Presidential Address:

Economics and Measurement

New measures to model decision making*

Ingvild Almås,[†] Orazio Attanasio,[‡] Pamela Jervis[§]

May 2, 2024

Abstract

Most empirical work in economics has considered only a narrow set of measures as meaningful and useful to characterize individual behavior, a restriction justified by the difficulties in collecting a wider set. However, this approach often forces the use of strong assumptions to estimate the parameters that inform individual behavior and identify causal links. In this paper, we argue that a more flexible and broader approach to measurement could be extremely useful and allow the estimation of richer and more realistic models that rest on weaker identifying assumptions. We argue that the design of measurement tools should interact with, and depend on, the models economists use. Measurement is not a substitute for rigorous theory, it is an important complement to it, and should be developed in parallel to it. We illustrate these arguments with a model of parental behavior estimated on pilot data that combines conventional measures with novel ones.

*This paper is based on the Presidential Address Attanasio presented at the World Congress of the Econometric Society in August 2020. Due to size limitations, we had to cut a large number of references. A version of the paper with a more extensive list of references (and additional tables) can be found as Almås et al. (2023). At different stages, we received very valuable feedback from many colleagues, including Joe Altonji, Alison Andrew, Manuel Arellano, Jere Behrman, Alberto Bisin, Richard Blundell, Margherita Borella, Agar Brugiavini, Andrew Caplin, Rossella Calvi, Sarah Cattan, Flavio Cunha, Maria Cristina De Nardi, Aureo de Paula, Jim Heckman, Mike Keane, Sonya Krutikova, Chuck Manski, Costas Meghir, Rohini Pande, Elena Pastorino, Larry Samuelson, Guglielmo Weber, Ken Wolpin. We also thank two referees for useful comments. Attanasio presented some of this material in seminars at the University of Chicago, Bonn University, the European University Institute, St Andrews. We thank Marianne Moreira and Diana Lopez Perez for capable research assistance. We are grateful to the British Academy and the European Research Center and the Research council of Norway for providing grants funding the data collection of data in Tanzania, including Attanasio's ERC Advanced Grant 695300 (HKADeC) and FAIR Project 262675. Jervis gratefully acknowledges financial support from the Institute for Research in Market Imperfections and Public Policy MIPP (ICS13_002 ANID) and the Center for Research in Inclusive Education, Chile (SCIA ANID CIE160009). The data set used in this paper was approved by the University College London Ethics Committee (2168/013) and the National Institute for Medical Research and the Ministry of Health, Community Development, Gender, Elderly & Children of Tanzania (NIMR/HQ/R.8c/Vol.I/1254).

[†]Institute for International Economics, Stockholm University and Norwegian School of Economics, e-mail: ingvild.almas@iies.su.se

[‡]Yale University, FAIR@NHH, NBER; e-mail: orazio.attanasio@yale.edu

[§]Universidad de Chile, Institute for Research in Market Imperfections and Public Policy, Institute for Fiscal Studies and Center for Research in Inclusive Education, e-mail: pjervisr@uchile.cl

1 Introduction

For many years, at least since Samuelson's (1938; 1948) and Arrow's (1959) contributions, most economists thought that good empirical work can only be based on a narrow set of measures. The prevalent perception, with a few exceptions, was that we can only meaningfully use measures of what people buy or do, their resources, prices, and (possibly) the markets they have access to. Samuelson (1938) motivates his contribution by arguing that properties of the utility function can be inferred from consumer choices rather than from introspection and psychological factors. This intuition gave rise to revealed preferences. Subsequently, the majority of the profession went further and became convinced that models of human behavior can *only* be studied from choice data.[1] While many economists have also proficiently used *objective* data, including biological and anthropometric data, the set of measures deemed acceptable and interesting by a large fraction of the profession has been limited. Data on stated preferences and their intensity, intentions to buy, stated actions in hypothetical situations, subjective expectations, and attitudes and tastes – as well as data on the possible drivers of individual choices, such as social norms, culture, or political attitudes – were seen as not particularly useful and outside the realm of economics.

From the measures perceived as meaningful and acceptable one could, under certain assumptions, infer and characterize preferences and other structural parameters that drive individual behavior, as well as some of the features of market structure and, maybe, identify empirically the causal links among different variables. To perform such exercises on a restricted set of measures, however, requires strong assumptions.

The reliance on restrictive and specific sets of measures, disregarding richer measures, such as data on rankings of different choices or the intensity of preference, data on stated intentions, or answers to questions about choices in hypothetical scenarios has been widespread in different economic fields and types of models, both static and dynamic. Such an approach has forced the use of restrictive models in empirical analysis.

In most static models, for instance, empirical research has used mostly choice-based data in combination with objectively observed variables (such as prices or other indi-

[1] This perception is clearly stated, for instance, by Varian (2014): "Up until now, we were thinking about what preferences could tell us about people's behavior. But in real life, preferences are not directly observable: we have to discover peoples' preferences from observing their behavior." (Chapter 7)

vidual and environmental data).[2] This approach has led researchers to focus on models that either imposed homogeneity assumptions or a very specific set of heterogeneous preferences. And even in models with heterogeneous preferences, such as the Random Utility Model, the use of choice-based data poses key identification problems. Such problems are particularly salient when individual choices are determined not only by preferences and resources but also by other factors that are typically deemed unobservable, such as individual beliefs.[3] In the case of dynamic models, where uncertain future variables play a role, researchers used assumptions such as adaptive, myopic or, in recent decades, rational expectations, which allowed an internally consistent solution of the model under study and its empirical applications.

The kind of measures researchers consider as meaningful and usable implicitly defines the models one can bring to data and, therefore, the domain of economics as a social science with empirical content. In this paper, we discuss the important role new measures can play in the study of economic behavior. The use of innovative measures, such as answers to hypothetical questions used in conjunction with choice data, can help the identification of causal links with less restrictive assumptions. It also limits the need to identify exogenous sources of variation in observational data.

We discuss what should be measured and how these measures relate to models of economic behavior. While we stress that rigorous and coherent models of human behavior and human interactions are key to understand economic reality, we do not think such an approach implies that empirical studies should exclusively use measures derived from choice data and easily and objectively measurable quantities (such as prices or anthropometric data). Instead, we argue that economic theory can and should inform the design of new measures that capture the factors salient for the models at hand. These measures can lead to the use of more flexible and richer models of individual behavior.

One of the main reasons a large part of the profession shied away from certain types of measures has been the challenge in designing appropriate tools to gather them. Therefore, a substantial amount of research effort should be devoted to the design and, importantly, the *validation* of new measurement tools, to ensure that they capture the

[2]Berry et al. (2004) is an exception, where rankings or 'second hypothetical choices' are used in the identification of demand models for cars.

[3]A very recent paper by Dardanoni et al. (2022) discusses the restrictions that the use of choice data requires for models with preference and other types of heterogeneity.

phenomena they are meant to. New measures, designed and devised by researchers, should be piloted, validated and shown to be correlated with individual choices. Appropriate techniques in psychology and survey design can (and should) be used to develop tools that could capture latent variables relevant for models of individual behavior.

Measurement research and progress are not new. An example is the development of the unfolding bracket technique (Juster et al., 2006), to measure variables, such as household wealth, which had been seen as impossible to assess accurately. New measures continue to emerge: Bloom and Van Reenen (2007), for instance, started a new research agenda to construct innovative measures of *management skills*, seen as inputs of a production function. Caplin (2021) has been proposing the design of new measures thorough *data engineering* with a logic close to what we argue for in this paper.

A focus on measurement does not remove the need for theoretical models. Indeed, the use and design of new measures should be driven by theory and the need to identify key parameters of the theoretical models under study, using direct measures instead of restrictive (structural) assumptions. Further, an approach to measurement and empirical strategies that goes beyond standard measures does not imply rejecting the modeling of individual behavior based on some sort of constrained optimization that satisfies a set of axioms. As we argue below, the use of new and innovative measures allows researchers to give empirical content to flexible and powerful models, which can replace restrictive models constrained by the lack of available measures. The development of new measures and new more flexible models should go hand in hand.

The rest of the paper is organized as follows. In Section 2, we provide some background material about measurement, its role and its interactions with economic theory. Given space limitations, we do not provide an exhaustive list of references and refer to Appendix A1 for a more complete list. In Section 3, we discuss various aspects of the relationship between theory and measurement, and present the case for going beyond standard measures. We discuss what measures are useful for different models, how the parameters of measurement systems can be identified and how they can be relevant for defining the metric of the relevant latent factors and how they can help the identification of causal links between the various latent factors. Having discussed these conceptual issues, we sketch, in Section 4, a model of household behavior where some of the issues on the relationship between theory and measurement are fleshed out and made explicit in terms of the latent factors and drivers that populate the model. In Section 5, we provide some examples of measurement work and its use, using a novel data set that was collected to pilot new measures of the abstract concepts that are used in the model considered in Section 4. In Section 6, we use the novel measures and suggest an empirical model of parental investment derived from a theoretical framework, to show how the new measures can be used in combination with standard ones. Section 7 concludes the paper.

2 The context

A stark statement of a restrictive approach to the study of preferences in economics is in Stigler and Becker (1977), entitled *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*." (emphasis added).

It is interesting to note the explicit reference to the stationarity of preference and its cross-sectional homogeneity. Although empirical studies based on choice data and objective measures do not necessarily require cross-sectional homogeneity – and indeed much empirical and econometric research has studied models which allow for heterogeneity – this statement implicitly asserts what are acceptable measures for economists and imposes important restrictions on the models researchers can eventually study.

Imposing the use of a limited set of measures implies that *identification* (that is, the possibility of retrieving empirically the fundamental features of the model under study) may only be achieved with strong assumptions on tastes, beliefs, expectations, and information individuals have access to. This is the price that is paid to assume that such variables and factors cannot be observed or measured. However, the profession has shown much skepticism towards novel measures that could provide information on these types of variables, such as questions that pose *hypothetical* situations and evidence from *stated* rather than *actual* choices.

These issues have been analyzed for a long time. An example of the arguments about what and how to measure can be found in the discussions of stated preferences and conjoint analysis in Luce (1956) (and other papers listed in Appendix A1). Likewise, the discussion of the Random Utility Model by Block and Marschak (1960) states that it may be beneficial to follow the practice in psychology of accepting subjects' ranking of objects and intensity of preferences, even if observed through a verbal statement (see also Caplin, 2021, for a discussion of these issues).

There are good reasons for the profession's skepticism about certain measures. Measuring hypothetical choices, preferences, and attitudes is fraught with difficulties. Framing effects, for instance, seem to be pervasive and introduce a number of potentially severe biases. Several studies, mentioned in Appendix A1, discuss common biases in answers to questions about hypothetical situations. An interesting debate in this respect regards the use of *contingent valuation*. While this type of measure is widely used in other disciplines, such as marketing, its use in economics has received a considerable amount of resistance. Hausman (1994) expressed doubts about its usefulness and Hausman (2012) labels the enterprise as hopeless. This skepticism is partly due to measurement difficulties, though it is also probably due to the ambiguities around what one is measuring when asking questions about hypothetical choices.

