By comparison, the skewness measure is equal to zero in the normal distribution, while extreme positive (negative) values are defined as values more than (less than) three times the inter-quartile range above the 75th percentile value (below the 25th percentile value).

differ qualitatively from the others in only two cases: the owner's willingness to take risks and the number of hours worked are both statistically insignificant in the OLS model, whereas they are significant in all of the other models. Only one of the estimated coefficients referring to schooling is statistically significant, although the estimated coefficient of cognitive ability is positive and significant in two statistical models (the log and robust regression models).

Table 2 reports the results of multiple regression analysis of primary business profits using data from the MS. The results are similar, but with fewer significant coefficients, reflecting not only the smaller sample but also the fact that the MS data were drawn from only 200 of the 401 villages in the BS and ES samples. One interesting difference is the inclusion of personal agency, a composite variable, among the right-side variables. This variable is positive and significant in all five statistical models.

Table 5 presents both adjusted and unadjusted estimates of the impact of the business training provided to randomly selected WBOs on reported primary business profits by alternative statistical model. The results indicate that the estimated impact of the training was positive and statistically significant in 4 of the 5 unadjusted models (including at the 0.01 level in the log regression model), but statistically significant only in the log regression model (at the 0.05 level) in the adjusted models. Levels of schooling and cognitive ability are insignificant in all five statistical models.

Primary business profits (calculated) (Table 3)

In addition to data on reported business profits, the MS collected data on average monthly primary business revenue and expenses (using 5 expense categories). These data can be used to calculate an alternative measure of primary business profits (i.e., revenue minus expenses). Table 3 shows the results when the same alternative regression models (with one exception)[37] are estimated with this calculated profits measure as dependent variable (cf Table 2 presents the results for reported primary reported profits with the MS sample for comparison). The results in Table 3 are similar in most respects to those in Table 2 (and Table 1). One difference is that calculated profits in Table 3 are positively and significantly related to the number of unpaid workers in three of the models, whereas they are consistently insignificant in Table 2. In addition, there are fewer significant coefficients in Table 3, and the results are less robust across the various statistical models. (Not shown) Estimates of the impact of the business training on calculated primary profits are statistically insignificant in all adjusted and unadjusted models. The idea that using a calculated measure of profits rather than the measure based on a single question, as in Table 1, increases the criterion validity of the profits measure is not supported by these results.

Primary business revenue (Table 5)

It is sometimes suggested that business revenue by itself may be a more reliable and valid measure of business performance than business profits, especially for very small businesses that may not have accurate accounts. Table 5 shows the results when the same alternative regression models are estimated with average monthly primary revenue from the MS as the dependent variable (in place of profits). The results are similar in most respects to those for reported profits in Table 2, including that there are fewer significant coefficients in the OLS model (only 2 versus 11 in the log regression model). However, there are some differences. First, the owner's schooling at the upper secondary and tertiary levels is negatively and significantly related to business revenue in several of the models, whereas

[37] Because 131 observations on the calculated profits measure are negative, an IHS transformation is used rather than the natural log transformation in column 3.

schooling is consistently insignificant in Table 2. Second, the number of unpaid workers is positively and significantly related to business revenue in Table 5, but is consistently insignificant in Table 2.

Wage and salary earnings (Table 6)

Data on wage and salary earnings in a typical month were collected for up to two jobs, along with data on the number of days and hours worked in each job. It is not possible to assess the test-retest reliability of these data because they were collected only in the BS, but their criterion validity can be assessed by estimating Mincer-like models with alternative statistical models in which the reported monthly earnings (or the log of monthly earnings) is regressed on the logs of the numbers of days and hours worked, dummy variables indicating the highest level of schooling completed, a measure of cognitive ability (using the IRT scale), age in years (as a proxy for experience), gender (as a proxy for labor market discrimination) and (not shown) regency of residence to control local differences in labor market opportunities. The results are presented in Table 6. The results are consistent with those obtained in other studies (Heckman and others 2003), for example, the significant positive effect of additional schooling and the significant linear (positive) and quadratic (negative) terms in age. The results also indicate that women business owners earn about 36-48% less than men business owners as wage and salary workers, other factors equal. The gender-specific results in columns 6 and 7 also indicate that the earnings of WBO, unlike those of MBOs, are not significantly related to age (as a proxy for experience) or even to additional schooling up to the tertiary level.

Personal assets

Data on the current value of business owners' accumulated savings (current savings balances) were collected by saving instrument in both the MS and ES (but not in the BS). The data on current savings balances include savings in kind (e.g., jewelry), which is an important savings instrument in this setting. In addition, data on business owners' outstanding loan balances by source were also collected in the MS and are used to calculate estimates of net financial assets. The data on current savings balances are also relevant to the dimension "vulnerability to shocks" (discussed below). (Data on *recent* savings and borrowing are discussed under Agency and Empowerment.)

Current savings balances (Table 11)

Both the MS and ES collected data on the total current savings balances of business owners. These data have very problematic distributions. Focusing on the ES data because of the larger sample size, they include 593 zero values (12.3% of the total), are highly skewed in a positive direction (skewness=23.4 and kurtosis=905.7, compared to values of zero and three respectively for the normal distribution), and 301 highly extreme values (defined as more than three times the interquartile range above the 75th percentile value). In terms of test-retest reliability, the correlation between the ES and MS values is only +0.268. However, this correlation is relatively low mainly as the result of the extreme outliers. If inverse hyperbolic sine (IHS) transformations are applied to the current total savings balance variables, skewness reduces to 0.574, kurtosis reduces to 2.588, and the correlation coefficient increases to +0.503.

Table 11 reports the estimates obtained using the total current savings balance as the dependent variable with six alternative statistical models.[38] The results indicate that the estimates are sensitive to the choice of statistical model. For example, the estimates obtained by applying OLS both to the

[38] A model with the IHS-transformed variable as dependent variable is not included in Table 11 because of problems with the IHS transformation when applied to non-negative variables with high concentrations of zeroes (Knowles 2020).

reported variable (column 1) and to a winsorized (highest 2% of values) version of the reported variable (column 2) do not obtain a significant gender coefficient, whereas the estimates obtained with the remaining four statistical models (columns 3-6) all positive and statistically significant. Similarly, significant relationship with schooling are limited to the tertiary level in columns 1 and 2, whereas they are found for lower levels of schooling in several of the other models. In contrast, the highly significant positive relationships with profits and HH income are robust with respect to all six of the statistical models. The results in Table 11 suggest that the best statistical model for this problematic variable is the two-part model in column 4, with 7 statistically significant variables versus only 3 with both the OLS and OLS winsorized models.

Net financial assets (Table 12)

The MS collected data on both current savings balances and current amounts of unpaid loans of business owners, making it possible to calculate the current value of their net financial assets as the difference between the two. This highly skewed variable (skewness=25.4, kurtosis=871.5) has a mean of 21.06 Rp. millions and a median of 1.08 Rp. millions, has positive, negative and zero values (1418, 816 and 81 respectively) with 90 very extreme negative values and 191 very extreme positive values. Unfortunately, there is no information on test-retest reliability because the data are limited to the MS.

Table 12 reports the results of multiple regression analysis of net financial assets as a function of selected criterion variables using alternative statistical models (the IHS-transformed variable is used as dependent variable in column 3 because the log transformation is undefined for non-positive values). OLS estimation applied to the reported value of net assets (column 1) yields four significant estimated coefficients, three of which have the same signs and are statistically significant across all five models. However, following winsorization of both the highest and lowest 2% of values (column 2), five additional coefficients become statistically significant, three of which have the same signs and are significant in all five statistical models. However, the owner's gender, although consistently positively related to net financial assets in all five models, is only statistically significant in columns 3-5 where the relationship becomes very strong.

In terms of their ability to provide the largest number of significant estimated coefficients (9), the winsorized OLS model in column 2 (with both the highest and lowest 2% of values winsorized), the model with the IHS transformed dependent variable in column 3, and the robust regression model in column 5 are preferred. In unreported regressions, however, multiplying the dependent variable by 1000 before applying the IHS transformation, as suggested by Bellemarre and Wichman (2019), has a large effect on the magnitude of the estimated relationships, as Knowles (2020) warns can occur with IHS transformations. Overall, the results in Table 12 underline the importance of estimating alternative statistical models when the dependent variable has such a highly problematic distribution.

Business assets

Total value of business capital (Tables 13-14)

Data on the current market value of six types of business capital were collected in all three survey rounds. The distribution of this variable is very problematic. Although there is only one reported zero value in the BS, the distribution is highly skewed (skewness=8.97, compared to zero for the normal distribution, while kurtosis=111.5, compared to 3 for the normal distribution). In addition, there are 666 extreme values (i.e., more than three times the interquartile range of 20.4 million Rupiah above the 75th

percentile value of 22.65 million Rupiah). In terms of test-retest reliability, the BS to ES correlation is 0.571, increasing to 0.662 between the logged values (which are approximately normally distributed).[39]

Table 13 reports the results of multiple regression analysis of the total business capital values using five alternative statistical models (columns 1-5). In addition, Table 13 reports the results of gender-specific analysis using the log regression model (columns 6-7). Several of the criteria variables are uniformly highly significant across the alternative statistical models, including the business owner's gender (lower among female owners), the owner's schooling (positive for completion of the tertiary level), the total profits from all businesses owned (positive) and the HH asset index (positive). Several other criterial are almost uniformly significant across the alternative models, including completion of upper secondary schooling (significantly positive except for the two OLS models in columns 1 and 2), the owner's cognitive ability (significantly positive, except in columns 5 and 6). These results are sufficient to establish the criterion validity of the total business capital variable, although they do not provide strong evidence favoring any one statistical model over another.

The results with respect to gender are noteworthy. First, it is clear from the results in columns 1-5 for both genders pooled that women business owners are significantly disadvantaged in terms of access to capital. For example, the results in column 3 with a logged dependent variable indicate that women's business capital is about 43% lower than men's, other factors equal.[40] The corresponding gender-specific results in columns 6-7 identify a weaker positive relationship between capital ownership and schooling below the tertiary level and a much weaker positive relationship between capital ownership and cognitive ability and the significantly lower values of total profits and HH assets among WBOs as factors contributing to the observed gender gap favoring men in business capital.

Because both baseline and endline data were collected on the value of business capital, it is possible to estimate the impact of business training provided to randomly selected WBOs, on the value of their business capital. Table 14 presents unadjusted and adjusted estimates of the impact of the business training using the same five alternative statistical models as Table 13. The unadjusted estimates are all positive, as expected, but only those for the robust regression statistical model (column 5) are statistically significant (at the 0.05 level). Among the adjusted estimates (i.e., those that include baseline values of the capital stock as covariates), the estimated impact is significant for three of the models (i.e., the OLS, log and robust regression models, with the latter two significant at the 0.05 level).

Number of primary business customers (Tables 33 and 34)

Data on the number of customers of both the primary and second (if present) business were collected in all three survey rounds. These data refer to the total number of customers, as reported by the owner at the time of interview, not to the number of customers actually purchasing goods and services from the business during a given period. As a measure of economic agency and empowerment, the number of customers reflects the efforts made by the owner to attract and retain customers over time, an important factor that is at least partially under the control of the owner. However, there is a problem of overlap in the number of primary and second business customers in some survey rounds.[41]

---

[39] For example, the logged baseline value of total business capital has a skewness of 0.14 and a kurtosis of 3.08.

