The Design and Use of New Measures

Orazio P. Attanasio
Yale & NBER

Expectations and Learning in Dynamic Structural Model

Econometric Society Summer School in Dynamic Structural Econometrics

University College London -July 2^{nd} 2025

Outline

Introduction

- For many years, most economists thought that good empirical work can only be based on a limited set of measures.

Introduction

- For many years, most economists thought that good empirical work can only be based on a limited set of measures.
- The prevalent perception has been that:
  - we can only credibly use measures of:
    - what people buy (or do);
    - their resources;
    - prices.

Introduction

- For many years, most economists thought that good empirical work can only be based on a limited set of measures.
- The prevalent perception has been that:
    - we can only credibly use measures of:
        - what people buy (or do);
        - their resources;
        - prices.
    - ... and from these measures we could possibly infer:
        - preferences and other structural parameters;
        - features of market structure.

Introduction

- For many years, most economists thought that good empirical work can only be based on a limited set of measures.
- The prevalent perception has been that:
    - we can only credibly use measures of:
        - what people buy (or do);
        - their resources;
        - prices.
    - ... and from these measures we could possibly infer:
        - preferences and other structural parameters;
        - features of market structure.
- However, *identification* could only be achieved with strong assumptions on:
    - tastes, beliefs, expectations, and information.

Introduction

- For many years, most economists thought that good empirical work can only be based on a limited set of measures.
- The prevalent perception has been that:
    - we can only credibly use measures of:
        - what people buy (or do);
        - their resources;
        - prices.
    - ... and from these measures we could possibly infer:
        - preferences and other structural parameters;
        - features of market structure.
- However, *identification* could only be achieved with strong assumptions on:
    - tastes, beliefs, expectations, and information.
- These *strong* assumptions were needed because preference and attitudes, beliefs and subjective expectations were largely perceived as unmeasurable.

Introduction

- For many years, most economists thought that good empirical work can only be based on a limited set of measures.
- The prevalent perception has been that:
  - we can only credibly use measures of:
    - what people buy (or do);
    - their resources;
    - prices.
  - ... and from these measures we could possibly infer:
    - preferences and other structural parameters;
    - features of market structure.
- However, *identification* could only be achieved with strong assumptions on:
  - tastes, beliefs, expectations, and information.
- These *strong* assumptions were needed because preference and attitudes, beliefs and subjective expectations were largely perceived as unmeasurable.
- Skepticism towards questions that pose *hypothetical* situations and evidence from *stated* rather than *actual* choices.

Introduction

- There were good reasons for relying on a *revealed preferences* approach.

Introduction

- There were good reasons for relying on a *revealed preferences* approach.
- Measuring hypotheticals, preferences, attitudes is fret with many difficulties.
  - Framing effects and several biases;
    - Debate on the use *Contingent Valuation*,
    - Hausman (1994, 2012b,a).... *from dubious to hopeless*.

Introduction

- There were good reasons for relying on a *revealed preferences* approach.
- Measuring hypotheticals, preferences, attitudes is fret with many difficulties.
  - Framing effects and several biases;
    - Debate on the use *Contingent Valuation*,
    - Hausman (1994, 2012b,a).... *from dubious to hopeless*.
  - What are we measuring? What are we modelling?
    - Stigler and Becker (1977): "De Gustibus Non Est Disputandum".
      "... tastes neither change capriciously nor differ importantly between people'. [...] one does not argue over tastes for the same reason that one does not argue over the Rocky Mountains - both are there, will be there next year, too, and are the same to all men."

Introduction

- The discussion of what to measure and how goes back a long time;

Introduction

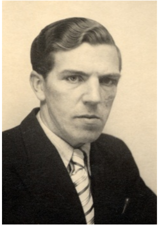
- The discussion of what to measure and how goes back a long time;
- Block and Marschak (1960) on RUM, cited by Caplin (2012);
  - *''Our particular way of defining the class of basic observations and, correspondingly, of the directly testable conditions is to some extent arbitrary. Depending on the range of possible experiments and other observations, it may be preferable to define the class more narrowly [..] [or] more broadly. Following the practice of psychologists, we might admit the ranking, by the subject, of three or more objects as an observable fact, although the subject observed action consists in this case of a verbal statement. [..] We might even admit as observable the subject verbal statements of the relative intensity of his preferences''.*
- Stated preferences and conjoint analysis:
  - Luce (1956, 1959); Luce and Tukey (1964); Luce and Suppes (1965).

Introduction

- The discussion of what to measure and how goes back a long time;

  Haavelmo (1958) presidential address is another important example:

I think most of us feel that if we could use *explicitly* such variables as, e.g., what people *think* prices or incomes are going to be, or variables expressing what people *think* the effects of their actions are going to be, we would be able to establish relations that could be more accurate and have more explanatory value. But because the statistics on such variables are not very far developed, we do not take the formulation of theories in terms of these variables seriously enough. It is my belief that if we can develop more explicit and a priori convincing economic models in terms of these variables, which are realities in the minds of people even if they are not in the current statistical yearbooks, then ways and means can and will eventually be found to obtain actual measurements of such data.

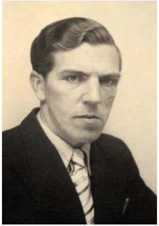


Introduction

- The discussion of what to measure and how goes back a long time;

  Haavelmo (1958) presidential address is another important example:

I think most of us feel that if we could use *explicitly* such variables as, e.g., what people *think* prices or incomes are going to be, or variables expressing what people *think* the effects of their actions are going to be, we would be able to establish relations that could be more accurate and have more explanatory value. But because the statistics on such variables are not very far developed, we do not take the formulation of theories in terms of these variables seriously enough. It is my belief that if we can develop more explicit and a priori convincing economic models in terms of these variables, which are realities in the minds of people even if they are not in the current statistical yearbooks, then ways and means can and will eventually be found to obtain actual measurements of such data.



- ... but the consensus went towards an almost exclusive revealed preference approach.

Introduction

- ... with the possible exception of experimental economics.
- Experimental economists have been trying a variety of different methods to measure *preferences*, *beliefs* and *attitudes*;
- Lab work on various mechanisms to elicit primitives.

Introduction

- ... with the possible exception of experimental economics.
- Experimental economists have been trying a variety of different methods to measure *preferences*, *beliefs* and *attitudes*;
- Lab work on various mechanisms to elicit primitives.
- More recently experiments have been brought to the field and collected together with observational data to measure:
  - preference for and attitudes towards redistribution and attitudes towards migrants;
  - bargaining and social preferences;
  - reciprocity in conflict areas;
  - willingness to compete.

References

Introduction

Things have been changing

- There are some interesting discussions about what we could and should measure:
  - Contributions in the volume edited by Caplin and Schott (2008) and in particular the discussion between Gul and Pesendorfer for *mindless economics* v Camerer for *mindful economics*.
- Several studies now use stated preferences to model consumption behavior;
  - ... and recent innovations go beyond using stated (in addition to) revealed preferences.

Introduction

Things have been changing

- There are some interesting discussions about what we could and should measure:
  - Contributions in the volume edited by Caplin and Schott (2008) and in particular the discussion between Gul and Pesendorfer for *mindless economics* v Camerer for *mindful economics*.
- Several studies now use stated preferences to model consumption behavior;
  - ... and recent innovations go beyond using stated (in addition to) revealed preferences.
- Researchers have been moving away from models that imply full rationality:
  - Robustness;
  - Non-rational beliefs;
  - Learning.

Introduction

Things have been changing

- There are some interesting discussions about what we could and should measure:
  - Contributions in the volume edited by Caplin and Schott (2008) and in particular the discussion between Gul and Pesendorfer for *mindless economics* v Camerer for *mindful economics*.
- Several studies now use stated preferences to model consumption behavior;
  - ... and recent innovations go beyond using stated (in addition to) revealed preferences.
- Researchers have been moving away from models that imply full rationality:
  - Robustness;
  - Non-rational beliefs;
  - Learning.
- .... and some innovative work has been done in terms of measurement.
  - Eliciting data on policy preferences;
  - Eliciting data on information.

References

Introduction

Things have been changing

- Measurement of subjective expectations.
  - Data on subjective expectations may allow avoiding strong assumptions.
  - These data are being used to estimate models of:
    - retirement choices;
    - education and occupation choices;
    - income and earnings dynamics.
- Measurement of beliefs and perceptions.
- Measurement of attitudes.

Introduction

Things have been changing

- Measurement of subjective expectations.
  - Data on subjective expectations may allow avoiding strong assumptions.
  - These data are being used to estimate models of:
    - retirement choices;
    - education and occupation choices;
    - income and earnings dynamics.
- Measurement of beliefs and perceptions.
- Measurement of attitudes.
- Answers to questions about choices and events in counterfactual situations make the *identification* of structural models of behavior easier.

References

Introduction

I will start by briefly discussing some of the conceptual issues are stake:

This discussion is largely based on Almås, Attanasio, and Jervis (2024)

- What to measure and how:
  - it should be informed by the economic models one works on.

Introduction

I will start by briefly discussing some of the conceptual issues are stake:

This discussion is largely based on Almås et al. (2024)

- What to measure and how:
    - it should be informed by the economic models one works on.
- It may be useful to go beyond standard concepts and attempt to measure:
    - Attitudes;
    - Subjective expectations;
    - Beliefs;
    - Choices in hypothetical situations: *thought experiments*

Introduction

I will start by briefly discussing some of the conceptual issues are stake:

This discussion is largely based on Almås et al. (2024)

- What to measure and how:
  - it should be informed by the economic models one works on.
- It may be useful to go beyond standard concepts and attempt to measure:
  - Attitudes;
  - Subjective expectations;
  - Beliefs;
  - Choices in hypothetical situations: *thought experiments*
  - Such measures allow the empirical study of more complex models and might achieve identification of the structural parameters and the establishment of causal links under weaker sets of assumptions (see Heckman and Pinto (2023))

Introduction

I will start by briefly discussing some of the conceptual issues are stake:

This discussion is largely based on Almås et al. (2024)

- What to measure and how:
  - it should be informed by the economic models one works on.
- It may be useful to go beyond standard concepts and attempt to measure:
  - Attitudes;
  - Subjective expectations;
  - Beliefs;
  - Choices in hypothetical situations: *thought experiments*
  - Such measures allow the empirical study of more complex models and might achieve identification of the structural parameters and the establishment of causal links under weaker sets of assumptions (see Heckman and Pinto (2023))
- Measurement systems and the importance of measurement error
  - Measurement error is pervasive and important and should be recognized as such;
  - Augmenting the theoretical models we consider with a *measurement system* could be useful to the design of survey strategies.
  - Methodological issues in using factor models as measurement systems:
    - Metric and anchoring;
    - Cardinality.

Introduction

I will then provide some examples of measurement work and its use:

- Subjective expectations and their use
  - Estimating flexible income processes in India and Colombia.

Introduction

I will then provide some examples of measurement work and its use:

- Subjective expectations and their use
  - Estimating flexible income processes in India and Colombia.
- The quality of information and risk sharing
  - Quality of information in extended families networks in Tanzania.

Introduction

I will then provide some examples of measurement work and its use:

- Subjective expectations and their use
  - Estimating flexible income processes in India and Colombia.
- The quality of information and risk sharing
  - Quality of information in extended families networks in Tanzania.
- Parental investment and beliefs about the process of child development
  - A static model applied to Colombian data;
  - The adaptability of beliefs: learning in India.

Outline

Outline

Measurement and Theory: the issues at stake

- *Much of the material in this section is taken from Almås et al. (2024)*

- *Economic models* often use abstract constructs that are not directly observable.

Measurement and Theory: the issues at stake

- *Much of the material in this section is taken from Almås et al. (2024)*
- *Economic models* often use abstract constructs that are not directly observable.
- The measures we collect and use in empirical work are (and should be) driven by the theoretical framework that organizes our thinking.

Measurement and Theory: the issues at stake

- *Much of the material in this section is taken from Almås et al. (2024)*
- *Economic models* often use abstract constructs that are not directly observable.
- The measures we collect and use in empirical work are (and should be) driven by the theoretical framework that organizes our thinking.
- Examples:
  - The work by Keynes, Kutznets, Stone, and others and the development of National Accounts: Keynes (1936); Kuznets et al. (1937); Kuznets (1941); Gilbert et al. (1949); Stone (1984).
  - Demand systems and price indexes: Stone (1954); Christensen et al. (1975); Deaton and Muellbauer (1980).

Measurement and Theory: the issues at stake

- *Much of the material in this section is taken from Almås et al. (2024)*

- *Economic models* often use abstract constructs that are not directly observable.
- The measures we collect and use in empirical work are (and should be) driven by the theoretical framework that organizes our thinking.
- Examples:
  - The work by Keynes, Kutznets, Stone, and others and the development of National Accounts: Keynes (1936); Kuznets et al. (1937); Kuznets (1941); Gilbert et al. (1949); Stone (1984).
  - Demand systems and price indexes: Stone (1954); Christensen et al. (1975); Deaton and Muellbauer (1980).
- For many years, the prevalent practice among most economists was the almost exclusive use of *choice* data, or objectively measureble variables
  - Consumption, income, prices, even anthropomteric....
  - ... but not attitudes, choices in hypothetical situations, beliefs, subjective expectations

Measurement and Theory: the issues at stake

- Interesting discussions in the 1950s:
  - Tobin (1959); Katona (1959) on buying intentions.

Measurement and Theory: the issues at stake

- Interesting discussions in the 1950s:
  - Tobin (1959); Katona (1959) on buying intentions.
- Even at that time some researchers tried out and justified alternative measures;
  - Juster (1966) on buying intentions and purchasing probabilities;
  - Katona's work on the Michigan survey and consumer sentiment Katona (1974);
    - Curtin (2016) provides a nice survey;
  - Juster and Shay (1964) hypothetical car loans to estimate the elasticity to maturity and interest rates.

Measurement and Theory: the issues at stake

- Interesting discussions in the 1950s:
  - Tobin (1959); Katona (1959) on buying intentions.
- Even at that time some researchers tried out and justified alternative measures;
  - Juster (1966) on buying intentions and purchasing probabilities;
  - Katona's work on the Michigan survey and consumer sentiment Katona (1974);
    - Curtin (2016) provides a nice survey;
  - Juster and Shay (1964) hypothetical car loans to estimate the elasticity to maturity and interest rates.
- Revealed preferences become the main approach in economics:
  - Samuelson (1938, 1948); Arrow (1959)

> "I propose, therefore, that we start anew in direct attack upon the problem, dropping off the last vestiges of the utility analysis. This does not preclude the introduction of utility by any who may care to do so, nor will it contradict the results attained by use of related constructs. It is merely that the analysis can be carried on more directly and from a different set of postulates", Samuelson (1938).

Measurement and Theory: The issues at stake

- An approach based *exclusively* on choice data can be very restrictive, as it precludes the analysis of important phenomena and of a wide sets of models.

Measurement and Theory: The issues at stake

- An approach based *exclusively* on choice data can be very restrictive, as it precludes the analysis of important phenomena and of a wide sets of models.
- The use of such measurements should complement the use of choice data and such measures should be *validated*:
  - Samuelson (1938) ends his paper with:
    "In concluding this exposition, it may be well to sound a warning. Woe to any who deny any one of the three postulates* here! For they are, of course, deducible as theorems from the conventional analysis. They are less restrictive than the usual set-up, and logically equivalent to the reformulation of Hicks and Allen. It is hoped however, that the orientation given here is more directly based upon those elements which must be taken as data by economic science, and is more meaningful in its formulation."

* I. confronted with a given set of prices and with a given income, our idealised individual will always choose the same set of goods.