Another interesting early discussion of what could and should be measured was the lively exchanges between Tobin (1959) and Katona (1959) on the usefulness of data on buying intentions; Tobin strongly criticized the usefulness of such data on the basis that they were not a good predictor of actual consumer choices. The reliability and predictive power of buying intentions and purchasing probabilities data were later discussed in Juster (1964) and then again in Manski (1990), who noticed that intention questions are not necessarily useless if formulated properly. Manski argues the issue is not *what* is being measured, but the *specific tools and questionnaires* being used.[4]

While these issues were being debated, some researchers used stated preferences and elicitation of hypothetical choices to estimate key parameters of economic models. Juster and Shay (1964), for instance, used the elicitation of stated choices in hypothetical situations to estimate the elasticity of the demand for loans to interest rates and loan maturities; the interest rate and maturity of these *hypothetical loans* were exogenously varied across respondents to the survey. Cross-sectional differences in loan demand

[4]Manski also remarked that intention data, while not been used much by economists, are instead widely used in other disciplines, including marketing.

elasticities were then used to discuss the importance of liquidity constraints. More recently, Lancaster and Chesher (1983) used, in conjunction with a model of employment search behavior, "the answers to two simple questions" which could be interpreted as providing information about the distribution of offer wages and reservation wages to "...*deduce* structural parameters rather than estimate them" (p. 1661).

What could and should be measured and its relation to theory was discussed in Haavelmo's (1958) Presidential Address to the Econometric Society. Haavelmo perceived measurement questions to be central to the development of economic theory:

> "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.*" (emph. added).

Haavelmo's address argues new measurement tools should be developed in reaction to theoretical models that are both rich enough to capture complex phenomena and require novel measurement tools. Such new measurement tools would allow the identification and characterization of richer models, where strong assumptions are relaxed. In what follows, we similarly argue that new measures for variables that play an important role in several theoretical models should be developed and validated.

These early efforts and discussions were not without consequence. For instance, Katona's work (see e.g., Katona, 1974) led to the establishment of the Michigan survey of consumer sentiment, which is still running (and used) today. In recent decades, however, the design and use of such surveys has become rare. From both a theoretical and measurement point of view, the consensus within most of the economic profession went towards an almost exclusive use of choice based and objectively observable data.

An exception is the field of experimental economics, where researchers construct controlled – and often artificial – settings in which subjects make choices and, in so doing, reveal their preferences, beliefs, and other key drivers of behavior (Plott and Smith, 2008). This approach allows researchers to control the situations presented to subjects, typically in a laboratory setting, so to control away many potential confounding factors to reveal features of human decision making. A considerable amount of such work has been done on mechanisms and protocols to elicit primitives of individual behavior. However, lab experiments often require participants to behave abstracting from their present circumstances, imposing a separation between experimental and actual behavior. Indeed, background information on experiment participants is rarely collected and experimental data are rarely used in conjunction with observational data.

Recently, experiments have been brought to the field to measure a variety of real life phenomena (e.g., preference for and attitudes towards redistribution, attitudes towards migrants, and willingness to compete (Almås et al., 2020b; Alesina et al., 2018; Buser et al., 2014)). The expanded use in the field of techniques and protocols developed in labs and the simultaneous collection of experimental and standard observational data are a sign that the economic profession has been changing its approach to what can and should be measured. At the same time, techniques used in empirical studies on standard data have been reproduced in the laboratory, in particular, for the analysis of auction models. Additional references on these issues are in Appendix A1.

Several other measurement novelties have proliferated in recent years. One important and relatively recent contribution, which has enlarged the set of variables that economists consider measurable, is the study of subjective expectations (e.g., about income, inflation, or rates of return) discussed in Manski (2004) and other papers mentioned in the Appendix A1. Subjective expectations data contained in the early NLS data, sometimes known as the *Parnes data* and discussed in Parnes (1975).[5] These types of data are now routinely collected; the Bank of Italy has collected subjective expectations data for a number of years and the Federal Reserve Bank of New York has been systematically collecting consumer expectations data since 2013 and similar data are also collected by the Bank of Spain and the European Central Bank. Research economists have also started to use subjective expectations data within structural models of economic behavior. An early use is Wolpin and Gonul (1985), while more recent examples are listed in Appendix A1. The availability of subjective expectations data al-

[5]We are grateful to Ken Wolpin for pointing us to these data.

lows researchers to avoid strong assumptions, such as rational expectations. Moreover, such data allow the identification of genuine measures of uncertainty, which, using data on actual realizations of the variable of interest is not easily disentangled from individual heterogeneity or variability that is known and deterministic to individuals.

A related topic is the measurement of beliefs about the return to specific investments, such as different types of investment in human capital and education. Rather than assuming that individuals have rational expectations about the returns to certain investments, researchers have started eliciting data on *beliefs* about returns. Several studies have started collecting and analyzing data on parental beliefs, following a practice that has been used for some time in psychology and child development, as surveyed, for instance, by Miller (1988). Examples of such studies include Cunha et al. (2013) and Attanasio et al. (2019) and others mentioned in Appendix A1.

Another example of new measurement tools being developed and used by economists is the study of stated preferences and answers to hypothetical questions. A good example of such a practice is Ameriks et al. (2020), which uses a combination of stated preferences and actual choice to identify complex structural models. Other papers along these lines are listed in Appendix A1. These approaches are analogous to questions about *intentions* (e.g., how respondents would allocate hypothetical resources among different potential uses), to elicit information about individual tastes and preferences.

While resources, preferences and tastes are obvious drivers of individual choices, other factors can also be important drivers of behavior. In certain contexts, for example, the quantity and quality of information available to individuals is important. Individuals' access to markets or networks can influence the allocation of resources across time and space. Additional factors, such as learning, risk sharing arrangements, preferences about different policy options, attitudes and social norms, might also affect individual choices (and preferences). When studying household-level choices, for instance, who controls resources and bargaining power within the household might be important. Recently, some new tools to analyze many of these phenomena have been developed, such as Almås et al. (2018) measuring individual spouses' willingness to pay for the control of resources and Jayachandran et al. (2021) trying alternative approaches.

These studies illustrate the active and ongoing discussions in economics about measurement and its relation to theory. Recent developments indicate that the profession is moving towards using choice data and directly observable variables in combination with *stated preferences* and answers to hypothetical questions.

An interesting and recent take on measurement and its relation to theory is discussed in Stantcheva (2022), where she says: “Surveys are not merely a research tool. They are also not only a way of collecting data. Instead, they involve creating the process that will generate the data.” We share this view. In what follows, we develop it to include the design and validation of new measures to be used in combination with traditional ones for the empirical study and characterization of economic models.

3 Measurement and Theory

The *economic models* economists work with often use abstract constructs that are not directly observable. The measures relating to these constructs that economists collect and use are (and should be) driven by the theoretical framework that organizes our thinking. Examples abound in several fields of economics. They include the work by Keynes, Stone, and others that led to the development of National Accounts and the analysis of consumer demand, which is behind most methods to construct price indices. The literature on the measurement of growth, consumption, and price indices has used theoretical models as a foundation for the construction of new measures. Analogously, the creative use of micro-data has informed the development and calibration of theoretical models. We report some of the relevant papers and references in Appendix A1.

We discuss the relation between theory and measurement in Section 3.1, and, in Section 3.2, argue that new measurement tools, beyond standard measures, are desirable and useful. In Section 3.3, we discuss some issues relevant for *measurement systems*.

3.1 Economic Models, Latent Factors and Measurement Systems

Many theoretical models in economics can be represented by:

$$F(\theta;\phi) = 0 \tag{1}$$

where θ is a matrix of variables of interest or factors, some of which are latent, in that they are not necessarily observed. The parameter vector ϕ characterizes the function

F, which represents individual behavior and interactions (such as markets), that is, the relevant economic model. F typically defines what the variables of interest are.

Within this framework it is easy to introduce a number of details about the features of the economy under study. For instance, one could include in model (1) uncertainty and imperfect information, and consider additional factors relating to the information available to the model's individuals. The dimension of the model's latent factors depends on the specific issues under study. Richer and more realistic F functions require a richer set of factors and, to be characterized empirically, a richer set of measures.

The *factors* that populate a theoretical model are often well-defined but unobserved variables, such as prices of very finely defined goods or the quality of family environment and schools. In practice, what is often available are *markers* corresponding to (some of) these theoretical constructs. To bring the theoretical models to data and to identify and estimate the parameters ϕ that define the causal links one is interested in, it is necessary to be explicit both about the theoretical model and about the relationship between the relevant latent factors and the available measures. In other words, to give empirical content to the function F in (1), one needs a *measurement system* that relates the *latent factors in F* to the available measures. A possible mapping is the following:

$$\mathbf{m} = g(\theta, \varepsilon) \tag{2}$$

where $\mathbf{m}$ are available measures related to the (potentially unobservable) factors θ through the function g. The vector ε is measurement error that, together with the possibility that the function g is not injective, prevents the direct observability of (some of the) θ. The model F defines the factors of interest, and, in turn, guides what measures to look for. The available measures and the measurement system in (2) define what latent factors one can study empirically and which models can be taken to data.

Such mapping between latent factors and constructs of interest for economic theory and a set of available measures resonates with Goldberger's (1972) description of the interplay of theory and measurement in his Fischer-Schulz lecture (*emphasis added*):

> "By structural equation models, I refer to stochastic models in which each equation represents a causal link, rather than a mere empirical association. The models arise in non-experimental situations and are characterized by simultaneity and/or errors in the variables. The errors in the variables may be due to measurement error in

the narrow sense, or *to the fact that measurable quantities are not the same as the relevant theoretical quantities*. Generally speaking the structural parameters do not coincide with coefficients of regressions among observable variables, but the model does impose constraints on those regression coefficients" .

Goldberger (1972) uses the Permanent Income model as an example, where *permanent income* is the interesting construct and the empirical measures potentially related to it are *current income and consumption*. In a similar vein, Griliches (1974) discusses the relationship between earnings, schooling, and ability. In a different context, the estimation of a production function with endogenous inputs could be viewed in a similar fashion (see Olley and Pakes (1992), and other papers listed in Appendix A1) . More recently, Cunha et al. (2010) use a relatively flexible version of (2) to estimate the production function of human capital. The early work on Multiple Indicators Multiple Causes (MIMIC) models and, more generally, studies on factor models in economics, psychology, sociology, and genetics, listed in Appendix A1 are relevant and important.