[40] i.e., exp(-0.566)=0.568

[41] There are no cases of overlap in the BS (but many zeroes for customers in the second business), but the same customer numbers are reported for *all* primary and second businesses in the MS and in 53 primary and second businesses in the ES.

Consequently, the analysis in this paper is limited to the reported number of customers in the primary business.

The number of primary business customers is nevertheless a problematic economic variable, with a BS mean of 251.8 and a standard deviation of 562.2, ranging from one customer to a maximum of 7,500 (N=4,819, skewness=15.169, kurtosis=355.045), with 417 very extreme values. Given consistent data were collected in all three survey rounds, it is possible to assess the test-retest reliability of the data on the number of primary business customers. However, the correlation between the BS and ES values is a statistically significant 0.328.

Table 33 reports the results of multiple regression analysis of the number of primary business customers using five alternative statistical models: OLS applied to the reported numbers (column 1), OLS applied to the winsorized reported numbers (highest 5% of reported numbers, column 2), log regression model (column 3), quantile (median) regression (column 4) and robust regression (column 5). The results indicate that WBOs report significantly fewer customers in all five statistical models (about 31% less in the log regression model). The number of customers is also positively and significantly related to the total value of business capital and negatively and significantly related to the number of paid workers in all five models (whereas the number of business customers is positively and significantly related to the number of unpaid workers in three models). Interestingly, the number of business owners is negatively and significantly related to mobile phone use in four models. Taken together, the results in Table 33 provide moderate support for the criterion validity of the reported number of primary business customers.

Table 34 presents both adjusted and unadjusted estimates of the impact of the business training on the number of primary business customers reported by women business owners using with the same five statistical models. According to the estimates from two of the five models (i.e., the log regression model and the robust regression model), the business training had a significant positive impact on the number of primary business customers. According to the log regression model, the adjusted (unadjusted) estimate indicates that the training increased the number of customers by 11.1% (13.3%).

Primary business is registered with the government (Table 72)

All three survey rounds collected data on whether the primary and second businesses were currently registered with the government at the time of the interview. At baseline, 13.1% of sample business owners' primary businesses were registered with the government, including 10.7% of women-owned businesses and 16.6% of men-owned businesses. At endline, the corresponding percentages were 19.7%, 17.2% (WBO) and 23.3% (MBO). Test-retest estimates indicate that this variable is relatively reliable (BS to ES r=0.420). Table 72 shows the results of multiple linear regression analysis of the dichotomous variable indicating that the primary business was registered at the time of the BS, including separate gender-specific estimates. The results indicate that this variable also has moderate criterion validity, with five highly significant estimated coefficients (all at the 0.001 level), including positive estimated relationships with the number of years worked in the primary business (limited to WBOs), adherence to recommended business practices, total business profits and the HH asset index and a negative estimated relationship with WBOs. Interestingly, this variable is not significantly related to schooling (except negatively at the tertiary level among WBOs) and is negatively related to cognitive ability (with the significant relationship limited to MBOs). (Not shown) The estimated impact of the business training provided to randomly selected WBOs is statistically insignificant in both adjusted and unadjusted models.

Household income

HH income (Table 7)

Data on HH income were collected only in the MS, using three questions. The first question asked for total HH income in the last month, including income from all sources but excluding the proceeds of any loans. This question was followed up by asking respondents to provide both a low-end and a high-end estimate of the total income earned by the household in a month. Responses to the first question can be used as a direct measure of HH income. A second, predicted measure of HH income can be obtained by regressing the natural logarithm of the reported measure referring to the last month on the low-end and high-end estimates. The predicted values from this regression may yield a more reliable and valid measure of HH income than the reported measure if the direct measure incorporates substantial measurement error.[42]

Table 7 presents estimates of linear regression models using natural logarithms of the reported and predicted measures of HH income as dependent variables (columns 1 and 2). Although the results are similar and suggest that both HH income measures have criterion validity, the model with predicted HH income has a higher R^2 and is therefore used as the HH income measure in this paper. One result that is at first surprising is that HH income is significantly higher in WBOs' HHs. This result implies that WBOs are more likely to come from higher-income households, based on the incomes of their spouse and possibly other HH members (because we know that from the results in Tables 1-6 that WBOs earn both lower profits and lower wages and salaries than men business owners, other factors equal). In order to further explore these gender differences, Table 7 re-estimates the model with predicted HH income separately for women and men business owners (columns 3 and 4). Although the overall models are significantly different (p=0.024), the estimated coefficients are qualitatively similar except that WBOs' cognitive ability is positively and significantly related to their HH income, whereas MBOs' is not.

Asset indexes (Tables 8-10)

In addition to the direct measures of HH income collected in the MS, data on the ownership of 20 household durable goods were collected in all three survey rounds. In addition, on the *numbers* of each asset owned were collected in the MS, while data on selected housing characteristics were collected in the BS. These types of data are considered relatively easy to collect and are considered relatively reliable. Indexes (composite variables) constructed from them are often used as a cost-effective alternative to measures of HH income based on detailed income or consumption data (Filmer and Scott 2012). However, the "face" validity of HH asset indexes as measures of HH income (as distinct from HH wealth or some other variable) is not obvious. One possible theoretical rationale is based on the idea that the demand for durable assets is positively related to HH income, with the strength of the relationship possibly stronger in the case of some more expensive assets (e.g., smart phones, refrigerators, washing machines, automobiles) than others (e.g., TVs, bicycles, simple cell phones).

Table 8 shows the results of simple regressions of both ownership (0-1) of the 20 durable goods in the MS data set (column 3) and the number of each item owned (column 4) on the natural log of the predicted measure of HH income discussed above (the sample means and standard deviations of the variables are shown in columns 1 and 2). The results indicate that most (but not all) of the simple regressions are highly significant (i.e., 30 of the 40 t-statistics). These results suggest that ownership of these goods probably contains a lot of information about HH income. The question is how best to distill

[42] This was the case in one study of the relationship between schooling outcomes and household income in Vietnam (Behrman and Knowles 1999).

this information into an estimate of HH income. The standard method is to use the loadings of each item in the eigenvector of the first principal component of the 20 variables (shown in columns 5 and 6) to calculate a predicted score that then becomes the asset index measure of HH income. Examining the values in columns 3-6, one sees that the estimated loadings in columns 5 and 6 are indeed closely correlated with the estimated t-statistics in columns 3 and 4, establishing the validity of the standard asset index as a measure of HH income. This is a very significant finding, given that no direct measure of HH income is involved in calculating the asset index.[43]

Although the results in Table 8 establish the validity of both asset indexes (one based on simple ownership of 20 durable items and the other based on the number of items owned), it remains to determine which of the two measures is "best" and whether the principal component scale (the most often used currently) is the best way to aggregate the 20 items into a single composite variable measuring HH income. Prior to the development of the principal components-based asset index, common practice was to simply use an unweighted count or proportional mean of the items owned. However, when the ownership measure is the number of items owned rather than a 0-1 measure of simple ownership, an aggregation formula based on item response theory (IRT) may perform better.

Table 9 compares the results of simple regression models estimated with MS data with the alternative scales of each of the two asset ownership variables (any item owned, number of items owned) as dependent variables and the log of predicted HH income as the explanatory variable. The results indicate that R^2 is generally higher for the asset indexes that are defined on the basis of the number of items owned (and not just whether any item is owned). Although the R^2s in Table 9 do not identify clearly a preferred scale for the composite variable based on ownership alone, reflecting the fact that they are highly correlated (r's vary only between 0.967 and 0.983 in the BS and between 0.957 and 0.983 in the ES) they do point to a slight preference for the IRT scale when the composite variable is based on the number of items owned instead of simple (yes-no) ownership.

Asset indexes based on yes-no ownership can be calculated with the data in all three survey rounds. The resulting estimates have relatively high test-retest reliability for all four alternative scales. For example, the correlations between the BS and ES values are all between 0.71 and 0.75. Test-retest reliability is higher for the weighted PC and IRT scales than for the unweighted PM and SM scales, but the differences are not large.

Table 10 compares both adjusted and unadjusted estimates of the impact of the business training on the HH asset indexes of WBOs by the alternative scales used to calculate the indexes. Although the unadjusted estimates are all statistically insignificant, all of the adjusted estimates are positive and statistically significant (although the index based on the IRT scale is significant at only the 0.10 level).

Quality of life

Subjective well-being (Table 17)

Data on the subjective well-being of the business owners were collected in the MS using the following two questions: (1) "Taking all things together, would you say you are very happy, somewhat happy, neither happy nor unhappy, somewhat unhappy or very unhappy?" (2) "How satisfied are you with your current work/job: very unsatisfied, somewhat unsatisfied, neither satisfied nor unsatisfied, somewhat satisfied, very satisfied? The responses to both questions were combined into a single composite

[43] The eigenvector loadings are the coefficients of the linear function that explains the highest proportion of the total variation in the 20 variables.

variable measuring subjective well-being.[44] The test-retest reliability of this subjective variable cannot be assessed because the data were collected only in the MS. However, the criterion validity of this composite variable can be assessed using multiple regression analysis with several alternative statistical models. Because the dependent variable has only 5 zero values, it is reasonable to include the log regression model among the alternative statistical models.

Table 17 reports the multiple regression estimates for the total MS sample (including both married and unmarried business owners) while Table 18 reports the results for the MS sample limited to married business owners. The two model specifications differ slightly, with the model for the total sample including as a criterion variable whether the business owner is currently married, whereas the model for married business owners substitutes a composite variable measuring marital agency for current marital status. These alternative specifications are used in order to explore the relationship of subjective well-being to agency (including personal agency, intra-HH agency, and agency within marriage). Both tables report results for the four alternative measurement scales (columns 1-4) as well as gender-specific models with the best-fitting IRT scale in columns 5 and 6. Because all four alternative variables were standardized prior to estimation, the estimated coefficients can be interpreted as estimated standard deviations.

The most striking finding in both models is a uniformly negative and statistically significant relationship between business owners' highest level of schooling completed and their subjective well-being. The relationships are uniformly strong across both scales and gender, between -0.2 and -0.3 standard deviations for upper secondary schooling and between - 0.3 and -0.4 standard deviations for tertiary schooling. The business owner's subjective well-being is positively and significantly related to the combined profits of her primary and any second business for all four scales, as expected, but although the well-being of MBOs is also positively related to their profits, the relationship is much stronger among WBOs and not even statistically significant among married MBOs. Well-being is also positively and significantly related to HH income for all four scales, as expected, but the relationship is only significant for all WBOs and not for MBOs or for only married women business owners.

Subjective well-being is positively and significantly related to the number of days worked in the primary business, but only for all WBOs (Table 17) and currently married WBOs (Table 18), while it is insignificantly related to the number of hours worked. Among all business owners (Table 17), well-being is positively related to being currently married, whereas among currently married business owners (Table 18), well-being is negatively and significantly related to marital agency with all scales, but the negative relationship is only significant among MBOs. Lastly, well-being is positively and significantly related to personal agency among both currently married and all business owners, and is even stronger among MBOs (including married MBOs).

