STAFF
REPORTS

NO. 1018
MAY 2022

# Expectations Data in Structural Microeconomic Models

Gizem Koşar | Cormac O'Dea

FEDERAL RESERVE BANK of NEW YORK

**Expectations Data in Structural Microeconomic Models**
Gizem Koşar and Cormac O'Dea
*Federal Reserve Bank of New York Staff Reports*, no. 1018
May 2022
JEL classification: D84, D15, C51, C13, D83

**Abstract**

A growing literature uses now widely available data on beliefs and expectations in the estimation of structural models. In this chapter, we review this literature, with an emphasis on models of individual and household behavior. We first show how expectations data have been used to relax strong assumptions about beliefs and outline how they can be used in estimation to substitute for, or as a complement to, data on choices. Next, we discuss the literature that uses different types of expectations data in the estimation of structural models. We conclude by noting directions for future research.

Key words: expectations data; beliefs, household surveys, structural models, hypotheticals, choice expectations, stated-preference data.

---

Koşar: Federal Reserve Bank of New York, CESifo (email: gizem.kosar@gmail.com). O'Dea: Yale University, IFS, NBER (email: cormac.odea@yale.edu). This chapter has been prepared for the *Handbook of Economic Expectations*, edited by Rüdiger Bachmann, Giorgio Topa, and Wilbert van der Klaauw. The authors thank Rory McGee, Moritz Mendel, Dana Scott, Lindsey Uniat, Basit Zafar, and the editors for helpful comments. Boryana Ilieva provided excellent research assistance.

# 1 Introduction

Across a wide range of applied research areas in economics, structural models are used both to understand the decision-making of economic agents and to evaluate the effects of counterfactual policies. In almost all cases, these models have been estimated using data on the choices agents make. A burgeoning literature, however, uses data on expectations instead of, or in addition to, data on observed choices. Expectations data can refer to data on how an economic agent believes some uncertain feature of reality will evolve or what choices the agents predict they will make in the future. This chapter discusses that literature.

There is no single accepted definition of a 'structural' model.[1] The scope of this chapter is to discuss the literature in which: i) the decision problem of the economic agent (usually an individual or household) is specified, ii) that model is explicitly estimated, and iii) that estimation involves the use of *micro* data on expectations. Our focus in this chapter is on structural models of individual and household decisions, with an emphasis on material not covered extensively elsewhere in this volume. For the literature on firms see the recent review articles by Aguirregabiria and Jeon (2020), Bachmann and Carstensen (2022), Candia et al. (2022), and Born et al. (2022). For the literature on financial market participants and investment and portfolio decisions see especially Baumeister (2022) and Diercks and Jendoubi (2022) in this volume.

We start our discussion by outlining a simple model in Section 2. This model, a stylized version of Van der Klaauw and Wolpin (2008), is of labor supply and savings decisions over the life-cycle. The purpose of the exposition of this model is threefold. First, it allows us to introduce the types of expectations data that have been used in the estimation of structural models. Second, it is used to illustrate *how* one can use expectations data in estimation and to highlight estimation issues particular to these data. Third, it allows us to motivate *why* expectations data can be valuable for identification and estimation of structural models.

The subsequent two sections then summarize the literature that uses expectations data in the estimation of structural models. We divide the literature according to a taxonomy provided by Manski (2004). In Section 3 we focus on papers in which data on expectations over future *states of nature* is used. These states of nature can be individual circumstances (e.g. future survival), economic conditions (e.g. earnings) or an aspect of policy (e.g. Social Security rules). In Section 4 we turn to the literature in which data on expectations over *choices* are used to estimate structural models. Choice expectations here encompass both statements of what choices agents expect to make in the future and choices that agents *would* make if faced with circumstances that are specified by the survey instrument. Section 5 concludes.

[1] See Haile (2020) for a recent treatment which clarifies the use of various descriptions of empirical work.

# 2 A model

To illustrate how and why one would use expectations data in the estimation of the parameters of a structural model, we first outline a simple life-cycle model. It is a simplified version of that in Van der Klaauw and Wolpin (2008) and considers the consumption and labor supply choices of individuals who face uncertainty over the future in three dimensions: a demographic characteristic (their longevity), a feature of the economic environment (their earnings), and an aspect of economic policy (Social Security rules). We do not impose here that individuals have rational expectations – they may have expectations over the future distribution of those uncertain objects that will not coincide with the ex-post distribution in the population.

## 2.1 Specification of a model

**Demographics** Agents live for up to $T$ periods. Time is discrete. Conditional on being alive in period $t$, each individual $i$ has a probability $s_{i,t+1}$ of surviving to period $t+1$.

**Choices** Each agent $i$ makes an extensive margin labor supply choice ($p_{it}$) each period until some retirement age $T^R < T$ and also decides how much of her resources to consume ($c_{it}$) each period. From age $T^R$, agents are retired and only make consumption decisions.

**Economic Environment** Wage offers ($y_{it}$) are assumed to be stochastic and are drawn from a distribution with mean $\mu$ and standard deviation $\sigma$. Wealth accrues interest at rate $r$. Agents enter each period $t$ with a stock of assets, $a_{it}$.

**Policy** Starting from the retirement age ($T^R$), all agents receive Social Security payments. Following Van der Klaauw and Wolpin (2008), Social Security payments are the product of two terms:

$$ss_i = \kappa f(AIME_i). \tag{1}$$

The function $f(AIME)$ represents current Social Security rules, which depend on 'average indexed monthly earnings' ($AIME$), a function of the agent's earnings history. $\kappa$ is a parameter which allows for the fact that Social Security rules may be different from those currently prevailing when the agents reach their claiming age. $\kappa$, a feature of future government policy, is not known; agents must form beliefs over it.

**Preferences** Agents discount the future geometrically at rate $\beta$ and have a utility function defined over consumption and leisure:

$$U(c_{it}, l_{it}) = \frac{(c_{it}^{\nu} l_{it}^{1-\nu})^{1-\gamma}}{1-\gamma}, \tag{2}$$

where $l_{it}$ is leisure, which takes a value $l_{it} = 1 - h\mathbb{1}[p_{it} = 1]$, where the endowment of leisure is normalized to 1 and $h$ is the fixed share of leisure that the agent forgoes in those periods in which she works. $\gamma$ is the coefficient of relative risk aversion and $\nu$ governs the relative importance to the agent, of consumption over leisure. Agents do not value wealth at death; so there is no bequest motive.

**Expectations** We must make an assumption about agents' knowledge of the model's environment and their expectations about the future. The most common assumption in the literature is that agents have full information over the deterministic features of the problem and that they have rational expectations over all stochastic features. The latter means that the agents' subjective belief distribution of a given stochastic feature will be equivalent to the 'objective' (according to the model) distribution of that feature.

The three stochastic model features here span three distinct types of "states of nature" that are relevant for decision-making under uncertainty. They are i) uncertain future personal or demographic states (here, survival), ii) uncertain features of the economic environment (here, wage offers) and iii) uncertain future policy features (here, Social Security). As we will discuss below, papers have used data on expectations in each of these domains to relax the assumption of rational expectations. To make clear that there can be a divergence between the objective distribution of some stochastic feature and agents' subjective beliefs, we will denote as $\bar{Z}_i$ the agent's subjective belief over any model object $Z$ and $\bar{\mathbb{E}}_i$ as the expectations operator with respect to the subjective belief distribution.

While we have outlined a model in which agents' expectations over model objects can differ from the objective distribution, we have not specified how those expectations are formed. This restriction is consequential: one of the advantages of explicitly modeling behavior is to be able to evaluate how those will evolve in counterfactual settings. Unless the expectation formation process is modeled, expectations will be assumed to stay unchanged in any counterfactual experiment. We will return to this issue in our discussion of the literature and in conclusion as a profitable direction for future research.

**Recursive Specification** The agent's decision problem in period $t$ (we suppress $i$ subscripts) can be expressed recursively through the value function:

$$V_t(\chi_t) = \max_{\{c_t, p_t\}} U(c_t, l_t) + \beta \bar{s}_{t+1} \bar{\mathbb{E}}_t V_{t+1}(\chi_{t+1}) \tag{3}$$

$$s.t. \begin{cases} a_{t+1} = (a_t + y_t \mathbb{1}[p_t = 1] - c_t)(1+r) & \text{if } t < T^R \\ a_{t+1} = (a_t + ss - c_t)(1+r) & \text{if } t \geq T^R \\ l_t = 1 - h\mathbb{1}[p_t = 1], \end{cases} \tag{4}$$

where $\chi_t = \{a_t, AIME_t\}$ collects the model's two state variables (assets and average indexed monthly earnings[2]), $\bar{s}_{t+1}$ represents the agents' perception of their survival probability, and $\bar{\mathbb{E}}$ indicates that the expectation is taken with respect to agents' subjective beliefs over other stochastic features.

**Parameters** A set of model parameters is likely to be unknown to the researcher. This will include preference parameters (e.g. $(\beta, \gamma, \nu)$) and potentially features related to expectations (e.g. the belief distribution over $\kappa$ and the earnings distribution). $\theta$ collects these unknown model features.

**Model Solution** If $\theta$ were known, the model, once fully specified, could be solved using standard methods (see, for example, Adda and Cooper (2003)). The solution would imply decision rules (or policy functions) which relate the state variables to optimal consumption and labor supply decisions. Let us denote the consumption and labor supply decision rules at time $t$ by

$$\tilde{c}(\chi_t, \theta), \quad \tilde{p}_t(\chi_t, \theta), \tag{5}$$

where we make it explicit that these depend on the model parameters ($\theta$) and the state variables ($\chi_t$). Let us further note that, just as the solution of the model implies trajectories of *choices* for agents, it also implies trajectories of *expectations of future choices* for those agents. A model parameterized by $\theta$ implies, for example, an expected value for any feature of behavior at all future ages $t+\tau$, or – to take a concept of expectations often measured in survey data – the probability of any single discrete outcome at age $t+\tau$. Examples of the latter that we highlight below include the probability of working at a particular age in the future or of having assets above a particular level at retirement, denoted by:

$$\mathbb{P}[\tilde{p}_{t+\tau} = 1 | t; \chi_t, \theta], \quad \mathbb{P}[\tilde{a}_{t+\tau+1} > \bar{a} | t; \chi_t, \theta] \tag{6}$$

[2]This is a function of lifetime earnings, on which Social Security payments in the U.S. are based. We do not include the law of motion here for it.

.

Before we turn to how expectations data are used in estimation, it is worth dwelling on the types of expectations data that are typically available.

## 2.2 Types of Expectations Data

A useful taxonomy of expectations data comes from Manski's (2004) seminal article on the measurement of expectations. Two distinct types are data on expectations over states of nature and data on expectations over choices. We discuss both in turn, giving some examples from the Health and Retirement Study (HRS).