Recognising explicitly that economists' theoretical models are often populated with abstract constructs is useful. It clarifies the research objectives and may motivate attempts to measure additional relevant variables, which, in turn, can motivate researchers to use more realistic models that are subject to less stringent assumptions.

Which latent factors to measure. The measurement system and the measurement tools used in empirical studies – as well as what should be measured (and possibly how) – should be informed by the specific questions researchers ask and by the theoretical models being used. Expanding the set of objects one measures allows the consideration of more flexible models and avoids strong and sometimes misleading assumptions.

The latent factors of interest depend on the complexity of the theoretical model being studied, which, in turn, might depend on the phenomenon being interpreted. In this regard, an explicit discussion of the restrictions imposed on the theory by data availability and measurement challenges is useful. In certain contexts, it might be apparent that these restrictions have negligible impacts for what is being studied; in other contexts, however, restrictive definitions of the relevant latent factors might substantially limit the ability of a model to explain observed phenomena.

When working with demand systems, researchers typically aggregate different commodities in coarse categories. However, available data (even when very detailed) might miss important components of the commodities considered, such as quality of certain commodities or the market structure faced by consumers.[6] When studying intrahousehold allocation, measures of individual-level consumption (in addition household-level expenditure) are often unavailable, forcing strong assumptions.

Consider, for instance, studies of production functions where output is the result of combining different inputs, such as human and physical capital. Until the late 1990s, studies of labor market inequality used a basic model where production is performed via a production function that uses two types of labor (skilled and unskilled). Such a simple model explained well a set of labor market facts.[7] That model, however, could not explain what happened in the first part of the 21st century. As a result, new models that disentangle *skills* from *tasks* have been developed, such as in Acemoglu and Autor (2011), where workers with the same skill level can perform different tasks. The empirical needs of these models require new types of data, such as the *Dictionary of Occupational Titles* or the more recent and detailed O*NET, which provide information on the task content of different occupations observed in Census and other data sources, (see, for instance, Autor and Dorn, 2013). Furthermore, to study empirically more complex production functions with flexible roles for different skills, such as *sociability*, *drive*, and *motivation*, in addition to *cognition*, it is key to measure these skills, how they differ across individuals, how different occupations might require different combinations, and how they are remunerated in the labor market. Likewise, the measurement of these latent factors, and the comparability of measurement tools, often used in different contexts, then becomes particularly challenging and key to the results one obtains.[8]

In summary, the questions a researcher is addressing, the theoretical models they use, and their empirical performance define the key latent factors of interest. This, in turn, defines which measures are needed to give empirical content and bite to the theo-

[6]Researchers typically use *price indices* for the aggregate measures of commodities that are used in analysis. However, in some contexts prices may not be linear and change with the quantity purchased, as in many models of price discrimination. Relevant citations in Appendix A1.

[7]As discussed, for instance, by Katz and Murphy (1992), which developed Tinbergen's original explanation of labor market inequality as the effect of the relative demand and supply of different skills.

[8]Similarly, in studying firms heterogeneity, Bloom and Van Reenen (2007), use an innovative survey, now deployed in several countries, to measure managerial skills as an input in the production function. Other studies relevant for the study of production functions are listed in Appendix A1.

retical framework at hand. While in some cases standard measures, possibly anchored by choice data, are sufficient, in many other cases they are not.

3.2 Beyond standard measures

The revealed preference approach has been very useful for structuring our thinking around agents' decision making and relating rigorous theory to data with relatively few assumptions. With observations about individual choices, the conditions under which these choices are made (such as the resources individuals control and the prices they face), and objectively measurable variables, preferences consistent with a set of axioms which defines a theory of economic behavior may be revealed. Identification of structural parameters can then be achieved *without explicit measures of tastes, beliefs, or attitudes, and without questions about behavior in hypothetical circumstances.*However, this approach often rests on strong assumptions.

Many interesting economic models routinely deviate from such assumptions and constructs, posing difficult questions if their empirical analysis relies exclusively on *choice data and variables objectively measured.* Often these challenges are tackled with strong theoretical or empirical assumptions. For instance, to study dynamic problems where future and uncertain variables are key, researchers frequently have used the assumption of rational expectations and of an efficient use of all the available information.

On the empirical side, to avoid endogeneity issues, researchers try to isolate or create sources of exogenous variation through 'natural' experiments, a strategy that often requires ad-hoc and un-testable assumptions. Even genuinely exogenous variation induced by randomization can often identify a narrowly defined set of parameters.

Expanding the set of measures one uses might make it possible to study more realistic, yet still rigorous, models. With new and more comprehensive measures, it may be possible to maintain consistency with economic theory while at the same time relax some strong assumptions as well as discriminate among alternative theoretical constructs. For instance, questions about behavior in hypothetical scenarios can introduce exogenous variation in a much richer way than natural experiments. While important contributions on new measures have been made in the literature – e.g. the aforementioned literature on subjective expectations – there is still a need to develop and validate new measures that can be used in the analysis of agents' decision making.

The need for and use of new measures: a few examples. In order to clarify the need for new measures that go beyond data on choices and to better understand the framework in which such new measures can be used, this section provides a few examples, some of which relate directly to the applications we present in Sections 4 and 5. These examples are relevant for: (i) the definition of decision units in models of choice; (ii) the separate identification and characterisation of preferences; and (iii) the characterization of the economic environment.

Decision units. A first step when modeling individual behavior is to *define* the decision unit. In standard consumer theory, the household has most often been considered as the relevant decision unit. In such a *unitary* model, the household as a whole is considered the relevant decision unit with well-defined preferences. In recent decades, however, researchers have focused on how resources are controlled and allocated *within* the household, developing alternative models where multiple decision makers, each with distinct preferences, interact within the household to arrive at household-level choices.

The collective model, first proposed in Chiappori (1988), is one such attractive alternative. Its key assumption is that choices, while resulting from interaction between decision-makers with potentially distinct preferences, are efficient. While a number of important theoretical results, listed in Appendix A1, have been derived, characterizing the parameters of the models used and testing their validity exclusively with *choice data* on household consumption and expendituresis challenging.

Much more can be learned by using additional information on *private* consumption. A number of researchers have used information at the individual level within households, such as Dercon and Krishnan (2000) and others listed in Appendix A1. However, even when *individual level* data are available, identifying the determinants of individual and eventually household behavior can only be achieved with strong assumptions. For instance, it is difficult to allow for *caring*-preferences, i.e., when one partner cares about the consumption of the other partner. Instead, nonstandard measures that do not rely on the exclusive use of choice-based data can generate important insights about the process of intrahousehold allocation of resources. For instance, hypothetical choice scenarios elicited separately from the individuals in the household can generate direct information on individual *tastes*. Likewise, it may be possible to derive information on the relative bargaining power within the household, a measure that goes beyond standard choice-based data. We discuss some of these measures in Section 5.1 and Appendix A1.

Disentangling beliefs and tastes. Agents in most models in economics make decisions to maximize an objective function, given the resources available to them. These decisions then depend on their preferences and on their perception of *the process that links actions to outcomes*. Often, the characterization of such a process is of key interest to researchers. Identifying the causal links that define it requires understanding how individual choices are made. It is often assumed that individual decision-makers *know* the process that determines the outcomes they care about, given their actions and other variables. However, in many situations, this assumption is a strong one.

The challenges related to disentangling individual perceptions or beliefs and tastes have been extensively discussed in several different settings (see e.g., Caplin, 2021). Possible approaches are: i) direct elicitation of beliefs (and retrieving preferences from choice data); ii) elicitation of preferences through experimental approaches and hypothetical choice scenarios, holding beliefs constant by giving surveys' respondents full information about the context; and iii) direct elicitation of beliefs *and* preferences.

Beliefs elicitation. Direct elicitation of beliefs has been used in a model of child development and parental investment in Cunha et al. (2013) and by Attanasio et al. (2019), eliciting the perceived productivity of parental investment on child development. We follow a similar approach in Section 5. In a different context, Mueller and Spinnewijn (2021), studying on search behavior among unemployed, *suggest using direct measures of beliefs while retrieving tastes as a residual*.

Preferences elicitation. Another approach is to elicit preferences directly *holding beliefs constant* in controlled situations with full information about the actual setting. This can be done either with experimental methods, revealing preferences through real choices, or posing different hypothetical scenarios to respondents and eliciting their (stated) preferences. An early example of such a strategy is the aforementioned Juster and Shay (1964). More recently, Ameriks et al. (2020) and Bernheim et al. (2021)

have used hypothetical questions to estimate parameters that characterize individual preferences. Some studies that use these approaches are discussed in Appendix A1.

The environment. To better understand individual (economic) behavior, it is useful to measure how wider constructs, such as institutions, communities, and society at large affect individual behavior via social norms, attitudes, or culture, and model their evolution (see e.g., Bisin and Verdier, 2000, for a discussion of how culture evolves).

Social norms and attitudes affect individuals' objective functions in significant ways. In a recent paper, for example, Field et al. (2021) study the effect of an intervention aimed at increasing female's control of resources and find that its impact resulted in an increase in female labor supply, contradicting the implications of a standard collective model with individual utility depending on consumption and leisure. The researchers argue that, in reality, social norms play an important role in determining choices and this type of intervention might have led to a shift in such norms. The challenge then is to determine appropriate and validated measures of such norms.

Along similar lines, interesting measures are those that attempt to capture what is sometimes defined as *social capital*, i.e. a set of norms that inform individual behavior and affect the ability of a society or community to provide public goods and internalize externalities, or other social attitudes that might affect individual interactions. Different studies that have looked at different approaches to measure social capital, ranging from measures of participation in certain activities –from blood donation to church attendance, (Guiso et al., 2004, 2006) — to data derived from field experiments (Attanasio et al. (2012) on determinants of group formation), to the effect of deterrence on preferences (Cavatorta and Groom, 2020).

In characterizing empirically certain markets (such as credit or insurance) and determining the model that best describes them, quantitative measures of specific frictions, jointly with *choice data*, can be very useful. In models of insurance with imperfect information, it may be useful to devise measures of the quality of information in a risk sharing group, as done, for instance, by Attanasio and Krutikova (2020).

What measures for what theories. We have discussed a few examples of theoretical models whose empirical analysis might need additional measures. One example is the identification of individual beliefs and preferences without strong assumptions.

In modeling parental behavior, for instance, it has been often assumed parents are fully aware of the nature of the process of child development. While such an approach can ease the analysis, a less restrictive model, embedding a more complex structure F in (1) that allows for distortions in parental beliefs, may be more realistic and avoid misleading conclusions. In Section 4, we sketch a model of parental investment, which we analyze empirically in Section 6, to illustrate how additional and somewhat unconventional measures could (and should) be used in conjunction with traditional ones.