Business owners' work time (Tables 19-20)

Data were collected in all three survey rounds on the number of days and hours worked in a typical month/day in the primary business and (if present) a second business. In addition, the BS collected data on up to two additional economic activities, which could be either additional businesses or wage and salary employment. These data provide a basis for assessing business owners' workloads and therefore their access to leisure time. In addition, business owners' work effort is one of the main factors under their control and therefore an important dimension of their agency and economic empowerment. Lastly,

[44] Preliminary analysis indicated that both measures were similarly related to the criteria specified in the multiple regression models.

because business owners' labor in their businesses is an important business input, data on labor inputs helps to evaluate the validity of business profits and other reported business outcomes.

Unfortunately, the data on business owners' labor inputs have some problems. The main problem is that there is considerable (inconsistent) overlap in the reported inputs into multiple activities. For example, business owners with only one source of earnings (their primary business), report working an average of 8.7 hours per day in a typical day, whereas the averages are 12.2, 15.3 and 17.6 hours per typical day respectively for business owners with 2, 3 or 4 jobs. Although these numbers are plausible, the corresponding numbers are nonsensical for the reported number of days worked in all jobs in a typical month (i.e., 47.1, 64.9 and 83.1 days worked per typical month for business owners with 2, 3 or 4 jobs) because 73.0% of the sample business owners report that they work 30 days in their primary business in a typical month (81.5% of women and 60.7% of men business owners), and because there is considerable overlap in the reported number of days worked per job.

Clearly, the raw data need to be adjusted in order to obtain meaningful measures of business owners' work time in all jobs. Accordingly, it is assumed that the maximum number of days worked in all jobs in a typical month is 30 days and that the maximum number of hours worked in all jobs in a typical day is 16 hours. After this adjustment, the mean total number of hours worked is more reasonable, i.e., 8.7, 11.7, 13.5 and 14.2 hours respectively for business owners with 1, 2, 3 or 4 jobs, while the mean total number of days worked ranges only from 28.6 with 1 job to 30 with 4 jobs. Because data were collected on labor inputs into the primary and second business (if present) in all three survey rounds, it is possible to assess test-retest reliability for the reported primary and second business labor inputs after making the same adjustments. The BS to ES correlation is equal to 0.457 for the number of days worked, 0.339 for the number of hours worked and 0.367 for the total number of hours worked in a typical month (i.e., the product of the number of days and worked).

Table 19 presents the results of multiple regression analysis of the total number of hours worked in a typical month, both for all business owners by alternative statistical models (columns 1-5) and for women and men business owners separately (columns 6 and 7). The total number of hours worked in a typical month (i.e., the product of the number of hours in a typical day and the number of days in a typical month) is used as the dependent variable because the distribution of the number of days worked per month is so heavily concentrated on 30 days, as discussed above. In contrast, the distribution of the number of hours worked in a typical month is not problematic (skewness=-0.046, kurtosis=2.177, no zero values). The results in Table 19 indicate that business owners' hours worked is significantly higher for women business owners, for business owners with higher total earnings, multiple earnings sources, a second business and additional business capital, as well as for business owners who receive help from other HH members.[45] Business owners' labor supply is significantly lower in HHs with higher income, as measured by the HH wealth index (i.e., the HH asset index including housing characteristics). These significant relationships are all consistent with prior expectations and therefore support the criterion validity of the number of hours worked by business owners in a typical month.

Apart from the significantly higher reported labor inputs of women business owners in column 1, the separate estimates for women and men business owners in columns 2 and 3 exhibit several gender differences (a test for overall gender homogeneity is rejected at the 0.001 level). For example, labor supply is negatively and significantly related to HH income only among women. There are also gender

[45] One might expect that receiving help from other HH members would enable business owners to work fewer hours, but any such tendency is apparently swamped by the willingness of other HH members to help business owners who are already working long hours.

differences in the help received from other HH members, with the exception that both women and men business receive help from a spouse. Other gender differences are that only women's hours worked is significantly related to the owner's total earnings and to the value of capital (both positively), whereas only men's hours worked is significantly related to having a second business (positively).

Because both baseline and endline data on business owner's labor inputs into their primary and second (if any) businesses were collected, it is possible to obtain both adjusted and unadjusted estimates of the impact of the business training given to randomly selected women business owners on their hours worked. These estimates are reported in Table 20. They indicate that all of the impact estimates are positive, but only those for the OLS winsorized (highest 5% of values) and the robust regression models are significant in both the adjusted and unadjusted models.

Vulnerability to shocks

Two measures of vulnerability to shocks are available in the data, one objective and the other subjective. The objective measure is the degree of instability in HH income (measured only in the MS as the difference between the high-end and low-end estimates of HH income as a proportion of HH income in the last month). The subjective data on vulnerability to shocks consists of yes-no responses to a single question in the BS: "In the last 7 days, did you worry that your household would not have enough food?"

Instability in HH income (Table 15)

This variable is problematic, with a highly skewed and concentrated distribution (skewness=11.0, kurtosis=171.0) and 161 very extreme values. Because data were collected on this variable only in the MS, it is not possible to assess its test-retest reliability. However, it is possible to use multiple regression analysis to assess its criterion validity.

Table 15 reports multiple regression analysis of the income instability measure using five alternative statistical models (columns 1-5). The results indicate that income instability is positively and significantly related to the number of earnings sources (including business profits and wage and salary earnings) in all five statistical models as well as among both women and men business owners (columns 6 and 7). A priori, one would expect that access to multiple earnings sources would lead to more stable HH income. Income instability is also negatively and significantly related to wage and salary earnings as one of the earnings sources in all but one of the alternative statistical models as well as among both women and men business owners. Income instability is also negatively and significantly lower among women business owners in three of the five statistical models. In contrast, income instability is not consistently related to either the combined profits of primary and second businesses or to the direct measure of HH income available in the MS (i.e., sign reversals among significant estimated coefficients across statistical models). The results in Table 15 provide only weak evidence supporting the criterion validity the income instability measure.

HH food insecurity (Table 16)

Multiple regression analysis (using the OLS regression model) of the subjective measures of HH food insecurity is reported in Table 16, both for all business owners (column 1) and for women and men business owners separately (columns 2 and 3). The results indicate that concern about food insecurity is negatively and significantly related, as expected, both to business owners' earnings from all sources (including profits from all businesses owned and any wage and salary earnings) and to the HH asset index (as a measure of HH income). Concern is also significantly lower among business owners with more schooling and higher cognitive ability. Concern is also positively and significantly related to the number of children in the HH. These results all support the criterion validity of this subjective variable.

The analysis of current financial savings balances and of net financial assets discussed above is also relevant to a business owner's vulnerability to shocks, as indicated by the fact that 76% of women business owners and 68% of men business owners in the BS listed saving for "health emergencies" or "other emergencies" as one of their three main saving goals.

Agency and Empowerment

Personal agency (Table 21)

The MS asked respondents five questions on "assertiveness" and ten questions on "attitudes" to obtain a personality profile of the business owners. The five questions on assertiveness concern comfort levels when speaking out at a meeting, talking to an employee about a disagreement, negotiating with a seller for a fair price, bargaining with a supplier to get a lower price, and talking to a spouse about a money issue. Responses were coded using a 5-point Likert scale registering the respondent's comfort level. The questions on attitudes cover a wide range of topics (i.e., planning ahead, saving, persistence, risk-taking, leadership, competition, optimism). Respondents are asked to indicate their degree of agreement with the statement using a different 5-point Likert scale.[46] The responses to these 15 questions were combined into a composite variable measuring the personal agency of business owners.

The principal components analysis indicates that the first component explains 16.2% of the total variation in the 15 responses and that the responses to all 15 questions are positively and significantly correlated with the first component, with r's ranging from 0.274 to 0.515. Cronbach's alpha is 0.619 and cannot be increased by dropping any item. However, the question assessing respondent's comfort level in talking to a spouse about a money issue was asked only to married business owners. Including that item would restrict the composite variable to married business owners. Accordingly, that item was dropped.

Because these data on assertiveness and attitudes were collected only in the MS, it is not possible to assess the test-retest reliability of this composite variable. However, it is possible to use linear regression analysis to assess its criterion validity and to identify a possible preference between the four alternative scales. The results are presented in Table 21. Unfortunately, they do not identify a clear preference among the alternative scales, as both the proportional mean and the principal component scales both achieve R^2s of 0.10, while the standardized mean and IRT scale have only slightly lower R^2s of 0.09. This is not surprising because the four measures are highly correlated (the six baseline r's range only between 0.939 and 0.990, with the lowest r's involving the IRT scales).

However, the results in Table 21 clearly support the criterion validity of the "personal agency" composite variable. For example, personal agency is positively and significantly related to the highest level of schooling completed (but not to cognitive ability). Personal agency is positively and significantly related to business profits, to HH income and to following recommended business practices (another composite variable). In contrast, is about 0.3 standard deviations lower among women business owners, while the personal agency of married business owners is about 0.17 standard deviations lower, other factors equal. These results support the criterion validity of the personal agency composite variable.

[46] It is usually recommended to include a mixture of positive and negative qualities in such "test batteries" in order to avoid response bias (Furr and Bacharach 2014). Most analytical software, including Stata's "pca" and "alpha" commands, automatically changes signs as needed to align the items. However, in this case, all of the questions concern positive qualities.

Intra-HH agency (Tables 22-24)

All three survey rounds collected comparable data on which household member(s) make the following five decisions: 1) whether or not to purchase an appliance for home use, 2) decisions related to work away from home by HH members, 3) whether to provide support to other family members, 4) whether to save for the future, and 5) whether to sign up for a new banking project. The pre-coded responses identifying the decision-makers include: 1) respondent only, 2) respondent and spouse, 3) spouse only, 4) respondent and other HH member (not spouse), and 5) another HH member only. Principal components analysis indicates that all five questions load strongly onto the first principal component, with loadings varying from 0.416 to 0.479. Cronbach's alpha is correspondingly high (0.756).

It is not clear how the responses should be coded to reflect different levels of intra-HH agency. For example, do married respondents who are sole decision-makers have more intra-HH agency than those who make the decisions jointly with their spouse, and if yes, how much more? And how much agency is lost if only the spouse decides? The IRT model used in this paper categorical (as distinct from dichotomous), the graded response model (GRM), requires the responses to be ranked.[47] It is assumed here that the ordering (in terms of the response numbering used above) is 1-2-4-3-5.

Because data on intra-HH agency were collected in all three survey rounds, it is possible to assess the test-retest reliability of the intra-HH agency variable. The test-retest correlations indicate that the IRT scale has the highest r (0.254), compared to values of 0.228 to 0.231 for the other three scales. The same results are obtained for the test-retest reliability between the ES and MS values (i.e., the IRT r is equal to 0.365, compared to values of 0.330 to 0.338 for the other three scales.

Table 22 presents the results of multiple regression analysis of the intra-HH agency composite variable by the four alternative scales. The results indicate that the proportional mean scale achieves the highest R^2 (0.17), followed by the closely related standardized mean scale (R^2=0.16). However, the results in Table 22 are most relevant to establishing the validity of the intra-HH agency composite variable. They indicate that intra-HH agency is significantly lower among business owners who are not heads of household (i.e., spouses or children of the HH head), increases with age (but only up to about age 45) and increases with the business owner's cognitive ability and total earnings. Although it is somewhat surprising that the intra-HH agency is not significantly related to gender or to the highest level of schooling completed, the other results support the criterion validity of the intra-HH agency composite variable.