II. behaviour is independent of the units in which prices are expressed.

III. In any two price and income situations and corresponding quantities of consumer's goods given by equations (1.0) the individual must always behave consistently in the sense that (5.12) and (5.22) cannot hold simultaneously.

Measurement and Theory: The issues at stake

- An approach based *exclusively* on choice data can be very restrictive, as it precludes the analysis of important phenomena and of a wide sets of models.
- The use of such measurements should complement the use of choice data and such measures should be *validated*:
    - Samuelson (1938) ends his paper with:
      "In concluding this exposition, it may be well to sound a warning. Woe to any who deny any one of the three postulates* here! For they are, of course, deducible as theorems from the conventional analysis. They are less restrictive than the usual set-up, and logically equivalent to the reformulation of Hicks and Allen. It is hoped however, that the orientation given here is more directly based upon those elements which must be taken as data by economic science, and is more meaningful in its formulation."
- Manski (1990) argued that the issue is not *what* is being measured, but *the specific tools and questionnaires* being used.
- In the same paper, Manski notices that intention data, while not used much by economists, are widely used in other disciplines.

Beyond standard measures

- Limiting the set of *admissable* measures leads to strong assumptions are forced on the empirical models that are estimated.
  - The lack of data on subjective expectations often implies assuming rational expectations;

Beyond standard measures

- Limiting the set of *admissable* measures leads to strong assumptions are forced on the empirical models that are estimated.
  - The lack of data on subjective expectations often implies assuming rational expectations;
  - The lack of data on subjective beliefs about the returns to investment opportunities often implies assuming perfect knowledge of the relevant processes;

Beyond standard measures

- Limiting the set of *admissable* measures leads to strong assumptions are forced on the empirical models that are estimated.
    - The lack of data on subjective expectations often implies assuming rational expectations;
    - The lack of data on subjective beliefs about the returns to investment opportunities often implies assuming perfect knowledge of the relevant processes;
    - The lack of data on the quality of information in networks or extended families often implies assuming complete information.

Beyond standard measures

- Limiting the set of *admissable* measures leads to strong assumptions are forced on the empirical models that are estimated.
  - The lack of data on subjective expectations often implies assuming rational expectations;
  - The lack of data on subjective beliefs about the returns to investment opportunities often implies assuming perfect knowledge of the relevant processes;
  - The lack of data on the quality of information in networks or extended families often implies assuming complete information.
- The use of data on subjective expectations and choices in counterfactual situations changes the nature of the residuals of the equations that are estimated and therefore the nature of identification.
- This is particularly relevant for the estimation of models with lagged dependent variables or selection.

Beyond standard measures

- One issue with measures that go beyond choices and objectively measured variables is that they might be difficult to obtain.
- What we often have are *markers* related to the *latent factors* that populate the theoretical models we use.

Beyond standard measures

- One issue with measures that go beyond choices and objectively measured variables is that they might be difficult to obtain.
- What we often have are *markers* related to the *latent factors* that populate the theoretical models we use.
- Data on:
  - Attitudes
  - Beliefs
  - Subjective expectations and counterfactuals

  need to be validated with choice data.

Outline

Measurement error and its modeling

- The best approach to the difficulties of designing and implementing new measurement tools is to recognise explicitely the presence of measurement error.
- In the literature this is explicitly recognised with the construction of a measurement system, see the fundamental contribution of Cunha, Heckman, and Schennach (2010) and Cunha and Heckman (2008).

Measurement error and its modeling

- The best approach to the difficulties of designing and implementing new measurement tools is to recognise explicitely the presence of measurement error.
- In the literature this is explicitly recognised with the construction of a measurement system, see the fundamental contribution of Cunha, Heckman, and Schennach (2010) and Cunha and Heckman (2008).
  - A Measurement System relates available measures to unobserved latent factors.

Measurement error and its modeling

- The best approach to the difficulties of designing and implementing new measurement tools is to recognise explicitely the presence of measurement error.
- In the literature this is explicitly recognised with the construction of a measurement system, see the fundamental contribution of Cunha, Heckman, and Schennach (2010) and Cunha and Heckman (2008).
  - A Measurement System relates available measures to unobserved latent factors.
- The constructions of a careful measurement system presents key challenges:
  - what is the nature of measurement error? is it uncorrelated across measures?

Measurement error and its modeling

- The best approach to the difficulties of designing and implementing new measurement tools is to recognise explicitely the presence of measurement error.
- In the literature this is explicitly recognised with the construction of a measurement system, see the fundamental contribution of Cunha, Heckman, and Schennach (2010) and Cunha and Heckman (2008).
  - A Measurement System relates available measures to unobserved latent factors.
- The constructions of a careful measurement system presents key challenges:
  - what is the nature of measurement error? is it uncorrelated across measures?
  - what is the metric of the unobserved latent factors?

Measurement error and its modeling

- The best approach to the difficulties of designing and implementing new measurement tools is to recognise explicitely the presence of measurement error.
- In the literature this is explicitly recognised with the construction of a measurement system, see the fundamental contribution of Cunha, Heckman, and Schennach (2010) and Cunha and Heckman (2008).
  - A Measurement System relates available measures to unobserved latent factors.
- The constructions of a careful measurement system presents key challenges:
  - what is the nature of measurement error? is it uncorrelated across measures?
  - what is the metric of the unobserved latent factors?
  - is it comparable across different contexts and different times?

Measurement error and its modeling

- The best approach to the difficulties of designing and implementing new measurement tools is to recognise explicitely the presence of measurement error.
- In the literature this is explicitly recognised with the construction of a measurement system, see the fundamental contribution of Cunha, Heckman, and Schennach (2010) and Cunha and Heckman (2008).
  - A Measurement System relates available measures to unobserved latent factors.
- The constructions of a careful measurement system presents key challenges:
  - what is the nature of measurement error? is it uncorrelated across measures?
  - what is the metric of the unobserved latent factors?
  - is it comparable across different contexts and different times?
  - do the latent factors have cardinality?

Measurement error and its modeling

- The best approach to the difficulties of designing and implementing new measurement tools is to recognise explicitely the presence of measurement error.
- In the literature this is explicitly recognised with the construction of a measurement system, see the fundamental contribution of Cunha, Heckman, and Schennach (2010) and Cunha and Heckman (2008).
  - A Measurement System relates available measures to unobserved latent factors.
- The constructions of a careful measurement system presents key challenges:
  - what is the nature of measurement error? is it uncorrelated across measures?
  - what is the metric of the unobserved latent factors?
  - is it comparable across different contexts and different times?
  - do the latent factors have cardinality?
- Some of these challenges can be tackled when designing questionnaires and their deployment.

A Measurement System for Child Development

- Suppose we are studying child development in the early years.
- Suppose we assume that child development has three dimensions:
  - Cognition and language;
  - Internalising Socio-emotional Skills;
  - Externalising Socio-emotional Skills.

A Measurement System for Child Development

- Suppose we are studying child development in the early years.
- Suppose we assume that child development has three dimensions:
  - Cognition and language;
  - Internalising Socio-emotional Skills;
  - Externalising Socio-emotional Skills.
- We can represent such a system as:

$$m_{i,t}^{jk} = \alpha_t^{j,k} + \beta_t^{j,k}\theta_{i,t}^{j} + \epsilon_{i,t}^{jk}, \quad j = 1, ...J; \;\; k = 1, ..., K.$$

- $\theta_{i,t}^{j}$ is factor j for individual i at time t;
- $m_{i,t}^{jk}$ is measure k for factor j;
- $\epsilon_{i,t}^{jk}$ is an *additive* measurement error ;
- $\alpha_t^{j,k}$ and $\beta_t^{j,k}$ are parameters representing the *discriminating* and *saliency* properties of a measurement item.

A Measurement System for Child Development

- **Assumptions**
  - At least 2 measures $m_{i,t}^{jk}$ per factor are available;
  - Measurement errors are independent across at least two measures;

A Measurement System for Child Development

- **Assumptions**
  - At least 2 measures $m_{i,t}^{jk}$ per factor are available;
  - Measurement errors are independent across at least two measures;
  - Although non-parametric identification might be possible with enough measures, assumptions about the distribution of the factors $\boldsymbol{\theta}$ are typically used.

A Measurement System for Child Development

- **Assumptions**
  - At least 2 measures $m_{i,t}^{jk}$ per factor are available;
  - Measurement errors are independent across at least two measures;
  - Although non-parametric identification might be possible with enough measures, assumptions about the distribution of the factors $\boldsymbol{\theta}$ are typically used.
  - In this example, each measure is determined by only one factor.
  - This is a *dedicated* system;
  - This assumption can be somewhat relaxed.

Existing measures and measurement systems

- A good measurement system could be interpreted as an effective way to aggregate and summarize the available measures and items.

Existing measures and measurement systems

- A good measurement system could be interpreted as an effective way to aggregate and summarize the available measures and items.
- This is analogous to the scoring algorithms that are often used in psychometrics, where a set of (often binary) variables are converted into a *score*.

Existing measures and measurement systems

- A good measurement system could be interpreted as an effective way to aggregate and summarize the available measures and items.
- This is analogous to the scoring algorithms that are often used in psychometrics, where a set of (often binary) variables are converted into a *score*.
- Often available measures use pre-defined scoring algorithms.
  - Examples of child development measures:
    - Bayley Scales of Infant Development; Woodcock Johnson; MacArthur-Bates Communicative Development Inventories (MB-CDIs).
  - These scoring algorithms were constructed calibrating on obsolete samples and/or are over-simplified.

Existing measures and measurement systems

- Using predefined scoring algorithms in all contexts is not necessary and can be very misleading and inefficient.
    - Different measurement systems and scoring algorithms should be used.
        - adapting to the context and the nature of the items (continuous, discrete, binary).

Existing measures and measurement systems

- Using predefined scoring algorithms in all contexts is not necessary and can be very misleading and inefficient.
  - Different measurement systems and scoring algorithms should be used.
    - adapting to the context and the nature of the items (continuous, discrete, binary).
- Estimating an explicit measurement system from the individual available items also allows flexibility about functional form assumptions on the distribution of latent factors.

Outline

Scaling: the importance of anchoring

- We need to fix the metric for the unobserved factors:
  - Normalise at least one of α's and one of the β's per factor;
  - Normalise the mean and variance of the $\boldsymbol{\theta}$.

Scaling: the importance of anchoring

- We need to fix the metric for the unobserved factors:
  - Normalise at least one of α's and one of the β's per factor;
  - Normalise the mean and variance of the $\boldsymbol{\theta}$.
- These normalisations define the scale and location of the unobservable factors;
  - e.g child development: height at 2 or wages at 22?

Scaling: the importance of anchoring

- We need to fix the metric for the unobserved factors:
  - Normalise at least one of α's and one of the β's per factor;
  - Normalise the mean and variance of the $\boldsymbol{\theta}$.
- These normalisations define the scale and location of the unobservable factors;
  - e.g child development: height at 2 or wages at 22?
- The metric used to evaluate the unobserved latent factors is important:
  - Comparability across different contexts;
    - e.g. in measuring child development, comparing across different ages and measuring growth;
  - Evaluating the size of the impact achieved by certain interventions.

Scaling: the importance of anchoring

- Some of the normalisations are not innocuous:
  - Normalising $\beta_t^{jk} = 1, \ \forall t$ for a specific measure k is a very strong assumption.
  - Analogous considerations apply for the α's.
    - Agostinelli and Wiswall (2016) have some interesting work on this.
- The lack of longitudinal data covering long periods with the same measure is an additional problem.

Scaling: the importance of anchoring

- Some of the normalisations are not innocuous:
  - Normalising $\beta_t^{jk} = 1, \ \forall t$ for a specific measure k is a very strong assumption.
  - Analogous considerations apply for the α's.
    - Agostinelli and Wiswall (2016) have some interesting work on this.
- The lack of longitudinal data covering long periods with the same measure is an additional problem.
- Issues when different items are available over different ages or different cohorts.
  - Link adjacent ages with similar items to establish bridges over the entire life cycle,
    - Attanasio et al. (2020) look at child development from age 6 to 72 months.

Scaling: the importance of anchoring

- Some of the normalisations are not innocuous:
  - Normalising $\beta_t^{jk} = 1, \ \forall t$ for a specific measure k is a very strong assumption.
  - Analogous considerations apply for the α's.
    - Agostinelli and Wiswall (2016) have some interesting work on this.
- The lack of longitudinal data covering long periods with the same measure is an additional problem.
- Issues when different items are available over different ages or different cohorts.
  - Link adjacent ages with similar items to establish bridges over the entire life cycle,
    - Attanasio et al. (2020) look at child development from age 6 to 72 months.
- This problem is common to many data sets.

Cognitive Skills in the UK Millenium Cohort Study

- The MCS is one of the best cohort studies and follows children born in 2000.
- We have several measures of development in different dimensions: cognition, internalising and externalising skills and more.

Cognitive Skills in the UK Millenium Cohort Study

- The MCS is one of the best cohort studies and follows children born in 2000.
- We have several measures of development in different dimensions: cognition, internalising and externalising skills and more.
- Children are observed at birth and ages 3, 5, 7, 11, 14 and 17.

Cognitive Skills in the UK Millenium Cohort Study

- The MCS is one of the best cohort studies and follows children born in 2000.
- We have several measures of development in different dimensions: cognition, internalising and externalising skills and more.
- Children are observed at birth and ages 3, 5, 7, 11, 14 and 17.

Cognitive Skills in the MCS	
Age 9 months	Denver Developmental Screening Test and MacArthur Communication Development Inventory
Age 3	Bracken Basic Concepts subscales (colours, etc) and British Ability Scales (BAS) Naming Vocabulary
Age 5	BAS Naming Vocabulary, Pattern Construction, and Picture Similarities scales
Age 7	BAS Pattern Construction, Word Reading scales, and NFER Number Skills
Age 11	BAS Verbal Similarities scale, Cambridge Gambling Task, and Spatial Working Memory
Age 14	Cambridge Gambling Task and Applied Psychology Unit Vocabulary test
Age 17	Number Analogies activity and GCSE grades by subject

Outline

Data collection strategies

- Specifying precisely the measurement system and the assumptions that will be required to use the available measures can inform the way data are collected.

Data collection strategies

- Specifying precisely the measurement system and the assumptions that will be required to use the available measures can inform the way data are collected.
- One can design survey methods to ensure that crucial assumptions hold.

Data collection strategies

- Specifying precisely the measurement system and the assumptions that will be required to use the available measures can inform the way data are collected.
- One can design survey methods to ensure that crucial assumptions hold.
- E.g. independence of measurement errors of different items is a key assumption.
- This assumption can be insured by appropriate survey features:
  - randomise evaluators assignment for different measures;
  - timing of data collection.

Data collection strategies

- Specifying precisely the measurement system and the assumptions that will be required to use the available measures can inform the way data are collected.
- One can design survey methods to ensure that crucial assumptions hold.
- E.g. independence of measurement errors of different items is a key assumption.
- This assumption can be insured by appropriate survey features:
  - randomise evaluators assignment for different measures;
  - timing of data collection.
- More generally, the economic model one uses should dictate and direct:
  - the type of measures collected;
  - how they are collected.

Outline

Outline

The design and use of subjective expectations data

- My first example is about **measures of subjective expectations**.
- This is based on work I have done with Manolo Arellano, Sam Crossman and Victor Sancibrian: Arellano, Attanasio, Crossman, and Sancibrián (2024)

The design and use of subjective expectations data

- My first example is about **measures of subjective expectations**.
- This is based on work I have done with Manolo Arellano, Sam Crossman and Victor Sancibrian: Arellano et al. (2024)
- Data on subjective expectations have become relatively common, following the path-breaking work of Manski (2004) and collaborators.
- It is now established that well designed questionnaires can be used to elicit the probability distribution of future uncertain variables.