In some cases, the additional measures that allow the empirical characterization of more general models are just finer and more detailed versions of existing ones (as in the case of individual rather than the household level of consumption). In others, the new measures try to capture new concepts that are specific to the model being analyzed, such as (distorted) beliefs about child development or bargaining power within the household. In general, the choices between alternative theoretical approaches should not be conditioned and limited by the availability (or lack thereof) of data and appropriate measures.

These considerations are relevant, we believe, for the debate between Gul and Pesendorfer (2011) and Camerer (2011). While both papers make some interesting and important points, they both take what we think is an over-restrictive approach. Gul and Pesendorfer (2011) insist that economic models describe the behavior and interactions of agents that are assumed to maximize a given objective function and that these models' features should be consistent with a set of axioms that help to frame them. However, to characterize empirically these models, they refrain from using data and measures different from data on actual choice. While the premise that the empirical models economists use should be theory consistent is a sound one, we believe that complex models are often better analyzed and characterized using data beyond those derived from observed choices. Symmetrically, such data could allow the use of richer models.[9]

Camerer (2011), on the other hand, points out that, again correctly in our opinion, new measures (such as the neurological data and biological markers he discusses) can be useful to better characterize individual behavior. However, Camerer (2011) seems to

[9] While the measures of certain variables might be affected by multiple biases, the problem, as clearly discussed by Manski (1990), is not with the measures *per se* but with the tools used to collect them.

want to characterize individual behavior in a way that abstracts from a set of theoretical axioms and to describe directly the relationship between biological mechanisms and behavior. Apart from the difficulty in pursuing such a strategy, such an approach goes beyond the realm of economics. The rejection of a specific model *in one context* is not a good reason to throw away the whole approach and work with models that are not consistent with a set of axioms.

Interestingly, Benhabib and Bisin (2011) argue that traditional decision theory, as advocated by Gul and Pesendorfer (2011), focuses only on the need to model *choices*, while other approaches, somewhat misleadingly labeled as *behavioral economics*, want to understand the *processes* that lead to a specific set of choices. To better understand *processes*, additional measures, such as biological and anthropometric ones but also measures of a wide variety of latent factors, can be useful, if such measures are integrated in well-defined decision models of individual choices.

As economists, we need models that focus on economic ideas. Such models should not aim to describe completely psychological processes or define what happiness is. Well-constructed economic models, whose main aim is to describe and understand individual choices and interactions, should be based on a set of axioms and be consistent with them. To give empirical content to such models additional measures that complement data on choices (and prices and resources), whose design and features should be driven by the needs of the theory, can be useful.

3.3 Measurement systems

In this section, we discuss the use of measurement systems and we focus attention on a situation where we have data on a set of measures that are reflecting, but not completing revealing the latent factors of interest. These issues have become particularly important as economists have started using more extensively data on variables that are not easily or routinely measured, such as soft skills or child development. Much has been learned from psychometrics, but a number issues that become salient when these new measures are used in quantitaive ecomic models, have to be tackled.

Equation (2) defines a *measurement system* in a fairly general way. Having discussed its relation to theory, we now move to a number of issues relevant for the specification and estimation of such measurement systems. To make the discussion concrete, we use a specific characterization of the measurement system (2), similar to that used by Cunha et al. (2010). We denote with θ_{it}^{j} the $j-th$ element of the vector θ for individual i at time (or age) t. Let $m_{i,t}^{jk_j}$ be measure k_j of the K_j available and relevant for factor j. We assume that factors and continuous measures are related by the following system.

$$m_{i,t}^{jk_j} = \alpha_t^{jk_j} + \beta_t^{jk_j}\theta_{it}^{j} + \varepsilon_{it}^{jk_j}, \quad j = 1,...J; \;\; k_j = 1,...,K_j. \tag{3}$$

For discrete measures, we assume an Item Response Theory (IRT) model, extensively used in the psychometric literature and that we discuss at length in Appendix A2.

The relation between factors and measures as written in equations (3) is, in many ways, restrictive. However, it is useful to discuss some of the issues with measurement systems, starting with what is needed to identify empirically such a model.

Identification of parameters of the measurement system. The identification of the parameters of the model above or, more generally, that in equation (2), requires the availability of several measures linked to each latent factor with measurement errors that are uncorrelated among at least two of them. An extensive discussion of the identification issues of these models is contained in Cunha et al. (2010), and other contributions listed in Appendix A1.[10]

We will not repeat the formal arguments here, but we do stress that the need of independent measurement errors should inform the way surveys are designed and data collected. It would be desirable, for instance, to randomly assign different interviewers to collect different measures targeted at the same variables or collect them on different (randomly allocated) days. A similar argument applies to attrition (an extreme form of measurement error). One could, for instance, allocate resources spent on minimizing attrition randomly across observations. These considerations about data collection make clear the existing trade-offs: certain strategies might maximize data quality, while others might provide information that could be used to deal with measurement error.

Metrics and cardinalization. Considering equations (3), it is clear that, even with specific assumptions about the distribution of the latent factors θ^j, not all parameters

[10]In some cases, non-parametric identification requires at least 3 variables. *Identification* here refers to the parameters of the function g in equation (2).

of these systems (that is, the $\alpha_t^{jk_j}$'s, $\beta_t^{jk_j}$'s, and the moments of the θ^j's), can be identified. It will be necessary to define a metric for the unobserved latent factors through appropriate normalization.

One possibility is to normalize the mean of the factors to 0 and the variance-covariance matrix to the unit one, an option often used in standard software packages, together with that of normality of the latent factors. Alternatively, one could normalize the $\alpha_t^{jk_j}$ and $\beta_t^{jk_j}$ of a specific measure to 0 and 1, respectively, effectively using that measure as the relevant metric for the unobserved factor. Both approaches are valid and effectively equivalent, with some important caveats.

Regardless of whether one normalizes the mean and variance of the factors or some of the parameters of the measurement system, when one has longitudinal data, it is necessary to establish whether one imposes these normalizations for the first t or for every t. Depending on the context, different approaches might be more useful. If one is interested in how the latent factors change over time,[11] it might be more useful to normalize only at one point in the observation sample. Such an approach, however, imposes the assumption that the relationship between the various measures and the latent factors is unchanged (i.e. measurement invariance). The imposition of these normalizations has implications for the interpretation of the evolution of different factors and their growth, as discussed in Agostinelli and Wiswall (2017).[12]

An issue, related to the normalization of the different measures and the identification of a metric for the unobserved latent factors, is that they often enter economic models as cardinal variables. In some cases, the relevant cardinal metric can be easily identified, in others that is not the case. This statement applies not only to models that consider, say, consumption or income but also, for instance, to the models of child development we consider below or in studies that consider the *value added* provided by schools. The issue is particularly difficult to deal with when the available measures are of an ordinal nature, using, for instance, Likert scales. The specification of the measurement system, in most contexts, should strive to obtain cardinal measures, which can be used in combination with ordinal measures but that provide the necessary anchor and metric that allows to obtain cardinal measures of the relevant factors.

[11] For child development, for instance, one might be interested in measuring *children growth.*

[12] A related issue, relevant for longitudinal data, is when the measures that are appropriate for a factor change with time. These issues are discussed in the case of child development in Attanasio et al. (2020a).

From measurable items to the measurement of latent factors. As written in equation (3),the measurement system assumes that each factor affects only one measure; that is, it is a *dedicated* measurement system. This assumption can be relaxed and have several factors affecting a single measure. However, as mentioned by Cunha et al. (2010), to achieve identification it will be necessary to have at least one measure per factor that is affected only by that factor.

In many contexts, researchers estimate unobserved latent factors using a variety of tests that have become standard practice in the academic community and beyond. In the context of child development, for instance, much work exists in psychometrics which has influenced and shaped the development of a number of tests designed to measure different dimensions of child development or its drivers. Most of them are made of a large number of individual items that are routinely aggregated using scoring algorithms that deliver estimated developmental scores or indexes of parental investment. These algorithms were typically developed using factor models similar to those we discussed on samples of children on which the original items were tested. In other cases, the scoring mechanism is very simple, like the sum of correct answers to a number of questions.[13] While the original algorithms were eventually validated in a number of samples, it is not necessarily the case that, several decades later and in contexts that might be different from those in which the scoring algorithm was constructed, the same scoring algorithm is necessarily the best way to aggregate the information from the individual items. A feasible and more effective use of the data collected would be to re-estimate the scoring algorithms (that is, the measurement system that relates latent factors to the available measures) in the new context where these data are collected. As we discuss in Section 5, this can also lead to the collection of different tests that could more easily be deployed in a given context. Along the same lines, new items could be piloted to complement the existing standard tests.

These issues are particularly relevant in developing countries, which are contexts very different from those where the tests were developed and the scoring algorithms designed, typically in developed countries. Many items might present flooring or ceiling effects, so that the specific tool is not able to capture any variability in the study

[13] An example in child development is the MacArthur Language Inventory test, a list of a few hundred words, with the child's caregiver informing whether the child understands or can say each of them. There is no reason to take the unweighted sum of such words as the estimated latent factor.

sample. The estimation of a measurement system to construct context specific scoring algorithms is an effective and simple way to summarize efficiently the available measures.

Constructing new measures. As mentioned above, some models involve the consideration of factors that are not easily identifiable from choice data, such as (possibly distorted) beliefs, subjective expectations, attitudes and social norms. In these situations, it can be extremely useful to construct new measurement tools that allow the identification of rich and flexible theoretical models. Such measures, however, need to be constructed with care and need to be validated, as the relevant constructs might be difficult to elicit.

New tools have been developed and studied in the literature. A good example of such studies are some of the papers that have designed instruments to elicit subjective expectations. While initial studies would elicit only point expectations, it is now clear that it is possible, in many contexts, to elicit information on the subjective probability distribution respondents hold, using appropriate tools, such as those discussed in Delavande and Rohwedder (2008) and Danz et al. (2022). And in the future, even more nuanced measures could be designed, maybe capturing individual attitudes towards uncertainty and ambiguity.

Validation of new tools is important. It is increasingly clear that some tools designed to measure a certain construct work well in a given context, while in other contexts, alternative tools, designed to capture the same construct, work better in in different contexts. Good examples, which we mention again below, are the attempts to measure bargaining power within couples used by Almås et al. (2018) and Jayachandran et al. (2021).

Identifying causal links between latent factors. Another important set of related considerations is about the identification of causal links *between latent factors*. While the issue here is in terms of the theoretical structure that links the variables of interest, whether such links can be identified empirically might depend on the nature of the data that are available and collected. Indeed, the need to identify specific causal links should and often does drive the design and collection of specific surveys and data sets.

A good example is the design of a Randomised Controlled Trial (RCT) where subjects are randomly exposed to a treatment or an intervention. In this particular example, the investigator creates exogenous variation (exposure to a treatment) on which information is collected to establish (if the experiment is performed correctly) a specific causal link. However, when the object of interest is more complex than the average impact of the intervention in a given context, the data collection strategy should be informed by the specific questions that researchers (or policymakers) might have. If one wants to extrapolate the results obtained in a given context to a different environment, or change details of an intervention, it will be useful and necessary to collect information on the drivers of individual behavior that might affect the outcome of interest.