1. **Expectations over states of nature** relate to features of the economic and policy environment or of personal characteristics. These measured expectations often relate to beliefs over *future* realizations of stochastic events, though surveys also measure the extent of respondents' (potentially imperfect) understanding of some feature of the current environment. In the case of the model we have outlined, each possible combination of survival, earnings, and future social security rules comprise the states of nature.

   The HRS, for example, has asked respondents the following:

   - What is the percent chance that you will live to be 75 or more?
   - About how much do you expect the [future Social Security] payments to be in today's dollars?

2. **Expectations over choices** concern decisions individuals anticipate that they will make in the future, or that they would make under specified circumstances. In the case of the model we have outlined, these decisions are labor supply and consumption.

   The HRS asks individuals about their expectations of making certain decisions in the future:

   - Thinking about work in general and not just your present job, what do you think the chances are that you will be working full-time after you reach age 62?
   - Including property and other valuables that you might own, what are the chances that you (and your [husband/wife/partner]) will leave an inheritance totaling $10,000 or more?

This distinction between two types of expectations data can be used to characterize, in a straightforward manner, any feature in our simple model. In richer models, there may be features which straddle both groups – it could be that there are *choices* which are only possible if a particular *state of nature* arises. For example, in a frictional labor market where agents may not receive a job offer every period, the probability of working in a future period depends both on the realization of a state of nature (whether a job offer arrives) and on the choice the individual makes (whether the individual accepts the offer). Conversely, there could be *states of nature* which depend on *choices*. For example, in a model in which individuals invest in health, their survival risk could depend on their investment decisions. Expectations data on either states of nature or choices in each of these cases will contain information on the subjective expectations about the joint likelihood of the state of nature arising and the choice the individual will make. A further distinction worth drawing is that some data on expectations may encompass both groups by asking for expectations of a future *state of nature* conditional on a *choice* being made (for example, future earnings conditional on college choice). Therefore, the distinction between expectations about future states of nature and expectations about future choices will not always cleanly categorize data. Nevertheless, we rely on this useful distinction in organizing our discussion of the literature below.

## 2.3 Identification and the role of expectations data

Before turning to estimation, we discuss the role expectations data can play in the identification of parameters. A set of parameters is *identified* if different parameter values would, under the model, lead to different distributions of the observables.[3] Generally, the more limited the set of observables at hand is, the more restrictive the model will need to be to identify the features of interest. More specifically in the case of the model outlined above, without data on expectations over future Social Security payments, the conditions needed to identify $\beta$ (patience) and $\kappa$ (beliefs about future Social Security) will be more restrictive than those that would be needed for identification if such expectations data were available.

While the formal definition of identification of a structural parameter is a binary property that a model, paired with a joint distribution of observables, can have, researchers who estimate model parameters often more loosely characterize particular aspects of the data as 'identifying' certain features of structure. Keane (2010) states that this (looser) notion of identification of model features relates to: ...

> "...what are the key features of the data, or the key sources of (assumed) exogenous variation in the data, or the key *a priori* theoretical or statistical assumptions imposed in the estimation, that drive the quantitative values of the parameter estimates, and strongly influence the substantive conclusions drawn from the estimation exercise?"[4]

[3] More formally, features of data generating processes are "said to be identified if among the set of observationally equivalent structures [data generating processes], the value of that feature does not vary" (Matzkin (2013), following Hurwicz (1950)).

[4] Andrews et al. (2017) argue that much of the discussion of identification of the parameters of structural models using features of the data could be reframed in terms of 'sensitivity' of the estimated parameters to features of the

In most estimated structural models, the observable variation that has been used to identify the parameters comes from data on *behavior*. However, if data on *expectations* over choices, the implications of those choices, or states of nature are measured, these can be used. Data on choices and data on choice expectations have related, but distinct types of empirical content. Data on behavior in period $t$ contain information about $i$'s optimal behavior conditional on the information available in period $t$; data on expectations at time $t$ of behavior in period $t+\tau$ contain information about optimal behavior in $t+\tau$ conditional on the information available in period $t$. In the next section, we will discuss how these two types of data can be used in estimation.

## 2.4 Estimation

In estimation, some parameters might be set with reference to the literature or might be credibly estimated without the solution of the model being used. In life-cycle models such as the one above, these could include the interest rate $r$, a parameter which is often directly estimated from time series data on interest rates. The survival curve $\{s_\tau\}_{\tau=1}^{T}$, might also be estimated using demographic data (if rational expectations are assumed) or by directly invoking expectations data on survival (which allows a relaxation of that assumption).

In general, however, most unknown parameters will be estimated by bringing implications of the model solution as close as possible, in some metric, to empirical analogues of those implications. In our discussion below, we specify what it means to bring chosen model implications 'as close as possible' to the data for different estimation methods, highlighting how expectations data can be incorporated.

### 2.4.1 Estimation Methods

**Maximum Likelihood Estimation** In the specification of the model, we have made no distributional assumptions on preferences. Suppose we augment the model's preferences with a stochastic feature that has an assumed distribution. A simple example would be to augment the single period utility function by adding a shock to preferences, $\epsilon$, with a distribution $F(\epsilon)$. The utility in period $t$ would then be $U(c_t, l_t, \epsilon_t)$.[5] Suppose further that we have data on choices (assets holdings) and expectations (stated beliefs on the probability of working at a particular age in the future) for a sample of $N$ individuals for $T$ periods. For any given vector of preference parameters $(\theta)$, we can evaluate the likelihood of observing

data.

[5]The effect of $\epsilon_t$ on utility will often be choice specific – it could represent a shock to the marginal utility of consumption, or of leisure, for example.

$\mathcal{D} = \{a_{it}, \mathbb{P}_{it}[p_{i\tau} = 1],\ i = 1 \ldots N,\ t = 1 \ldots T\}$ as the outcome of the model endowed with those parameters $\theta$. That is, we can form the likelihood function:

$$L(\mathcal{D}, \theta) = \prod_{i=1}^{N} L(\mathcal{D}_i, \theta). \tag{7}$$

The maximum likelihood estimate of $\theta$ is that parameter vector which maximizes the likelihood function. This is simply the usual likelihood estimator, applied in a context where some of the data are expectations. If the agent's problem lacks an analytical solution, and can only be solved numerically, or if there are missing state variables in the data for some or all observations, the evaluation of the likelihood function involves the solution of the model at $\theta$ and the simulation of behavior using implied decision rules. In that case, $\hat{\theta}$ is the simulated maximum likelihood estimate.

**Method of Simulated Moments (MSM)** When evaluation of the likelihood function is infeasible or its computation is prohibitively time-consuming, non-likelihood based simulation methods, such as MSM or indirect inference, might prove useful. In the case of MSM, a set of moments, which summarizes the behavior simulated from the model, is chosen. These moments, collected in $\tilde{\mathbf{m}}(\theta)$, depend on unknown parameters $\theta$. Parameter estimates are chosen such that the simulated moments are as close to the data moments ($\mathbf{m}$) as possible. That is, $\hat{\theta}$ is chosen to minimize the criterion function:

$$\hat{\theta} = \arg\min_{\theta} \ (\mathbf{m} - \tilde{\mathbf{m}}(\theta))' \mathbf{W} (\mathbf{m} - \tilde{\mathbf{m}}(\theta)). \tag{8}$$

where $\mathbf{W}$ is a symmetric positive definite matrix.

Moments should be chosen such that their values are informative about the parameters to be estimated. As an example: given a desire to estimate the relative weight of consumption in the utility function ($\nu$), moments on labor supply would be candidates for inclusion in $\mathbf{m}$. All else equal, the higher is $\nu$, the higher labor supply will be. Data on labor supply, or expectations of future labor supply, can therefore help discriminate between settings where $\nu$ is low (people place a relatively low value on consumption and a relatively high value on leisure) and settings where $\nu$ is high (the converse). Data on wealth over the life-cycle (or expectations of future wealth holdings) would similarly be informative for the discount factor ($\beta$).

**Indirect Inference** Indirect inference involves specifying an 'auxiliary model', which relates the observables together in a computable manner. The auxiliary model need not be the true data generating process, but should be easily-computable. Indirect inference estimation involves choosing the parameters by minimizing the distance between the auxiliary parameters estimated using the observed data and those estimated using the model predictions (i.e., the data simulated using the model solution). Formally, let $\hat{\beta}_A$ be the estimated parameters of the auxiliary model using observed data $y$, such that $\hat{\beta}_A = \arg\max_{\beta_A} L_A(y, \beta_A)$, and $\hat{\beta}_S$ be the estimated parameters of the auxiliary model using model predictions $y^S(\theta)$, such that $\hat{\beta}_S(\theta) = \arg\max_{\beta_S} L_A(y^S(\theta), \beta_S)$. Then the indirect inference estimate of preference parameters, $\hat{\theta}$ is defined as

$$\hat{\theta} = \arg\min_{\theta} \ (\hat{\beta}_S(\theta) - \hat{\beta}_A)'\mathbf{W}(\hat{\beta}_S(\theta) - \hat{\beta}_A), \tag{9}$$

where $\mathbf{W}$ is again a symmetric positive definite matrix. In this setup, both types of expectations data can be used in the estimation of the auxiliary model parameters. See Smith Jr (1993) and Gourieroux et al. (1993) for the introduction of this method and Van der Klaauw and Wolpin (2008) and Adda et al. (2022) for applications that use expectations data.

**Non-full solution methods of estimation** The estimation methods we have outlined so far involve repeated solution of the model at many candidate parameter vectors as a particular function is either minimized (MSM criterion function) or maximized (a likelihood function). These methods are computationally expensive, especially in complex models where solving the model is time-consuming. Methods have been developed that allow for the estimation of parameters of dynamic models while avoiding the repeated solution of the model. These methods, developed by Hotz and Miller (1993), leverage 'conditional choice probabilities' – the probability of a choice conditional on model states. We do not review this line of literature here (see the review article by Aguirregabiria and Mira (2010)), but we note that the key empirical input leveraged by these methods – measures of the probability that an individual will choose a particular discrete option in the future – are very much aligned with the concept of expectation measures often collected by surveys.

## 2.5 Issues Particular to Structural Estimation with Expectations Data

### 2.5.1 Constructing a model counterpart to expectations data

Bringing expectations data to bear on a structural model requires a precise determination of the model counterparts of what the expectations data measure. Often the data reflect objects for which the model has no single natural analogue. Van der Klaauw (2012), studying occupational choice and using data from the National Longitudinal Study of the High School Class of 1972, considers the response to the following question:

> "What kind of work will you be doing when you are 30 years old? (circle one that comes closest to what you expect to be doing)."