  $\rightarrow$ it is possible to go beyond point expectations and obtain subjective probability distributions or subjective CDFs.

The design and use of subjective expectations data

- It is also possible to elicit *conditional* expectations of future events.
  - The NYFed data are a good example:
  - Expectations of receiving an offer... and
  - Expectations of best offer conditional on receiving an offer;

The design and use of subjective expectations data

- It is also possible to elicit *conditional* expectations of future events.
  - The NYFed data are a good example:
  - Expectations of receiving an offer... and
  - Expectations of best offer conditional on receiving an offer;
- It is then possible to use these data to estimate complex models of individual behavior.

The design and use of subjective expectations data

- It is also possible to elicit *conditional* expectations of future events.
  - The NYFed data are a good example:
  - Expectations of receiving an offer... and
  - Expectations of best offer conditional on receiving an offer;
- It is then possible to use these data to estimate complex models of individual behavior.
- However, some important challenges are still present, even for variables such as income:
  - Anchoring: elicited or current income?

The design and use of subjective expectations data

- It is also possible to elicit *conditional* expectations of future events.
  - The NYFed data are a good example:
  - Expectations of receiving an offer... and
  - Expectations of best offer conditional on receiving an offer;
- It is then possible to use these data to estimate complex models of individual behavior.
- However, some important challenges are still present, even for variables such as income:
  - Anchoring: elicited or current income?
  - Unit of observation: individual or household?

The design and use of subjective expectations data

- It is also possible to elicit *conditional* expectations of future events.
  - The NYFed data are a good example:
  - Expectations of receiving an offer... and
  - Expectations of best offer conditional on receiving an offer;
- It is then possible to use these data to estimate complex models of individual behavior.
- However, some important challenges are still present, even for variables such as income:
  - Anchoring: elicited or current income?
  - Unit of observation: individual or household?
  - If several individuals within the household: joint or marginals?

The use of subjective expectations data

- A first use of the data is in the validation of the question obtained.
- This can be done by relating the answers to observable variables including actual realizations of the variables.

The use of subjective expectations data

- A first use of the data is in the validation of the question obtained.
- This can be done by relating the answers to observable variables including actual realizations of the variables.
- It is also possible to devise questions that embed consistency tests.

The use of subjective expectations data

- A first use of the data is in the validation of the question obtained.
- This can be done by relating the answers to observable variables including actual realizations of the variables.
- It is also possible to devise questions that embed consistency tests.
- The existing evidence is that it is possible to obtain consistent and meaningful answers if questions posed properly.

The use of subjective expectations data

- A first use of the data is in the validation of the question obtained.
- This can be done by relating the answers to observable variables including actual realizations of the variables.
- It is also possible to devise questions that embed consistency tests.
- The existing evidence is that it is possible to obtain consistent and meaningful answers if questions posed properly.
- The 'right' way to ask questions may be context specific.
- Some preliminary questions can serve as 'training' of the respondents.

Using subjective expectations data

- Once the data are available, they constitute an important opportunity to estimate more flexible models.
- From an econometric point of view, they can provide identification of simple and complex models under much weaker conditions.

Using subjective expectations data

- Once the data are available, they constitute an important opportunity to estimate more flexible models.
- From an econometric point of view, they can provide identification of simple and complex models under much weaker conditions.
- The additional information comes from the fact that *future* and *current* variables are *different objects*, changing the nature of the residuals.

Using subjective expectations data

- Once the data are available, they constitute an important opportunity to estimate more flexible models.
- From an econometric point of view, they can provide identification of simple and complex models under much weaker conditions.
- The additional information comes from the fact that *future* and *current* variables are *different objects*, changing the nature of the residuals.
- This same logic applies when questions about *conditional expectations* are available.

Using subjective expectations data

- Once the data are available, they constitute an important opportunity to estimate more flexible models.
- From an econometric point of view, they can provide identification of simple and complex models under much weaker conditions.
- The additional information comes from the fact that *future* and *current* variables are *different objects*, changing the nature of the residuals.
- This same logic applies when questions about *conditional expectations* are available.

- I will provide an example on a simple model of household income taken from some recent work from Arellano et al. (2024)

Data - India and Colombia

- For both India and Colombia, we use data on total household income expectations for poor households.
    - For India, we use data collected for the evaluation of a *microfinance intervention* (loans for cow or buffalo) in Andhra Pradesh.

Data - India and Colombia

- For both India and Colombia, we use data on total household income expectations for poor households.
  - For India, we use data collected for the evaluation of a *microfinance intervention* (loans for cow or buffalo) in Andhra Pradesh.
  - Two waves of data (collected in Jan/Feb 2008, April/June 2009); 774 households in 65 villages in the balanced panel.

Data - India and Colombia

- For both India and Colombia, we use data on total household income expectations for poor households.
  - For India, we use data collected for the evaluation of a *microfinance intervention* (loans for cow or buffalo) in Andhra Pradesh.
  - Two waves of data (collected in Jan/Feb 2008, April/June 2009); 774 households in 65 villages in the balanced panel.
  - Attanasio & Augsburg (2016) confirm the soundness of responses to the income expectations module.

Data - India and Colombia

- For both India and Colombia, we use data on total household income expectations for poor households.
  - For India, we use data collected for the evaluation of a *microfinance intervention* (loans for cow or buffalo) in Andhra Pradesh.
  - Two waves of data (collected in Jan/Feb 2008, April/June 2009); 774 households in 65 villages in the balanced panel.
  - Attanasio & Augsburg (2016) confirm the soundness of responses to the income expectations module.
  - The income expectations questions refer to total household income *a year from now*.

Data - India and Colombia

- For both India and Colombia, we use data on total household income expectations for poor households.
    - For India, we use data collected for the evaluation of a *microfinance intervention* (loans for cow or buffalo) in Andhra Pradesh.
    - Two waves of data (collected in Jan/Feb 2008, April/June 2009); 774 households in 65 villages in the balanced panel.
    - Attanasio & Augsburg (2016) confirm the soundness of responses to the income expectations module.
    - The income expectations questions refer to total household income *a year from now*.

    - For Colombia, we use data collected for the evaluation of a *conditional cash transfer* (CCT) programme in Colombia (*Familias en Acción*).
    - Households in SISBEN 1 (the lowest socioeconomic classification).
    - We use two waves collected in July/Nov.2003, Nov.2005/Mar.2006; 2,230 households in the balanced panel.

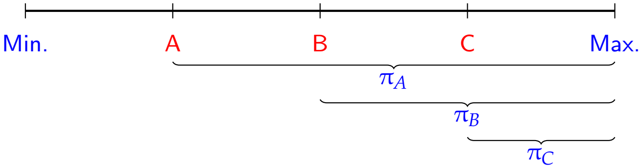
Data - India and Colombia

- For both India and Colombia, we use data on total household income expectations for poor households.
  - For India, we use data collected for the evaluation of a *microfinance intervention* (loans for cow or buffalo) in Andhra Pradesh.
  - Two waves of data (collected in Jan/Feb 2008, April/June 2009); 774 households in 65 villages in the balanced panel.
  - Attanasio & Augsburg (2016) confirm the soundness of responses to the income expectations module.
  - The income expectations questions refer to total household income *a year from now*.

  - For Colombia, we use data collected for the evaluation of a *conditional cash transfer* (CCT) programme in Colombia (*Familias en Acción*).
  - Households in SISBEN 1 (the lowest socioeconomic classification).
  - We use two waves collected in July/Nov.2003, Nov.2005/Mar.2006; 2,230 households in the balanced panel.
  - Income expectations refer to total household income *next month*.

Eliciting expectations

- In each wave of the surveys, subjective expectations about future income were elicited.
- First by asking the max. & min. possible income the household might earn over the next period.
- Then 3 probabilities are elicited using a ruler, π_A, π_B, π_C (as in Dominitz & Manski 1997 and others):

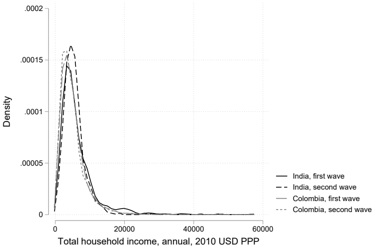


$$A = 0.75 * Min. + 0.25 * Max.; \quad B = 0.5 * (Min + Max.);$$

$$C = 0.25 * Min. + 0.75 * Max.;$$

Data

- Similar methods were used in India and Colombia.
- The main difference is that the expectations of future income refers to next month in Colombia and next year in India.
- In both cases these are very poor households.



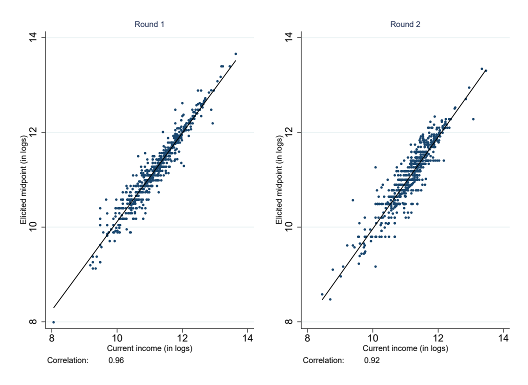
Note. The Figure shows the distribution of total household income in the two study populations, in 2010 PPP USD. Monthly income in Colombia is annualized for comparability.

Figure: Household income across study populations.

Validation : India

- To check whether the subjective expectations make sense we plot the subjective mean against actual income.

Figure: India: current income and reported midpoint

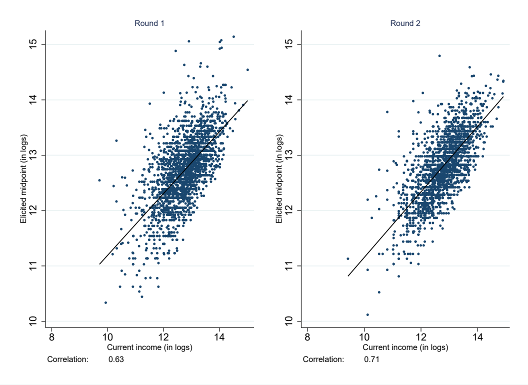


Note. The solid black corresponds to the linear regression fit.

Validation : Colombia

- To check whether the subjective expectations make sense we plot the subjective mean against actual income.

Figure: Colombia: current income and reported midpoint



Note. The solid black corresponds to the linear regression fit.

Modeling approach

- Subjective expectations data give us information on perceived conditional *cdf* which we match to the parameters of a structural model.

Modeling approach

- Subjective expectations data give us information on perceived conditional *cdf* which we match to the parameters of a structural model.
- A household's subjective probability distribution of (log) future income $y_{i,t+1}$ is $F_{it}(r) = \Pr\left(y_{i,t+1} < r \mid I_{it}\right)$; I_{it} is the information available to household i at t.

Modeling approach

- Subjective expectations data give us information on perceived conditional *cdf* which we match to the parameters of a structural model.
- A household's subjective probability distribution of (log) future income $y_{i,t+1}$ is $F_{it}(r) = \Pr\left(y_{i,t+1} < r \mid I_{it}\right);\ I_{it}$ is the information available to household i at t.
- The survey elicitation process yields noisy measurements p_{jit} of $F_{it}\left(r_{jit}\right)$ for $r_{jit} = r_{it}^{\mathsf{min}} + \left(r_{it}^{\mathsf{max}} - r_{it}^{\mathsf{min}}\right) j/4\ (j = 1, 2, 3)$

Modeling approach

- Subjective expectations data give us information on perceived conditional *cdf* which we match to the parameters of a structural model.
- A household's subjective probability distribution of (log) future income $y_{i,t+1}$ is $F_{it}(r) = \Pr\left(y_{i,t+1} < r \mid I_{it}\right);\ I_{it}$ is the information available to household i at t.
- The survey elicitation process yields noisy measurements p_{jit} of $F_{it}\left(r_{jit}\right)$ for $r_{jit} = r_{it}^{\text{min}} + \left(r_{it}^{\text{max}} - r_{it}^{\text{min}}\right) j/4\ (j = 1, 2, 3)$ or of the subjective cumulative odds $\ell_{jit}^{*} = logit\left[F_{it}\left(r_{jit}\right)\right]$, where $logit\,(p) = \ln\left[p/\left(1-p\right)\right]$.

Modeling approach

- Subjective expectations data give us information on perceived conditional *cdf* which we match to the parameters of a structural model.
- A household's subjective probability distribution of (log) future income $y_{i,t+1}$ is $F_{it}(r) = \Pr\left(y_{i,t+1} < r \mid I_{it}\right)$; I_{it} is the information available to household i at t.
- The survey elicitation process yields noisy measurements p_{jit} of $F_{it}\left(r_{jit}\right)$ for $r_{jit} = r_{it}^{\min} + \left(r_{it}^{\max} - r_{it}^{\min}\right) j/4$ $(j = 1, 2, 3)$ or of the subjective cumulative odds $\ell_{jit}^{*} = logit\left[F_{it}\left(r_{jit}\right)\right]$, where $logit(p) = \ln\left[p/(1-p)\right]$.
- We only observe 3 points of F_{it} for each unit, but many across units.
- We assume (plausibility) that elicitation errors are additive:

$$\ell_{jit} = \ell_{jit}^{*} + \varepsilon_{jit}$$

  where $\ell_{jit} = logit\left(p_{jit}\right)$ are the observed cumulative odds and ε_{jit} is an elicitation measurement error independent of I_{it}.
- The set I_{it} consists of time-varying and time-invariant characteristics.
- The time varying variables include observable current income y_{it} and indicators

Information set: time invariant variables

- We adopt a latent variable approach assuming that time-invariant characteristics are captured by an unobservable individual effect α_i.

Information set: time invariant variables

- We adopt a latent variable approach assuming that time-invariant characteristics are captured by an unobservable individual effect α_i.
- This effect is intended to capture both household-level characteristics and geographical (e.g. village level) characteristics.

Information set: time invariant variables

- We adopt a latent variable approach assuming that time-invariant characteristics are captured by an unobservable individual effect α_i.
- This effect is intended to capture both household-level characteristics and geographical (e.g. village level) characteristics.
- Therefore, our models take the form

$$\ell^*_{jit} = g\left(r_{jit}, y_{i,t}, x_{i,t}, \alpha_i\right) \qquad (i = 1, ..., n; j = 1, 2, 3; t = 1, 2)$$

where g is a nondecreasing function in its first argument.

Information set: time invariant variables

- We adopt a latent variable approach assuming that time-invariant characteristics are captured by an unobservable individual effect α_i.
- This effect is intended to capture both household-level characteristics and geographical (e.g. village level) characteristics.
- Therefore, our models take the form

$$\ell^*_{jit} = g\left(r_{jit}, y_{i,t}, x_{i,t}, \alpha_i\right) \qquad (i = 1, ..., n; j = 1, 2, 3; t = 1, 2)$$

  where g is a nondecreasing function in its first argument.
- We note that the individual effect α_i may be correlated with $\left(r_{jit}, y_{i,t}, x_{i,t}\right)$.