In Attanasio et al. (2020b), for instance, the objective is to establish what mechanism generates the observed impact of a stimulation intervention on child development, with particular emphasis on the hypothesis that parental investment (in time and materials) played an important role. To find an answer to this question it is necessary to identify the causal link from parental investment to child development, in addition to the overall impact of the intervention on child development. One cannot use the treatment as a valid instrument to identify such a causal link, as the hypothesis of interest is whether the intervention has (or does not have) an impact on child development (directly or through other mechanisms). Attanasio et al. (2020b) use variation across different towns in prices of toys and other items as well as exposure to violence of mothers at the time when they were adolescents as determinants of parental investment that are assumed not to affect child development directly. This assumption (and the availability of the relevant measures), allow then to identify the relevant causal links and perform a *mediation analysis* that takes into account the fact that some of candidate mediators are variables determined by choice.

The general point we want to make is that the design of surveys should be informed by the specific needs and research questions that are being addressed. Survey instruments should use a variety of measurement tools and should not be restricted to collect information about individual choices. Identification of structural models could be achieved with much weaker assumptions by combining standard measures with, for instance, the elicitation of respondents' responses in hypothetical situations.[14] In our

[14]The example of Juster and Shay (1964) is again a good one here.

application in Section 6, we discuss how the empirical study of the model we present in Section 4 is made easier and more interesting by the construction of a different set of variables, some of which require the development of new measurement tools.

4 Modeling parental investment

In this section, we consider a specific application of the ideas we have discussed so far, sketching a theoretical model of household decision making where some of its key components are often not directly observable. The main aim is to show how a number of variables that are not typically measured in standard surveys can be useful to identify the determinants of parental investment.

The framework we present is based on the collective model (Chiappori, 1988), which suggests that the household utility function is a weighted sum of the mothers utility function, $U^m(\cdot)$, and the father's utility function, $U^f(\cdot)$:

$$U = \mu(\mathbf{p}, Y; \mathbf{z})U^m(\mathbf{c}, H) + (1 - \mu(\mathbf{p}, Y; \mathbf{z}))U^f(\mathbf{c}, H), \tag{4}$$

where $\mathbf{c}$ is a vector of consumption of both private and public goods, $\mathbf{p}$ is the corresponding vector of prices, and H is the level of child development. $\mu(\cdot)$, often referred to as Pareto weights, represents the relative importance given to the mother's utility in the problem in equation (4).

The household faces two constraints: a standard (static) budget constraint and a *production function* of child development. We assume that the level of child development depends on initial conditions, H_0, parental financial investment in education, x, and other (unobserved) shocks η, which are not observed by the researchers but might be observed by the decision makers:

$$H = f(H_0, x, \eta). \tag{5}$$

The defining feature of this model is that household choices, resulting from the interactions of different decision makers with possibly different preferences, are efficient, in that they maximize the function in (4), given existing constraints and a set of Pareto weights. The model is silent about what determines the Pareto weights. They are allowed to depend both on variables that enter the problem through the budget constraint, such as prices and income, and, crucially, other variables $\mathbf{z}$, often labelled *distribution factors*. These are variables that matter for the sway in household decision making but do not influence the budget constraint or utility functions directly (Browning et al., 2013).

If the 'household' maximizes the utility function in equation (4), subject to the constraints we mention, parental investment will depend on household resources,on *individual* preferences as aggregated into *household* preferences by the Pareto weights $\mu(\cdot)$, and the properties of the assumed production function. We notice, however, that for the determination of parental investment, the properties of the production function of child development are not necessarily key, unless they coincide with those of the production function as *perceived by the parents*. In the presence of potentially distorted beliefs about the process of child development, what matters is the perceived productivity of parental investment, which can be different for husband and wives.[15]

A simple parametric example can make this framework clear and help us to relate it to the empirical exercises in Section 6. We assume that there are q private goods for husband and wife and that the only public good is child human capital. We also assume that both the individual utility functions and the production function of human capital are Cobb-Douglas (CD). While the commodities we consider are private, we allow the consumption of each spouse to affect the utility of the other spouse. Therefore, we have:

$$\ln U^i = \sum_{j=1}^{q} \left(\alpha^i_{jm} \ln C_{jm} + \alpha^i_{jf} \ln C_{jf} \right) + \alpha^i_k \ln H^k, \quad i = \{m, f\}; \tag{6}$$

where C_{jw} is the consumption of commodity j consumed by the wife, and C_{jh} is the consumption of commodity j consumed by the husband. α^i_{jw} and α^i_{jh} are the parameters of the CD function corresponding to the consumption of commodity j by each of the two household members in the utility function of member $i, i = w, h$. Analogously, α^i_k determines the utility derived by member i from the child's development. We assume that the utility function for the collective household is:

$$U^F = \mu U^m + (1 - \mu) U^f. \tag{7}$$

[15]Note also that this framework can easily be extended to child development being a function of both material investment, x, and *time* investment, e.g., reading or talking to the child, see e.g., Attanasio et al. (2020b,c); Cunha and Heckman (2008); Heckman et al. (2013, 2020); and Todd and Wolpin (2003).

The *perceived* production function for human capital is given by:

$$\ln H^k = \gamma_0^i \ln H_0 + \gamma^i \ln X + \eta, \quad i = \{m, f\}; \tag{8}$$

where γ_0^i and $\gamma_i, i = w, h$ are the parameters of the production function as perceived by the husband and wife. Finally, the budget constraint, with prices normalized to 1 for notational simplicity, is:

$$X + \sum_{j=1}^{q} (C_{jw} + C_{jh} + C_{jk}) = Y. \tag{9}$$

It is easy to show that the household parental investment function is given by:

$$X = \frac{\mu U^m \alpha_k^m \gamma^m + (1-\mu) U^f \alpha_k^f \gamma^f}{A} Y, \tag{10}$$

where the denominator is given by:

$$A = \mu U^m \left(\sum_j (\alpha_{jm}^m + \alpha_{jf}^m) + \alpha_k^m \gamma^m \right) + (1-\mu) U^f \left(\sum_j (\alpha_{jm}^f + \alpha_{jf}^f) + \alpha_k^f \gamma^f \right). \tag{11}$$

While the expression for parental investment in equation (10) is particularly simple because of the CD assumption (which, for instance, implies unit elasticity and constant shares for all commodities), the expression illustrates the role played by the preference parameters (the $\alpha_{ji'}^i,\ i, i' = m, f$), the Pareto weights, and the 'perceived' production function γ^i's. We note that mother's and father's individual preferences for each private commodity and child human capital are mediated by the Pareto weights given by μ.

The Engel curve of parental investment. While the CD assumption makes derivations very simple, the implied homotheticity is obviously not a plausible assumption. One possibility is to use a generalization of Deaton and Muellbauer (1980)'s AIDS system for Engel curves, as done, for instance, by Browning and Chiappori (1998). In this case, household i's budget shares for investment in human capital ($s_i = X_i / Y_i$) is a function of parental preference parameters, parental distribution factors, prices and total household expenditure, allowing the expenditure elasticity to be different from 1:

$$s_i = G(\tau_{\mathbf{i}}, \mu_i, \gamma_{\mathbf{i}}, \mathbf{p}) + \beta(\ln Y_i - a(\mathbf{p})) + u, \tag{12}$$

where G is a function that depends on the main determinants of the parental problem above: parents' tastes for child human capital relative to other commodities, represented by two latent factors $\tau_{\mathbf{i}}$, parental beliefs about the productivity of parental investment in the production of human capital (again represented by two latent factors $\gamma_{\mathbf{i}}$), the relative bargaining power within the couple μ_i, the vector of relative prices, and, as in the AIDS system, log real income $lnY_i - a(\mathbf{p})$ and an unobserved shock u. It is also possible to make both the slope coefficient, β, and the price index, $a(\mathbf{p})$, functions of preference, τ_i, beliefs, $\gamma_{\mathbf{i}}$, and bargaining power, μ_i, and to extend this model to a Quadratic Almost Ideal Demand System (QUAIDS), as in Banks et al. (1997).

We will get back to these issues when, in Section 6, we estimate a simple version of equation (12) for investment in child human capital using measures on parental allocation preferences, parental decision making power, and beliefs, in conjunction with actual choice data we collected in Tanzania. Some implications of this model, however, are clear. The share of expenditure on child development will be larger the higher the perceived returns to investment, and the more utility parents derive from child human capital over private consumption. If wives have a stronger preference for child development over private consumption than husbands, the investment share should be increasing in the weight wives have in decision making. The opposite would hold true if husbands have a stronger preference for human capital than their wives.

5 New and old measures: using them jointly

In this section, we briefly describe a novel survey that was collected explicitly to pilot a number of new measures. In Section 6, we use (some of) these measures to get an empirical characterization of the model of parental investment we have sketched in Section 4. Here we provide a description of the different measures and how they were obtained. While the sample we use is relatively small, this data set is indicative of the possible use of a combination of new measures and choice-based data to estimate models of individual behavior. The description we provide in this section is also indicative of the difficulties that come with collecting new measures and the importance of a direct link between theory and the abstract constructs used in researchers' theoretical frameworks.

5.1 A pilot survey in Tanzania

The data we use in this and the next section were collected in two districts of the largely rural Kagera region in Tanzania. The overall goal of this data collection was to improve the measurement of child development and its drivers and design new measures. The survey included three samples: in the first, the respondents were mothers, in the second, fathers and, in the third, couples. The different samples are useful to characterize differences across mothers and fathers and how joint (couple) decisions are made. We describe the data in detail in Appendix A3, here we focus on the novel measures.

Bargaining Power. The power that women have within households has received much attention, especially when analyzing models of intrahousehold allocations. In our Tanzania sample, a first measure related to the bargaining power within the couple replicated the approach used by Almås et al. (2018), who conducted a controlled ("laboratory") experiment in a sample collected within an RCT in North Macedonia. The measure was designed to capture the potential impact of targeting women rather than men with a Conditional Cash Transfer, which was given to women in some villages and to men in others. To capture the *bargaining power* latent factor, after the initial data collection, the wives were called to an office to run an incentivized experiment. They were told: "Here are 100 Denars that we will give to your husband. How much are you willing to pay to have them paid to you?" This amount, which is completely independent of the government-administered cash transfer, was actually paid out (to the husband or the wife, depending on her choice).[16] The idea is that women willing to sacrifice a higher proportion of the amount offered are less powerful within the couple.

Almås et al. (2018) show that such a measure of bargaining power, which we label *Willingness To Pay* (WTP), correlates with a number of observables in a predictable manner. Moreover, in villages where the government grant was targeted at wives, the WTP decreased significantly; in these villages, women were willing to pay less to get control of any additional transfers.