Tables 23 and 24 assess the impact of the business training on the intra-HH agency of women business owners, using follow-up data from both the MS (Table 23) and the ES (Table 24) to obtain both adjusted and unadjusted estimates. The MS results show a very strong positive impact of the business training on the intra-HH agency of women business owners in both adjusted and unadjusted models, independent of the scale used to calculate the composite variable (although the highest R^2 is obtained with the IRT scale, consistent with its relatively high test-retest reliability). However, surprisingly, there is no significant impact when measured with the ES data. Either the impact of the business was short-lived, or the responses in the MS were biased.

Marital agency (Tables 25-27)

Data were also collected in all three survey rounds on eight questions related to married business owners' agency, including: 1) What percentage of your earnings is your spouse aware of? 2) What

[47] There are many alternative IRT models, including those developed for the analysis of qualitative (unordered) responses (Nering and Ostini 2010).

percentage of your earnings are used to cover HH expenses (not including personal goods or business expenses)? 3) What percentage of HH expenses are covered by your spouse? 4) Did your spouse ask for money during the past 12 months that you did not want to provide? 5) Who has access to the money you make from your business? 6) Who decides how the money from your business will be spent? 7) Is there some money that you have sole control over? 8) Who decides how the money earned by your spouse is spent?

One complication is that most of the questions are only applicable to married business owners. As a result, the BS sample size for the marital agency questions is limited to 4,337 business owners (2,596 women business owners and 1,777 men business owners). A second complication is that the responses vary in type, from simple yes-no responses (questions 4 and 7), responses in percentages (questions 1-3), 5-point Likert scales (question 8) and qualitative (unordered) responses (questions 5 and 6).[48] Principal components analysis of the eight responses indicates that they are multidimensional, with responses to four of the questions having relatively small loadings on the first principal component (i.e., questions 2, 4, 7 and 8). These questions are not included in the composite variable.[49] Responses to the remaining four questions all have acceptably high positive loadings on the first principal component (ranging from 0.386 to 0.539) and a Cronbach's standardized alpha of 0.492. The percentage responses to questions 1 and 3 were recoded to a Likert scale (e.g., 0-20%=5, 30-40%=4, etc. for question 1 and 0-20%=1, 30-40%=2, etc. for question 3). Because questions 5 and 6 allowed multiple responses with ambiguous meaning for agency, the responses to these two variables were coded as a dichotomous variable (1 if only the respondent decides, 0 otherwise). Without these simplifications, a more complicated hybrid IRT model would be needed (Stata 2021).

Calculation of the composite variable using the four alternative scales discussed in section 4.2 yields a set of four highly correlated composite variables (r's range between 0.953 and 0.995 in the BS data). Because consistent data were collected in all three survey rounds, it is possible to assess the test-retest reliability of the composite variable. The correlations between the ES and BS values of the composite variable exhibit a narrow range from 0.490 for the IRT scale to 0.523 for the standardized mean scale, while the correlations between the MS and BS values range from 0.496 for the IRT scale to 0.541 for the standardized mean scale. Although the test-retest data do not indicate a clear preference for one (or even two) of the four scales, they do identify the IRT scale as clearly the least reliable scale.

Table 25 presents the results of multiple regression analysis of the marital agency composite variable using the alternative scales. The standardized mean scale achieves the highest R^2 (0.25) among the four alternative scales, consistent with the test-retest reliability, clearly identifying it as the most reliable scale for the marital agency composite variable. However, the other estimates in Table 25 are surprising in several respects. First, and most importantly, women business owners report significantly higher marital agency than men business owners (about one standard deviation higher, other factors equal), while marital agency is negatively and significantly related to business owners' total earnings and cognitive abilities. Marital agency is also unrelated to business owners' age and schooling. These results are counter-intuitive and cast doubt on the criterion validity of the marital agency composite variable. They also encourage further thinking about the meaning of marital agency in this context, as well as experimentation with more complex IRT models.

[48] The detailed responses are described in Appendix 1.

[49] The PCA also indicates that the sign of responses to question 3 needs to be reversed (i.e., business owners whose spouses cover a larger share of HH expenses have less, not more agency).

Tables 26 and 27 assess the impact of the business training on the marital agency of women business owners, using follow-up data from both the MS (Table 26) and the ES (Table 27) to estimate both adjusted and unadjusted models. The MS results show a significant positive impact of the business training on the marital agency of women business owners in both adjusted and unadjusted models, independent of the composite variable's scale (although the highest R^2 is obtained with the principal component scale). However, as with intra-HH agency impact analysis reported in Tables 23 and 24, there is no significant impact when measured with the ES data. Again, either the impact of the business training on marital agency was short-lived, or the responses in the MS were biased.

Personal savings (Tables 35-39)

Personal savings is an important dimension of agency/empowerment, particularly in relation to the business owner's level of earnings. Data on the recent savings of business owners were collected in all three survey rounds (either savings during the last 12 months in the BS or ES or savings during the last three months in the MS). Separate questions were asked, i.e., whether any money was saved during the last 12 months (BS and ES) or last 3 months (MS), and if yes, how much was saved in various saving instruments, including bank accounts, ROSCAs, in cash at home, with family or other relatives, and in physical assets. Although the analysis in this paper focuses on the total amount of money saved in all saving instruments, a problematic variable, Table 35 reports multiple regression analysis of the BS data on whether any money was saved during the last 12 months for all business owners and separately by gender. The results indicate that WBOs were significantly more likely than MBOs to report any savings during the last 12 months (about 14% more likely, other factors equal).

Total savings during the applicable reference period is a very problematic variable in all three survey rounds. The BS data indicate that it is highly skewed (skewness=50.615, kurtosis=2853.2), with 1,072 zero values and 257 very extreme values. However, the natural log of the nonzero values of total savings (BS) has an approximately normal distribution (skewness=-0.302, kurtosis=3.771). Table 36 presents the results of multiple regression analysis of total savings using six alternative statistical models.[50] The results indicate that the two-part and robust regression model estimates (columns 4 and 6) include eight significant estimated coefficients, compared to only three in the OLS regression model estimates (column 1). The results also indicate that total savings is positively and significantly related to the intensity of bank use, total earnings and the HH asset index (a measure of HH income) in all six statistical models. Although these are the only significant relationships in the OLS model (column 1), the other statistical models show positive and significant relationships with completion of tertiary schooling (5 models), and both gender and completion of upper secondary school (4 models each), and negative and significant relationships with subjective time preference (3 models).

Table 37 shows the results of multiple regression analysis of the same total savings variable as a ratio to annualized total earnings. The results are similar in most respects to those in Table 36, except that only the robust regression model estimates include 8 significant estimated coefficients (compared to only one in the OLS regression model estimates). The conclusion is simple: applying OLS estimation to problematic dependent variables can give quite misleading results, including with respect to such important variables as gender and schooling.

---

[50] The Tobit and two-part models (columns 3 and 4) are often used for dependent variables with a large concentration of zero values (Wooldridge 2010). It is not possible to use a log regression model because the log of zero is not defined. An inverse hyperbolic sine (IHS) transformation, which is defined for both zero and negative values, is frequently used in such cases (Ravallion 2017, Bellemarre and Wichman 2019). However, Knowles (2020) shows that this transformation can yield misleading results in models with large concentrations of zero values.

Table 38 reports the results of gender-specific analysis of both savings variables using the robust regression model. These estimates show that the strong positive relationship between both savings variables and schooling and cognitive ability is mainly limited to MBOs and that the significant negative relationship between the ratio of total savings to total earnings and the level of total earnings is limited to WBOs (columns 4-6).

Table 39 reports both adjusted and unadjusted estimates of the impact of the business training provided to randomly selected women business owners on any savings (column 1) and on total savings during the last 12 months using the same six alternative statistical models. The results show that the training had a significant positive impact on any saving during the past 12 months, according to both the adjusted and unadjusted estimates (column 1), and a positive and significant impact on total savings in three of the unadjusted models and in two of the adjusted models (not including either the OLS or OLS winsorized regression models).

Borrowing

In addition to savings, investment in both business assets and non-business assets (housing, consumer durables) can also be financed by borrowing. The ability to borrow may also reduce vulnerability to shocks, an important dimension of Achievements. Detailed data on recent borrowing and on the value of outstanding loans from banks were collected only in the MS. In addition, data on whether the owner currently has any bank loans were obtained in all three survey rounds, while limited data on borrowing from money lenders and from non-bank financial institutions were collected in the ES. Impact estimates are not shown for any of the borrowing outcomes because none showed any significant effects of the business training provided to randomly selected women business owners.

Any current bank loans (Table 40)

Almost one-third of business owners (30.8%) reported outstanding bank loans in the BS, increasing to 35.5% in the ES. In terms of test-retest reliability, the correlation between this outcome in the BS and ES is 0.522 (relatively higher for this outcome possibly because it is a cumulative measure). Table 40 reports the results of multiple regression analysis of this dichotomous (non-problematic) variable at baseline. The results indicate that women business owners are 7.3% less likely to have any outstanding bank loans, other factors equal, that having any outstanding bank loans is positively and significantly related to the number of banks in which accounts are held, the intensity of their use and to total earnings and that having any outstanding bank loans is negatively and significantly related to completed schooling (but significantly only at the tertiary level). There are also some interesting gender differences in Table 40, specifically, having any outstanding bank loans is negatively related to the HH asset index (but significantly for MBOs only) and positively to cognitive ability (but significantly only for WBOs) and to being currently married (but significantly only for MBOs).

Borrowing during the past 3 months (Tables 41-42)

The MS collected more detailed data on borrowing, including whether any money was borrowed during the past 3 months and if yes (29.3%), the amount borrowed by source (mean total amount borrowed=3.091 Rp. millions, including zero values). Table 41 reports the results of multiple regression analysis of any borrowing during the past 3 months, including gender-specific models, and separate models for the source of any loans, i.e., whether any money was borrowed from a bank, friends or

family, a ROSCA[51] or other source.[52] The results indicate that WBOs were significantly more likely (7.1% more likely, but significant only for borrowing from a ROSCA) to have reported borrowing during the past three months, other factors equal. The results also indicate that any borrowing is negatively and significantly related to completion of tertiary schooling among both women and men business owners, a relationship that is significant only for borrowing from friends or family. Any borrowing is also positively and significantly related to borrowing for emergencies, a relationship that is also limited to borrowing from friends or family. Any borrowing (and especially from banks) is also positively and significantly related to the number of banks in which a business owner has accounts and to HH income (but not to the HH asset index (a composite variable), measures included in the model). The negative estimated relationship with the HH asset index is surprising given that the HH asset index is most sensitive to the ownership of relatively expensive assets that would likely be associated with borrowing.

The *amount* of money borrowed during the past 3 months is a problematic variable (skewness=10.516, kurtosis=139.062), with 1,642 zero values and 406 very extreme values (i.e., more than three times the interquartile range above the 75th percentile value). Table 42 (columns 1-4) reports the results of multiple regression analysis of the amount borrowed using four alternative statistical models that are often used for dependent variables with outliers and/or large concentrations of zero vales.[53] Gender specific estimates obtained with Tobit models are presented in columns 5 and 6.[54] The results indicate that the amount of money borrowed during the last 3 months is positively and significantly related to the same three variables (borrowing for emergencies, number of banks in which accounts are held and HH income) with all four statistical models (columns 1-4) as well as with the Tobit model used for men and women business owners separately (columns 5 and 6).