Linear model

- We first consider a linear autoregressive model with logistic shocks (ignoring $x_{i,t}$ for notational simplicity):

$$
\begin{aligned}
&y_{i,t+1} = \rho y_{i,t} + \alpha_i + \sigma v_{i,t+1} \\
&v_{i,t+1} \sim \mathit{Logistic} \text{ independent of } \left(y_{i,t}, \alpha_i\right).
\end{aligned}
$$

Linear model

- We first consider a linear autoregressive model with logistic shocks (ignoring $x_{i,t}$ for notational simplicity):

$$\begin{aligned} y_{i,t+1} &= \rho y_{i,t} + \alpha_i + \sigma v_{i,t+1} \\ v_{i,t+1} &\sim \mathit{Logistic} \text{ independent of } \left(y_{i,t}, \alpha_i\right). \end{aligned}$$

- The corresponding conditional *cdf* is:

$$\begin{aligned} \Pr\left(y_{i,t+1} \leqslant r \mid y_{i,t}, \alpha_i\right) &= \Pr\left(v_{i,t+1} \leqslant \frac{r - \rho y_{i,t} - \alpha_i}{\sigma} \mid y_{i,t}, \alpha_i\right) \\ &= \Lambda\left(\frac{r - \rho y_{i,t} - \alpha_i}{\sigma}\right). \end{aligned}$$

Linear model

- We first consider a linear autoregressive model with logistic shocks (ignoring $x_{i,t}$ for notational simplicity):

$$y_{i,t+1} = \rho y_{i,t} + \alpha_i + \sigma v_{i,t+1}$$
$$v_{i,t+1} \sim \mathit{Logistic} \text{ independent of } \left(y_{i,t}, \alpha_i\right).$$

- The corresponding conditional *cdf* is:

$$\begin{aligned}\Pr\left(y_{i,t+1} \leqslant r \mid y_{i,t}, \alpha_i\right) &= \Pr\left(v_{i,t+1} \leqslant \frac{r - \rho y_{i,t} - \alpha_i}{\sigma} \mid y_{i,t}, \alpha_i\right) \\ &= \Lambda\left(\frac{r - \rho y_{i,t} - \alpha_i}{\sigma}\right).\end{aligned}$$

- Therefore, in this case g is linear:

$$\ell_{jit} = \ell^*_{jit} + \varepsilon_{jit} = \beta_0 r_{jit} + \beta_1 y_{i,t} + \eta_i + \varepsilon_{jit},$$

where $\beta_0 = 1/\sigma$, $\beta_1 = -\rho/\sigma$ and $\eta_i = -\alpha_i/\sigma$.

Linear model: Using subjective expectation vs observed income

- This is a linear panel model with fixed effects and strictly exogenous regressors that can be consistently estimated by within-groups to produced unbiased estimates.

Linear model: Using subjective expectation vs observed income

- This is a linear panel model with fixed effects and strictly exogenous regressors that can be consistently estimated by within-groups to produced unbiased estimates.
- Despite superficial similarities there are profound differences between the two approaches.
    1. An AR(1) model without fixed effects can be estimated on a single cross-section, with subjective expectations., Two waves are needed for fixed effects.
    2. Estimation of subjective expectation models does not suffer from Nickell bias.
        - The reason is that outcomes are not future incomes but points in the predictive distribution;
        - Therefore the error term does not contain future shocks but only measurement error in predictive probabilities.

Linear model: Using subjective expectation vs observed income

- This is a linear panel model with fixed effects and strictly exogenous regressors that can be consistently estimated by within-groups to produced unbiased estimates.
- Despite superficial similarities there are profound differences between the two approaches.
  1. An AR(1) model without fixed effects can be estimated on a single cross-section, with subjective expectations., Two waves are needed for fixed effects.
  2. Estimation of subjective expectation models does not suffer from Nickell bias.
     - The reason is that outcomes are not future incomes but points in the predictive distribution;
     - Therefore the error term does not contain future shocks but only measurement error in predictive probabilities.
  3. The subjective expectation approach does not force households to have rational expectations in the sense of optimal statistical forecasts.

Non linear models

- This model can be generalised to include non linearities and interactions with observables.
- In the paper we consider a version of the Arellano Blundell and Bonhomme (2017) model:

$$\ell_{jit} = \beta_0(r_{jit}) + \beta_1(r_{jit})\psi\left(y_{i,t}\right) + \beta_2(r_{jit})\eta_i + \varepsilon_{jit}$$

Non linear models

- This model can be generalised to include non linearities and interactions with observables.
- In the paper we consider a version of the Arellano Blundell and Bonhomme (2017) model:

$$\ell_{jit} = \beta_0(r_{jit}) + \beta_1(r_{jit})\psi\left(y_{i,t}\right) + \beta_2(r_{jit})\eta_i + \varepsilon_{jit}$$

- A similar approach can be used in estimation.

Non linear models

- This model can be generalised to include non linearities and interactions with observables.
- In the paper we consider a version of the Arellano Blundell and Bonhomme (2017) model:

$$\ell_{jit} = \beta_0(r_{jit}) + \beta_1(r_{jit})\psi\left(y_{i,t}\right) + \beta_2(r_{jit})\eta_i + \varepsilon_{jit}$$

- A similar approach can be used in estimation.
- Given the nonlinearity, the model properties will vary with the current income realisation, the quantile of the shocks and the distribution of fixed effects.
- We report:
    - Interquartile ranges;
    - The Bowley-Kelley measure of skewness for different quantiles;
    - The ABB measure of persistence:

$$\rho_{it}\left(\tau\right) = \frac{\partial q_{it}\left(\tau\right)}{\partial y_{i,t}} = -\frac{\partial g\left(q_{it}\left(\tau\right), y_{i,t}, x_{i,t}, \alpha_i\right)}{\partial y_{i,t}} \Big/ \frac{\partial g\left(q_{it}\left(\tau\right), y_{i,t}, x_{i,t}, \alpha_i\right)}{\partial r}.$$

Results: linear model

Models with or without fixed effects

- When the model includes household fixed effects, we also report how much of this variation is explained by village effects

Results: linear model

Models with or without fixed effects

- When the model includes household fixed effects, we also report how much of this variation is explained by village effects

Indian data

- Measured persistence by ρ is around unity without fixed effects and 0.94 when fixed effects are included.
- The inclusion of fixed effects absorbs quite a bit of the variability in the estimated σ.

India: *linear AR(1): with and without fixed effects*

	No FE	FE
ρ	0.97	0.93
	(0.94, 1.00)	(0.90, 0.96)
σ	0.56	0.31
	(0.51, 0.60)	(0.29, 0.33)
$IQR_{0.75}$	1.22	0.69
	(1.13, 1.33)	(0.64, 0.74)
$IQR_{0.90}$	2.44	1.38
	(2.25, 2.65)	(1.29, 1.47)
σ^2_η		0.22
		(0.18, 0.27)
σ^2_η village		0.14
		(0.14, 0.19)
σ^2_ε	1.24	1.14
	(1.21, 1.27)	(1.10, 1.18)

Note. $n = 2230 \times 6$; 95% block bootstrap CI

Results: linear model

Models with or without fixed effects

- When the model includes household fixed effects, we also report how much of this variation is explained by village effects

Colombian data

- Estimated ρ is close to unity without fixed effects, but the estimates are halved when fixed effects are included.

Colombia: *linear AR(1) with and without fixed effects*

	No FE	FE
ρ	0.71	0.50
	(0.67, 0.74)	(0.46, 0.55)
σ	0.98	0.65
	(0.93, 1.03)	(0.63, 0.67)
$IQR_{0.75}$	2.16	1.43
	(2.05, 2.26)	(1.38, 1.48)
$IQR_{0.90}$	4.31	2.86
	(4.10, 4.52)	(2.75, 2.96)
σ^2_η		0.48
		(0.44, 0.52)
σ^2_η village		0.12
		(0.12, 0.17)
σ^2_ε	1.46	1.09
	(1.42, 1.49)	(1.05, 1.12)

Note. $n = 2230 \times 6$; 95% block bootstrap CI

India: linear model with observable controls

- We first generalise the model adding controls on the sources of income and shocks.

$$\ell_{jit} = \beta_0 r_{jit} + \beta_1 y_{it} + \delta_0' x_{it} + \delta_1' x_{it} y_{it} + \eta_i + \varepsilon_{jit}.$$

ρ	$\leqslant$2 sources	3 sources	4+ sources
No shock	0.87	0.91	0.83
	(0.79, 0.95)	(0.84, 0.99)	(0.74, 0.93)
Health	0.92	0.97	0.89
	(0.86, 0.98)	(0.91, 1.04)	(0.81, 0.97)
Agricultural	0.90	0.97	0.87
	(0.86, 0.95)	(0.92, 1.01)	(0.81, 0.94)
Other	0.99	1.04	0.97
	(0.88, 1.09)	(0.93, 1.14)	(0.84, 1.08)
σ		0.30	
		(0.28, 0.32)	
$IQR_{0.90}$		1.34	
		(1.23, 1.42)	
σ^2_η		0.25	
		(0.22, 0.32)	
σ^2_η village		0.15	
		(0.15, 0.21)	
σ^2_ε		1.13	
		(1.08, 1.16)	

Note. n parenthesis we report 90% block bootstrap CI (1000 repetitions).

India: linear model with observable controls

- We first generalise the model adding controls on the sources of income and shocks.

$$\ell_{jit} = \beta_0 r_{jit} + \beta_1 y_{it} + \delta_0' x_{it} + \delta_1' x_{it} y_{it} + \eta_i + \varepsilon_{jit}.$$

ρ	$\leqslant$2 sources	3 sources	4+ sources
0% farm	0.87	0.90	0.83
	(0.83, 0.92)	(0.85, 0.95)	(0.76, 0.90)
50% farm	0.91	0.94	0.87
	(0.87, 0.95)	(0.90, 0.98)	(0.80, 0.93)
75% farm	0.93	0.96	0.89
	(0.88, 0.98)	(0.92, 1.01)	(0.82, 0.96)
σ		0.30	
		(0.28, 0.32)	
$IQR_{0.90}$		1.33	
		(1.24, 1.41)	
σ^2_η		0.23	
		(0.19, 0.28)	
σ^2_η village		0.13	
		(0.13, 0.18)	
σ^2_ε		1.12	
		(1.07, 1.16)	

Note. n parenthesis we report 90% block bootstrap CI (1000 repetitions).

Table: India — linear model augmented with household characteristics (shocks and income

Colombia: linear model with observable controls

- We first generalise the model adding controls on the sources of income and shocks.

$$\ell_{jit} = \beta_0 r_{jit} + \beta_1 y_{it} + \delta_0' x_{it} + \delta_1' x_{it} y_{it} + \eta_i + \varepsilon_{jit}.$$

ρ	1 earner	2 earners	3+ earners
0% regular	0.34	0.36	0.48
	(0.21, 0.48)	(0.24, 0.47)	(0.34, 0.63)
75% regular	0.51	0.52	0.61
	(0.43, 0.58)	(0.43, 0.59)	(0.51, 0.71)
100% regular	0.56	0.57	0.65
	(0.49, 0.63)	(0.48, 0.66)	(0.55, 0.76)
σ		0.64	
		(0.62, 0.67)	
$IQR_{0.90}$		2.83	
		(2.72, 2.92)	
σ^2_η		0.48	
		(0.45, 0.52)	
σ^2_η village		0.11	
		(0.12, 0.16)	
σ^2_ε		1.08	
		(1.04, 1.11)	

Note. n parenthesis we report 90% block bootstrap CI (1000 repetitions).

Colombia: linear model with observable controls

- We first generalise the model adding controls on the sources of income and shocks.

$$\ell_{jit} = \beta_0 r_{jit} + \beta_1 y_{it} + \delta_0' x_{it} + \delta_1' x_{it} y_{it} + \eta_i + \varepsilon_{jit}.$$

ρ	1 earner	2 earners	3+ earners
No shock	0.54	0.55	0.58
	(0.47, 0.62)	(0.47, 0.63)	(0.48, 0.68)
Health	0.64	0.65	0.67
	(0.50, 0.77)	(0.51, 0.79)	(0.53, 0.81)
Other	0.44	0.46	0.48
	(0.32, 0.56)	(0.33, 0.58)	(0.34, 0.61)
σ		0.65	
		(0.62, 0.67)	
$IQR_{0.90}$		2.84	
		(2.73, 2.93)	
σ^2_η		0.48	
		(0.45, 0.53)	
σ^2_η village		0.11	
		(0.12, 0.16)	
σ^2_ε		1.09	
		(1.04, 1.11)	

Note. n parenthesis we report 90% block bootstrap CI (1000 repetitions).

Results on non-linear models: Colombia

- In the Colombian data introducing observables does matter for deviations from linearity.

Results on non-linear models: Colombia

- In the Colombian data introducing observables does matter for deviations from linearity.
- High persistence is concentrated among households with 3+ earners with low income and negative shocks.

Results on non-linear models: Colombia

- In the Colombian data introducing observables does matter for deviations from linearity.
- High persistence is concentrated among households with 3+ earners with low income and negative shocks.
- As in India, we observe skewness decreasing with income (ABB-like), but with values in the positive range.

Results on non-linear models: Colombia

- In the Colombian data introducing observables does matter for deviations from linearity.
- High persistence is concentrated among households with 3+ earners with low income and negative shocks.
- As in India, we observe skewness decreasing with income (ABB-like), but with values in the positive range.
- The linear AR(1) model is not as strongly rejected on the Colombian data as it was for India, but it can still be rejected.

Results on non-linear models: India

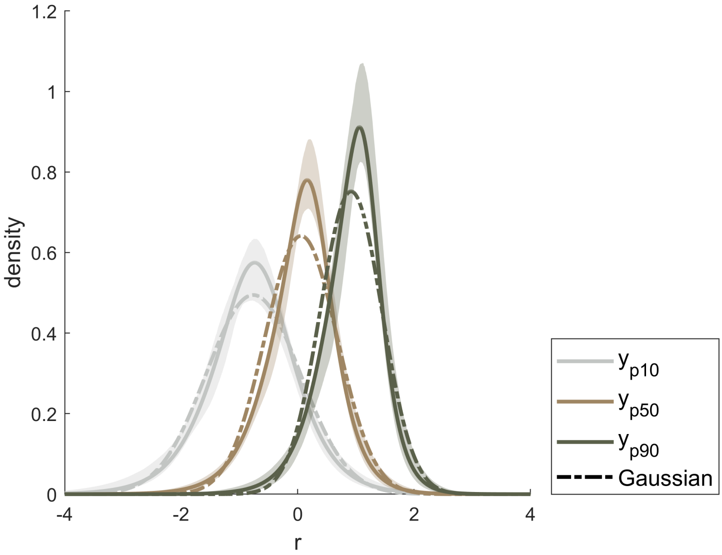
- We now consider a non-linear flexible model with additive fixed effects:

$$\ell_{jit} = \beta_0^{\dagger}(s_{jit}) + \beta_1^{\dagger}(s_{jit})\psi\,(y_{it}) + \eta_i + \varepsilon_{jit},$$

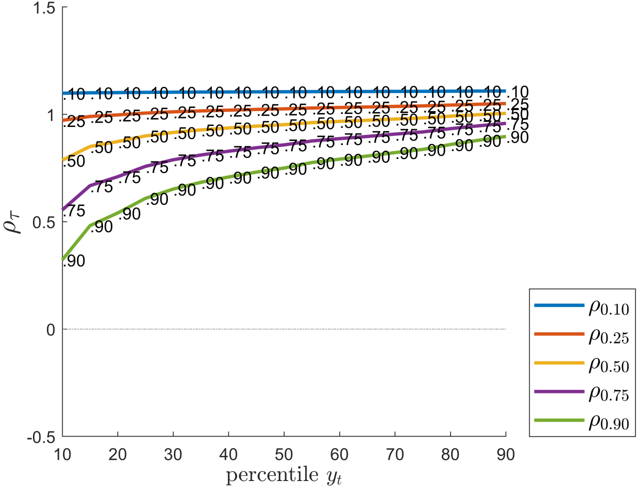
	$yp10$	$yp50$	$yp90$
$IQR_{0.75}$	0.56	0.46	0.42
	(0.49, 0.79)	(0.39, 0.56)	(0.33, 0.48)
$IQR_{0.90}$	1.31	1.04	0.90
	(1.04, 3.32)	(0.83, 1.50)	(0.70, 1.12)
$SK_{0.90}$	−0.25	−0.29	−0.29
	(−0.70, −0.04)	(−0.50, −0.11)	(−0.45, −0.12)
$\rho\ \tau 0.25$	1.00	1.05	1.07
	(0.93, 1.11)	(1.01, 1.10)	(1.03, 1.10)
$\rho\ \tau 0.50$	0.93	1.01	1.04
	(0.83, 0.97)	(0.95, 1.03)	(0.99, 1.06)
$\rho\ \tau 0.75$	0.82	0.97	1.02
	(0.63, 0.88)	(0.89, 0.99)	(0.95, 1.04)
σ^2_η		0.49	
		(0.38, 0.63)	
σ^2_η village		0.19	
		(0.18, 0.29)	
σ^2_ε		1.10	
		(1.01, 1.26)	

Note. In parenthesis we report 90% block bootstrap CI (1000 repetitions).