The incentivized and the hypothetical exercises in Macedonia yield similar results, indicating that hypothetical formulations of these questions could also be used in sur-

[16] 100 Denars corresponds, for this sample, to two days of paid work. Based on the wife's answer, a second hypothetical question was asked considering much larger amounts.

veys. We use such a hypothetical version in the Tanzanian samples; the WTP question was asked to husbands in the *fathers sample* as well as mothers in the *mothers sample*. The resulting WTP is extremely skewed, with a few observations with very high values and, as a consequence, large difference between median and mean of the distribution (see Table A2 in Appendix A3.1 for the distribution of WTP).

We observe a considerable difference between husbands' and wives' WTP, with wives willing to sacrifice on average 32% of the transfer to get control over it. The median wife, both in the mothers and couples samples, is willing to pay just over 6% of the transfer. Husbands, on the other hand, are willing to pay only 10% on average, with the median being 0. The difference in the WTP is very marked and probably reflects different bargaining positions within the marriage, indicating that men have more control over resources than women.

Translating the WTP measure into a measure of bargaining power within the couple is not simple. To perform such an exercise, it would be necessary to identify a mapping from the theoretical constructs of the collective model, such as the Pareto weights, and individual preferences with the possibility of altruism and public goods or some of its intermediate outcomes, such as *sharing rules*, to the WTP measure.

Existing surveys provide alternative information on resource control within the household through a number of questions that explicitly ask who in the households is responsible for decisions on expenditure on various commodities and other decisions, with the possible answers being 'the husband' 'the wife' or 'both'. In the Tanzania survey we are using, six questions of this kind, taken from the Tanzania Demographic and Health Survey (DHS), were also posed to the respondents. The questions were about who is mainly responsible for decisions about major household expenditures, children's education, own and children health expenditures, what food to cook, and whether the wife can go out.

Using the answers to these questions we estimate a measurement system and, from that, a latent factor of interest, which should reflect women's decision power in the couple, as measured by the standard survey questions mentioned above. We estimate these measurement systems separately in the mothers and couples samples (where the wife is the respondent) and the fathers sample (where the husband is the respondent). The results of this exercise are in Table A3 in Appendix A3.1, which shows that for four of the questions the loading factors in the measurement systems estimated in the different samples are similar, the two exception being the questions about major household purchases and cooking, which seem to be more salient in the fathers sample. We notice that the variance of the estimated factor is much larger in the mothers and couple sample, which has also a much larger mean.

Comparing this factor to our WTP based measure (as done in the regression reported in Table A4 of Appendix A3.1) we find that, as one would expect, for mothers, the two variables are significantly and negatively related (the higher the share mothers are willing to forfeit to get control of the payment, the lower is mothers' decision making power within the household).[17] The R-squared of this regression, however, is very low at 0.03, indicating a considerable amount of variation in WTP not related to the traditional measures of control. In the fathers' sample the relation is negative but not significant.

Jayachandran et al. (2021) have used an approach similar to the one we have used to derive the WTP described above in India. They also use a machine learning algorithm to identify questions that accurately reflect women's decision power in the couple. They conclude that, in the Indian context, the latter approach seems to work better, which confirms our claim that some measures may work in some contexts but not other, and hence it is important to validate measures properly and in the relevant contexts before applying them.

Beliefs on returns to parental investment. Another important driver of individual choices that we consider in our application is parental perception about the process of child development. As is clear from the model presented in Section 4, parental investment is driven by parental perception of the return on child development. While much of the existing literature assumes that parents know the process of child development and how it depends on the child's current development, parental investment, and possibly other factors, it is increasingly clear that these perceptions might be distorted.

Several studies have started eliciting beliefs about the relationship between parental behavior and child development. Cunha et al. (2013), for instance, have elicited beliefs about the process of child development from disadvantaged mothers in Philadelphia;

[17] The couples sample shift is not significant.

other studies are listed in Appendix A3.2. Attanasio et al. (2019) develop the approach by Cunha et al. (2013) to measure mothers' beliefs about the process of child development within a survey of an RCT evaluating a parenting intervention in Colombia.

The strategy used by Attanasio et al. (2019) consists of presenting mothers with *scenarios* in terms of initial conditions and investment and asking them to map these scenarios into child development outcomes. The implicit assumption is that all mothers use the same mapping between latent factors and observable markers, so the scenarios proposed in the questionnaires have a relation to the latent factors that researchers want to capture. This approach allows researchers to estimate perceived rates of return to parental investments under different initial conditions.

In the Tanzania sample, we use this approach to elicit beliefs – for both fathers and mothers – about different aspects of the developmental process and the importance of certain parental inputs. We provide additional details about the elicitation strategy in Appendix A3.2; here it suffices to say that the returns we consider for high and low levels of initial development are measured as the difference in the expected outcomes between high and low levels of investment for the two levels of initial conditions.

Consistent with the evidence reported by Attanasio et al. (2019) in Colombia, the returns to parental investment are perceived to be higher for low than for high initial conditions in the whole sample and for mothers: the difference in means for the entire sample is equal to 0.140 (p-value=0.000) and to 0.200 (p-value=0.000) for mothers. For the fathers sample there is no significant difference between the returns with high or low initial conditions: the point estimate of the difference is 0.030 (p-value=0.570). Similar patterns are present for the other dimensions, such as socio-emotional development (see Appendix A3.2 for more details). In the same Appendix, in Table A8, we report the measurement system we estimate using the returns to parental investment on Language and Socio-emotional skills elicited in our survey.

Measuring preferences. As discussed above, an important determinant of individual choices is preferences. While, under a number of assumptions, the structural parameters that characterize them can be estimated from choice data, in some contexts, answers to hypothetical questions can be particularly useful. In a recent paper, Ameriks et al. (2020) use stated hypothetical choices elicited with Strategic Survey Questions

(SSQs) to estimate some preference parameters in a fully specified structural model. The authors note that to develop direct measures of preferences, it is desirable to develop survey instruments that allow respondents to provide information that identifies preference parameters in a language they are comfortable with as well as in a format that allows for a precise mapping to the structural parameters of interest.

In studying the model of parental investment in Section 4, in which preferences of different decision makers interact to determine household choices, it is important to elicit not only preferences on the allocations of scarce resources among different commodities but also between different individuals within the family. Cherchye et al. (2021) use a similar approach in Kenya to elicit tastes for couples' allocation between child investment and own consumption.

In the Tanzania sample, before the standard survey, a key respondent or respondents (the wife, the husband, or the couple jointly) were asked to allocate a hypothetical amount, represented by a pile of beans, between different expenditure categories and between three individuals (the husband, the wife, or the child in the household).[18] The six possible expenditure categories considered were: clothing, food, learning materials (such as books, notebooks, and pens), health expenditures, transportation, and school expenditures. These categories were chosen to be able to match the information collected on actual expenditure. While the question was not explicit about this issue, we interpret the answers to the hypothetical questions as referring to the allocation of additional resources that the household would normally have access to.

In Appendix A3.3 Table A6, we report the share of total additional resources allocated to each individual in the family (spouse, child, and self) for each of the samples (couples, mothers, and fathers). A number of interesting results emerge from this exercise. First, both mothers and fathers allocate some resources to their spouse, indicating that the participants care about their spouse's consumption. Second, mothers allocate more than fathers to children, but allocate the same share to themselves, which implies mothers allocate less to their spouses than fathers. These differences, although

[18]The question was posed as: "We would now like to understand how you would prefer to spend 300k Shillings, if we were to give this money to you. Use these 60 beans, each representing 5k Shillings, and this cardboard card with 3 different expenditure options (mother, father, and your child); for each question distribute the beans according to your preferences. Imagine that your child is 5 for this exercise." See Almås et al. (2020a) for a full description of the protocol followed.

relatively small, are statistically significant. Third, the couple's decisions seem much closer to those of fathers than those of mothers. The similarity between the fathers and the couple allocations might indicate large differences in decision making power between spouses as fathers, consistent with the evidence on the WTP, hold considerably more decision power within the household.

Next, we look at the allocation among different commodities, and in particular, for the resources allocated to the child. There are some differences between mothers and the other samples, particularly in clothing and health (where the mothers' shares are significantly higher) and in learning materials (where the fathers' share is marginally higher), see Appendix A3.3 Table A7 for details. Once again, the couple's decisions are more similar to those of fathers than those of mothers.

In the model we discussed in Section 4, it was clear that the relative parental taste for child human capital and alternative allocations of resources is a key determinant of parental investment. While we do not estimate a structural version of that model, as it would be implied, for instance, by the AIDS version of Engel curves in equation (12), we use the answers about stated preferences in the couple samples to derive some information about individual couples' tastes and relate them to parental investment. As is evident from equation (10), which was derived under homothetic preferences, parental investment can depend on the shares of resources allocated to adult goods relative to the resources allocated to child consumption and parental investment.

To estimate a factor representing *the taste for child human capital*, we use the answers to the allocation questions to construct eight variables as the ratios of the resources allocated to each spouse for four adult commodities (food, clothing, health, and transportation) to the resources allocated to the child and estimate a factor model to extract a latent factor we label *relative taste for child human capital* from these eight variables. We perform this analysis in each of the samples and report the loading factors for each of the variables, as well as the intercept for the equations of the measurement system corresponding to each observable variable, in Appendix A3.3 Table A8.

The factor estimated from this analysis is what we use in Section 6 to model parental investment. From Table A8 two facts are immediately apparent. First, several variables seem to be important markers of the taste for child development. We find important differences in the three samples, with the preferences in the couples sample seeming more similar to the fathers' rather than the mothers'. In the latter, measures of the mother's and father's health expenditures and father's food expenditures are not particularly important.

Having presented some evidence on the new measures that were collected in Tanzania, one important issue and challenge is their validation. In particular, we check if these measures co-vary in a sensible way with choice data or, more generally, with standard measures. This would be a first step towards a systematic use of these measures within models of individual behavior. We turn to this in the next section.

5.2 Using available measures efficiently

Even when the latent factors that one wants to analyze are reasonably standard, it may not be easy to obtain meaningful measures from the available raw data that can be used in empirical analysis. In many contexts, certain algorithms that aggregate available measures become the accepted standard and are widely used, even when alternative approaches may be more efficient and meaningful. As discussed in Section 3.3, the explicit specification of a measurement system can be interpreted as the construction of a scoring mechanism that efficiently uses the available data to estimate the latent factors of interest. The study of child development and its drivers, such as parental investment or school quality, provides a particularly salient example that has received substantial attention in recent years.