Currently outstanding loans (Tables 43-44)

The MS collected data on the total amount of current outstanding loans by source, including whether any money was borrowed during the last 3 months and if yes (30.8%), the amount still owed (mean=11.207 Rp. millions, including zero values). The amount of outstanding loans is a problematic variable (skewness=10.823, kurtosis=176.014), with 761 zero values and 151 very extreme values. Table 43 reports the results of multiple regression analysis of both whether there are currently any outstanding loans (column 1) and of the total amount of outstanding loans (including zero values) with four alternative statistical models (columns 2-5) that are often used for dependent variables with outliers and/or large concentrations of zero values.[55] The results obtained using gender-specific two-part models are reported in columns 6 and 7.[56] The results in Table 43 indicate that both the presence of any currently outstanding loans and the total amount of currently outstanding loans (including zero values) are significantly lower among women business owners but that the estimated coefficients are

---

[51] ROSCA=Rotating savings and credit association, a community savings mechanism used widely in rural areas.

[52] Other sources include cooperatives, BMTs (Islamic financial institutions) and other unidentified sources.

[53] Unfortunately, the dependent variable has too many zero values to support estimates of standard errors by bootstrapping clusters, as is necessary with the quantile (median) regression or robust regression models.

[54] The Tobit model is selected for the gender-specific analysis because it yields the most highly significant results (i.e., at the 0.001 level) in the full sample.

[55] Again, there are too many zero values to support bootstrapped estimates using the quantile regression or robust regression models.

[56] In this case, there is no clear basis for choosing one of the four alternative statistical models, as they all yield the same number of significant estimates (6).

significant only for OLS applied to the reported variables (columns 1 and 2). The gender-specific results in columns 6 and 7 indicate that a significant positive relationship with primary and second profits and a significant negative relationship with marital agency are limited to WBOs, whereas a significant negative relationship with tertiary schooling is limited to MBOs.

The results in Table 43 also indicate that there are some differences between the estimates in column 1 (any outstanding loans) and those in columns 2-5 (the total amount of currently outstanding loans using four alternative statistical models), including that the former is significantly related to both the highest completed levels of schooling (negatively and systematically) and to cognitive ability (positively), whereas the amount borrowed is not. Instead, the total amount of currently outstanding loans is positively and significantly related to business owners' responses indicating that they borrow for emergencies, to HH income and the HH asset index, whereas the presence of any outstanding loans is not. However, both variables are positively and significantly related to the number of banks in which accounts are held and to the log of primary and second profits combined. Lastly, it is noted that the total amount of currently outstanding loans is negatively and significantly related to marital agency (composite variable) in two of the four models and that this relationship is only statistically significant for WBOs (columns 6-7).

Table 44 shows the results of multiple regression analysis of the sources of any currently unpaid loans (dichotomous variables, not problematic). The results indicate that WBOs and business owners with higher levels of completed schooling are significantly less likely to report any outstanding loans with a bank, whereas business owners with higher cognitive ability, with accounts in more banks and higher primary and second profits are significantly more likely to report any outstanding loans with a bank. The results also indicate that business owners who report that they borrow for emergencies are significantly more likely to have outstanding loans with friends or family, while women business owners are significantly more likely to have outstanding loans from a ROSCA.

Borrowing from money lenders and non-bank financial institutions (Table 45)

The ES collected data from business owners on whether or not they had ever borrowed from a money lender or from a nonbank financial institution and if yes (8.9% and 36.2% respectively), how many times they had borrowed from each source during the past 12 months. The responses to the second question are count variables, with values 0, 1, 2, … , N. The reported number of loans from money lenders range from 0 to 100, with 4,435 zero values and 199 positive values (including two extreme responses of 25 and 100). In contrast, the reported number of more frequently reported loans from nonbank financial institutions range only from 0 to 10, with 3,699 zero values and 935 positive values.

Table 45 shows the results of multiple regression analysis of these data. The analysis in column 1 shows that those who report that they have ever borrowed from a money lender are significantly older, have significantly less schooling, have significantly lower values of the HH asset index, and are significantly less likely to report having had a bad experience with a bank. The number of loans from money lenders during the past 12 months is also significantly higher among those with less schooling and with lower values of the HH asset index in all three of the alternative statistical models (including the preferred Poisson regression model, which is most commonly used to analyze count variables). Business owners who report having borrowed from a nonbank financial institution are also significantly older, significantly less likely to be women, have significantly higher cognitive ability and primary and second business profits, are significantly more likely to report that they borrow for emergencies, have accounts in significantly more banks and are significantly less likely to have had a bad experience with a bank in the past.

Business investment (Tables 46-49)

Business investment during the past 12 months (Tables 46-49)

Data on business investment were collected separately for the primary and (if present) second business in the ES and were obtained using the following two questions: (1) was any investment in "equipment, furniture and fixtures, land, buildings (including renovations)" made in the last 12 months and, if yes, how much was spent? (2) was any investment made in the last 12 months in the form of increases in product stock and, if yes, what is value of these increases in product stock? In addition, the level of investment during approximately the same time period can be estimated indirectly as the change in the reported total market value of business capital between the BS and the ES, thereby creating a unique opportunity to obtain an independent assessment of the reliability of the reported investment data.

Reported investment is a highly problematic variable, with a mean of Rp. 4.666 million, a standard deviation of 29.072, and 2,081 zero values, highly skewed (skewness=845.239, kurtosis=66.948), and with 394 very extreme values (i.e., values that are more than three times the interquartile range above the 75th percentile value). Investment estimated indirectly from the changes in the reported total business capital values between the BS and the ES is also problematic, with a mean of Rp. 10.256 millions, standard deviation of 101.287, highly skewed (skewness=4.954, kurtosis=113.615), with 18 zero values, 1,533 negative values and 472 (407) positive (negative) very extreme values.

Although they are based on different survey questions and do not refer to exactly the same time periods, the reported and indirectly estimated investment variables are significantly correlated (r=0.205) when appropriately adjusted for their different time periods.[57] Assuming that both variables are measures of the same unobserved true level of investment, measurement error would account for the observed differences between these two variables.

The multiple regression analysis of investment reported in Table 46 is done with the full sample of business owners and using five alternative statistical models.[58] Importantly, the estimates are relatively robust across all five statistical models, implying that in this case the problematic nature of the reported investment variable does not significantly bias the results even when OLS is applied directly to the reported value. The estimated coefficients of three of the right-side variables (the HH asset index, primary/second business profits and the total current savings balance) are positive and significant in all five statistical models (with all but three of the estimated coefficients significant at the 0.001 level). Although these relationships are not necessarily causal, they are consistent with the idea that business investment is financed out of all three sources (i.e., HH income, business profits and accumulated savings). In addition, reported investment is positively and significantly related to the presence of any current bank loans in all five models, suggesting that bank loans are also an important source of financing for business investment. These results, with the OLS winsorized model yielding the largest

[57] The mean number of days between the ES and BS household interviews (D) is 514.7 days, ranging from 353 to 759. Accordingly, the formula for the calculated investment variable (CI) is: $CI=(365/D)*(K_{ES}-K_{BS})$, where K_{ES} and K_{BS} refer to the reported total business capital values in the ES and BS respectively.

[58] The models include OLS applied to the reported investment variable (column 1), OLS applied to the reported investment variable with the top 2% of values winsorized (column 2), OLS applied to the IHS (inverse hyperbolic sine) transformation of the reported variable (column 3), Tobit model with the reported dependent variable (column 4), two-step model estimation with the reported dependent variable (column 5, using a probit model in the first step and the log regression model in the second step), and the quantile (median) regression model with the reported dependent variable (column 6). The robust regression model is not used because of the large number of zeroes in the investment variable.

number of significant estimated coefficients (8), are sufficient to establish the criterion validity of the investment data. Investment is also significantly lower among WBOs in three of the models, negatively and significantly related to age in two of the models, and positively and significantly related to the number of children in the HH in three models and to owners' completion of tertiary schooling in two models.

Table 47 presents the results of multiple regression analysis of reported business investment during the past 12 months using the same model as in Table 46 but with 5% of the least reliable observations dropped from the estimation sample (i.e., the observations with the largest absolute differences between the reported and indirectly estimated investment variables). The results are generally consistent with those in Table 46. The main difference is that the estimated standard errors are smaller in Table 47, translating into more significant estimated coefficients in some models (for example, in the case of age, gender, and schooling). The implication is that the measurement error removed from the sample used to obtain the estimates in Table 47 is mainly random measurement error that does not bias the estimates. Large measurement errors do not have the same effects as extreme outliers, skewed distributions and large concentrations of zero values, all of which may bias the estimates (although the results for investment are relatively robust across all five of the statistical models).

Although there are no baseline data on business investment, it is possible to obtain unadjusted estimates of the impact of the business training on the investment of women business owners. The unadjusted impact estimates obtained with the full sample are reported in Table 48. They show a positive impact in all five models that is significant (at the 0.01 level) in only the quantile (median) regression model. The corresponding unadjusted estimates for the trimmed sample (with the 5% of least reliable observations dropped) are presented in Table 49. Significant estimated impact is again limited to the quantile (median) regression model (column 6), with a slightly lower estimated positive impact than in Table 48 and significant only the 0.05 level.

Personal work effort

This is discussed above under Achievements/Quality of life: Business owners' work time.

Mobilization of additional labor

In addition to their own work effort, business owners are able mobilize additional labor by employing both unpaid and paid workers.

Employment of unpaid and paid workers (Tables 30, 31 and 32)

All three survey rounds collected data on the number of unpaid and paid workers in both the primary and second (if present) businesses. The number of unpaid workers in the data set includes business owners. When these are dropped, the mean number of unpaid workers in the primary business in the BS is 1.025 (N=4,825), with a standard deviation of 1.152, ranging from zero to a maximum of 43. The number of unpaid workers is highly skewed (skewness=13.412, kurtosis=438.585) with 15 very extreme values (i.e., more than 3 times the inter-quartile range higher than the 75th percentile value).

The number of paid workers in the primary business is similarly problematic, with a BS mean of 0.436 and a standard deviation of 1.952, ranging from zero (N=4,825) to a maximum of 60 (skewness=15.169, kurtosis=355.045). According to the definition of a very extreme value that is used in this paper (i.e., exceeds the 75% percentile value by 3 times the interquartile range), all 782 of the non-zero values of this variable are very extreme values because the 75th percentile value is 0.

Because data were collected on the number of employees in all three survey rounds, it is possible to assess the test-retest reliability of these two problematic variables. The correlations between BS and ES

numbers of unpaid workers (excluding the business owner) and paid workers are 0.293 and 0.586 respectively. The results of multiple regression analysis of the numbers of unpaid and paid workers in the primary business are reported in Table 30, both for all business owners combined and separately by gender. The results are interesting and suggest that the employment data have criterion validity. Of particular interest is that the results for all business owners combined (columns 1 and 4) show that the number of unpaid workers is closely related to the demographic features of the business owners' HHs, including current marital status, household size and the number of working-age female HH members, whereas the number of paid workers is most closely related to economic variables, including the total value of business capital, primary business profits, and HH income (as measured by the HH assets index). Although these relationships are similar between women and men business owners, there is a striking gender difference in that WBOs employ significantly more unpaid workers, whereas MBOs employ significantly more paid workers.