Density at different levels of current income



Persistence in India at different quantiles and at different current income



Results on non-linear models: Colombia

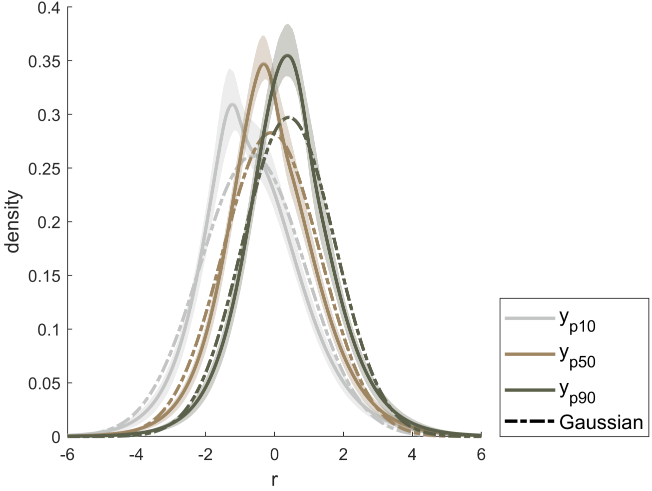
- We now consider a non-linear flexible model with additive fixed effects:

$$\ell_{jit} = \beta_0^{\dagger}(s_{jit}) + \beta_1^{\dagger}(s_{jit})\psi\,(y_{it}) + \eta_i + \varepsilon_{jit},$$

	$yp10$	$yp50$	$yp90$
$IQR_{0.75}$	1.91	1.62	1.52
	(1.22, 4.19)	(1.11, 3.00)	(1.10, 2.41)
$IQR_{0.90}$	3.85	3.57	3.48
	(2.49, 8.13)	(2.39, 6.74)	(2.37, 6.18)
$SK_{0.90}$	0.37	0.27	0.16
	(0.21, 0.56)	(0.13, 0.50)	(0.05, 0.37)
$\rho\ \tau 0.25$	0.59	0.49	0.38
	(0.46, 0.69)	(0.26, 0.65)	(−0.07, 0.62)
$\rho\ \tau 0.50$	0.50	0.58	0.49
	(−0.31, 0.68)	(0.41, 0.68)	(0.20, 0.65)
$\rho\ \tau 0.75$	0.19	0.26	0.39
	(−1.24, 0.53)	(−0.72, 0.57)	(−0.39, 0.63)
σ^2_η		0.47	
		(0.41, 0.58)	
σ^2_η village		0.11	
		(0.11, 0.17)	
σ^2_ε		1.10	
		(1.05, 1.23)	

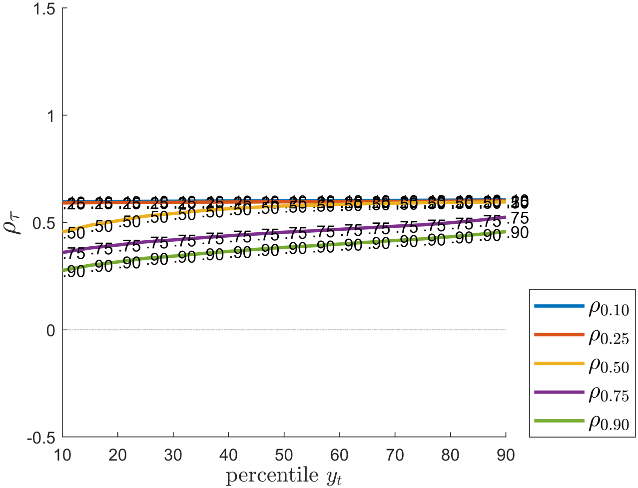
Note. In parenthesis we report 90% block bootstrap CI (1000 repetitions).

Density in Colombia at different level of current income



Note. The figure reports estimates of nonlinear persistence for Colombia for the flexible model with

Persistence in Colombia at different quantiles and at different current income



Outline

Information frictions in extended families in Tanzania

- The third example is about a novel **measure of information quality**.
- This material is based on Attanasio and Krutikova (2020)

Information frictions in extended families in Tanzania

- The third example is about a novel **measure of information quality**.
- This material is based on Attanasio and Krutikova (2020)
- Information issues have received a considerable amount of attention in different contexts
  - contract theory;
  - financial markets;
  - insurance.

Information frictions in extended families in Tanzania

- The third example is about a novel **measure of information quality**.
- This material is based on Attanasio and Krutikova (2020)
- Information issues have received a considerable amount of attention in different contexts
  - contract theory;
  - financial markets;
  - insurance.
- In the context we study the focus is on'imperfect' insurance of idiosyncratic risk.

Information frictions in extended families in Tanzania

- The third example is about a novel **measure of information quality**.
- This material is based on Attanasio and Krutikova (2020)
- Information issues have received a considerable amount of attention in different contexts
  - contract theory;
  - financial markets;
  - insurance.
- In the context we study the focus is on'imperfect' insurance of idiosyncratic risk.
- I will therefore :
  - Present a data set with information on extended families in Tanzania;
  - Propose a measure of information asymmetry.

Information frictions in extended families in Tanzania

- The third example is about a novel **measure of information quality**.
- This material is based on Attanasio and Krutikova (2020)
- Information issues have received a considerable amount of attention in different contexts
  - contract theory;
  - financial markets;
  - insurance.
- In the context we study the focus is on'imperfect' insurance of idiosyncratic risk.
- I will therefore :
  - Present a data set with information on extended families in Tanzania;
  - Propose a measure of information asymmetry.
  - Sketch a model of the effect of the specific friction we consider on risk sharing.

Information frictions in extended families in Tanzania

- The third example is about a novel **measure of information quality**.
- This material is based on Attanasio and Krutikova (2020)
- Information issues have received a considerable amount of attention in different contexts
  - contract theory;
  - financial markets;
  - insurance.
- In the context we study the focus is on'imperfect' insurance of idiosyncratic risk.
- I will therefore :
  - Present a data set with information on extended families in Tanzania;
  - Propose a measure of information asymmetry.
  - Sketch a model of the effect of the specific friction we consider on risk sharing.
  - Relate information asymmetry to risk sharing and vulnerability to shocks.

Information frictions in extended families in Tanzania

- The third example is about a novel **measure of information quality**.
- This material is based on Attanasio and Krutikova (2020)
- Information issues have received a considerable amount of attention in different contexts
  - contract theory;
  - financial markets;
  - insurance.
- In the context we study the focus is on'imperfect' insurance of idiosyncratic risk.
- I will therefore :
  - Present a data set with information on extended families in Tanzania;
  - Propose a measure of information asymmetry.
  - Sketch a model of the effect of the specific friction we consider on risk sharing.
  - Relate information asymmetry to risk sharing and vulnerability to shocks.
  - Consider a new measure of network centrality constructed from the information quality measure and relate it to risk sharing and vulnerability to shocks.

The Tanzania KHDS data

- We use the Kagera Health and Development Survey (KHDS) - a unique longitudinal data set from Tanzania.
- Kagera is a relatively isolated region, far from the capital, where agriculture remains main source of income

The Tanzania KHDS data

- Study follows individuals from baseline (1991-1994 wave) sample of 915 households from 51 communities in Kagera region for 20 years
    - Wave 1991-94 (4 rounds)
    - Wave 2004
    - Wave 2010

The Tanzania KHDS data

- Study follows individuals from baseline (1991-1994 wave) sample of 915 households from 51 communities in Kagera region for 20 years
  - Wave 1991-94 (4 rounds)
  - Wave 2004
  - Wave 2010
- Very rich and high quality data collected, including:
  - Demographics;
  - Consumption;
  - Income (including transfers) and wealth;
  - Reciprocal information on wealth indicators.
- Exceptionally low attrition due to huge tracking effort - 2010 sample includes at least one individual from 92% of baseline households

Figure: Sample 1991-1994

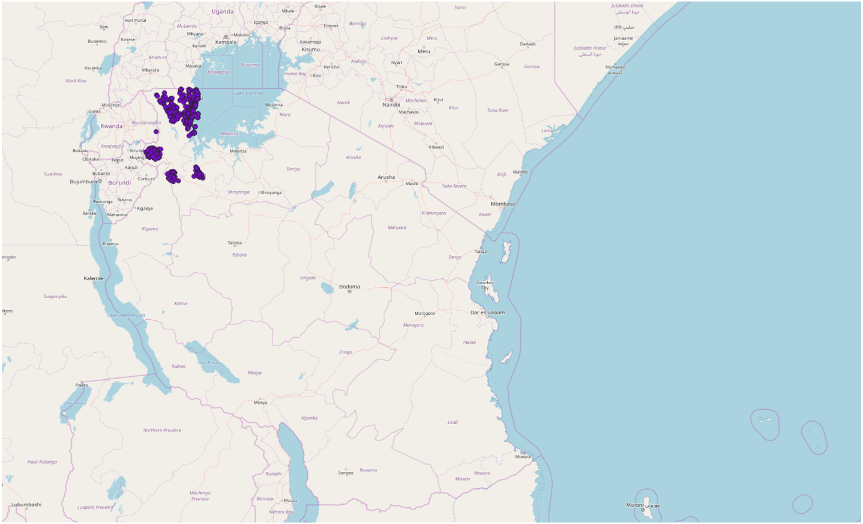


Figure: Sample 2004

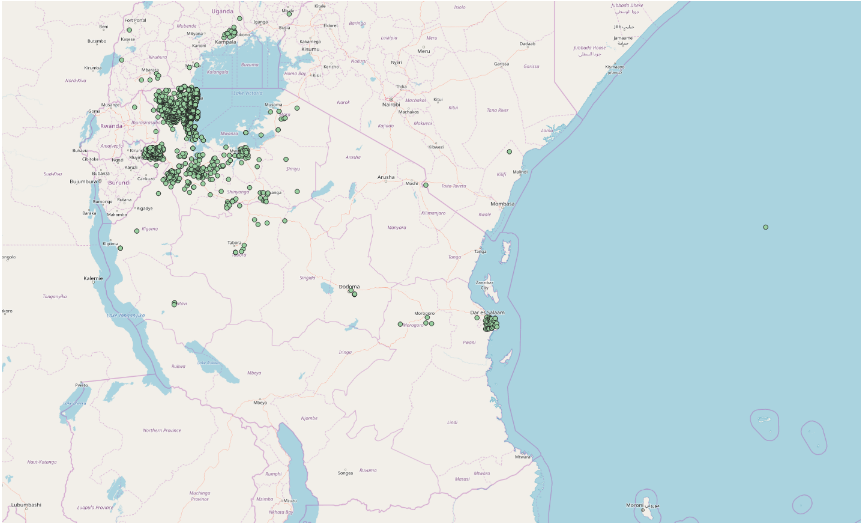
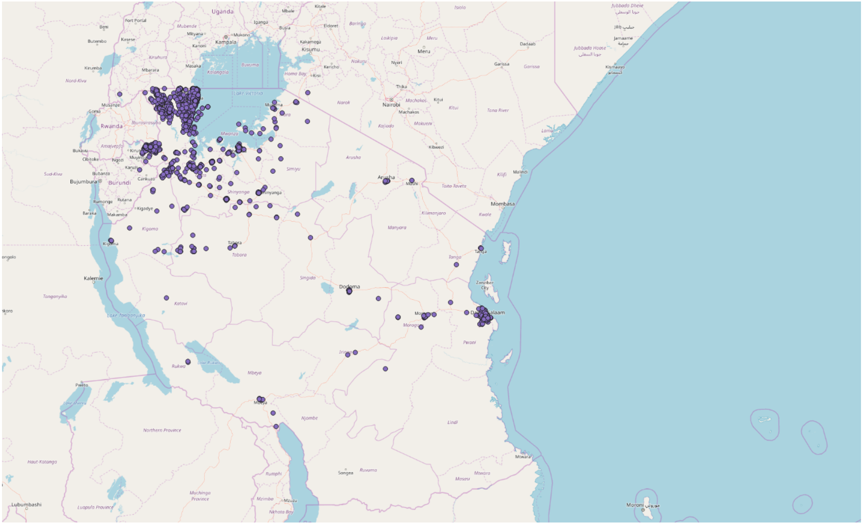


Figure: Sample 2010



Sample

Table: KHDS Sample

Round	HH	Ext HH	HH per Ex	Mean dist w/in HH
1991-1994	915	915	1	0
2004	2,774	831	3.34 (1.99)	74.74km (152.15)
2010	3,314	817	4.06 (2.38)	137.96km (185.63)

Wealth information

- All individuals within an extended family are asked a number of questions about the ownership of a variety of assets.
- We are interested in questions about asset ownership of those individuals as a measure of their wealth and income.

Wealth information

- All individuals within an extended family are asked a number of questions about the ownership of a variety of assets.
- We are interested in questions about asset ownership of those individuals as a measure of their wealth and income.

- Each household member is asked whether they own:
  1. house;
  2. land;
  3. oxen/bulls, dairy cows, non-dairy cows, other big livestock;
  4. phone (mobile or landline);
  5. video-equipment, TV, camera;
  6. Car, motorbike, other vehicle

Information on asset holding by other extended household members.

- All individuals within an extended family are also asked the same questions about the asset ownership by all other household members.
- It is therefore possible to compare:
  - actual asset ownership (as reported in the questionnaire)
  - asset ownership as reported by all oher household members.

Information on asset holding by other extended household members.

- All individuals within an extended family are also asked the same questions about the asset ownership by all other household members.
- It is therefore possible to compare:
  - actual asset ownership (as reported in the questionnaire)
  - asset ownership as reported by all oher household members.
- This information can be used to construct measures of asymmetric information.

Measuring asymmetric information

- We judge the amount and quality of information among two individuals looking at the quality of information about assets.

Measuring asymmetric information

- We judge the amount and quality of information among two individuals looking at the quality of information about assets.
- The assumption is that discrepancies and imprecisions about asset ownership information reflect information about income flows and shocks.

Measuring asymmetric information

- We judge the amount and quality of information among two individuals looking at the quality of information about assets.
- The assumption is that discrepancies and imprecisions about asset ownership information reflect information about income flows and shocks.
- We use the available information on asset ownership to construct a ‘wealth’ index using an IRT model.

Measuring asymmetric information

- We judge the amount and quality of information among two individuals looking at the quality of information about assets.
- The assumption is that discrepancies and imprecisions about asset ownership information reflect information about income flows and shocks.
- We use the available information on asset ownership to construct a 'wealth' index using an IRT model.
- This delivers, for each individual j belonging to household h, $\hat{\theta}_j^{jh} = \hat{f}(\mathbf{X}_j^{jh})$ where
    - the function $\hat{f}()$ is estimated from an IRT on asset ownership data
    - $\mathbf{X}_j^{jh}$ is the vector of ownership of different assets referring to individual j as declared by j themselves.