It is recognized that measuring child development is difficult, especially in the early years and when one wants to assess different dimensions of development, including socio-emotional skills. Analogous considerations apply to measures of the drivers of child development, such as parental investment and school quality. These difficulties are even more serious in developing country contexts, both because the administration of some of the most frequently used tests is often difficult and requires specialized testers and because many of these tests were developed and validated in what has been termed Western, Educated, Industrialised, Rich, and Democratic (WEIRD) samples (Henrich et al., 2010).[19] While a number of sophisticated tests have been developed and validated in developed countries, the relevance and effectiveness of such tests in completely dif-

[19]For concepts such as parental investment and school quality, the application of tests developed in different contexts can easily result in flooring and ceiling effects.

ferent contexts might be limited.

Because of these considerations, a number of efforts to develop a new generation of child development tests are under development, which are listed in Appendix A1. Within the Tanzania project we have been discussing, a new test was also developed, described in Attanasio et al. (2023), which combined elements from different well-established tests (such as the Bayley Scales of Infant and Toddler Development, Third Edition - Bayley-III, the Caregiver Reported Early Development Instruments (CREDI), and others) to construct an efficient and easy to administer test with a limited number of items. The approach consisted of estimating a measurement system such as in equations (2) and (3), relating the dimensions of interest to the various elements that make up these tests. This procedure, now widely used (see, for instance, Cunha et al. (2010), Agostinelli and Wiswall (2017), Heckman et al. (2020) among others) can be seen as an effective alternative to the use of standard algorithms that typically come with these tests. The construction of a new scoring algorithm through the estimation of a measurement system obtained in a given context is an effective way to adapt the existing tests to new realities and make different contexts comparable.

A similar set of issues is relevant for measures of latent factors that are key drivers of child development, such as parental investment. It is not always clear how many dimensions of investment to consider and how to adapt available measures to different contexts.[20] A number of standardized measures exist and have been widely used, such as the Home Observation Measurement of the Environment (HOME) index or the Family Care Indicators (FCI). However, given the set of items that make up these tests, it is not clear that the same scoring algorithm should be used in different contexts, as different items might be differently salient and relevant depending on the context.

In the Tanzania context considered here, we follow an approach to measuring child development similar to that taken by Attanasio et al. (2023) and estimate a factor model that identifies a single latent factor. We use this factor as our measure of parental investment in Section 6. In particular, we use a number of items as markers of parental investment, including: (i) time spent in activities with children; (ii) play materials present in the house; (iii) material investments for the child (including food, clothing, footwear,

[20]Researchers use different strategies to measure parental investment. Attanasio et al. (2020b), for instance, consider time and material investment separately and show that they might have different impacts on child development. Others, such as Cunha et al. (2010), consider a single dimension.

confectioneries, among others); (iv) share of expenditure on children items over the total expenditures of the household; (v) items from the *social scale* from the Parental Style Questionnaire (PSQ; Bornstein (1996)); and (vi) items from the *didactic scale* from the PSQ. In Appendix A3.4 we define (i)-(vi).

In Appendix A3.4 Table A10, we report some of the estimates of the parameters of equations (3) for the sample of couples. We notice that several markers are relevant to the factor we are considering. Different measures of parenting skills, for instance, do not seem particularly relevant for the parenting investment factor, while material investment plays an important role.

6 A model of parental investment

In the model we presented in Section 4, the Engel curve for parental investment can be seen as an approximation similar to the one used by Browning and Chiappori (1998). In this section, we estimate a version of equation (12) for parental investment using both choice data on investment and total expenditure, as well as the measures of preferences, beliefs, and bargaining power that we collected and discussed in Section 5. Rather than estimating a detailed version of the model that incorporates tastes for a variety of different commodities, we focus on parental investment and summarize the information on couples' preferences for child development relative to other uses of resources. We want to investigate how our novel measures co-vary with actual behavior in terms of parental investment in a way which is consistent with the model we presented.

We presume that in the sample where the non-conventional questions are directed to the couple, the answers reflect the 'aggregated' couple preferences and beliefs about child development. Therefore, for this exercise, we use only the couple samples, where questions are directed to the couple, as it would be hard to model observed investment by the couple on the basis of the preferences and beliefs of only one partner.

We estimate different measurement systems to extract from the available measures information about factors that enter the model in equation (12). In Section 5, we discuss the systems we use to extract the latent factors representing couples' relative tastes for child development, their beliefs about the productivity of parental investment, and bargaining power within the couple as well as actual parental investment. The estimates of these measurement systems are reported in Appendix A3.

In equation (12), the share of parental investment in total expenditure depends on (the log of) total expenditure and the latent factors representing both spouses' preferences and beliefs about the productivity of parental investment, as well as a bargaining power factor, which aggregates the spouses' preferences and beliefs. As we interpret the preferences and beliefs factors elicited in the couples sample as reflecting the *couples' preferences and beliefs*, one can modify equation (12) as:

$$s_i = G(\tau_i(\mu_i), \gamma_i(\mu_i), \mathbf{p}) + \beta(\ln Y_i - a(\mathbf{p})) + u_i \tag{13}$$

where τ_i and γ_i are now unidimensional factors that aggregate (through the bargaining power factor μ_i) the preferences and beliefs about the effectives of parental investment of the two spouses.[21] If we interpret the responses from the couples sample as reflecting these aggregate factors, the bargaining power should not enter the parental investment equation once we control for the aggregated factors. However, it is possible that the linear specification that approximates equation (12) is too restrictive so that the bargaining power measure could enter it significantly.

In Table 1, we report the results of regressing the parental investment factor, as estimated via the measurement system discussed in Appendix A3 on: log total expenditure, the beliefs and taste factors (and their interactions), and our measure of bargaining power. As parental investment is a factor with an arbitrary scale, the size of the coefficients on total expenditure in Table 1 is difficult to interpret. However, we notice that (log) total expenditure attracts a positive and significant coefficient in all specifications.

In column 1, we only use 'standard' data, in that we regress the investment factor (which is determined by investment in time spent with children and expenditure in commodities targeted to children) on the log total expenditure and a set of observable child and mother specific control variables. In column 2, we introduce our estimates of taste and beliefs questions to capture the determinants of parental investment as the arguments of function $G()$ in equation (13). We observe that the relative taste for commodities other than children human capital is strongly significant and with the expected negative sign. This result can be partly interpreted as a validation of the measure of taste

[21] We note that while the preferences and beliefs factors are derived from a number of measurements, for bargaining power we have a single measure (that is likely to be affected by measurement error).

Table 1: Modeling parental investment: Couples sample

	(1) Parental inv.	(2) Parental inv.	(3) Parental inv.	(4) Parental inv.	(5) Parental inv.
Log of total expenditure	0.082*	0.093**	0.140***	0.144***	0.140***
	(0.047)	*(0.045)*	*(0.044)*	*(0.044)*	*(0.044)*
Relative taste for child human capital (RC)		-0.967***	-1.159***	-1.694***	-1.170*
		(0.305)	*(0.294)*	*(0.539)*	*(0.693)*
Beliefs		0.369*	0.394*	0.093	0.394*
		(0.197)	*(0.199)*	*(0.323)*	*(0.200)*
Bargaining Power (BP)			-0.237**	-0.248**	-0.239
			(0.104)	*(0.104)*	*(0.173)*
Beliefs*RC				2.155	
				(1.823)	
BP*RC					0.014
					(0.802)
R-squared	0.152	0.248	0.378	0.385	0.378
Observations	142	142	126	126	126

Note: The table displays a regression of the parental investments factor (estimated with the measurement system discussed in Appendix A3) on: i) Log of total expenditure from the household survey; ii) Bargaining power is a measure of female bargaining power, and more precisely, it is 1 minus the share that the woman is willing to pay to gain control over a fixed amount of money; iii) Relative taste for child human capital (a factor of ratios of the resources allocated to each spouse for four adult commodities to the resources allocated to the child in our allocation experiment); and iv) Beliefs is a factor of the returns to parental investment on Language and Socio-emotional skills elicited in our survey. All regressions include child and mothers characteristics: child's age and gender, number of siblings in the household, dummy for mother's secondary education and cognition (as measured by the Raven's test). Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.Source: Tz Pilot.

that we use, which is derived from an experiment on hypothetical allocations; where the specific experiment does not refer to actual investment choices at all. Analogously, the factor measuring beliefs about the productivity of investment is also significant (albeit only at the 10% level) with the expected positive sign in the equation for parental investment.

In column 3, we add to the variables in column 2 our measure of the wife's bargaining power within the couple. As mentioned above, if we interpret the answers given by the couple's questions on their tastes and beliefs as representing the aggregation of the individual tastes and beliefs, (and if the equation is linear), such a variable should not be a significant determinant of investment. We find that, while the other coefficients do not change much with the introduction of this variable takes a negative coefficient which is strongly significantly different from zero. Without a further detailed analysis of the various factors at play, it is difficult to interpret this coefficient. A possible avenue for future research is to map this variable to the parameters of a model of intrahousehold decision making, such as the collective model introduced by Chiappori (1988) and collaborators.

As the significance of the bargaining power measure might be signal the presence of nonlinearities in the function $G()$ in equation (13), in columns 4 and 5 of Table 1, we introduce interactions of the taste factor with beliefs (in column 4) and with bargaining power (in column 5). In neither of the two columns do we find any significant interactions between taste and the two variables considered.

This exercise is a first step in utilizing novel measures in a unifying framework that combines elicited beliefs, preferences, and decision making power with observational data. The next step of this research, as alluded above, might be the use of these measures into a structural model of parental behavior and map them more directly onto its parameters. To achieve such a goal, it may be necessary to develop finer measurement tools than those used here. Such a model could quantify the importance of different determinants of parental investment (parental beliefs, parental resources or preferences) and child development. This conceptual framework could then guide policy makers in the design of effective policies by explicitly identifying how certain interventions obtain the observed effects and what are the mediating factors that such policies should target it.

7 Conclusions

In this paper, we have analyzed the role that measurement does and should play in economics and its relation to economic theory. We argued that measurement issues – including what should be measured, how to construct effective measures of the latent factors that populate economic models, and how to use such measures – should be informed by economic theory. Economic models are attempts to describe certain aspects of human behavior in specific contexts in a coherent fashion that allows generalizations, extrapolation, and ultimately the identification of causal links between different variables. Depending on what is being modeled or 'explained,' bringing the relevant latent factors to data might require the measurement of different variables.

Academic economists have, for a long time, shied away from measuring certain variables. With some exceptions, economists have relied almost exclusively on data on choices, prices, and resources, or, more generally, objectively observable variables.

This approach refrains from using measures of attitudes, intentions, stated preferences, beliefs, subjective expectations, social norms, etc. There are good reasons to treat such measures with caution and even skepticism. These measures might be difficult to collect and can be affected by different types of bias. However, empirical work that relies exclusively on choice data, and supposedly objective measures, imposes strong restrictions to the economic theories and models that can be brought to data, which typically take the form of strong assumptions on the structure of the models one works with. There are many important models, some of which we discussed, that could be analyzed with much more substance and empirical bite were they to use a wider set of measures. It is interesting to note that such measures, while not widely used by economists, have been extensively used in other disciplines, from marketing to psychology and child development.