There is no statistical analogue to which the answer to this question can be mapped, though the form of the question suggests that respondents might select the option with the highest choice probability (the mode), and this is how Van der Klaauw (2012) interprets the answers. Delavande and Rohwedder (2011) similarly interpret the answer to the question "At what age do you expect to start collecting these [Social Security] benefits?" as the age at which claiming probability is highest.

A similar issue arises with questions that ask individuals what they expect for some continuous quantity. In their study of human capital and the return migration decisions of migrants, Adda et al. (2022) make use of a question that asks respondents "How long do you want to live in Germany?". One would not expect this to be an exact prediction of exactly how long the migrant will stay in Germany, but rather some summary measure which takes into account uncertainty. In constructing a model analogue, they assume that this reflects the median duration that a migrant will stay. Van der Klaauw and Wolpin (2008) face a similar issue and interpret the number of dollars the respondents 'expect' for [future Social Security] payments in today's dollar as expected values.

One type of data where a natural mapping to a model object does exist is when respondents are asked for choice probabilities. These are well defined objects in models of the type we are discussing. Surveys such as the HRS, its sister studies internationally, and the Federal Reserve Bank of New York's Survey of Consumer Expectations place a large emphasis on collecting expectations data in this form.[6] An additional advantage of probabilistic questions is in the richness of their empirical content. In the context of stated-choice experiments, Manski (1999) shows that the stated-choice approach may lead to different results than actual choices if the respondents are not provided with the full information that would be available to them when facing the actual choice problems. Elicited choice probabilities, on the other hand, can address the incompleteness of the scenarios by allowing respondents to express uncertainty over their choices. Moreover, they provide more information than choice experiments which elicit a single preferred option, as probabilities allow respondents to provide a ranking for their choices as well.

[6] In addition to being readily interpretable, choice probabilities are richer in empirical content than questions that solicit looser concepts of expectations and intentions. Juster (1966) notes that: "Intentions seem to have no informational content that a probability survey does not also have, and the probability survey is able to extract information that is not obtainable from intentions surveys".

### 2.5.2 Use of data on choice expectations in Maximum Likelihood Estimation

The use of maximum likelihood techniques together with data that record respondents' stated choices or choice probabilities can lead to settings with discontinuous and non-differentiable likelihood functions. During the repeated evaluation of the likelihood function in estimation, if, for a given trial parameter vector, the reported expected choice probability is not equal to the model-generated choice probability or the elicited 'most-likely' choice is not equal to the one predicted by the model, numerical optimization can become difficult and standard inferential techniques may not be used. This issue, of course, is not limited to the use of expectations data. See Van der Klaauw (2012) for a fuller discussion of a case where it occurs with expectations data and a proposed approach, which involves the assumption that individuals report their expectations (in this case, the most likely future choice) with an error, to deal with this issue.

### 2.5.3 Focal Point Responses to Probabilistic Expectation Questions

A robust feature of data on probabilistic expectations is that there tends to be an excess mass of responses at certain focal points, e.g. 0%, 50% and 100%. These masses in some cases display an implausible degree of certainty (0% or 100%), and in other cases may represent some bias in reporting (e.g. an excess mass at 50% could represent either rounding or a lack of understanding of the question or inability to formulate an answer; see Fischhoff and Bruine De Bruin (1999)).[7] Whatever the reason for such responses, it is unlikely that the underlying behavioral model to which the data will be applied will imply probabilities of future events that accord with these distributions. Using data contaminated by focal point biases to estimate a model which does not account for them means the model is not a correctly-specified data-generating process for the data in hand. Gan et al. (2005) develop a method for estimating individual-level survival curves in the presence of focal point biases, and subsequently use survival curves estimated through such an approach in a life-cycle model to study the interplay between subjective mortality risk and bequests. Blass et al. (2010) and Wiswall and Zafar (2018) show how to deal with these focal points in estimating preferences in a random utility model using choice expectations. This issue is also tackled by Hendren (2013), who uses subjective probability elicitations to study the role of private information in insurance rejections.

Our focus in this Section has been on *how* expectations data might be used in the estimation of structural models. In Sections 3 and 4, we will turn to *why* one would use it, and we will discuss the applied literature that has done that.

[7] See the discussions in Bruine De Bruin et al. (2022); Giustinelli (2022); Hudomiet et al. (2022); Delavande (2022) and Gábor (2022) on how to deal with focal points in different contexts.

# 3 Literature I: Expectations over the states of nature

As we noted in Section 2.2, Manski's (2004) taxonomy for data on expectations draws a distinction between data on expectations over states of nature and data on expectations over choices. In this section, we discuss the literature that uses the former in the estimation of structural models. In Section 4, we discuss the literature on choice expectations. This taxonomy does not, of course, perfectly bisect the literature, and some papers will be referenced in both sections. We discuss most such papers in this section and defer some details on their use of choice expectations data to Section 4.

Dynamic structural models consider the behavior of agents making decisions today that yield payoffs both today and in the future. The future payoff will often depend on the realized state of nature: this could be the health of an individual, their earnings, the state of the economy, the state of economic policy, or some other unknown feature about the future state of the world. Agents make their decisions today based on the current payoff and their expectations about the future state of nature and, thus, about future payoffs. A classic identification problem (see Manski (1993) and Manski (2004)) involves separating the role of preferences from expectations over the future states of nature, since observed choices might be compatible with several combinations of preferences and expectations. The conventional resolution to this identification problem in estimating choice models has been to use data on observed choices and realizations of that state of nature, together with an assumption of rational expectations. However, this approach does not allow for the possibility of subjective expectations being different than objective measures. An alternative approach to overcome this identification problem is to combine data on observed choices with subjective expectations to make inference on preferences. This alternative approach facilitates making weaker assumptions on the expectations formation rule. In fact, as Manski (2004) states, "...it is enough to assume that elicited expectations faithfully describe persons' perceptions of their environments." The literature we review in this section follows this second approach and combines data on observed choices with subjective expectations to estimate choice models.

We organize our discussion of this literature according to the way the data are incorporated to the model. Section 3.1 discusses the literature that combines subjective expectations data with observed choices without a particular focus on modeling the belief formation process. Section 3.2 then discusses papers that emphasize the modeling of these beliefs.

## 3.1 Allowing for subjective expectations

An early example of using subjective expectations data in a choice model is by Nyarko and Schotter (2002), who elicit players' subjective expectations of opponents' behavior in a two-person game. The paper uses these elicited expectations in the estimation of this simple game, where the best response function of each player depends on that player's beliefs on the opponent's move. Comparing different assumptions on players' expectations, the authors find that the model that uses subjective expectations best predicts observed behavior. Other early examples include Lochner (2007), who studies the link between individuals' beliefs about own arrest probabilities and criminal behavior, and Delavande (2008), who studies contraception choice using data on elicited expectations of choice-specific outcomes (such as pregnancy, side effects, and protection from STDs), observed choice data, and a random utility model. The paper shows that there is substantial heterogeneity in choice-specific outcome expectations and that taking these into account rather than assuming everyone has full-information rational expectations (FIRE)[8] and homogeneous expectations matters for parameter estimates. In recent work, Miller et al. (2020) use similar data on beliefs about contraceptive attributes and a structural model to show the role of biased beliefs about pregnancy risk in driving the unmet need for contraception in Sub-Saharan Africa.

Education decisions often involve a trade-off between a cost that is to be incurred today and some future (uncertain) returns. Expectations over those returns, therefore, are central in models of education decisions and the literature using expectations data in structural models is larger than in other domains.[9] Early contributions by Arcidiacono et al. (2012) and Zafar (2013) implement surveys which directly measure expectations of how future earnings depend on major choice. Both studies also elicit measures of perceived ability or enjoyment of study under different choices of major. These data are used together with data on observed (and intended) major choices to estimate models of major choice taking into account subjective expectations of future returns. Wiswall and Zafar (2015) combine a survey eliciting students' expectations with an information treatment to estimate a structural model of major choice.[10] Patnaik et al. (2020) also study major choice, but this time using a life-cycle model. Their aim is to separately identify the role of expectations over future earnings conditional on each major from the role of preferences (risk tolerance and patience), while allowing both beliefs and preferences to be heterogeneous.

Turning to different features of the college experience, Gong et al. (2019) quantify the consumption value of attending college. Using data on consumption in college and expectations of earnings post-college from

[8]In what follows, we refer to full-information rational expectations when we use the term rational expectations.

[9]In this section we summarize only the papers which have used expectations data in the estimation of structural models. The broader literature using these data in education research is summarized in Giustinelli (2022).

[10]See Section 4.2.2 for a detailed discussion of this paper.

the Berea Panel Survey (a survey which will be discussed further below), the authors, making use of the Euler equation, find a large consumption value of college. Delavande et al. (2020) model the time allocation of college students. They collect data on perceived academic and pecuniary returns to different time investments (e.g. studying, volunteering, internships) as well as measures of students' enjoyment of those activities. Estimating a model of time allocation during college, they find that differences in expectations play some role in explaining the heterogeneity in investments in college, but that the differences in the constraints students face (e.g. access to internships, work or family responsibilities) are more influential.

There are well-documented socio-economic gradients in undergraduate and postgraduate education and differences in perceived returns to college programs have been shown to be relevant in explaining these gradients. Attanasio and Kaufmann (2014), for example, investigate the role of own and parental expectations about future labor market risks in schooling decisions and how these links differ by the gender of the student. Kaufmann (2014) shows, with similar data, that children from poorer households require a greater perceived rate of return to attend college than do children from richer households. Hastings et al. (2015) show that lower-income students over-estimate the returns to low-earning college degree programs and bring together a field experiment and a model of college demand to study the impact of providing accurate information to college students. Boneva et al. (2021) elicit the beliefs of undergraduates about returns to postgraduate education, expectations over college graduation, and the probability of getting a postgraduate degree. They estimate a choice model for postgraduate study using this data, to study the role of differences in beliefs.

Giustinelli (2016) brings together expectations data and structural models to study *group* decision-making by parents and children over the choice of high school track. The challenge here is to separate the roles of each group member's preferences, their beliefs over uncertain choice-specific outcomes, and how the group comes to a decision. A unique data set, which brings together rich data on student and parental expectations over the long-run outcomes of each track choice as well as on the student's perceived enjoyment of each track, facilitates the estimation of a variety of models of decision-making: unilateral, bilateral, and non-strategic.