Asymmetric information

- Each household member is asked the same ownership questions about other household members;

Asymmetric information

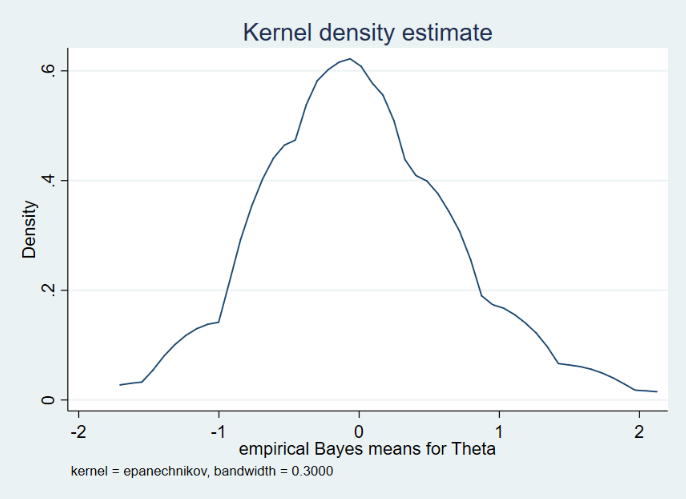
- Each household member is asked the same ownership questions about other household members;
- Using parameters from "true" model (i.e. own assets) we construct an index for reported asset ownership.

Asymmetric information

- Each household member is asked the same ownership questions about other household members;
- Using parameters from "true" model (i.e. own assets) we construct an index for reported asset ownership.
- $\hat{\theta}_j^{kh} = \hat{f}(\mathbf{X}_j^{kh})$ is the wealth index for individual j as perceived by individual k in household h.

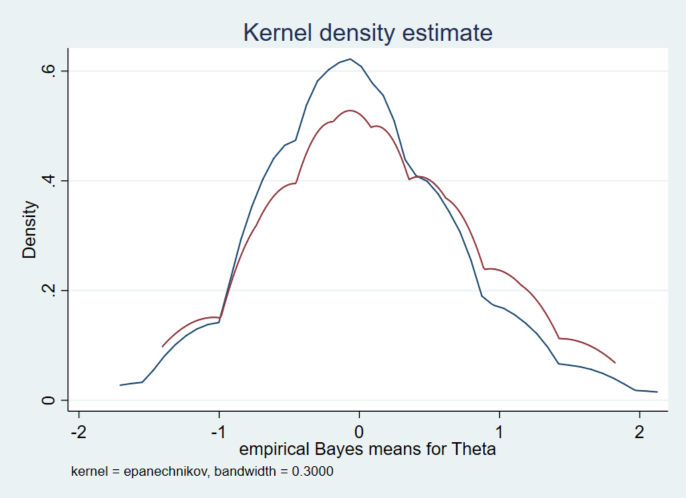
Distribution of wealth index from asset ownership

Figure: Kdensity of "true" asset ownership θ



Distribution of 'true' and 'perceived wealth index

Figure: Kdensity of "true" & "reported" asset ownership θ)



Asymmetric information

- Asymmetric information between j and k belonging to extended household g is approximated by the difference between the wealth indicator of individual j as estimated using own report and that estimated from the report of individual k about j's asset ownership.

Asymmetric information

- Asymmetric information between j and k belonging to extended household g is approximated by the difference between the wealth indicator of individual j as estimated using own report and that estimated from the report of individual k about j's asset ownership.
- In particular we consider 3 different measures:

$$\begin{aligned} q_{i,j}^{g,1} &= |\hat{\theta}_{i,g}^{i} - \hat{\theta}_{i,g}^{j}| \\ q_{i,j}^{g,2} &= |e^{\hat{\theta}_{i,g}^{i}} - e^{\hat{\theta}_{i,g}^{j}}| \\ q_{i,j}^{g,3} &= e^{|\hat{\theta}_{i,g}^{i} - \hat{\theta}_{i,g}^{j}|} \end{aligned}$$

Asymmetric information

- Asymmetric information between j and k belonging to extended household g is approximated by the difference between the wealth indicator of individual j as estimated using own report and that estimated from the report of individual k about j's asset ownership.
- In particular we consider 3 different measures:

$$
\begin{aligned}
q_{i,j}^{g,1} &= |\hat{\theta}_{i,g}^{i} - \hat{\theta}_{i,g}^{j}| \\
q_{i,j}^{g,2} &= |e^{\hat{\theta}_{i,g}^{i}} - e^{\hat{\theta}_{i,g}^{j}}| \\
q_{i,j}^{g,3} &= e^{|\hat{\theta}_{i,g}^{i} - \hat{\theta}_{i,g}^{j}|}
\end{aligned}
$$

- For each of these measures, we construct an index that varies between 0 and 1:

$$
\alpha_{ij}^{g,\ell} = \frac{1}{1+q_{i,j}^{g,\ell}}, \quad \ell = 1,2; \quad \alpha_{ij}^{g,3} = \frac{2}{1+q_{i,j}^{g,3}}.
$$

Properties of our measure of asymmetric information

	Mean	SD
Quality of information by distance btw households		
Q1 (0.4km)	0.97	0.11
Q2 (4.2km)	0.81	0.23
Q3 (17.2km)	0.75	0.24
Q4 (81.6km)	0.70	0.25
Q5 (613.5km)	0.67	0.24

Properties of our measure of asymmetric information

	Mean	SD
Quality of information by distance btw households		
Q1 (0.4km)	0.97	0.11
Q2 (4.2km)	0.81	0.23
Q3 (17.2km)	0.75	0.24
Q4 (81.6km)	0.70	0.25
Q5 (613.5km)	0.67	0.24
Quality of information by last time households spoke		
Less Than A Month Ago	0.86	0.21
Less Than A Year Ago	0.74	0.23
Less Than 2 Years Ago	0.71	0.25
Less Than 5 Years Ago	0.66	0.25
More Than 5 Years Ago	0.64	0.26
Don't Remember	0.59	0.26
N	12,693	

Network structures and position in the network

- Given that we are studying networks, we can consider the structure of the network and the position of each member in the network.

Network structures and position in the network

- Given that we are studying networks, we can consider the structure of the network and the position of each member in the network.
- We consider each pair of individual members and determine if a meaningful relation exists between them using the data on the quality of information.

Network structures and position in the network

- Given that we are studying networks, we can consider the structure of the network and the position of each member in the network.
- We consider each pair of individual members and determine if a meaningful relation exists between them using the data on the quality of information.
- Typically, in network analysis 'adjacency matrixes' are made of 0 and 1 indicating whether two individuals know each other.
  - $\alpha_{ij} = 1$ if individual i knows individual j.

Network structures and position in the network

- Given that we are studying networks, we can consider the structure of the network and the position of each member in the network.
- We consider each pair of individual members and determine if a meaningful relation exists between them using the data on the quality of information.
- Typically, in network analysis 'adjacency matrixes' are made of 0 and 1 indicating whether two individuals know each other.
  - $\alpha_{ij} = 1$ if individual i knows individual j.
  - Such matrices are then used to compute various properties of networks.

Network structures and position in the network

- Given that we are studying networks, we can consider the structure of the network and the position of each member in the network.
- We consider each pair of individual members and determine if a meaningful relation exists between them using the data on the quality of information.
- Typically, in network analysis 'adjacency matrixes' are made of 0 and 1 indicating whether two individuals know each other.
    - $\alpha_{ij} = 1$ if individual i knows individual j.
    - Such matrices are then used to compute various properties of networks.
- In our analysis we construct adjacency matrices with $\alpha_{ij}^{g,\ell}$, $\ell = 1, 2, 3$.
    - We note that $\alpha_{ij}^{g} \in [0, 1]$
    - .... and these matrices can be asymmetric.

Network structures and position in the network

Asymmetric weighted adjacency matrixes

- Given the asymmetric information matrixes for each extended family, we can construct different properties for each network.

Network structures and position in the network

Asymmetric weighted adjacency matrixes

- Given the asymmetric information matrixes for each extended family, we can construct different properties for each network.
- Given an adjacency matrix $A^{g,\ell}$, we construct measures of their position in the network.
- We use measures of degree centrality, which can be obtained averaging, for each household, the elements of the row or the columns of the adjacency matrix.
- As the matrices are not symmetric, the measures obtained averaging the rows or the columns are different.

Network structures and position in the network

Asymmetric weighted adjacency matrix

- Averaging over the rows of the adjacency matrix $A^{g,\ell}$ we get the in-degree centrality, that is the average quality of the information the network has about the wealth of household i

$$InQ_i^{g,\ell} = \frac{1}{K_g - 1} \sum_{k \neq i} \alpha_{ik}^{g,\ell}$$

where K_g is the number of households in family g.

Network structures and position in the network

Asymmetric weighted adjacency matrix

- Averaging over the rows of the adjacency matrix $A^{g,\ell}$ we get the in-degree centrality, that is the average quality of the information the network has about the wealth of household i

$$InQ_i^{g,\ell} = \frac{1}{K_g - 1} \sum_{k \neq i} \alpha_{ik}^{g,\ell}$$

  where K_g is the number of households in family g.

- Analogously, we can construct the out-degree centrality measure for household i averaging the elements of the matrix $A^{g,\ell}$ corresponding to column i.

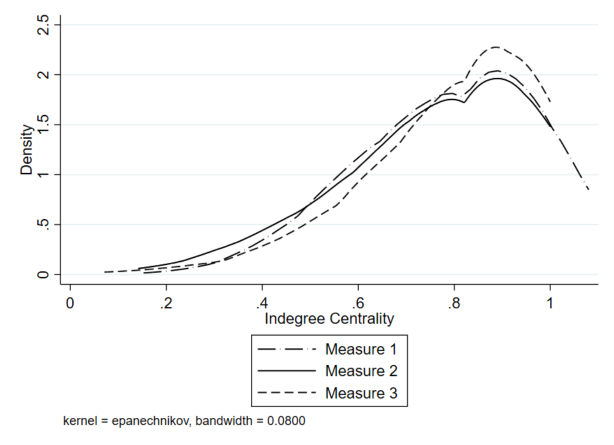
$$OutQ_i^{g,\ell} = \frac{1}{K_g - 1} \sum_{k \neq i} \alpha_{ki}^{g,\ell}$$

- Finally, we can also define the quality of information in family g averaging the individual measures as:

$$IQ^{g,\ell} = \frac{1}{K_g} \sum_{j}^{K_g} InQ_j^{g,\ell}$$

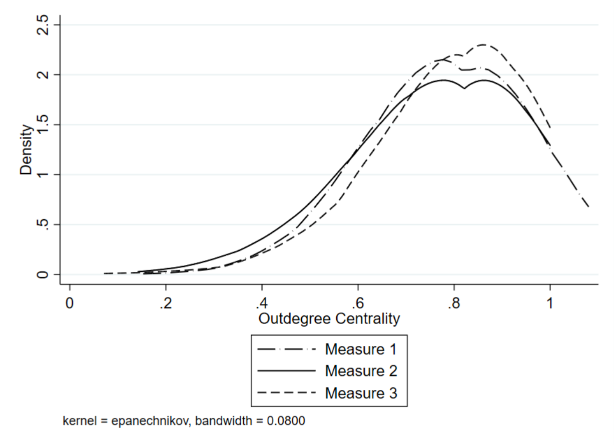
Descriptive statistics on the quality of information

Figure: Kdensity of $InQ_i^{h,\ell}$



Descriptive statistics on the quality of information

Figure: Kdensity of $OutQ_i^{h,\ell}$



Descriptive statistics on the quality of information

Table: Summary statistics and correlation matrix: network centrality measures

	Mean	SD	$InQ_i^{h,1}$	$InQ_i^{h,2}$	$InQ_i^{h,3}$	$OutQ_i^{h,1}$	$OutQ_i^{h,2}$	$OutQ_i^{h,3}$
Household level								
$InQ_i^{h,1}$	0.79	0.17	1.000					
$InQ_i^{h,2}$	0.77	0.20	0.944	1.000				
$InQ_i^{h,3}$	0.81	0.18	0.983	0.929	1.000			
$OutQ_i^{h,1}$	0.78	0.16	0.318	0.313	0.303	1.000		
$OutQ_i^{h,2}$	0.77	0.18	0.316	0.322	0.303	0.956	1.000	
$OutQ_i^{h,3}$	0.80	0.16	0.305	0.301	0.294	0.983	0.943	1.000
N	2,780							
Family network								
$IQ^{h,1}$	0.79	0.13						
$IQ^{h,2}$	0.77	0.15						
$IQ^{h,3}$	0.80	0.13						
N	709							

Risk sharing: a conceptual framework

- We start considering risk sharing within extended family g.
- Individual j belonging to household g receives a stochastic endowment $y_t^{j,g}$.

$$y_t^{j,g} = \bar{y^g}_t + \epsilon_t^j.$$

  - y is perishable: this can be easily relaxed, (we do not consider saving for simplicity here).

Risk sharing: a conceptual framework

- We start considering risk sharing within extended family g.
- Individual j belonging to household g receives a stochastic endowment $y_t^{j,g}$.

$$y_t^{j,g} = \bar{y^g}_t + \epsilon_t^j.$$

  - y is perishable: this can be easily relaxed, (we do not consider saving for simplicity here).
  - $\bar{y^g}_t$ may include transfers received as a part of a risk sharing agreement with other families.
- Individual j receives utility from consumption c_t^j, which equals to their endowment plus a transfer τ_t^j, which can be negative :

$$c_t^j = y_t^{j,g} + \tau_t^j$$

Perfect risk sharing

- The implications under perfect risk sharing can be derived considering a social planner problem, as in Townsend (1994).

$$\max_{\{\tau_t^{i,g}\}^{i=1,..K_g}} \sum_{j=1}^{K_g} \lambda_{j,g} \sum_{t=0}^{\infty} \beta^t \int_Y u(y_t^{j,g} + \tau_t^{j,g}) d\mu^t(y_t^{j,g}) \qquad s.t.$$

$$\sum_{j=1}^{K_g} y_t^{j,g} = \sum_{j=1}^{K_g} c_t^{j,g} \quad \forall t \qquad\qquad c_t^{j,g} = y_t^{j,g} + \tau_t^{j,g} \quad \forall t, j$$

- $\lambda_{j,g}$ is the Pareto weight given to individual j, which allows for inequality and is assumed to be constant.
- $\mu^t()$ is a probability measure of the stochastic endowment y_t^j, which reflects the available (and public) information.
- $y_t^{j,g}$ is completely observable *(ex-post)* and can be contracted upon.

Perfect risk sharing

- The first order conditions for this problem, in the absence of frictions (information, enforceability) after taking logs are:

$$\lambda_{i,g} u'(c_t^{i,g}) \beta^t = \nu_t^g$$

  with ν_t^g the multiplier of the aggregate resource constraint for group g.
- Note:
  - the right hand side, ν_t does not depend on j;
  - $\lambda_{i,g}$ does not depend on t.
  - the f.o.c. does not depend on $\mu^t()$: the condition applies in any state of the world and any history.

Perfect risk sharing

- The first order conditions for this problem, in the absence of frictions (information, enforceability) after taking logs are:

$$\lambda_{i,g} u'(c_t^{i,g})\beta^t = \nu_t^g$$

  with ν_t^g the multiplier of the aggregate resource constraint for group g.
- Note:
  - the right hand side, ν_t does not depend on j;
  - $\lambda_{i,g}$ does not depend on t.
  - the f.o.c. does not depend on $\mu^t()$: the condition applies in any state of the world and any history.