Because the dependent variables are problematic, it is useful to assess how robust the estimates in Table 30 are with respect to the use of alternative statistical models. Tables 31 and 32 compare the results of multiple regression analysis of the number of unpaid workers (Table 31) and the number of paid workers (Table 32) across four alternative statistical models, including the OLS regression model, OLS winsorized regression model, the two-part regression model, and the Poisson regression model (a statistical model that is commonly used for count variables, i.e., 0,1,2,3, …, N). Although the results in both Tables 31 and 32 are broadly similar across the alternative statistical models, the OLS winsorized regression model (column 2) yields the largest number of statistically significant estimated coefficients in both tables (10 and 7 respectively).

(Not shown) Both adjusted and unadjusted estimates of the impact of the business training provided to randomly selected WBOs on the number of both unpaid and paid workers are statistically insignificant with all four statistical models.

Mobilization of HH labor (Table 70)

All three survey rounds asked business owners who among HH members helps them in their primary and (if present) second business. The different HH members were grouped into the following five categories: spouse, daughter, son, other female HH member, other male HH member. These five variables referring to the primary business were combined into a single composite variable "HH labor mobilization," with the responses aggregated into a single value using the four alternative scales discussed in Section 4.2. The correlations between the alternative scales of this variable in the ES are only moderately high, ranging from r=0.811 to 0.965, with the highest correlations observed between the principal component and IRT scales. In terms of test-retest reliability, the correlations between the BS and ES values range from 0.496 to 0.535, with the highest r observed between the proportional mean scale values.

Table 70 presents the results of multiple regression analysis of the HH labor mobilization composite variables by its four alternative scales as well as separately by gender (for the IRT scale). The results indicate that HH labor mobilization is positively and significantly related to the owner's highest level of completed schooling (linearly) is negatively and significantly related to age (but only among women business owners), to HH size and to the number of working-age women in the HH, while it is negatively and significantly related to the owner's age and to the number of children in the HH. The mobilization of HH labor is also significantly related (at the 0.001 level) to the number of unpaid workers in the primary business, but the relationship is negative for the unweighted proportional and standardized mean scales having negative estimated coefficients, whereas it is positive for the weighted principal component and

IRT scales.[59] Despite this irregularity, the results in Table 70 are moderately supportive of the criterion validity of the "HH labor mobilization" composite variable. However, it is noted that primary business profits are not significantly related to this composite variable when it is added to the model in Table 1.

The business training provided to randomly selected women business owners does not have any significant impact on this composite variable in either adjusted or unadjusted models (not reported).

Business practices (Tables 28 and 29)

Data on self-declared adherence to 15 recommended practices were collected in both the BS and ES (but not in the MS). Nine of the questions concerned record-keeping and record-use practices, while four others concerned marketing practices and the remaining questions focused on negotiating with suppliers for a lower price. The responses are all limited to yes-no answers (as opposed to seeking additional information about the frequency or regularity of the practices). The responses to all 15 questions in the BS are positively correlated with the first principal component (Cronbach's alpha is 0.811). However, three of the 15 questions were not included in the "business practices" composite variable for the following reasons. Two of them permitted responses of "Not applicable" and including them would have reduced the sample size from 4,820 to 4,529, a price that was considered too high for the additional information provided. A third question was dropped because the response was too closely correlated with that of a related question (r=0.994).

The business practices composite variable was calculated from the 12 remaining responses using the four alternative scales discussed in section 4.2. At baseline, the four alternative measures had correlations ranging from 0.898 to 0.991. In terms of test-retest reliability, the correlations between the BS and ES values ranged from highs of 0.504 and 0.500 for the proportional mean and standardized mean scales to a low of 0.433 for the IRT scale. Table 28 presents the results of multiple regression analysis of the business practices composite variables based on the alternative scales. The results are consistent with the test-retest reliability data, i.e., the proportional mean and standardized mean scales are the most reliable scales for the business practices composite variable, while the IRT scale is least reliable.

The regression results in Table 28 also provide strong evidence that the business practices composite variable has criterion validity. Adherence to recommended business practices is significantly higher among women business owners and is positively and significantly related (linearly) to business owners' highest level of schooling completed and willingness to take risks (but surprisingly, not to the owner's cognitive ability). Adherence to recommended business practices is also positively and significantly related to the number of paid workers in the primary business, to total profits from all businesses owned and to the total value of business capital owned. In contrast, adherence to recommended business practices is inversely related to business owners' age.

Table 29 reports adjusted and unadjusted estimates of the impact of the business training on the adherence to recommended business practices by women business owners. The results indicate that the business training had a positive and significant impact (at the 0.01 level for three scales and at the 0.05 level for the IRT scale) of about 0.10 standard deviations on adherence to recommended business practices, with the R^2 highest with the proportional mean and standardized mean scales, consistent with the results of the test-retest assessment of reliability.

[59] This pattern is robust with respect to the winsorization of the highest 2% of values, so the sign reversal is not due to the presence of outliers.

Use of financial services

Intensity of use of banking services (Tables 50-52)

All three survey rounds asked which of a list of 11 banking services were being used by business owners (with multiple responses permitted). The services listed include: savings account, checking account, home mortgage, certificate of deposit (CD), Letter of credit, business loan, vehicle loan, personal loan, health/life insurance, micro credit, and other. Principal components analysis of the BS data shows that all 11 of the services load positively on the first component. However, the first component only accounts for 19.4% of the total variation in the 11 measures, indicating that the set of 11 measures is multi-dimensional to some extent, but with no other dimension standing out as a possible secondary dimension. Accordingly, it was decided to construct a composite variable "intensity of bank use" that includes all 11 measures. The composite variable was calculated using the four alternative scales discussed in section 4.2.

At baseline, the correlations between the four alternative scales used for this composite variable have correlations ranging from 0.817 to 0.996, with the lowest r's observed for pairwise correlations involving the standardized mean scale. In terms of test-retest reliability, the pairwise correlations between the BS and ES values are all close to 0.57 for three of the four scales, but is only 0.43 for the standardized mean scale, with a similar pattern observed for the pairwise correlations between the BS and MS values.

Table 50 presents the results of multiple regression analysis of the intensity of bank use composite variable using the four alternative scales. These results are consistent with the test-retest reliability estimates, i.e., the proportional mean, the principal component and IRT scales achieve the highest R^2s (0.11-0.12), with the standardized mean scale achieving a lower R^2 of 0.10. The results in Table 50 also clearly establish the validity of the intensity of bank use composite variable, which increases significantly with age, but at a decreasing rate. Women business owners have significantly lower intensity of bank use, other factors equal, whereas intensity of bank use is positively related to completion of higher levels of schooling (linearly) and to both cognitive ability and trust in banks. Intensity of bank use is also positively and significantly related to both total earnings and to the HH wealth index (the HH asset index that includes housing characteristics). Table 51 shows similar results obtained with the ES data. However, the estimated gender gap favoring men is smaller in the ES than in the BS. In addition, data on the distance from the owner's home to selected banking services (i.e., the nearest bank branch, the nearest ATM and the nearest branchless banking agent) are available in the ES and a composite variable based on these data is negatively and significantly related to the intensity of bank use, as expected.

The smaller gender gap in the intensity of bank use in the ES versus the BS raises the question of whether the business training provided to randomly selected women business owners made a significant contribution to this change. Table 52 presents both adjusted and unadjusted estimates of the impact of the business training on intensity of bank use for all four scales. The unadjusted estimates, although positive, are all very small and statistically insignificant, but the adjusted estimates for both the proportional mean and principal component scales are both positive and significant at the 0.05 level.

Use of mobile banking services (Tables 53-56)

As discussed in the introduction, the social experiment for which the survey data were collected included both supply- and demand-side interventions designed to support the use of mobile money (MM) available through village-based branchless banking services. The supply-side intervention included both uniform training provided to all branchless banking agents working for the partner bank as well as additional incentives paid to randomly selected agents of the partner bank. The demand-side intervention included business training provided exclusively to randomly assigned *women* business

owners in all villages, including training in the use of the partner bank's branchless banking services. However, unlike the supply-side intervention that was targeted exclusively to the agents of the partner bank, the demand-side intervention may have encouraged the use of MM services from other banks. In the BS, only 1.68% of business owners reported having ever used any MM services (1.58% of WBOs and 1.82% of MBOs). In the ES, however, 4.23% of business owners reported ever use of any MM services (4.73% of WBOs and 3.49% of MBOs).

Because data on ever use of MM were collected in all three survey rounds, it is possible to assess the test-retest reliability of the data on MM ever use, although the length of the time intervals between measurement and re-measurement and the fact that so many changes were occurring during this period may also affected the (unobserved) true values. The correlation between reported ever use of MM at baseline and endline is statistically significant (r=0.253, N=4828).

All three survey rounds also collected data on the use of mobile phones for "banking transactions" and for "mobile money" transactions. These two variables (which are not included in the "intensity of mobile phone use" composite variable discussed below), were combined with the variable on ever use of MM into a single composite variable on "MM use." This composite variable has higher test-retest reliability than the MM ever use measure alone, with the baseline-endline r ranging between 0.311 and 0.390, depending on the scale.

Tables 53 (based on BS data) and 54 (based on ES data) present the results of multiple regression analysis of the MM use composite variable using the four alternative scales to construct the dependent variable. The results indicate that the MM use composite variable has criterion validity and that both the proportional mean and IRT scales are preferred on the basis of their higher R^2s in both Tables 53 and 54. The results indicate that MM use is positively and significantly related to completion of tertiary schooling (and even to completion of upper secondary schooling in the ES data), to the composite variable measuring intensity of bank use (with the relationship stronger in the ES data) and to the HH asset index (again, with the relationship stronger in the ES data). MM use is not significantly related to the distance to banking services composite variable (unlike intensity of bank use in Table 51), which is expected since MM is designed to compensate for limited physical access to branch-based bank services. MM use is also positively and significantly related to the "trust in banks" composite variable (discussed below) in Table 54 (but not in Table 53).

MM use is consistently lower among women business owners in both Tables 53 and 54, but the relationship is only significant in both tables with the IRT scale. Table 55 compares the results of multiple regression analysis of MM use (using the proportional mean scale) by gender using data from both the BS and ES. The results indicate that the positive relationship between MM use and schooling strengthens among women business owners between the BS and ES (cf columns 2 and 5) and may account for the narrowing of the gender gap in MM use between the BS and ES.

Table 56 shows adjusted and unadjusted estimates of the impact of the business training provided to women business owners on their reported ever use of MM (column 1) and on the MM use composite variable by the four alternative scales (columns 2-5). The results indicate that the unadjusted estimate of impact on ever use of MM (column 1) is positive and significant (at the 0.10), whereas the unadjusted estimate of impact on the "MM use" composite variable is positive and significant at the 0.01 level with all four scales (columns 2-5). The results also indicate that the adjusted estimate of impact on ever use of MM (column 1) is positive but not significant, whereas the adjusted estimate of impact on the ""MM use" composite variable is positive and significant (at the 0.05 level) with all four scales (columns 2-5). These results illustrate the value of using composite variables to measure outcomes whenever possible.