Recent papers, by bringing together data on perceived admittance probabilities to schools or programs and models of applicant behavior highlight how allocation mechanisms can interact with subjective expectations. Kapor et al. (2020), for example, study the interplay between school assignment mechanisms and parental beliefs over students' acceptance probabilities. Allocation mechanisms that imply returns to strategic behavior can improve welfare relative to strategy-proof mechanisms by allowing participants to express the intensity of their preferences. However, if applicants (parents in their case) are misinformed about admittance probabilities, such allocations can be inefficient.[11] The authors survey households in a U.S. school district and estimate a model of school choice in which households are allowed to have erroneous subjective beliefs on the admittance probability. Using their estimated model they find that a switch from the baseline (non-strategy proof) allocation mechanism to a strategy-proof allocation mechanism would *increase* welfare, whereas a planner who assumed parents have rational expectations would incorrectly believe it would *decrease* welfare. Tincani et al. (2021) also study admissions, but at the college level. They study a Chilean experiment, which guaranteed college admittance to the top 15% of students in treated high school classes. Survey data show that students overestimate their chances of guaranteed admission under this program. Treated students were found to respond to this experiment, which increased their admittance chances, by reducing their effort. To evaluate how decisions and outcomes (effort, application, admissions, and enrollments) would change if students had correct beliefs, the paper estimates a structural model of student decisions.

A large literature has investigated the extent to which stated survival probabilities differ systematically from objective survival probabilities.[12] Building on this literature, several recent papers bring estimated subjective survival curves into life-cycle models, relaxing the assumption that individuals' have rational expectations over their survival. Gan et al. (2015) estimate individual survival curves using reported survival probabilities from the HRS and use these in a model of wealth decumulation and bequests. They find that a model with subjective expectations better fits the observed decumulation and bequest behavior than a model with life table survival probabilities. Using methods developed by Gan et al. (2005), they account for the tendency of individuals to report probabilities at focal points (0 or 1).[13] Bissonnette et al. (2017) use the panel of survival expectations in the HRS and a life-cycle model to study the welfare losses associated with a divergence of subjective and objective survival probabilities, which they find are large. Heimer et al. (2019) find that young individuals are overly pessimistic about their survival, which, in a life-cycle model, causes them to undersave relative to what would be the case if they had accurate expectations. In contrast, the old are overly optimistic about their survival prospects, which causes them to decumulate wealth slower than they would if they had accurate expectations. de Bresser (2021) evaluates whether a life-cycle model can predict the retirement response to a pension reform in the Netherlands and finds that a model with subjective and heterogeneous survival probabilities can better explain the reform's impact than

[11] See also Arteaga et al. (2022) which highlights that, when information is costly to acquire, beliefs are central to the welfare of school applicants even when the allocation mechanism is strategy-proof.

[12] Early examples include Hamermesh (1985), Hurd and McGarry (1995), Hurd and McGarry (2002), and Hurd et al. (2005). See Hudomiet et al. (2022) and Gábor (2022) for in-depth discussions.

[13] See also Comerford (2019) on modes of asking questions to mitigate biases in elicited survival expectations.

models with life table survival curves. Bairoliya and McKiernan (2021) also study retirement and Social Security claiming decisions, estimating a model using expectations data from the HRS. O'Dea and Sturrock (2021) study the implications of biases in subjective survival probabilities for the 'annuity puzzle' – the fact that annuity demand is modest, despite the longevity insurance that it provides. They start from the fact (also shown by Teppa and Lafourcade (2013) and Wu et al. (2015)) that an annuity that is priced fairly can appear actuarially-unfair to an individual who is pessimistic about their survival chances. Then, using subjective survival curves estimated from the English Longitudinal Study of Ageing and a life-cycle model, they show that, in their setting, survival pessimism is one quantitatively important explanation for low annuity demand. Focusing on the implications of heterogeneity in survival expectations, Hosseini (2015) studies the role of adverse selection in annuity markets and estimates the heterogeneity in life expectancy using subjective expectations data from the HRS, while Foltyn and Olsson (2021) use subjective survival curves in an overlapping-generations model to show the role of expectations in driving wealth inequality.

In measuring the nature and extent of risks over the life-cycle, it has been important to separately identify permanent and transitory shocks to income. Pistaferri (2001) shows how subjective expectations over future income, as well as data on income realizations, can be used to separately identify these different shocks. Attanasio et al. (2020) apply this approach to study the extent to which households are able to smooth consumption in the face of income shocks. A crucial parameter in life-cycle models relates to the patience of the agents. Mahajan et al. (2020) provide results for the identification of the time-preference of potentially present-biased agents using an exclusion restriction on a variable that affects utility only through the perceived value of future states. Using data on the perceived malaria risk conditional on usage of insecticide-treated nets (ITNs), the paper examines the role of time inconsistency in the demand for ITNs.

A number of papers have used subjective expectations data in models of labor market decisions.[14] Arcidiacono et al. (2020) study occupational choice. They use a survey which elicited earnings beliefs conditional on major and occupational choice probabilities from undergraduates and then followed them after graduation as they made their actual choices. They document heterogeneity in earnings beliefs by occupation and find evidence of sorting on gains. Mueller et al. (2021) allow workers to have biased beliefs about the job finding rate and estimate a model of labor market transitions that incorporates a mapping between these beliefs and the actual rates, by targeting moments that include perceptions of the job finding rate at different points in an unemployment spell. Using these moments allows the authors to uncover the

[14]See Mueller and Spinnewijn (2022) for a detailed discussion of papers on the labor market that use expectations data.

heterogeneity in the true job finding rates. Ilieva et al. (2021) document that women in Germany are overly optimistic about human capital accumulation in part-time work and use the estimated model to show that if agents had accurate beliefs, there would be a decline in part-time work and an increase in wages. Jäger et al. (2022) show that workers tend to anchor their beliefs about outside options to their current wage. This phenomenon, in an equilibrium model of labor supply, can provide one mechanism which sustains wage markdowns.[15]

A growing literature has shown the importance of investments in children (both of time and of resources), in developing their skills (see Cunha et al. (2010)). A series of recent papers has shown that parents often do not appreciate that the returns to these investments are large. Cunha et al. (2013) show that parents underestimate the productivity of investments and Boneva and Rauh (2018) document that parents believe the returns to late investments are greater than earlier investments, Attanasio et al. (2019) show that there is significant heterogeneity in beliefs across mothers and that they tend to underestimate the returns on investments, and Attanasio et al. (2020) compare beliefs about returns to different types of investments. Embedding these beliefs into models of parents' decision-making is a profitable area for future research.

While a very large literature uses structural models to investigate the role of policy in shaping behavior, the literature which confronts uncertainty in the policy environment is much smaller. Contributing to this is the fact that the policy environment is a highly multi-dimensional object, which brings with it measurement challenges, as well as the fact that the non-stationarity in the policy environment precludes the use of realizations as estimates of what might happen in the future. As outlined in Section 2, Van der Klaauw and Wolpin (2008) is an example of a paper that does introduce policy uncertainty, over Social Security benefits, into a model estimated using expectations data. Since then, evidence has accumulated documenting how agents consider the policy environment to be uncertain. Delavande and Rohwedder (2011) and Luttmer and Samwick (2018) both show this in the case of US Social Security. Ciani et al. (2019) complement this evidence by documenting that expectations are revised in advance of, as well as following, reform announcements and by showing that there is substantial heterogeneity in expectations, even after reform announcements.

[15]The papers discussed here use expectations data in the estimation of model features. Schneider (2020) uses data on choices and policy variation to identify expectations of beliefs over re-employment prospects.

## 3.2 Modeling subjective expectations

We now turn to the papers which emphasize the process by which expectations over states of nature are formed.[16] An early example is by Bellemare et al. (2008), who combine choice data in an ultimatum game with data on the proposer's expectations over the opponent's acceptance probability, by allowing reported beliefs to have measurement errors and to depend on preferences. Their results indicate that estimating the model using subjective expectations data leads to better in- and out-of-sample fits compared to those achieved assuming rational expectations.

A set of papers study the formation of expectations for college students using the Berea Panel Survey (BPS), an ongoing high-frequency panel in which expectations about future academic performance and future earnings are regularly measured.[17] Stinebrickner and Stinebrickner (2014a) study the choice students face between staying in college and dropping out. The paper specifies a discrete-choice model in which students decide, each semester, whether to stay in college or to drop out. In this setup, students learn about their ability, a process which the authors can estimate given their repeated measures of expected future academic performance. They find that learning about one's ability explains a large share of college drop-outs – poor academic performance makes the experience of being in college less enjoyable and also lowers the expected financial return to remaining in college. In a similar manner, Stinebrickner and Stinebrickner (2014b) estimate a model of major choice. They use data from the BPS on entering students' expectations about their major choice and data on their evolving expectations and performance through college, to study the reasons for the gap between the number of students who intend to major in science and those who ultimately do so. They find that overoptimism over own-aptitude for science is an important factor.

Delavande and Zafar (2019) study the choice between different types of universities in Pakistan. Their aim is to separate the roles played by the expectations of pecuniary returns to different choices from non-pecuniary features, such as the ideology of a school or parental approval. Their survey collects data on students' preference orderings over schools as well as their expectations about future outcomes conditional on school choice. The estimated model shows that non-pecuniary factors dominate expected earnings in driving college choice in this setting. In addition to collecting preference orderings given the actual cost of attending each college, survey respondents were asked for preference orderings assuming a counterfactual world where there were no financial constraints. Using this data, the authors validate their model by

[16] As elsewhere in this chapter, our focus is on models of where the expectations formation process is brought into structural microecononomic models. See Baley and Veldkamp (2022) for a dedicated treatment of learning models.

[17] The BPS has been an unparalleled resource for studying decisions made by college students (see Giustinelli (2022)). Of particular relevance to the relaxation of rational expectations over earnings, a feature of many of the papers in this section, see Crossley et al. (2021) and Crossley et al. (2022). They find, as do d'Haultfoeuille et al. (2021) using other data, that departures from rational expectations are common.

generating the predictions of the estimated model with the technology altered to relax financial constraints. They compare these predictions with the elicited preferences under the assumption of this counterfactual world and find a close correspondence between the two.

Learning models estimated using subjective expectations data are commonly used to explain labor market transitions. Conlon et al. (2018), for example, use data on expectations over future wage offers to estimate a job search model that allows for heterogeneous and potentially biased beliefs as well as learning about the wage offer distribution. They use elicited expectations on wage offers to identify the model's learning rule. They then show that incorporating expectations data in the estimation allows the estimated model to better fit the reported reservation wages, relative to a complete information model. Cortés et al. (2021) also focus on learning about job offers, but study gender differences in risk aversion, optimism, and updating of beliefs over the expected offer distribution in the labor market. They estimate a model of job search that incorporates these gender differences and allows beliefs to change over time. Their results show that the gender gap can be significantly reduced by providing accurate information to students. Hoffman and Burks (2020) study the quitting decision of truckers, using a structural model which embeds a model of how they learn about their productivity. They make use of high-frequency survey data that include subjective expectations of truck drivers about their own future productivity (in terms of paid miles to be driven). The results indicate that drivers learn their true productivity over time, and that this learning is slower than what is predicted by Bayesian updating.