Perfect risk sharing

- The first order conditions for this problem, in the absence of frictions (information, enforceability) after taking logs are:

$$\lambda_{i,g} u'(c_t^{i,g}) \beta^t = \nu_t^g$$

  with ν_t^g the multiplier of the aggregate resource constraint for group g.
- Note:
  - the right hand side, ν_t does not depend on j;
  - $\lambda_{i,g}$ does not depend on t.
  - the f.o.c. does not depend on $\mu^t()$: the condition applies in any state of the world and any history.
- Taking logs:

$$ln(\lambda_{i,g}) + ln(u'(c_t^{i,g})) + tln(\beta) = ln(\nu_t^g)$$

Perfect risk sharing

- The first order conditions for this problem, in the absence of frictions (information, enforceability) after taking logs are:

$$\lambda_{i,g} u'(c_t^{i,g})\beta^t = \nu_t^g$$

  with ν_t^g the multiplier of the aggregate resource constraint for group g.
- Note:
  - the right hand side, ν_t does not depend on j;
  - $\lambda_{i,g}$ does not depend on t.
  - the f.o.c. does not depend on $\mu^t()$: the condition applies in any state of the world and any history.
- Taking logs:

$$ln(\lambda_{i,g}) + ln(u'(c_t^{i,g})) + tln(\beta) = ln(\nu_t^g)$$

- Taking differences across time periods:

$$\Delta_s ln(u'(c_t^{i,g})) + sln(\beta) = \Delta_s ln(\nu_t^g)$$

Testing perfect risk sharing

- Both the level and difference specifications do not depend on individual resources, controlling for group resources.
- These restrictions can be tested as in Townsend (1994):

$$ln(u'(c_t^{i,g})) = \tilde{\kappa}^{i,g} + ln(\nu_t^g) + \tilde{\gamma} ln(y_t^{i,g}) + \tilde{\epsilon}_t^{i,g}$$

$$\Delta_s ln(u'(c_t^{i,g})) = \kappa^g + \Delta_s ln(\nu_t^g) + \gamma \Delta_s ln(y_t^{i,g}) + \epsilon_t^{i,g}$$

  where $\tilde{\epsilon}_t^{i,g}$ and $\epsilon_t^{i,g}$ reflect measurement error and other unobservables.
- The coefficients $\tilde{\gamma}$ and γ measure the *vulnerability* of a single individual to idiosyncratic shocks; they should be 0 under perfect risk sharing.
- This test is based only on consumption and endowment data and does not require information about the decentralization.

Testing perfect risk sharing

- Both the level and difference specifications do not depend on individual resources, controlling for group resources.
- These restrictions can be tested as in Townsend (1994):

$$ln(u'(c_t^{i,g})) = \tilde{\kappa}^{i,g} + ln(\nu_t^g) + \tilde{\gamma} ln(y_t^{i,g}) + \tilde{\epsilon}_t^{i,g}$$

$$\Delta_s ln(u'(c_t^{i,g})) = \kappa^g + \Delta_s ln(\nu_t^g) + \gamma \Delta_s ln(y_t^{i,g}) + \epsilon_t^{i,g}$$

  where $\tilde{\epsilon}_t^{i,g}$ and $\epsilon_t^{i,g}$ reflect measurement error and other unobservables.
- The coefficients $\tilde{\gamma}$ and γ measure the *vulnerability* of a single individual to idiosyncratic shocks; they should be 0 under perfect risk sharing.
- This test is based only on consumption and endowment data and does not require information about the decentralization.
- In the absence of income information, one can use data on idiosyncratic shocks.

Risk sharing and imperfect information

- We first estimate a version of the equation in differences.

$$\Delta_s ln(u'(c_t^{j,g})) = \nu_t^g + \gamma_{01} BS_t^{j,g} + \gamma_{02} GS_t^{j,g} + \epsilon_t^{j,g}$$

where γ_{01} and γ_{02} measure the effect of 'bad' and 'good' idiosyncratic shocks on changes in log consumption, which should be 0 under perfect risk sharing.

Risk sharing and imperfect information

- We first estimate a version of the equation in differences.

$$\Delta_s ln(u'(c_t^{j,g})) = \nu_t^g + \gamma_{01} BS_t^{j,g} + \gamma_{02} GS_t^{j,g} + \epsilon_t^{j,g}$$

  where γ_{01} and γ_{02} measure the effect of 'bad' and 'good' idiosyncratic shocks on changes in log consumption, which should be 0 under perfect risk sharing.
- We then allow these coefficients depend on the *average quality of information* in group g.

$$\Delta_s ln(u'(c_t^{j,g})) = \nu_t^g + (\gamma_{01} + \gamma_{11} IQ^{g,\ell}) BS_t^{j,g} + (\gamma_{02} + \gamma_{12} IQ^{g,\ell}) GS_t^{j,g} + \epsilon_t^{j,g}$$

Risk sharing and imperfect information

- We first estimate a version of the equation in differences.

$$\Delta_s ln(u'(c_t^{j,g})) = \nu_t^g + \gamma_{01} BS_t^{j,g} + \gamma_{02} GS_t^{j,g} + \epsilon_t^{j,g}$$

  where γ_{01} and γ_{02} measure the effect of 'bad' and 'good' idiosyncratic shocks on changes in log consumption, which should be 0 under perfect risk sharing.
- We then allow these coefficients depend on the *average quality of information* in group g.

$$\Delta_s ln(u'(c_t^{j,g})) = \nu_t^g + (\gamma_{01} + \gamma_{11} IQ^{g,\ell}) BS_t^{j,g} + (\gamma_{02} + \gamma_{12} IQ^{g,\ell}) GS_t^{j,g} + \epsilon_t^{j,g}$$

- Finally, we consider the quality of the information available to individual j and of the about individual j: inward and outward degree centrality.

$$\Delta_s ln(u'(c_t^{j,g})) = \nu_t^g + (\gamma_{01} + \gamma_{21} IP_i^{g,\ell}) BS_t^{j,g} + (\gamma_{02} + \gamma_{22} IP_i^{g,\ell}) GS_t^{j,g} + \epsilon_t^{j,g}$$

$$\Delta_s ln(u'(c_t^{j,g})) = \nu_t^g + (\gamma_{01} + \gamma_{21} OP_i^{g,\ell}) BS_t^{j,g} + (\gamma_{02} + \gamma_{22} OP_i^{g,\ell}) GS_t^{j,g} + \epsilon_t^{j,g}$$

Risk sharing and network information quality

$$\Delta_s ln(u'(c_t^{j,g})) = \nu_t^g + (\gamma_{01} + \gamma_{11} IQ^{g,\ell}) BS_t^{j,g} + (\gamma_{02} + \gamma_{12} IQ^{g,\ell}) GS_t^{j,g} + \epsilon_t^{j,g}$$

Table: Sensitivity of risk-sharing to quality of information within family network

	(1)
Inf. quality measure	*none*
Bad shock in 2010	-0.134***
	(0.035)
Good shock in 2010	0.00289
	(0.036)
Good shock X mean degree cent $IQ^{h,\ell}$	
Bad shock X mean degree cent $IQ^{h,\ell}$	
Constant	0.434***
	(0.0125)
Observations	2,780

Risk sharing and network information quality

$$\Delta_s ln(u'(c_t^{j,g})) = \nu_t^g + (\gamma_{01} + \gamma_{11} IQ^{g,\ell}) BS_t^{j,g} + (\gamma_{02} + \gamma_{12} IQ^{g,\ell}) GS_t^{j,g} + \epsilon_t^{j,g}$$

Table: Sensitivity of risk-sharing to quality of information within family network

	(1)	(2)	(3)	(4)
Inf. quality measure	*none*	$IQ^{h,1}$	$IQ^{h,2}$	$IQ^{h,3}$
Bad shock in 2010	-0.134***	-0.420	-0.324	-0.324
	(0.035)	(0.256)	(0.223)	(0.223)
Good shock in 2010	0.00289	0.679***	0.561**	0.561**
	(0.036)	(0.262)	(0.227)	(0.227)
Good shock X mean degree cent $IQ^{h,\ell}$		-0.863***	-0.729**	-0.729**
		(0.332)	(0.293)	(0.293)
Bad shock X mean degree cent $IQ^{h,\ell}$		0.365	0.248	0.248
		(0.324)	(0.287)	(0.287)
Constant	0.434***	0.434***	0.434***	0.434***
	(0.0125)	(0.0240)	(0.0241)	(0.0241)
Observations	2,780	2,780	2,780	2,780

Standard errors in parentheses
Dep var = change in lnpcconsumption btw 2004-2010 (2010 prices); Family network FE
Shock = 1 if reported by anyone in household
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Risk sharing and asymmetric information

- This evidence shows that networks with better information quality are 'closer to perfect risk sharing' than networks with worse information.
- Notice that the variation in quality of information is *across* networks/ extended families.

Risk sharing and asymmetric information

- This evidence shows that networks with better information quality are 'closer to perfect risk sharing' than networks with worse information.
- Notice that the variation in quality of information is *across* networks/ extended families.
- This evidence is silent about possible risk sharing happening with other households or families *outside the network*.

Risk sharing and asymmetric information

- This evidence shows that networks with better information quality are ‘closer to perfect risk sharing’ than networks with worse information.
- Notice that the variation in quality of information is *across* networks/ extended families.
- This evidence is silent about possible risk sharing happening with other households or families *outside the network*.
- The next step is to look at evidence about the quality of information regarding individual households *in each network*
    - Inward degree centrality (how much the rest of the network knows about the household receiving a shock)
    - Outward degree centrality (how much an individual household knows about the rest of the network.

Risk sharing and asymmetric information

- This evidence shows that networks with better information quality are 'closer to perfect risk sharing' than networks with worse information.
- Notice that the variation in quality of information is *across* networks/ extended families.
- This evidence is silent about possible risk sharing happening with other households or families *outside the network*.
- The next step is to look at evidence about the quality of information regarding individual households *in each network*
    - Inward degree centrality (how much the rest of the network knows about the household receiving a shock)
    - Outward degree centrality (how much an individual household knows about the rest of the network.
- These measures can be distinguished because the adjacence matrices are possibly asymmetric.

Risk sharing and in-degree centrality

$$\Delta_s ln(u'(c_t^{j,g})) = \nu_t^g + (\gamma_{01} + \gamma_{21} IP_i^{g,\ell})BS_t^{j,g} + (\gamma_{02} + \gamma_{22} IP_i^{g,\ell})GS_t^{j,g} + \epsilon_t^{j,g}$$

Table: Risk-sharing and quality of information family network has about affected households

	(1)	(2)	(3)	(4)
In-degree cent. measure	*none*	$InQ_i^{h,1}$	$InQ_i^{h,2}$	$InQ_i^{h,3}$
Bad shock in 2010	-0.134***	-0.457***	-0.426***	-0.419**
	(0.0350)	(0.164)	(0.146)	(0.164)
Good shock in 2010	0.00289	0.428**	0.279*	0.417**
	(0.0363)	(0.166)	(0.144)	(0.166)
HH indegree cent $InQ_i^{h,\ell}$		-0.337**	-0.252*	-0.327**
		(0.152)	(0.139)	(0.148)
Good shock X HH indegree cent $InQ_i^{h,\ell}$		-0.535***	-0.360**	-0.511**
		(0.206)	(0.182)	(0.202)
Bad shock X HH indegree cent $InQ_i^{h,\ell}$		0.413**	0.379**	0.357*
		(0.203)	(0.184)	(0.199)
Constant	0.434***	0.697***	0.627***	0.695***
	(0.0241)	(0.121)	(0.109)	(0.121)
Observations	2,780	2,780	2,780	2,780

Standard errors in parentheses
Dep var = change in lnpcconsumption btw 2004-2010 (2010 prices); Family network FE
Shock = 1 if reported by anyone in the household
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Risk sharing and out-degree centrality

$$\Delta_s ln(u'(c_t^{j,g})) = \nu_t^g + (\gamma_{01} + \gamma_{21} OP_i^{g,\ell}) BS_t^{j,g} + (\gamma_{02} + \gamma_{22} OP_i^{g,\ell}) GS_t^{j,g} + \epsilon_t^{j,g}$$

Table: Risk-sharing and quality of information affected households have about family network

	(1)	(2)	(3)	(4)
Out-degree centr. measure	*none*	$OutQ_i^{h,1}$	$OutQ_i^{h,2}$	$OutQ_i^{h,3}$
Bad shock in 2010	-0.134***	-0.139	-0.126	-0.0997
	(0.0350)	(0.185)	(0.160)	(0.186)
Good shock in 2010	0.00289	0.116	0.0852	0.162
	(0.0363)	(0.191)	(0.165)	(0.195)
HH outdegree cent $OutQ_i^{h,\ell}$		0.124	0.0523	0.195
		(0.183)	(0.161)	(0.180)
Good shock X HH outdegree cent $OutQ_i^{h,\ell}$		-0.144	-0.107	-0.200
		(0.239)	(0.209)	(0.238)
Bad shock X HH outdegree cent $OutQ_i^{h,\ell}$		0.00741	-0.0104	-0.0424
		(0.233)	(0.204)	(0.230)
Constant	0.434***	0.336**	0.393***	0.278*
	(0.0241)	(0.146)	(0.126)	(0.146)
Observations	2,780	2,780	2,780	2,780

Standard errors in parentheses
Dep var = change in lnpcconsumption btw 2004-2010 (2010 prices); Family network FE
Shock = 1 if reported by anyone in the household
$^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$

Risk sharing and the quality of information

- This evidence shows that the amount of risk sharing within extended families observed in our sample depends on the quality of information as measured.

Risk sharing and the quality of information

- This evidence shows that the amount of risk sharing within extended families observed in our sample depends on the quality of information as measured.
- Furthermore, it shows that what seems to matter is the quality of the information that the extended family *on average* has about the individual affected by a shock, whether positive or negative.
- Outward centrality does not matter.

Risk sharing and the quality of information

- This evidence shows that the amount of risk sharing within extended families observed in our sample depends on the quality of information as measured.
- Furthermore, it shows that what seems to matter is the quality of the information that the extended family *on average* has about the individual affected by a shock, whether positive or negative.
- Outward centrality does not matter.
- The next steps of this research is to provide a more structural interpretation to these results.

Risk sharing in networks: a more structural approach

- In a recent paper, Ambrus, Gao, and Milán (2021) consider risk sharing in a network.
- In their set-up, two members of a (potential) risk sharing group are either connected or not.
- If a connection exists, there is no asymmetric information, and endowment are fully observed.
- Transfers can only be contracted among members who are connected.

Risk sharing in networks.

- Ambrus et al. (2021) derive several important and interesting results, considering endowment processes with different levels of connections.
    - Pareto efficient allocations equalise *expected* ratios of marginal utility for each connected pair of individuals (conditional on local information).
    - The variability of consumption of a given individual depends on their position in the network.
    - More central individuals, in many situations, end up with more variable consumption.

Risk sharing in networks.

- Ambrus et al. (2021) derive several important and interesting results, considering endowment processes with different levels of connections.
  - Pareto efficient allocations equalise *expected* ratios of marginal utility for each connected pair of individuals (conditional on local information).
  - The variability of consumption of a given individual depends on their position in the network.
  - More central individuals, in many situations, end up with more variable consumption.
- Our data suggest an important extension: connections might not be 0/1 but could be stronger and weaker.

Risk sharing in networks.

- Ambrus et al. (2021) derive several important and interesting results, considering endowment processes with different levels of connections.
  - Pareto efficient allocations equalise *expected* ratios of marginal utility for each connected pair of individuals (conditional on local information).
  - The variability of consumption of a given individual depends on their position in the network.
  - More central individuals, in many situations, end up with more variable consumption.
- Our data suggest an important extension: connections might not be 0/1 but could be stronger and weaker.
- This makes the theory more complex but, probably, more realistic.