Networking

Connections with other sample women business owners (Table 58)

The BS collected data on the relationships between the respondents (both men and women) and the seven sample *women* business owners (only six, in the case of WBO respondents). The information collected includes whether the respondent knows each of the other sample WBOs (yes-no) and, if yes, how the respondent knows each of the sample WBOs (i.e., as an acquaintance, as a distant family member or friend, as a close family member or friend, or as a business partner). WBO respondents reported that they knew a mean number of 3.08 of the other sample WBOs, while MBO respondents reported that they knew a mean number of 2.80 of the sample women business owners. In addition, the BS asked each respondent (both men and women) how frequently they talk with the sample WBOs (i.e., every day, a few times per week, once per week, a few times per month, once per month, every 3 months, every 6 months, once per year) and whether they talk about their businesses when they meet (yes-no). Lastly, the respondents (both men and women) were asked whether they belong to any business-related associations or groups (yes-no).

The responses to all of these questions were combined into a single composite variable ("business connections") using the same four alternative scales to calculate the composite variable. This composite variable is limited to WBOs because the data for MBOs is incomplete (no information was collected on the relationships of either women or men business owners with the sample men business owners). Because these data were collected only in the BS, it is not possible to assess the test-retest reliability of the resulting composite variable.[60] However the variable is highly correlated between the alternative scales, with r's ranging only from 0.941 to 0.998. Table 58 reports the results of multiple regression analysis of the "business connections" variable by alternative scale for the sample of WBOs. Although the results do not provide any basis for choosing between the alternative scales, the results are broadly consistent with prior expectations. Business connections increase significantly with age (at a decreasing rate) and with the owner's cognitive ability, adherence to recommended business practices, the number of customers of the primary business, total earnings and with the HH wealth index. In contrast, business connections are significantly lower for business owners who have completed tertiary-level schooling. In light of the negative relationship with tertiary-level schooling, the strong positive relationship with cognitive ability is puzzling (it may be that business owners who have completed tertiary schooling draw on a geographically wider set of peer contacts). However, the overall results in Table 58 support the criterion validity of the business connections variable. (Not shown) When the business connections composite variable is added to the primary business profits model in Table 1, the relationship is consistently positive but statistically insignificant in all five of the alternative statistical models.

Community participation

Participation in community activities (Table 59)

Data on participation in community organizations were collected in the BS, including whether the respondents participated in any voluntary activities during the past year and, if yes, how many hours were volunteered in a typical month, whether any charitable or social contributions were made in the

[60] It would be informative to compare the corresponding responses of pairs of respondents (i.e., how each characterized their relationship with the other). Unfortunately, the ID numbers of the six sample WBOs are not reported (only their names were entered in the BS questionnaire). It would also be informative to ask the same questions in the MS and ES to see if any of the responses had changed in the interim as well as to assess the extent of any spillover effects from the treated to the untreated.

last year and, if yes, the amount contributed. The responses to these four questions were combined into a composite variable "community participation," which is highly correlated across the four alternative scales (r's vary from 0.899 to 0.999). Because these data were collected only in the BS, it is not possible to assess the test-retest reliability of this composite variable.

Table 59 shows the results of multiple regression analysis of the community participation composite variable by the four alternative scales. The results do not provide a clear basis for choosing among the alternative scales (although the IRT scale yields a slightly lower R^2). However, the results for all four scales indicate that participation in voluntary activities increases significantly with age (at a decreasing rate), that WBOs have significantly lower rates of participation (about 0.17 standard deviations lower in columns 1-3 and about 0.13 lower with the IRT scale in column 4), that participation in voluntary activities is positively and significantly related (linearly) to the highest level of schooling completed and positively and significantly related both to cognitive ability and to the HH wealth index. The gender-specific results using the proportional mean scale in columns 5 and 6 indicate that the significant positive relationships with age and total earnings and the significant negative relationship with number of children in the HH are limited to MBOs, whereas the significant positive relationship with completion of lower secondary schooling and the significant negative relationship with household size are limited to WBOs. Taken together, these results support the criterion validity of the community participation composite variable.

Resources

Age and experience

Age (Table 60)

Age is an important variable, mainly as a proxy for experience but also for life-cycle effects relevant to saving, borrowing and investment. Data on age in years of all respondents was obtained in all three survey, presumably in part as a check on the accuracy of the re-interview process.[61] In addition, data on the age of all household members age 15 and older were obtained in the ES, together with their gender, their relationship to the HH head and the number of mobile phones owned.

Unfortunately, age misreporting is a problem. There is no indication in the BS questionnaire that the respondent was asked to show any document (e.g., ID card) to document age. In the MS questionnaire, respondents are asked their national ID number, but there is no indication that the respondent's age was checked against the ID card (or even if the ID card itself was examined). In the ES questionnaire, both the age, date of birth and ID Number was obtained from all HH members age 15 and older. However, the date of birth information is not included in the data file (presumably to ensure confidentiality). The only information available in the data set is the respondent's age in years.

In the absence of the date of birth it is difficult to assess the reliability of the age data. However, it is clear that the reported ages of business owners in each survey round should be at least as high as their reported ages in previous survey rounds (since the HH interview dates of individual business owners did not overlap between survey rounds). Unfortunately, the age data for several business owners does not meet even this minimal consistency check. For example, the reported ages of 176 business owners in the MS (7.6% of the total) are lower than the ages reported for the same business owners in the BS, while the reported ages of 231 business owners in the ES (5.0%) are lower than the reported ages for

[61] The only other socio-demographic characteristics for which data were collected in all survey rounds are the gender of the business owners and their relationships to heads of household.

the same business owners in the BS. Only 5 of the 176 and 19 of the 231 are ages reported by respondents other than the business owners themselves, so age misreporting by proxy respondents is not the source of the errors. Overall, 9.9% of the ES sample (N=4,666) included one or more apparent errors in the reported ages. Fortunately, most of the errors involve differences of only a few years that do not affect the signs or significance levels of the results. Table 60, for example, reports regression results for two age-related dependent variables, i.e., the cognitive ability score (N=4,812) and wage and salary earnings (N=646), both for the full sample (columns 1 and 3) and for smaller samples in which the obvious errors in age reporting have been deleted (columns 2 and 4). Although the results differ by small amounts, they are quite similar.

Business experience (Table 61)

Separate data on the numbers of years of experience working in the primary and second business were collected in all three survey rounds. In terms of test-retest reliability, the correlations between the reported numbers of years of experience working in the primary business from the three different survey rounds range only between 0.781 and 0.796, indicating a high degree of test-retest reliability in these data. Table 61 presents the results of multiple regression analysis of primary business profits as the dependent variable. The model is the same as that reported in Table 1, including the same five alternative statistical models, but with the addition of both linear and quadratic terms in the reported number of years worked in the primary business as explanatory variables. Primary profits are positively and significantly related to the number of years of experience in all five models (at the 0.001 level in the OLS model with winsoriziation and in the log regression model), whereas the quadratic variable in experience is negatively related to primary profits but significantly in only the log regression model (at only the 0.05 level). Interestingly, both the linear and quadratic terms in age remain statistically significant in all five statistical models in Table 61, with their estimated coefficients reduced only slightly in magnitude from their values in Table 1, suggesting that age and experience, although often used interchangeably as measures of "experience," may reflect somewhat different latent factors.

Education

Highest completed level of schooling (Table 60)

Data on the highest level of schooling completed were collected in the BS (data on the highest grade completed within levels were not obtained). Because the data on schooling were collected only in the BS it is not possible to assess their test-retest reliability, but it is likely that the schooling data are relatively reliable. Dummy variables defined on the basis of the reported highest level of completed schooling are included in many of the regression models designed to assess criterion validity in this paper and are often statistically significant, including several models in which the estimated coefficients indicate a nearly linear positive or negative relationship between the highest level of schooling completed and the dependent variable. When the relationship is even approximately linear, as in Table 60, it does not make much difference how schooling is represented, for example, as a set of dummy variables referring to each level of schooling, as a simple count of the highest level completed (e.g., 1-4) or by an IRT scale. When the relationship is not linear, however, it is preferable to use a dummy variable representing each level of completed schooling to allow for maximum flexibility (as is routinely done in this paper). Although schooling is often significantly related to dependent variables, this not always the case. For example, schooling is not consistently related significantly to primary business profits with other factors held constant (Table 61 and Tables 1-3), and relationships with schooling are sometimes significantly negative (e.g., models of subjective well-being in Tables 17 and 18, with the relationship clearly nonlinear in Table 18).

Cognitive ability (Table 62)

The respondent's cognitive ability is assessed by administering a very brief test consisting of four questions assessing math skills, including two open-ended problems (one involving multiplication, the other involving division) and two multiple-choice questions involving calculations of compound interest. A test score based on the (right-wrong) answers to the four questions was calculated (a composite variable measuring "cognitive ability") using the four alternative scales. Because the test was only administered in the BS, no assessment of test-retest reliability of the cognitive ability composite variable is possible. Table 62 presents the results of multiple regression analysis of the cognitive ability composite variable as a function of age, gender, and the highest level of schooling completed. The results indicate that the IRT scale yields the highest R^2 and is therefore the preferred scale, that cognitive ability is positively and significantly related to age (but with the relationship reaching a maximum at about age 42), to the highest level of schooling completed (linearly), and not significantly related to gender, with the highest level of schooling completed held constant (completed schooling is sharply lower among WBOs). Although these results establish the criterion validity of the cognitive ability composite variable, the percentage of the total variation in cognitive ability accounted for by differences in schooling is very small in this case (as indicated by the R^2s, which range between 0.02 and 0.04). Some of the unexplained variation in cognitive ability is undoubtedly due to errors of measurement (no test is perfect), but it also reflects the effects of unobserved factors such as "native intellectual ability."

Subjective characteristics

Willingness to take risks (Table 63)

Data on willingness to take risks were collected in the BS and MS. In the BS, business owners were asked "In general, would you say that you are someone who takes risks, or do you try to avoid risk?" Respondents were asked to rate their willingness to take risks on a scale of 1 to 10 with 10 meaning that "you are always ready to take risks." According to the questionnaire, if the respondent had difficulty understanding this question, a one paragraph explanation of risk-taking was read to them before obtaining their response. In addition, the battery of questions on "Attitudes" in the MS included asking respondents to express their degree of agreement on a scale of 1-5 (with 5=strongly agree) with the statement "A person can get rich by taking risks."[62] The responses to these two questions are positively and significantly correlated (r=0.222).

Table 63 reports the results of multiple regression analysis of these two standardized measures of willingness to take risks. The results in columns 1 and 2 are based on the responses in the BS on willingness to take risks using two alternative scales (proportional mean and IRT scales). The results do not indicate a preference for either scale, which is not surprising given that the two measures are highly correlated (r=0.981). However, the results show that WBOs are significantly less willing to take risks (about 0.26-0.27 standard deviations less willing, other factors equal) and that willingness to take risks is positively and significantly related to the completion of tertiary-level schooling, to cognitive ability and to the level of the owners profits from primary and (if present) second businesses, but not to owners' ages or to their HH incomes (as measured by the HH asset index).