In Section 3.1 we discussed several papers that used survival expectations, where the focus was more on understanding the implications of survival than on modeling these expectations. Wang (2014) finds that smokers underestimate the implications of smoking for their longevity. Incorporating this channel in a model of smokers' choice of if and when to quit, the paper finds that smokers are estimated to be substantially less patient under rational expectations than they are found to be when subjective expectations are taken into account. Groneck et al. (2016) model the formation of survival beliefs using a model of Bayesian learning with cognitive limitations. They show that their model, calibrated with subjective expectations data from the HRS, can explain several regularities in the old-age savings behavior. See also Caliendo et al. (2020) who study the implications of ambiguity in survival expectations and Ludwig and Zimper (2013), de Bresser (2019), and Grevenbrock et al. (2021) on modeling the formation of survival expectations. Hentall-MacCuish (2021) estimates a life-cycle model that takes into account the belief formation process through agents' knowledge of their public pension entitlements. Using data on expectations of future entitlements to the UK's State Pension, the paper finds that accommodating endogenous, heterogeneous, and potentially erroneous beliefs helps the model explain bunching of labor market exits at the normal retirement age.

Hamilton et al. (2011) use data on the subjective expectations of a marketing manager to estimate a structural model of advertising decisions. By jointly estimating the manager's preferences with the actual and expected demand functions, they show that managers are overly-optimistic about advertising effectiveness. Studying financial decision-making, Bellemare et al. (2020) estimate a model of portfolio choice under uncertain return, incorporating ambiguity, loss-aversion preferences, and belief updating rules for how investors update their ambiguity, using elicited data on portfolio choices and stated beliefs over return distributions. Their results support dominant ambiguity aversion and belief updating and provide evidence against the hypothesis that loss aversion dominates ambiguity aversion for financial decisions.

In this section, we reviewed the literature that uses data on expectations of the states of nature in the estimation of structural models. In discussing these papers we emphasized whether the belief formation process was modeled. Modeling these beliefs explicitly is an important feature, as it allows researchers to evaluate the extent to which those beliefs would change in a counterfactual setting. We return to this in concluding the chapter when pointing to directions for future research.

# 4 Literature II: Data on choice expectations

The second category of expectations data in Manski's (2004) taxonomy comprises data on choice expectations. These have many uses in the estimation of structural models. First, data on actual choices may not be available, and thus, data on choice expectations might be used as a substitute. Second, even when data on actual behaviors are available, researchers may still prefer eliciting choice expectations conditional on an explicitly specified economic environment, one which maps directly into the model's setting.[18] This allows the choice set and the characteristics of the individual choices to be specified in detail while eliciting choice expectations, making observable some characteristics which would be unobservable in revealed preference data.[19] Third, while eliciting choice expectations researchers can experimentally vary the choice attributes of interest, in doing so generating exogenous variation that can facilitate identification. Finally, in settings where data on choice expectations and revealed choices are both available, they can either be used together in estimation, improving efficiency as in Van der Klaauw (2012), or if one is used in estimation, the other can be used for validating the model.

[18] Such environments can be specified through future hypothetical scenarios. See Koşar et al. (2021) for an extensive discussion on the design of such choice experiments using hypotheticals.

[19] An example to this would be the unobserved components of jobs when the agent is choosing among different job options in a labor supply model, or the unobservable location attributes in a model of migration.

In the remainder of this section, we discuss how choice expectations data have been used in the literature to estimate structural microeconomic models. We will organize our discussion by distinguishing between data on unconditional choice expectations and data on conditional choice expectations. The difference between these two is whether the expectations are elicited conditional on an explicitly specified future environment. Section 4.1 discusses contributions that use data elicited through questions on what the respondent *will* do in the future without specifying any contingency about that future. Such questions are often found in large, general purpose household surveys. Section 4.2 discusses the set of papers that solicits choice expectations conditional on some future circumstances – what the respondent *would* do in those specified circumstances. These questions have often been developed in bespoke surveys developed for particular papers. Finally, Section 4.3 discusses the literature that uses 'strategic survey questions', a type of design that elicits beliefs with the specific goal of identifying a particular feature of a given model.

## 4.1 Unconditional Choice Expectations

Unconditional choice expectations data take the form of a respondent's expected choices in the future, without the survey specifying any details of the environment the individuals will be facing when this actual decision will be made (other than perhaps their age or the time period in which the decision will be made). These choice expectations can be considered to be conditional only on the current information set of the individual.

A very early contribution by Wolpin and Gonul (1985) tests whether data on expected retirement ages are consistent with those predicted by a labor supply model. They find that expectations data contain valuable information and highlight their use in estimating models of labor supply. Another early example is by Van der Klaauw (2012), who studies the decision to become a teacher. He develops a dynamic model of career decisions under uncertainty, and shows how data on observed choices can be used together with choice expectations data to estimate parameters using simulated MLE. The data come from questions on the expected age 30 occupation of a sample with an average age of 25. The paper first estimates the model parameters using only choice data and then shows that the predictions from that estimated model are consistent with elicited choice expectations. The paper then estimates parameters using both realized choice data and choice expectations data and shows that these estimates are very close to those obtained using only realized choice data, but that the standard errors from the estimation that incorporates both types of data are smaller.[20]

[20] Somewhat similarly, but studying college major choice rather than occupational choice, both Arcidiacono et al. (2012) and Zafar (2013) combine data on observed choices with intended choices to estimate preferences. See Section

The research design by Van der Klaauw (2012) does not rely on expectations data for identification, but instead for efficiency and validation. However, choice expectations data are often particularly valuable for identifying certain parameters in a structural model. Adda et al. (2022) provide one such example, by studying the interplay between human capital choices and the return migration decisions of migrants. One decision migrants face is how much to invest in host-country-specific human capital (e.g. language), for which the returns vary with the intended duration of migration. To study this issue, the authors specify a model in which migrants, who differ in their preferences for living in their origin and host countries, are faced with human capital and return migration choices. Using repeated observations on the intended duration of stay from the German Socio-Economic Panel, the authors estimate the extent of heterogeneity in location preferences and the persistence of such preferences. In the absence of expectations data in this setting the most valuable observable that might identify heterogeneity and persistence in location preferences would be the level of language skills. However measured language skill has significant limitations for the purpose at hand as it is slow moving and, as language skills typically improve rather than deteriorate, it cannot be used to measure negative shocks to location preferences.

The papers we reviewed in this section so far use discrete choice expectations. An example of using probabilistic unconditional choice expectations is by Van der Klaauw and Wolpin (2008). This was discussed in Section 3, with a focus there on their use of expectations data on policy uncertainty. In their paper, the authors also use data on choice expectations, specifically the probabilities of working at older ages and leaving a bequest. The value-added of expectations data in that context is to avail of measurements of model implications that cannot be available at the time of writing (since most of the sample were younger than 62 at the time of observation). McGee (2021) also leverages probabilistic data on bequest intentions – to identify the strength of bequest motives in a structural model developed to study how old-age wealth shocks are transmitted into savings, consumption, and bequests. The paper shows that, conditional on a rich set of observables, there is substantial heterogeneity in stated probabilities of leaving bequests among survey respondents in England. Expectations data are then used to separately identify precautionary saving motives from bequest motives.

## 4.2 Conditional Choice Expectations

Surveys often ask people what they expect to do in hypothetical future scenarios. The empirical content of these elicited responses is different from that of unconditional choice expectations. Consider a ques-

3 for a more detailed discussion of those papers.

tion which inquires about work expectations. Suppose, as in the HRS, individuals are asked about their probability of being in work at a future age, but without specifying anything about their circumstances at that age. The analogue of this, in the context of the simple model outlined in Section 2.1, would be $\mathbb{P}_i[\tilde{p}_{it+\tau} = 1]$, an individual's probability of being in work at age $t + \tau$, *integrating over all possible states of the world*. Suppose that, alternatively, individuals are asked to consider the probability that they will be in work conditional on some future state of the world (for example based on their health status, as in Giustinelli and Shapiro (2019) who study the interplay between health and retirement). Denoting the state variable(s) which are considered in the question's conditioning as $\mathbf{x}$ and the realized future state as $\mathbf{x}^j$, the model analogue to this question would be $\mathbb{P}_i[\tilde{p}_{it+\tau} = 1|\mathbf{x}_{t+\tau} = \mathbf{x}^j_{t+\tau}]$. This distinction naturally has implications for use of such data in estimation.

Choice expectations data elicited using hypothetical scenarios provide variation of a type not often available with data on observed choices. Whereas revealed preference data is observed in only one state of the world, creating selection problems that abound in economics, choice expectations can be assessed in many different states of the world. Similarly, whereas conditional choice expectations can be measured under conditions which have not (and may never) prevail, choice behavior can only be observed under conditions which actually have prevailed – as Ameriks et al. (2020) note, there is "no obvious behavioral imprint of frustrated desires". Finally, soliciting choice expectations under multiple scenarios for the same individual leads to repeated observations that can be used to exploit within-person variation.

In the remainder of this section, we will classify studies based on whether they use stated discrete choice data (in Section 4.2.1) or choice probabilities (in Section 4.2.2) in the estimation of structural models. With stated choice expectations data, researchers elicit the most preferred choice or the preference ranking across different options. With probabilistic conditional choice expectations data, researchers instead elicit the probability of each option being chosen.[21]

### 4.2.1 Stated Discrete Choice Data

A growing literature uses data on how employment decisions vary with job characteristics in the estimation of labor supply models. In this setting, observed data on choices are the outcomes of the interplay between individual preferences and market conditions – that is, the interplay between labor supply and labor demand. An advantage of using conditional choice expectations data here is that, by specifying carefully the terms of

[21] As noted in Section 2.5.1, probabilistic expectations data are richer in empirical content than stated choice data, in that a researcher needs some mapping to go from the former to the latter. Probabilistic choice expectations data are also more likely to be informative about rankings of options than are stated choice data.

hypothetical job offers, labor demand can be held constant, allowing for identification of parameters which govern labor supply.[22]

Focusing on retirement decisions, Van Soest and Vonkova (2014) study how individuals would trade off different combinations of retirement trajectories and income in retirement. Using conditional choice expectations data, which display respondents' most likely choices among different retirement scenarios that vary replacement rates and retirement ages, allows the authors to estimate preferences for pension plans that either may not exist or may be inaccessible to the respondents. The paper estimates the parameters of a labor supply model using simulated MLE. They find that the effect of financial incentives on retirement age are, in many cases, larger than those estimated with revealed preference data. The paper argues that this is due to the fact that more flexible choice options are presented to the respondents in the hypotheticals. Ameriks et al. (2020) also study the transition into retirement and ask whether older individuals would work for longer if employment opportunities with more flexible schedules were available to them. The authors use a labor supply model and responses to the choices individuals would make if faced by hypothetical employment offers to resolve an identification problem – whether the abrupt fall in labor supply near retirement is due to supply-side factors (e.g., a high intertemporal elasticity of labor supply) or demand-side factors (e.g., non-convexities in production technologies that make it unproductive for firms to hire part-time workers). They find that a latent desire to smooth leisure, identified using the data on conditional choice expectations, implies that demand-side factors play a substantial role.