References I

Agostinelli, F. and M. Wiswall (2016). Estimating the technology of children's skill formation. Technical report, National Bureau of Economic Research.

Alan, S. and S. Ertac (2019). Mitigating the gender gap in the willingness to compete: Evidence from a randomized field experiment. *Journal of the European Economic Association 17*(4), 1147–1185.

Alesina, A. and G.-M. Angeletos (2005). Fairness and redistribution. *American economic review 95*(4), 960–980.

Alesina, A., S. Stantcheva, and E. Teso (2018). Intergenerational mobility and preferences for redistribution. *American Economic Review 108*(2), 521–54.

Almås, I., O. Attanasio, and P. Jervis (2024, July). Presidential address: Economics and measurement: New measures to model decision making. *Econometrica 92*(4), 947–978.

References II

Almås, I., L. Berge, K. Bjorvatn, V. Somville, and B. Tungodden (2020). Adverse selection into competition: Evidence from a large-scale field experiment in tanzania. *NHH Working paper*.

Almås, I., A. Cappelen, E. Sørensen, and B. Tungodden (2020). Fairness across the world: preferences and beliefs. Technical report, NHH working paper.

Almås, I., A. W. Cappelen, and B. Tungodden (2020). Cutthroat capitalism versus cuddly socialism: Are americans more meritocratic and efficiency-seeking than scandinavians? *Journal of Political Economy 128*(5), 1753–1788.

Ambrus, A., W. Gao, and P. Milán (2021, 12). Informal Risk Sharing with Local Information. *The Review of Economic Studies 89*(5), 2329–2380.

Ameriks, J., J. Briggs, A. Caplin, M. Lee, M. D. Shapiro, and C. Tonetti (2020). Older americans would work longer if jobs were flexible. *American Economic Journal: Macroeconomics 12*(1), 174–209.

References III

Arcidiacono, P., V. J. Hotz, A. Maurel, and T. Romano (2017). Ex ante returns and occupational choice. *Unpublished manuscript*.

Arellano, M., O. Attanasio, S. Crossman, and V. Sancibrián (2024, September). Estimating flexible income processes from subjective expectations data: Evidence from india and colombia. Working Paper 32922, National Bureau of Economic Research.

Arrow, K. J. (1959). Rational choice functions and orderings. *Economica 26*(102), 121–127.

Attanasio, O., R. Bernal, M. Giannola, and M. Nores (2020). Child development in the early years. Technical report, NBER, WP No 27812.

Attanasio, O., T. Boneva, and C. Rauh (2019). Parental beliefs about returns to different types of investments in school children. NBER Working Papers 25513, National Bureau of Economic Research, Inc.

References IV

Attanasio, O. and S. Krutikova (2020, 07). JEEA-FBBVA LECTURE 2019: Consumption Insurance in Networks with Asymmetric Information: Evidence from Tanzania. *Journal of the European Economic Association 18*(4), 1589–1618.

Attanasio, O. P. and K. M. Kaufmann (2014). Education choices and returns to schooling: Mothers' and youths' subjective expectations and their role by gender. *Journal of Development Economics 109*, 203–216.

Baranov, V., S. Bhalotra, P. Biroli, and J. Maselko (2020). Maternal depression, women's empowerment, and parental investment: Evidence from a randomized controlled trial. *American Economic Review 110*(3), 824–59.

Becker, A., B. Enke, and A. Falk (2020). Ancient origins of the global variation in economic preferences. In *AEA Papers and Proceedings*, Volume 110, pp. 319–23.

Ben-Akiva, M. E., D. McFadden, K. Train, et al. (2019). *Foundations of stated preference elicitation: Consumer behavior and choice-based conjoint analysis*. Now.

References V

Bianchi, F., S. C. Ludvigson, and S. Ma (2020). Belief distortions and macroeconomic fluctuations. Technical report, National Bureau of Economic Research.

Blass, A. A., S. Lach, and C. F. Manski (2010). Using elicited choice probabilities to estimate random utility models: Preferences for electricity reliability. *International Economic Review 51*(2), 421–440.

Block, H. D. and J. Marschak (1960). Random orderings and stochastic theories of response. In I. Olkin, S. Ghurye, W. Hoeffding, W. Madow, and H. Mann (Eds.), *Contributions to Probability and Statistics*. Stanford University Press.

Boneva, T. and C. Rauh (2018). Parental beliefs about returns to educational investments—the later the better? *Journal of the European Economic Association 16*(6), 1669–1711.

References VI

Cappelen, A., J. List, A. Samek, and B. Tungodden (2020). The effect of early-childhood education on social preferences. *Journal of Political Economy 128*(7), 000–000.

Cavatorta, E. and B. Groom (2020). Does deterrence change preferences? evidence from a natural experiment. *European Economic Review*, 103456.

Charness, G., U. Gneezy, and A. Imas (2013). Experimental methods: Eliciting risk preferences. *Journal of Economic Behavior & Organization 87*, 43–51.

Chen, X., L. P. Hansen, and P. G. Hansen (2020). Robust identification of investor beliefs. Technical report, National Bureau of Economic Research.

Christensen, L. R., D. W. Jorgenson, and L. J. Lau (1975). Transcendental logarithmic utility functions. *The American Economic Review 65*(3), 367–383.

Cunha, F., I. Elo, and J. Culhane (2013). Eliciting maternal beliefs about the technology of skill formation. *NBER Working Paper 19144*.

References VII

Cunha, F. and J. H. Heckman (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *The Journal of Human resources 43*(4), 738–782.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica 78*(3), 883–931.

Curtin, R. (2016). Katona: a founder of behavioral economics. In *Handbook of Behavioral Economics*, pp. 30–47. Routledge.

Danz, D., L. Vesterlund, and A. J. Wilson (2020). Belief elicitation: Limiting truth telling with information on incentives. Technical report, National Bureau of Economic Research.

de Bresser, J. (2019). The role of heterogeneous expectations in life cycle models: Evaluating the accuracy of counterfactuals.

References VIII

Deaton, A. and J. Muellbauer (1980). An almost ideal demand system. *The American economic review 70*(3), 312–326.

Dizon-Ross, R. (2019). Parents' beliefs about their children's academic ability: Implications for educational investments. *American Economic Review 109*(8), 2728–65.

Dohmen, T., A. Falk, D. Huffman, and U. Sunde (2018). On the relationship between cognitive ability and risk preference. *Journal of Economic Perspectives 32*(2), 115–34.

Falk, A., A. Becker, T. Dohmen, B. Enke, D. Huffman, and U. Sunde (2018). Global evidence on economic preferences. *The Quarterly Journal of Economics 133*(4), 1645–1692.

Falk, A. and J. Hermle (2018). Relationship of gender differences in preferences to economic development and gender equality. *Science 362*(6412).

References IX

Ghisolfi, S. (2020). Requiring contributions during bargaining increases inequality in outcomes. *Working paper*.

Gilbert, M., C. Clark, J. Stone, F. Perroux, D. Lieu, Evelpides, F. Divisia, Tinbergen, Kuznets, Smithies, et al. (1949). The measurement of national wealth: discussion. *Econometrica: Journal of the Econometric Society*, 255–272.

Giustinelli, P., C. F. Manski, and F. Molinari (2019). Precise or imprecise probabilities? evidence from survey response on late-onset dementia. Technical report, NBER.

Haavelmo, T. (1958). The role of the econometrician in the advancement of economic theory. *Econometrica 26*(3), 351–357.

Hansen, L. P. and T. J. Sargent (1981). A note on wiener-kolmogorov prediction formulas for rational expectations models. *Economics Letters 8*(3), 255–260.

References X

Hausman, D. M. (1994). *The philosophy of economics: An anthology*. Cambridge University Press.

Hausman, J. (2012a). Contingent valuation: from dubious to hopeless. *Journal of Economic Perspectives 26*(4), 43–56.

Hausman, J. A. (2012b). *Contingent valuation: A critical assessment*. Elsevier.

Heckman, J. J. and R. Pinto (2023, December). Econometric causality: The central role of thought experiments. Working Paper 31945, National Bureau of Economic Research.

Hossain, T. and R. Okui (2013). The binarized scoring rule. *Review of Economic Studies 80*(3), 984–1001.

Juster, F. T. (1966). Consumer buying intentions and purchase probability: An experiment in survey design. *Journal of the American Statistical Association 61*(315), 658–696.

References XI

Juster, F. T. and R. P. Shay (1964). *Consumer Sensitivity to Finance Rates: An Empirical and Analytical Investigation*. NBER.

Karadja, M., J. Mollerstrom, and D. Seim (2017). Richer (and holier) than thou? the effect of relative income improvements on demand for redistribution. *Review of Economics and Statistics 99*(2), 201–212.

Katona, G. (1959). On the predictive value of consumer intentions and attitudes: A comment. *The Review of Economics and Statistics*, 317–317.

Katona, G. (1974). Understanding consumer attitudes. *Surveys of Consumers 1976*, 203–219.

Kesternich, I., F. Heiss, D. McFadden, and J. Winter (2013). Suit the action to the word, the word to the action: Hypothetical choices and real decisions in medicare part d. *Journal of Health Economics 32*(6), 1313–1324.

Keynes, J. M. (1936). The general theory of interest, employment and money.

References XII

Kindermann, F., J. Le Blanc, M. Piazzesi, and M. Schneider (2019). Learning about housing cost-survey evidence from german house price booms. Technical report, mimeo, Stanford University.

Kuziemko, I., M. I. Norton, E. Saez, and S. Stantcheva (2015). How elastic are preferences for redistribution? *American Economic Review 105*(4), 1478–1508.

Kuznets, S. (1941). Statistics and economic history. *The Journal of Economic History 1*(1), 26–41.

Kuznets, S. et al. (1937). National income and capital formation, 1919-1935. *NBER Books*.

List, J., J. Pernaudet, and D. . Suskind (2020). It all starts with beliefs: Addressing the roots of educational inequities by changing parental beliefs. Technical report.

Luce, R. and P. Suppes (1965). Preference, utility, and subjective utility. *Handbook of Mathematical Psychology, III, New York: Wiley*, 249–409.

References XIII

Luce, R. D. (1956). Semiorders and a theory of utility discrimination. *Econometrica, Journal of the Econometric Society*, 178–191.

Luce, R. D. (1959). *Choice Behavior. A Theoretical Analysis*. New York: Wiley.

Luce, R. D. and J. W. Tukey (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of mathematical psychology 1*(1), 1–27.

Manski, C. F. (2004). Measuring expectations. *Econometrica 72*(5), 1329–1376.

Manski, C. F. and F. Molinari (2010). Rounding probabilistic expectations in surveys. *Journal of Business & Economic Statistics 28*(2), 219–231.

Miller, G., Á. De Paula, and C. Valente (2020). Subjective expectations and demand for contraception. Technical report, National Bureau of Economic Research.

Nyarko, Y. and A. Schotter (2002). An experimental study of belief learning using elicited beliefs. *Econometrica 70*(3), 971–1005.

References XIV

Potter, S. (2016). The advantages of probabilistic survey questions: remarks at the it forum and rcea bayesian workshop, keynote address, rimini, italy, may 2016. Technical report, Federal Reserve Bank of New York.

Potter, S., M. Del Negro, G. Topa, and W. Van der Klaauw (2017). The advantages of probabilistic survey questions. *Review of Economic Analysis 9*(1), 1–32.

Samuelson, P. A. (1938). A note on the pure theory of consumer's behaviour. *Economica 5*(17), 61–71.

Samuelson, P. A. (1948). Consumption theory in terms of revealed preference. *Economica 15*(60), 243–253.

Stigler, G. J. and G. S. Becker (1977). De gustibus non est disputandum. *The American Economic Review 67*(2), 76–90.

Stone, R. (1954). Linear expenditure systems and demand analysis: an application to the pattern of british demand. *The Economic Journal 64*(255), 511–527.

References XV

Stone, R. (1984). Richard stone-prize lecture: The accounts of society. *Nobelprize. org. Nobel Media AB*.

Tobin, J. (1959). On the predictive value of consumer intentions and attitudes. *The review of economics and statistics*, 1–11.

Van der Klaauw, W. and K. I. Wolpin (2008). Social security and the retirement and savings behavior of low-income households. *J. of Econometrics 145*(1-2), 21–42.

Wiswall, M. and B. Zafar (2015). Determinants of college major choice: Identification using an information experiment. *The Review of Economic Studies 82*(2), 791–824.

Zheng, Y., J. Pantano, et al. (2012). Using subjective expectations data to allow for unobserved heterogeneity in hotz-miller estimation strategies. In *2012 Meeting Papers*, Number 940. Society for Economic Dynamics.

Experimental literature

- Lab work on various mechanisms to elicit primitives.
  - Nyarko and Schotter (2002);
  - Hossain and Okui (2013);
  - Danz, Vesterlund, and Wilson (2020);
  - Charness, Gneezy, and Imas (2013).
- Lab in the field:
  - Alesina, Stantcheva, and Teso (2018) Almås, Cappelen, and Tungodden (2020) ;
  - Ghisolfi (2020) and Cappelen, List, Samek, and Tungodden (2020);
  - Cavatorta and Groom (2020) ;
  - Alan and Ertac (2019) and Almås, Berge, Bjorvatn, Somville, and Tungodden (2020) .

Back

References: stated preferences, deviation from rational expectations and beliefs and policy preferences

- Stated preferences:
  - Ben-Akiva, McFadden, Train, et al. (2019);
  - Kesternich, Heiss, McFadden, and Winter (2013);
  - Blass, Lach, and Manski (2010);
  - Ameriks, Briggs, Caplin, Lee, Shapiro, and Tonetti (2020).
- Deviations from rational expectations:
  - Hansen and Sargent (1981);
  - Chen, Hansen, and Hansen (2020);
  - Kindermann, Le Blanc, Piazzesi, and Schneider (2019);
  - Bianchi, Ludvigson, and Ma (2020).
- Beliefs and policy preferences:
  - Kuziemko, Norton, Saez, and Stantcheva (2015);
  - Karadja, Mollerstrom, and Seim (2017);
  - Alesina and Angeletos (2005).

Back

References: subjective expectations

- Collecting subjective expectations data:
  - Manski (2004);
  - Manski and Molinari (2010);
  - Giustinelli, Manski, and Molinari (2019);
  - NY Fed work on collecting subjective expectations data systematically (Potter, Del Negro, Topa, and Van der Klaauw, 2017; Potter, 2016).
- ...and using it.:
  - Van der Klaauw and Wolpin (2008);
  - de Bresser (2019);
  - Arcidiacono, Hotz, Maurel, and Romano (2017);
  - Wiswall and Zafar (2015);
  - Zheng, Pantano, et al. (2012);
  - Attanasio and Kaufmann (2014);
  - **?**.

References: beliefs, attitudes, preferences

- Cunha, Elo, and Culhane (2013), **?**, Baranov, Bhalotra, Biroli, and Maselko (2020), List, Pernaudet, and Suskind (2020) on beliefs about child development;
- Dizon-Ross (2019); **?** on perceptions of ability;
- Boneva and Rauh (2018); Attanasio, Boneva, and Rauh (2019) on returns to education.
- Miller, De Paula, and Valente (2020) on beliefs about contraception effectiveness.
- **?** collect and use data about self-perceptions about academic achievement among high school students in Mexico;
- Andrew Caplin and collaborators have been working with Vanguard samples, engineering new measures and questions:
- Ameriks et al. (2020).
- Attitudes
  - Falk, Becker, Dohmen, Enke, Huffman, and Sunde (2018); Becker, Enke, and Falk (2020); Falk and Hermle (2018); Dohmen, Falk, Huffman, and Sunde (2018);
  - Almås, Cappelen, Sørensen, and Tungodden (2020);