Obviously, given the difficulties in collecting the innovative measures we are advocating, they should be validated properly. Moreover, as we have argued, these measures are particularly useful when utilized in combination with choice data and other standard measures. This combination could be used both to validate the new measures, possibly via simple correlations and to estimate and test richer models.

In the second part of the paper, we have put in practice some of these ideas, using a set of new measures collected in rural Tanzania to estimate a model of parental investment, which we relate to measures of parental preferences, elicited from stated preferences, bargaining power within couples, and parental beliefs about the process of child development. While we do not estimate a fully structural model, we show how these data can be used in combination with standard measures to quantify the importance of different factors affecting parental behavior. In a next step, these data can be used to identify important causal links. Eliciting from respondents' information about choice under different counterfactual and hypothetical scenarios to gather information about preferences, one can, in some situations, solve *through measurement* some of the endogeneity and identification issues that make empirical work challenging.

Every day, new data are being created and used, most noticeably administrative data from a variety of sources and contexts. This is obviously a positive development. However, we believe that well-designed and innovative survey measures of variables and constructs that are theoretically relevant can be just as useful. Indeed, many new measures are being developed in this direction, including those we have cited and others listed in Appendix A1. Future research should devote substantial efforts to develop, design, implement, and validate new measurement tools that can provide useful evidence for economic theory and, ultimately, public policy.

References

Acemoglu, D. and D. H. Autor (2011). "Skills, tasks and technologies: Implications for employment and earnings." In "Handbook of Labor Economics Volume 4," (edited by Ashenfelter, O. and D. E. Card). Amsterdam: Elsevier.

Agostinelli, F. and M. Wiswall (2017). "Identification of dynamic latent factor models: The implications of re-normalization in a model of child development." Tech. rep., NBER, WP No 22441.

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

Almås, I., A. Armand, O. Attanasio, and P. Carneiro (2018). "Measuring and changing control: Women's empowerment and targeted transfers." *The Economic Journal* **128**(612), F609–F639.

Almås, I., O. Attanasio, and P. Jervis (2023). "Economics and measurement: New measures to model decision making." WP 30839, NBER.

Almås, I., O. Attanasio, P. Jervis, and C. Ringdal (2020a). "Targeted cash transfers and children's welfare: Should women be targeted?" Tech. rep., NHH working paper.

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

Ameriks, J., J. Briggs, A. Caplin, M. D. Shapiro, and C. Tonetti (2020). "Long-term-care utility and late-in-life saving." *Journal of Political Economy* **128**(6), 2375–2451.

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

Attanasio, O., A. Barr, J. C. Cardenas, G. Genicot, and C. Meghir (2012). "Risk pooling, risk preferences, and social networks." *American Economic Journal: Applied Economics* **4**(2), 134–67.

Attanasio, O., R. Bernal, M. Giannola, and M. Nores (2020a). "Child development in the early years." Tech. rep., NBER, WP No 27812.

Attanasio, O., S. Cattan, E. Fitzsimons, C. Meghir, and M. Rubio-Codina (2020b). "Estimating the production function for human capital." *American Economic Review* **110**(1), 48–85.

Attanasio, O., F. Cunha, and P. Jervis (2019). "Subjective parental beliefs. their measurement and role." Tech. rep., National Bureau of Economic Research.

Attanasio, O. and S. Krutikova (2020). "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., C. Meghir, and E. Nix (2020c). "Human capital development and parental investment in india." *The Review of Economic Studies* **87**(6), 2511–2541.

Attanasio, O. P., M. Bornstein, and P. Jervis (2023). "Constructing effective measures of child development: ages 6 to 36 months." Tech. rep., In progress.

Autor, D. H. and D. Dorn (2013). "The growth of low-skill service jobs and the polarization of the us labor market." *American Economic Review* **103**(5), 1553–97.

Banks, J., R. Blundell, and A. Lewbel (1997). "Quadratic engel curves and consumer demand." *Review of Economics and statistics* **79**(4), 527–539.

Benhabib, J. and A. Bisin (2011). "Choice and process: Theory ahead of measurement." In "*The Foundations of Positive and Normative Economics: A Hand Book*," (edited by Caplin, A. and A. Schott). Oxford University Press.

Bernheim, B. D., D. Bjorkegren, J. Naecker, and M. Pollmann (2021). "Causal inference from hypothetical evaluations." Tech. rep., NBER, WP No 29616.

Berry, S., J. Levinsohn, and A. Pakes (2004). "Differentiated products demand systems from a combination of micro and macro data: The new car market." *Journal of Political Economy* **112**(1), 68–105.

Bisin, A. and T. Verdier (2000). "Beyond the melting pot: cultural transmission, marriage, and the evolution of ethnic and religious traits." *Quarterly Journal of Economics* **115**(3), 955–988.

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

Bloom, N. and J. Van Reenen (2007). "Measuring and Explaining Management Practices Across Firms and Countries*." *Quarterly Journal of Economics* **122**(4), 1351–1408.

Bornstein, M. H. (1996). "Ideas about parenting in argentina, france, and the united states." *International Journal of Behavioral Development* **19**(2), 347–368.

Browning, M. and P. A. Chiappori (1998). "Efficient intra-household allocations: A general characterization and empirical tests." *Econometrica* **66**(6), 1241–1278.

Browning, M., P.-A. Chiappori, and A. Lewbel (2013). "Estimating consumption economies of scale, adult equivalence scales and household bargaining power." *Review of Economic Studies* **80**, 1267–1303.

Buser, T., M. Niederle, and H. Oosterbeek (2014). "Gender, competitiveness, and career choices." *The quarterly journal of economics* **129**(3), 1409–1447.

Camerer, C. (2011). "The case for mindful economics." In "*The Foundations of Positive and Normative Economics*," (edited by Caplin, A. and A. Schott). Oxford Univ. Press.

Caplin, A. (2021). "Economic data engineering." Working Paper 29378, National Bureau of Economic Research.

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

Cherchye, L., P.-A. Chiappori, B. De Rock, C. Ringdal, and F. Vermeulen (2021). "Feed the children." Tech. rep., CEPR Discussion Paper No. DP16482.

Chiappori, P. A. (1988). "Rational household labor supply." *Econometrica* **56**(1), 63–90.

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

Cunha, F. and J. J. Heckman (2008). "Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation." *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.

Danz, D., L. Vesterlund, and A. J. Wilson (2022). "Belief elicitation and behavioral incentive compatibility." *American Economic Review* **112**(9), 2851–83.

Dardanoni, V., P. Manzini, M. Mariotti, H. Petri, and C. Tyson (2022). "Mixture choice data: revealing preferences and cognition." *Journal of Political Economy* ***forth***.

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

Delavande, A. and S. Rohwedder (2008). "Eliciting Subjective Probabilities in Internet Surveys." *Public Opinion Quarterly* **72**(5), 866–891.

Dercon, S. and P. Krishnan (2000). "In sickness and in health: Risk sharing within households in rural ethiopia." *Journal of Political Economy* **108**(4), 688–727.

Field, E., R. Pande, N. Rigol, S. Schaner, and C. Troyer Moore (2021). "On her own account: How strengthening women's financial control impacts labor supply and gender norms." *American Economic Review* **111**(7), 2342–75.

Goldberger, A. S. (1972). "Structural equation methods in the social sciences." *Econometrica* **40**(6), 979–1001.

Griliches, Z. (1974). "Errors in variables and other unobservables." *Econometrica:* **42**(6), 971–998.

Guiso, L., P. Sapienza, and L. Zingales (2004). "The role of social capital in financial development." *American Economic Review* **94**(3), 526–556.

Guiso, L., P. Sapienza, and L. Zingales (2006). "Does culture affect economic outcomes?" *Journal of Economic Perspectives* **20**(2), 23–48.

Gul, F. and W. Pesendorfer (2011). "The case for mindless economics." In "*The Foundations of Positive and Normative Economics*," (edited by Caplin, A. and A. Schott). Oxford Univ. Press.

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

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

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

Heckman, J., R. Pinto, and P. Savelyev (2013). "Understanding the mechanisms through which an influential early childhood program boosted adult outcomes." *American Economic Review* **103**(6), 2052–86.

Heckman, J. J., B. Liu, M. Lu, and J. Zhou (2020). "Treatment effects and the measurement of skills in a prototypical home visiting program." Tech. rep., NBER.

Henrich, J., S. J. Heine, and A. Norenzayan (2010). "The weirdest people in the world?" *Behavioral and Brain Sciences* **33**(2-3), 61–135.

Jayachandran, S., M. Biradavolou, and J. Cooper (2021). "Using machine learning and qualitative interviews to design a five-question survey module for women's agency." Tech. rep., Northwestern University Working Paper.

Juster, F., H. Cao, M. Perry, and M. Couper (2006). "The effect of unfolding brackets on the quality of wealth data in hrs." *SSRN Electronic Journal* .

Juster, F. T. (1964). *Anticipations and Purchases: An Analysis of Consumer Behavior*. Princeton University Press.

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

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.

Katz, L. F. and K. M. Murphy (1992). "Changes in Relative Wages, 1963–1987: Supply and Demand Factors*." *The Quarterly Journal of Economics* **107**(1), 35–78.

Lancaster, T. and A. Chesher (1983). "An econometric analysis of reservation wages." *Econometrica* **51**(6), 1661–76.

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

Manski, C. F. (1990). "The use of intentions data to predict behavior: A best-case analysis." *Journal of the American Statistical Association* **85**(412), 934–940.

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

Miller, S. A. (1988). "Parents' beliefs about children's cognitive development." *Child Development* **59**(2), 259–285.

Mueller, A. and J. Spinnewijn (2021). "Expectations data, labor market and job search." *Handbook Chapter (Draft)* .

Olley, G. S. and A. Pakes (1992). "The dynamics of productivity in the telecommunications equipment industry." Tech. rep., National Bureau of Economic Research.

Parnes, H. S. (1975). "The national longitudinal surveys: New vistas for labor market research." *The American Economic Review* **65**(2), 244–249.

Plott, C. R. and V. L. Smith (2008). *Handbook of experimental economics results*, vol. 1. Elsevier.

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.

Stantcheva, S. (2022). "How to run surveys: A guide to creating your own identifying variation and revealing the invisible." Working Paper 30527, NBER.

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

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

Todd, P. E. and K. I. Wolpin (2003). "On the specification and estimation of the production function for cognitive achievement." *The Economic Journal* **113**(485), F3–F33.

Varian, H. R. (2014). *Intermediate microeconomics with calculus: a modern approach*. WW norton & company.

Wolpin, K. I. and F. Gonul (1985). "On the use of expectations data in micro surveys: The case of retirement." Tech. rep., Ohio State University.