Studying workers across the age distribution, Maestas et al. (2018) estimate the willingness-to-pay for job attributes using discrete stated-choice experiments that vary a broad set of job characteristics. Amenities are often correlated with other job characteristics in observational data. Hypothetical scenarios enable the authors to generate variation in these amenities (holding everything else constant), which can be used to recover preferences. Estimating a labor supply model with this data, they find substantial heterogeneity in preferences for different non-wage amenities across demographic groups and, that accounting for this preference heterogeneity increases the measured wage inequality. Similarly, Koşar et al. (2021) tackle the problem of identifying heterogeneity in preferences, but for leisure. They design and implement a survey in which hypothetical wage-hours pairs are presented to respondents, who are asked to choose one of the job offers or unemployment. This data facilitate the estimation of preference heterogeneity in a canonical life-cycle model by providing exogenous variation in wages and weekly hours for the same individual – a richness seldom available with observational data. Using the estimated model, the authors show that the

[22] There is also a large literature in marketing, environmental, and natural resource economics that use discrete stated-choice experiments to recover preferences. We do not review that literature here – see Manski (1999, 2004) for two comprehensive reviews.

preference heterogeneity identified with the choice expectations data has important implications for the predicted responses to changes in tax policy and childcare subsidies.

Just as stated-choice data can be used to disentangle the role of labor demand and supply, they can similarly be used to isolate the workings of each side of the marriage market. Andrew and Adams-Prassl (2021) study the mechanisms driving school drop-out and early marriage for young girls in India. To separate the role of preferences of girls' families, the beliefs of those families about marriage market outcomes, and the preferences of grooms (the other side of the market), the authors design and implement a survey which elicits the stated choices of brides' families when presented with hypothetical vignettes that vary the outcomes over brides' completed education, brides' age of marriage, and grooms' characteristics. Using this induced exogenous variation in bride's and groom's characteristics, they estimate a structural model of parental choice of daughter's education and marriage-market behavior and show how the elicited choice data can be used to identify the parameters of parental utility function.

Finally, Lagakos et al. (2022) use discrete choice expectations data to validate the findings of a structural model of migration choices of rural migrants, estimated using data from a randomized field experiment where landless households in Bangladesh were offered one-time migration subsidies. The estimated model generates a large non-monetary utility cost of migration and the authors validate this result using responses from a discrete stated-choice experiment administered to the same sample. Bossavie et al. (2021) also study migration, and use data on the expectations of migrants prior to migration in the estimation of their model.

### 4.2.2 Probabilistic conditional choice expectations data

The seminal empirical application of using elicited choice probabilities to estimate preferences is by Blass et al. (2010), who measure the willingness to pay for the reliability of electricity services using a hypothetical choice methodology. Following the theoretical discussion by Manski (1999), Blass et al. (2010) assume preferences follow a random utility model with random coefficients. The utility of individual $i$ from alternative $j$ is:

$$U_{ij} = X_{ij}\beta_i + \varepsilon_{ij}, \tag{10}$$

where $\beta_i = b + \eta_i$ and $X_{ij}$ denotes the observed characteristics of choice alternatives and personal attributes. The respondents observe $X_{ij}$, the characteristics of the choice alternatives stated in the scenarios. The probabilistic choice expectations they report reflect their uncertainty about their choice, which is captured by $\varepsilon_{ij}$. Manski (1999) calls this *resolvable uncertainty*, referring to the uncertainty respondents face about the unknown components of the environment that are not specified in the scenarios, but that would be known in an actual choice setting. Accounting for this type of uncertainty in a model, whether using data on probabilities of choices or data on a single most likely choice (such as in Van der Klaauw (2012)) is pivotal, as the responses individuals give will conflate both information on their preferences and their assessments of how their environment will evolve.[23]

The identifying assumption to recover preferences here is that the scenario-specific unobserved terms, $\varepsilon_{ij}$ for all $j \in \{1, \ldots, J\}$, are *iid* and independent of the scenario attributes. This is generally achieved through the scenario design, where respondents are instructed that the choice alternatives vary only in the characteristics specified in the scenarios and are otherwise identical. It is common to further assume $\varepsilon_{ij}$ for each attribute $j$ are *iid* with Type I extreme value distribution. With this assumption, the choice probabilities implied by (10) reduce to a multinomial logit form. Taking the log odds ratio leads to the linear mixed-logit model:

$$\ln\left(\frac{q_{ij}}{q_{i1}}\right) = (X_{ij} - X_{i1})b + u_{ij}, \quad \forall j \in \{2, \ldots, J\}, \tag{11}$$

where $u_{ij} = (X_{ij} - X_{i1})\eta_i$ and $q_{ij}$ refers to the probability of respondent $i$ choosing alternative $j$. Probabilistic expectations data (as discussed in Section 2.5.3) suffer from a rounding problem, where respondents tend to report expectations in increments of 5% or 10%. However, when choice probabilities are rounded to 0 or 1, the log odds ratio becomes undefined and thus, the least squares estimation can not be implemented. To overcome this issue, Blass et al. (2010) assume $\eta_i$ are normally distributed with mean 0, which implies that $u_{ij}$ are also normally distributed around 0 conditional on $X_{ij}$ and have a median of 0 conditional on $X_{ij}$. With this assumption, the median becomes:

$$M\left[\ln\left(\frac{q_{ij}}{q_{i1}}\middle|\mathbf{X}\right)\right] = (X_{ij} - X_{i1})b, \tag{12}$$

and the parameters can be estimated using the Least Absolute Deviation (LAD) estimator. Since the median of a random variable is not affected by transformations that do not affect the ordering of values relative to the median, any zeros and ones can be suitably transformed and estimation can proceed.[24] Following Blass et al. (2010), researchers have used elicited choice probabilities to understand preferences for political candidates and the voting behavior (Delavande and Manski, 2015), to estimate consumers' willingness to pay for electric power generated from different sources (Morita and Managi, 2015), and to estimate preferences for long-term care insurance products (Boyer et al., 2017), land-use scenarios (Shoyama

[23]With data on observed choices there is no resolvable uncertainty to account for.

[24]Given the symmetry assumption, Blass et al. (2010) refer to the parameter estimates as mean preferences.

et al., 2013), and postgraduate education (Boneva et al., 2021).

Wiswall and Zafar (2018) extend this methodology and apply it to study how preferences for workplace attributes affect the gender gap in labor market earnings. Instead of assuming $\beta_i = b + \eta_i$ as in equation (10), they allow $\beta_i$ to freely vary in the population. This, together with the Type I extreme assumption on $\varepsilon_{ij}$, leads to the following linear mixed-logit model for the preferences in equation (10):

$$\ln\left(\frac{q_{ij}}{q_{i1}}\right) = (X_{ij} - X_{i1})\beta_i. \tag{13}$$

The authors use job choice probabilities elicited from undergraduates through hypothetical job scenarios, that vary in attributes such as expected earnings, earnings growth, workplace flexibility, and dismissal probability, to estimate their model. Wiswall and Zafar (2018) assume these choice probabilities are reported with an error, which has a zero median conditional on the observed characteristics $X_{ij}$ and takes a linear-in-log form. The final log odds ratio of the observed choice probabilities ($\widetilde{q_{ij}}$) becomes:

$$\ln\left(\frac{\widetilde{q_{ij}}}{\widetilde{q_{i1}}}\right) = (X_{ij} - X_{i1})\beta_i + \omega_{ij}, \tag{14}$$

This can be estimated using the LAD estimator. Note that this formulation does not impose any distribution on the measurement errors ($\omega_{ij}$), which, together with the rich variation created by the hypothetical scenarios, enables the estimation of this model separately for each individual to recover individual preferences. Koşar et al. (2021) follow the approach introduced by Wiswall and Zafar (2018) to recover the distribution of individual-level preferences for location characteristics and estimate the non-monetary costs of moving. Gong et al. (2022) use a similar methodology and elicit location premiums to estimate the importance of non-pecuniary benefits in location decisions of college graduates from low-income backgrounds.

The papers discussed above use surveys that collect data on, but do not attempt to perturb, conditional choice expectations. Rather than using multiple hypothetical scenarios to vary the attributes and to identify their impact on choices, it is also possible to use information experiments. These involve shifting beliefs on certain choice-specific outcomes by providing information to the respondents. An early example of combining informational interventions, subjective conditional choice probabilities, and a model of decision-making is by Wiswall and Zafar (2015), who study college major choice. In their survey, the authors first elicit beliefs about respondents' own expected future earnings, labor market status, and marital status at age 30 conditional on receiving a degree in different majors, as well as their perceptions of the population distribution of these outcomes. Next, they give respondents accurate information on the population characteristics of the graduates of each major. The revision in respondents' beliefs about their conditional, choice-specific outcomes as a result of the experimental design leads to a panel data, which are then used to estimate a dynamic model of major choice, leveraging the within-individual variation in beliefs. The results show that heterogeneous tastes are the most important factor in driving major choice. The authors also show that estimation of the model using only cross-sectional expectations data, without taking into account the correlation of tastes with earnings expectations, overestimates the role of earnings in major choices. For more details and references on this estimation method, see Fuster and Zafar (2022).

Following the approach by Wiswall and Zafar (2015), Ruder and Van Noy (2017) analyze how information on the population estimates of the earnings risk for college majors affect the preferences over these college majors and Baker et al. (2018) estimate the preferences of community college students for different college majors, by experimentally manipulating their expected labor market outcomes through an information experiment.

## 4.3 Strategic Survey Questions

Conditional choice expectations data elicited using questions that are designed 'strategically' to identify the features of a particular model have been termed 'strategic survey questions' (SSQs). Utilizing this type of data to estimate model parameters goes back to Barsky et al. (1997), who obtained direct measurements of parameters governing risk tolerance, time preference, and intertemporal substitution using a survey using hypothetical scenarios. SSQs are designed alongside the parametric specification of preferences, since the hypothetical questions are designed with the identification of the parameters of the model in mind.

An early example of using SSQs to identify preference parameters in a structural model is by Ameriks et al. (2011), who study the reasons why annuities are rarely purchased, despite the fact that they can provide valuable longevity insurance. Two candidate explanations for this lack of demand have been a desire to avoid publicly-provided nursing home care, should it be necessary, and bequest motives, both of which have similar implications for wealth accumulation. To identify the parameters governing the strength of these motives, the authors develop two SSQs and use the mapping between the preference specification and the responses to these SSQs to estimate the preference parameters using simulated MLE. In a similar manner, Ameriks et al. (2020) tackle the question of why many households retain wealth late into life, using a model that incorporates precautionary saving against health risks, the potential need for long-term care, and an uncertain lifespan. The authors combine observed wealth data and SSQs to jointly estimate risk aversion, aversion to publicly-provided long term care, and the strength of the bequest motive. They find that the risk of needing long-term care and bequest motives each have quantitatively similar roles in determining late-in-life saving. Using the same set of SSQs, Ameriks et al. (2016) estimate preferences in a life-cycle model with incomplete markets and stochastic health and mortality risks, to analyze why individuals hold little long-term care insurance even though they face significant late-in-life risks.