The results in columns 3 and 4 are based on responses to the statement in the MS "Attitudes" battery that "A person can bet rich by taking risks." Although these two measures are only moderately

[62] The questions on "Attitudes" were combined with the questions on "Assertiveness" in the "personal agency" composite indicator discussed above.

correlated (r=0.222), as mentioned above, the results are quite similar (and are directly comparable because both variables are standardized). The main difference is that the stated willingness to take risks is more closely (linearly) related to schooling in columns 3 and 4 than in columns 1 and 2. The model in column 4 includes the direct measure of HH income available in the MS instead of the HH asset index included in column 3. Both measures of HH income in columns 3 and 4 are insignificant, again consistent with the estimates in columns 1 and 2. Taken together, these results indicate that both variables are likely valid indicators of willingness to take risks and that the differences between them probably consist mainly of random measurement error. It is also noted that the BS measure is positively and significantly related to primary business profits in Table 1 (in four of the five alternative statistical models, the exception being the OLS model).

Subjective time preference (Tables 64 and 65)

Business owners with higher subjective time preference focus on their well-being in the present or near future relative to their well-being in the more distant future. The BS asked business owners the same three questions in order to assess the degree of their subjective time preferences.[63] The first question asked them whether they would prefer to receive Rp. 500,000 in 6 months or in 7 months. Not surprisingly, about 94% percent of respondents indicated that they would prefer to receive the payment in 6 months instead of waiting an additional month to receive the same payment. The second question asked them whether they would prefer to receive Rp. 750,000 in 7 months or Rp. 500,000 in 6 months (i.e., receive an additional Rp. 250,000 by waiting a month longer). The third question asked the respondents whether they would prefer to receive Rp 500, 000 in six months or Rp. 1,000,000 in 7 months (i.e., receive an additional Rp. 500,000 by waiting a month longer). Respondents who still preferred to receive the smaller payment of Rp. 500.000 in six months even when they could receive twice the amount (Rp. 1,000,000) by waiting an additional month are assumed to have the strongest subjective time preference.

The responses to the three questions are included as the individual items in a composite variable "subjective time preference" using the four alternative scales described in section 4.2 to aggregate the individual responses to the single-valued composite variable. Lower values of "subjective time preference" indicate that respondents are more willing to wait to receive a higher future payment. The correlations between the four alternative scales ranged from 0.864 to 0.968, with the lowest r's observed in the pairwise correlations that include the standardized mean scale. Because data on subjective time preferences were collected only in the BS, an assessment of the test-retest reliability of this composite variable is not possible.

Table 64 presents the results of multiple regression analysis of the subjective time preference composite variable by the four alternative scales. The results indicate that subjective time preference is significantly lower among WBOs, implying that they are more willing than MBOs to defer future rewards to gain more, other factors equal. The results also indicate that subjective time preference is also negatively and significantly related both to the highest level of schooling completed and to cognitive ability, but not to age or to either personal earnings or HH income. The R^2s in Table 64 are very low (ranging only from 0.01 to 0.02) and do not indicate a clear basis for preferring one scale over another. However, a low R^2 does not necessarily imply low criterion validity (it may instead signal the presence of important unobserved idiosyncratic factors). Table 65 shows the results of adding the subjective time preference measure to the model for total savings in Table 36. The results indicate that subjective time

[63] This widely used method of eliciting an individual's subjective time preference is referred to in the literature as the "money earlier or later" (MEL) method (Cohen, Ericson, Laibson and White 2020).

preference is negative and significant (at the 0.05 level) in three of the six statistical models. Although these results provide some support for the criterion validity of the time preference composite variable, they fall short of providing strong support. However, it is noted that many studies find similarly weak relationships between empirical measures of subjective time preference and economic behavior such as saving (Cohen, Ericson, Laibson and White 2020).

Access to financial services

Access to bank accounts (Tables 66 and 68)

All three survey rounds asked business owners if they currently have a bank account in their name (yes-no), and if yes, when it was first opened (recoded to the number of years with a bank account and to zero for those without a bank account), the names of the banks in which they have accounts (recoded to the number of banks in which they report having accounts), and, if they do not have a bank account, whether anyone else in the HH currently has a bank account (yes-no, recoded to "yes" if the respondent also has a bank account). These four recoded variables were included as separate items in a composite variable "access to bank accounts" with the values aggregated to a single value using the four alternative scales described in section 4.2. The values of the "access to bank account" composite variable are highly correlated across the four scales (r's range only between 0.982 and 0.999). Test-retest correlations between BS and ES values of "access to bank accounts" are also relatively high (ranging between 0.563 and 0.613, with the highest value observed for the principal component scale), indicating that this variable is relatively reliable.

Table 66 reports the results of multiple regression analysis of the "access to bank accounts" composite variable by the four alternative scales (columns 1-4) and by gender (columns 5-6) based on the BS data. The results indicate that the highest R^2 is obtained with the IRT scale and that access to bank accounts increases significantly with age (but at a decreasing rate) and that access is significantly lower (by about 0.13 standard deviations) among WBOs, other factors equal. The results also indicate that access to bank accounts increases significantly with the highest level of schooling completed (linearly), with cognitive ability, with total earnings and with the HH wealth index and that it is significantly lower among currently married owners. Table 68 presents the results of similar analysis of the "access to bank accounts" composite variable using the ES data, including among the right-side variables the "distance to banking services" composite variable (discussed below) that is only available in the ES data and which the results in Table 68 indicate is negatively and significantly related to bank account access, as expected. The results in Tables 66 and 68 strongly support the criterion validity of the access to bank accounts composite variable.

In addition, the gender-specific results in columns 5 and 6 of Table 66 indicate that the relationship with age, cognitive ability and current marital status are limited to WBOs. The results in Tables 66 and 68 also suggest that the gender gap favoring men in access to bank accounts decreased in magnitude between the BS and the ES. (Not shown) However, both the adjusted and unadjusted estimates of the impact of the business training randomly provided to WBOs are consistently insignificant across all four alternative scales.

Distance to banking services (Tables 67-68)

The ES asked business owners the distance from their homes to the nearest bank branch, bank ATM and branchless banking agent, using 6 pre-coded distance ranges (0.5 km or less, more than 0.5 km to 1 km, more than 1 km to 5 km, more than 5 km to 10 km, more than 10 km to 15 km, and more than 15 km). The responses to these three questions were combined into a single "distance to banking services" composite variable using the alternative scales described in section 4.2. The correlations between the

values of this composite variable across the alternative scales range from 0.833 to 0.997. Assessment of its test-retest reliability is not possible because the data on distance to the nearest banking services were collected only in the ES.

Table 67 shows the results of multiple regression analysis of the distance to banking services composite variable by the four alternative scales. The explanatory variables are limited to the village-level dummy variable indicating that the village is semi-urban (as distinct from rural) and to 28 dummy variables representing the 29 sub-districts in which the sample villages are located (the estimated coefficients of which are not shown in the table). The results indicate that the distance to banking services is significantly higher (at the 0.001 level) in semi-urban village for all scales, but with the highest R^2 obtained using the IRT scale. Apparently, banks expect residents of semi-urban villages to use the services of their urban branches.

Table 68 reports the results of adding the "distance to banking services" composite variable to the right side of a model explaining the previously discussed "access to bank accounts" composite variable by alternative scales and by gender. The results indicate that "access to bank accounts" is negatively and significantly related to the distance of the owner's place of residence from banking services (at the 0.001 level), thereby establishing the criterion validity of the distance to banking services composite variable.

Access to infrastructure (Communications)

Intensity of mobile phone use (Table 57)

Data on access to and the use of mobile phones were collected in all three survey rounds. The eight possible uses identified in the question include: phone calls, messaging, use of the *Whatsapp* app, browsing the internet, playing games, and "use of social media (e.g., facebook, etc.)," mobile money and banking transactions. Principal components analysis indicates that all eight possible uses load positively on the first principal component, which explains 38.2% of the total variation in all eight reported uses. Because almost all mobile phone users reported phone calls and messaging as uses while relatively few reported using their phones for banking or mobile money (2.4% and 7.2% respectively), the "intensity of mobile phone use" composite variable is limited to the remaining four uses (the first principal component of which explained 47.4% of the variation in the four retained uses). The resulting mobile use composite variable is highly correlated across the four alternative scales (from 0.980 to 0.9997 at baseline). It also exhibits relatively high test-retest reliability across survey rounds (from 0.607 to 0.614 between BS and ES measures, with the highest r's observed with the IRT scale).

Table 57 shows the results of multiple regression analysis of the mobile phone use composite variable with the four alternative scales. As expected from the very high correlations across the alternative scales, the results do not provide a clear basis for preferring one scale over another. However, the results in Table 57 clearly establish the criterion validity of the mobile phone use composite variable. As expected, mobile phone use is negatively and significantly related to a business owner's age (with the strength of the negative relationship decreasing with age) and positively and significantly related (linearly) to the highest level of completed schooling. Mobile phone use is also significantly lower among WBOs (by about 0.3 standard deviations), other factors equal. Separate gender analysis (not shown) indicates that the very significantly lower use of mobile phones by WBOs is almost completely due to their lower levels of completed schooling (48.3% of MBOs completed at least upper secondary schooling, compared to only 38.8% of WBOs). (Not shown) Both the adjusted and unadjusted estimates of the impact of the business training provided to randomly selected WBOs on mobile phone use are statistically insignificant across all four scales.

Social capital (trust)

Trust in banks (Table 69)

The BS asked business owners how much confidence they had (1) in the enforcement of contracts between state-owned banks and their customers, (2) in the enforcement of contracts between non-state-owned banks and their customers, and (3) in being able to get their savings back If they were to deposit them in a bank. The responses ranged from 1 to 5, where 1 indicates "no confidence at all" and 5 indicates "extremely confident." PCA was applied to the three responses, finding that they all load positively on the first principal component with loadings varying from 0.385 to 0.665. Accordingly, the three responses were combined into a single composite variable "trust in banks" using the four alternative scales described in section 4.2. Because these data were collected only in the BS, it is not possible to assess their test-retest reliability. However, it is noted that the lowest pairwise correlations between the values using the different scales are those involving the IRT scale.

Table 69 shows the results of multiple regression analysis of the "trust in banks" composite variable by alternative scales. Although the results do not provide a basis for choosing among the alternative scales (the R^2s are all about 0.02), as many as 8 of the relationships are statistically significant. For example, trust in banks is positively and significantly related to age (but decreasing with age) and is significantly higher among women business owners (although this relationship is not significant with the IRT scale), other factors equal, and with the number of different banks in which the owner has accounts, and is negatively and significantly related to the business owner reporting a prior bad experience with banks. The low R^2s do not necessarily indicate that the "trust in banks" measure incorporates a lot of measurement error and is therefore unreliable. Instead, the low R^2s may indicate that an individual business owner's "trust in banks" depends mainly on unobserved idiosyncratic factors. In fact the regression results in Tables 50 and 51 indicate that the "trust in banks" composite variable is positively and significantly related to the intensity of bank use in both the BS and ES and is therefore probably a reliable and valid measure. The conclusion is that the trust in banks composite variable has moderately strong criterion validity and is probably also reliable, despite the low R^2s in Table 69.