SSQs are commonly used to identify time preference and risk aversion parameters. Patnaik et al. (2020) recover parameters corresponding to constant relative risk aversion preferences and geometric discounting, using two separate games. The authors then use these parameters, along with probabilistic choice expectations and subjective beliefs on choice-specific outcomes, to estimate a model of college major choice. Ameriks et al. (2020) use SSQs to jointly estimate *individual-level* risk tolerance as well as individuals' perceptions for the mean and variance of stock returns. Similarly, SSQs can also be used to analyze the state-dependence in marginal utility of consumption across health and disabled states, as in Brown et al. (2016).

# 5 Conclusion

This chapter discusses how expectations data can be, and have been, used in the estimation of structural microeconomic models. Data on individual or household expectations over future states of the world can be used to relax strong (and often-untested) assumptions on how expectations are formed, while data on expected future choices can be used to substitute for, and to complement, behavioral data, which have been the primary data source for estimating these models.

We conclude by noting two directions for future research. The first relates to the importance of measurement. The past two decades have seen a proliferation in the measurement of expectations, with the development of several dedicated high-quality surveys and with expectations questions now embedded in several general-use surveys.[25] Using this data, individuals' expectations have been shown to correlate with other features of their decision problems (preferences and constraints) in various areas. To the extent that measurement of those other features of decision problems do not have the same rich measurement base as do expectations, there is a risk of ascribing to expectations some role which should be shared with those (omitted) factors. Several recent papers that we review have made strides on measuring distributions of expectations alongside other dimensions of individual heterogeneity.

A second direction relates to the modeling of expectations formation. Many of the papers we review in this chapter do not model how individuals form their expectations. A common approach is to assume a

[25] See Bruine De Bruin et al. (2022).

specific form for expectations (e.g. rational expectations, if expectations data are not used) or to use elicited data to directly measure agents' expectations over future events, which are then assumed to be invariant to policy changes. However, the extent to which the process governing the formation of expectations varies with policy and which factors influence their formation are still very much open questions. In order to identify such processes, one would require expectations data with repeated observations and, ideally, a long time series. Several data sets that pioneered the elicitation of expectations, such as the Health and Retirement Study and the Federal Reserve Bank of New York's Survey of Consumer Expectations, have, by now, lengthy time series. These might facilitate the modeling of the expectations process in estimated structural models.

# References


Adda, J. and R. Cooper (2003). *Dynamic economics: quantitative methods and applications.* MIT press.

Adda, J., C. Dustmann, and J.-S. Görlach (2022). The dynamics of return migration, human capital accumulation, and wage assimilation. *Review of Economic Studies*.

Aguirregabiria, V. and J. Jeon (2020). Firms' beliefs and learning: Models, identification, and empirical evidence. *Review of Industrial Organization 56*(2), 203–235.

Aguirregabiria, V. and P. Mira (2010). Dynamic discrete choice structural models: A survey. *Journal of Econometrics 156*(1), 38–67.

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

Ameriks, J., J. Briggs, A. Caplin, M. D. Shapiro, and C. Tonetti (2016). The long-term-care insurance puzzle: Modeling and measurement. Working Paper 22726, National Bureau of Economic Research.

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.

Ameriks, J., A. Caplin, S. Laufer, and S. V. Nieuwerburgh (2011). The Joy of Giving or Assisted Living? Using Strategic Surveys To Separate Bequest and Precautionary Motives. *Journal of Finance 66*(13105), 519–561.

Ameriks, J., G. Kézdi, M. Lee, and M. D. Shapiro (2020). Heterogeneity in Expectations, Risk Tolerance, and Household Stock Shares: The Attenuation Puzzle. *Journal of Business & Economic Statistics 38*(3), 633–646.

Andrew, A. and A. Adams-Prassl (2021). Revealed beliefs and the marriage market return to education. Working paper.

Andrews, I., M. Gentzkow, and J. M. Shapiro (2017). Measuring the sensitivity of parameter estimates to estimation moments. *Quarterly Journal of Economics 132*(4), 1553–1592.

Arcidiacono, P., V. J. Hotz, and S. Kang (2012). Modeling college major choices using elicited measures of expectations and counterfactuals. *Journal of Econometrics 166*(1), 3–16.

Arcidiacono, P., V. J. Hotz, A. Maurel, and T. Romano (2020). Ex ante returns and occupational choice. *Journal of Political Economy 128*(12), 4475–4522.

Arteaga, F., A. J. Kapor, C. A. Neilson, and S. D. Zimmerman (2022). Smart matching platforms and heterogeneous beliefs in centralized school choice. *Quarterly Journal of Economics*.

Attanasio, O., T. Boneva, and C. Rauh (2020). Parental beliefs about returns to different types of investments in school children. *Journal of Human Resources*, 0719–10299R1.

Attanasio, O., F. Cunha, and P. Jervis (2019). Subjective parental beliefs. their measurement and role. Working Paper 26516, National Bureau of Economic Research.

Attanasio, O. and K. Kaufmann (2014). Education choices and returns to schooling: Intrahousehold decision making, gender and subjective expectations. *Journal of Development Economics 109*, 203–216.

Attanasio, O., A. Kovacs, and K. Molnar (2020). Euler equations, subjective expectations and income shocks. *Economica 87*(346), 406–441.

Bachmann, R. and K. Carstensen (2022). Firm surveys. In *Handbook of Economic Expectations*, Chapter 2.

Bairoliya, N. and K. McKiernan (2021). Revisiting retirement and social security claiming decisions. SSRN Working Paper 3896031.

Baker, R., E. Bettinger, B. Jacob, and I. Marinescu (2018, August). The Effect of Labor Market Information on Community College Students' Major Choice. *Economics of Education Review 65*, 18–30.

Baley, I. and L. Veldkamp (2022). Bayesian learning. In *Handbook of Economic Expectations*, Chapter 25.

Barsky, R. B., F. T. Juster, M. S. Kimball, and M. D. Shapiro (1997, May). Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study. *Quarterly Journal of Economics 112*(2), 537–579.

Baumeister, C. (2022). Measuring market expectations. In *Handbook of Economic Expectations*, Chapter 6.

Bellemare, C., S. Kröger, and K. M. Sossou (2020). Optimal frequency of portfolio evaluation in a choice experiment with ambiguity and loss aversion. *Journal of Econometrics*.

Bellemare, C., S. Kröger, and A. Van Soest (2008). Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities. *Econometrica 76*(4), 815–839.

Bissonnette, L., M. D. Hurd, and P.-C. Michaud (2017). Individual survival curves comparing subjective and observed mortality risks. *Health economics 26*(12), e285–e303.

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

Boneva, T., M. Golin, and C. Rauh (2021). Can perceived returns explain enrollment gaps in postgraduate education? *Labour Economics*, 101998.

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

Born, B., Z. Enders, G. J. Müller, and K. Niemann (2022). Firm expectations about production and prices: Facts, determinants, and effects. In *Handbook of Economic Expectations*, Chapter 15.

Bossavie, L., J.-S. Görlach, Ç. Özden, and H. Wang (2021). Temporary migration for long-term investment. Policy Research Working paper 9740, World Bank Group.

Boyer, M., P. De Donder, C. Fluet, M.-L. Leroux, and P.-C. Michaud (2017). Long-term care insurance: Knowledge barriers, risk perception and adverse selection. Working Paper 23918, National Bureau of Economic Research.

Brown, J. R., G. S. Goda, and K. McGarry (2016). Heterogeneity in state-dependent utility: Evidence from strategic surveys. *Economic Inquiry 54*(2), 847–861.

Bruine De Bruin, W., A. Chin, J. Dominitz, and W. van der Klaauw (2022). Household surveys and probabilistic questions. In *Handbook of Economic Expectations*, Chapter 1.

Caliendo, F. N., A. Gorry, and S. Slavov (2020). Survival ambiguity and welfare. *Journal of Economic Behavior & Organization 170*, 20–42.

Candia, B., O. Coibion, and Y. Gorodnichenko (2022). The macroeconomic expectations of firms. In *Handbook of Economic Expectations*, Chapter 14.

Ciani, E., A. Delavande, B. Etheridge, and M. Francesconi (2019). Policy uncertainty and information flows: Evidence from pension reform expectations. CEPR Discusssion Paper DP13988.

Comerford, D. A. (2019). Asking for frequencies rather than percentages increases the validity of subjective probability measures: Evidence from subjective life expectancy. *Economics Letters 180*, 33–35.

Conlon, J. J., L. Pilossoph, M. Wiswall, and B. Zafar (2018). Labor market search with imperfect information and learning. Working Paper 24988, National Bureau of Economic Research.

Cortés, P., J. Pan, L. Pilossoph, and B. Zafar (2021). Gender differences in job search and the earnings gap: Evidence from business majors. Working Paper 28820, National Bureau of Economic Research.

Crossley, T., Y. Gong, T. R. Stinebrickner, and R. Stinebrickner (2021). Examining income expectations in the college and early post-college periods: New distributional tests of rational expectations. Working Paper 28353, National Bureau of Economic Research.

Crossley, T. F., Y. Gong, R. Stinebrickner, and T. Stinebrickner (2022). The ex post accuracy of subjective beliefs: A new measure and decomposition. *Economics letters 210*, 110160.

Cunha, F., I. Elo, and J. Culhane (2013). Eliciting maternal expectations about the technology of cognitive skill formation. Working Paper 19144, National Bureau of Economic Research.

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

de Bresser, J. (2019). Measuring subjective survival expectations–do response scales matter? *Journal of Economic Behavior & Organization 165*, 136–156.

de Bresser, J. (2021). The role of heterogeneous expectations in life cycle models. CentER Discussion Paper 2021-034.

Delavande, A. (2008). Pill, patch, or shot? subjective expectations and birth control choice. *International Economic Review 49*(3), 999–1042.

Delavande, A. (2022). Expectations in development. In *Handbook of Economic Expectations*, Chapter 11.

Delavande, A., E. Del Bono, and A. Holford (2020). Academic and non-academic investments at university: The role of expectations, preferences and constraints. *Journal of Econometrics*.

Delavande, A. and C. F. Manski (2015). Using elicited choice probabilities in hypothetical elections to study decisions to vote. *Electoral Studies 38*, 28–37.

Delavande, A. and S. Rohwedder (2011). Individuals' uncertainty about future social security benefits and portfolio choice. *Journal of Applied Econometrics 26*(3), 498–519.

Delavande, A. and B. Zafar (2019). University choice: The role of expected earnings, nonpecuniary outcomes, and financial constraints. *Journal of Political Economy 127*(5), 2343–2393.

d'Haultfoeuille, X., C. Gaillac, and A. Maurel (2021). Rationalizing rational expectations: Characterizations and tests. *Quantitative Economics 12*(3), 817–842.

Diercks, A. M. and H. Jendoubi (2022). Expectations of financial market participants. In *Handbook of Economic Expectations*, Chapter 13.

Fischhoff, B. and W. Bruine De Bruin (1999). Fifty–fifty= 50%? *Journal of Behavioral Decision Making 12*(2), 149–163.

Foltyn, R. and J. Olsson (2021). Subjective life expectancies, time preference heterogeneity, and wealth inequality. SSRN Working Paper 3895119.

Fuster, A. and B. Zafar (2022). Survey experiments on economic expectations. In *Handbook of Economic Expectations*, Chapter 5.

Gábor, K. (2022). Retirement expectations. In *Handbook of Economic Expectations*, Chapter 12.

Gan, L., G. Gong, M. Hurd, and D. Mcfadden (2015). Subjective mortality risk and bequests. *Journal of Econometrics 188*(2), 514–525.

Gan, L., M. Hurd, and D. McFadden (2005). Individual survival curves. In D. Wise. (Ed.), *Analyses in Economics of Aging I*, pp. 377–411. The University of Chicago Press.

Giustinelli, P. (2016). Group decision making with uncertain outcomes: Unpacking child–parent choice of the high school track. *International Economic Review 57*(2), 573–602.

Giustinelli, P. (2022). Expectations in education. In *Handbook of Economic Expectations*, Chapter 9.

Giustinelli, P. and M. D. Shapiro (2019). Seate: Subjective ex ante treatment effect of health on retirement. Working Paper 26087, National Bureau of Economic Research.

Gong, Y., L. Lochner, R. Stinebrickner, and T. R. Stinebrickner (2019). The consumption value of college. Working Paper 26335, National Bureau of Economic Research.

Gong, Y., T. Stinebrickner, R. Stinebrickner, Y. Yao, et al. (2022). The role of non-pecuniary considerations: Locations decisions of college graduates from low income backgrounds. Working Paper 20221, University of Western Ontario, Centre for Human Capital and Productivity.

Gourieroux, C., A. Monfort, and E. Renault (1993). Indirect inference. *Journal of Applied Econometrics 8*(S1), S85–S118.

Grevenbrock, N., M. Groneck, A. Ludwig, and A. Zimper (2021). Cognition, optimism, and the formation of age-dependent survival beliefs. *International Economic Review 62*(2), 887–918.

Groneck, M., A. Ludwig, and A. Zimper (2016). A life-cycle model with ambiguous survival beliefs. *Journal of Economic Theory 162*, 137–180.

Haile, P. (2020). Structural vs. reduced form: Language, confusion, and models in empirical economics. `http://www.econ.yale.edu/~pah29/intro.pdf`.

Hamermesh, D. S. (1985). Expectations, life expectancy, and economic behavior. *Quarterly Journal of Economics 100*(2), 389–408.

Hamilton, B. H., T. Y. Chan, and C. Makler (2011). Using expectations data to infer managerial preferences. SSRN Working Paper 1830285.

Hastings, J., C. A. Neilson, and S. D. Zimmerman (2015). The effects of earnings disclosure on college enrollment decisions. Working Paper 21300, National Bureau of Economic Research.

Heimer, R. Z., K. O. R. Myrseth, and R. S. Schoenle (2019). Yolo: Mortality beliefs and household finance puzzles. *Journal of Finance 74*(6), 2957–2996.

Hendren, N. (2013). Private information and insurance rejections. *Econometrica 81*(5), 1713–1762.

Hentall-MacCuish, J. (2021). Costly attention and retirement. SSRN Working Paper 3363188.

Hoffman, M. and S. V. Burks (2020). Worker overconfidence: Field evidence and implications for employee turnover and firm profits. *Quantitative Economics 11*(1), 315–348.

Hosseini, R. (2015). Adverse selection in the annuity market and the role for social security. *Journal of Political Economy 123*(4), 941–984.

Hotz, V. J. and R. A. Miller (1993). Conditional choice probabilities and the estimation of dynamic models. *Review of Economic Studies 60*(3), 497–529.

Hudomiet, P., M. Hurd, and S. Rohwedder (2022). Mortality and health expectations. In *Handbook of Economic Expectations*, Chapter 10.

Hurd, M. D. and K. McGarry (1995). Evaluation of the Subjective Probabilities of Survival in the Health and Retirement Study. *Journal of Human Resources 30*(1995), S268.

Hurd, M. D. and K. McGarry (2002). The Predictive Validity of Subjective Probabities of Survival. *Economic Journal 112*(482), 966–985.

Hurd, M. D., S. Rohwedder, and J. Winter (2005). Subjective probabilities of survival: An international comparison. Working paper, RAND.

Hurwicz, L. (1950). Generalization of the concept of identification. In *Statistical inference in dynamic economic models*. Cowles Commission.

Ilieva, B., P. Eisenhauer, P. Haan, A. Schrenker, and G. Weizsäcker (2021). Biased wage expectations and female labor supply. Working paper.

Jäger, S., C. Roth, N. Roussille, and B. Schoefer (2022). Worker beliefs about outside options. Working Paper 29623, National Bureau of Economic Research.

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

Kapor, A. J., C. A. Neilson, and S. D. Zimmerman (2020). Heterogeneous beliefs and school choice mechanisms. *American Economic Review 110*(5), 1274–1315.

Kaufmann, K. M. (2014). Understanding the income gradient in college attendance in mexico: The role of heterogeneity in expected returns. *Quantitative Economics 5*(3), 583–630.

Keane, M. P. (2010). Structural vs. atheoretic approaches to econometrics. *Journal of Econometrics 156*(1), 3–20.

Koşar, G., A. Şahin, and B. Zafar (2021). The work-leisure tradeoff: Identifying the heterogeneity. Working paper.

Koşar, G., T. Ransom, and W. van der Klaauw (2021). Understanding migration aversion using elicited counterfactual choice probabilities. *Journal of Econometrics*.

Lagakos, D., A. M. Mobarak, and M. E. Waugh (2022). The Welfare Effects of Encouraging Rural-Urban Migration. Staff Report 635, Federal Reserve Bank of Minneapolis.

Lochner, L. (2007). Individual Perceptions of the Criminal Justice System. *American Economic Review 97*(1), 444–460.

Ludwig, A. and A. Zimper (2013). A parsimonious model of subjective life expectancy. *Theory and Decision 75*(4), 519–541.

Luttmer, E. F. and A. A. Samwick (2018). The Welfare Cost of Perceived Policy Uncertainty: Evidence from Social Security. *American Economic Review 108*(2), 275–307.

Maestas, N., K. J. Mullen, D. Powell, T. Von Wachter, and J. B. Wenger (2018). The Value of Working Conditions in the United States and Implications for the Structure of Wages. Working Paper 25204, National Bureau of Economic Research.

Mahajan, A., C. Michel, and A. Tarozzi (2020, May). Identification of time-inconsistent models: The case of insecticide treated nets. Working Paper 27198, National Bureau of Economic Research.

Manski, C. F. (1993). Adolescent econometricians: How do youth infer the returns to schooling? In *Studies of Supply and Demand in Higher Education*, pp. 43–60. University of Chicago Press.

Manski, C. F. (1999). Analysis of Choice Expectations in Incomplete Scenarios. *Journal of Risk and Uncertainty 19*, 49–66.

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

Matzkin, R. L. (2013). Nonparametric identification in structural economic models. *Annu. Rev. Econ. 5*(1), 457–486.

McGee, R. (2021). Old age savings and house price shocks. Working paper.

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

Morita, T. and S. Managi (2015). Consumers' willingness to pay for electricity after the Great East Japan Earthquake. *Economic Analysis and Policy 48*, 82–105.

Mueller, A. I. and J. Spinnewijn (2022). Expectations data, labor market and job search. In *Handbook of Economic Expectations*, Chapter 21.

Mueller, A. I., J. Spinnewijn, and G. Topa (2021). Job seekers' perceptions and employment prospects: Heterogeneity, duration dependence, and bias. *American Economic Review 111*(1), 324–63.

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

O'Dea, C. and D. Sturrock (2021). Survival pessimism and demand for annuities. *Review of Economics and Statistics*.

Patnaik, A., J. Venator, M. Wiswall, and B. Zafar (2020). The role of heterogeneous risk preferences, discount rates, and earnings expectations in college major choice. *Journal of Econometrics*.

Pistaferri, L. (2001). Superior information, income shocks, and the permanent income hypothesis. *Review of Economics and Statistics 83*(3), 465–476.

Ruder, A. I. and M. Van Noy (2017, April). Knowledge of earnings risk and major choice: Evidence from an information experiment. *Economics of Education Review 57*, 80–90.

Schneider, U. (2020). Identifying and estimating beliefs from choice data–an application to female labor supply. Working paper.

Shoyama, K., S. Managi, and Y. Yamagata (2013). Public preferences for biodiversity conservation and climate-change mitigation: A choice experiment using ecosystem services indicators. *Land Use Policy 34*, 282–293.

Smith Jr, A. A. (1993). Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics 8*(S1), S63–S84.

Stinebrickner, R. and T. Stinebrickner (2014a). Academic performance and college dropout: Using longitudinal expectations data to estimate a learning model. *Journal of Labor Economics 32*(3), 601–644.

Stinebrickner, R. and T. R. Stinebrickner (2014b). A major in science? initial beliefs and final outcomes for college major and dropout. *Review of Economic Studies 81*(1), 426–472.

Teppa, F. and P. Lafourcade (2013). Can longevity risk alleviate the annuitization puzzle? Empirical evidence from survey data. Working paper.

Tincani, M. M., F. Kosse, and E. Miglino (2021). Subjective beliefs and inclusion policies: Evidence from college admissions. UCL Working Paper.

Van der Klaauw, W. (2012). On the Use of Expectations Data in Estimating Structural Dynamic Choice Models. *Journal of Labor Economics 30*(3), 521–554.

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

Van Soest, A. and H. Vonkova (2014). How sensitive are retirement decisions to financial incentives? a stated preference analysis. *Journal of Applied Econometrics 29*(2), 246–264.

Wang, Y. (2014). Dynamic implications of subjective expectations: Evidence from adult smokers. *American Economic Journal: Applied Economics 6*(1), 1–37.

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

Wiswall, M. and B. Zafar (2018). Preference for the workplace, investment in human capital, and gender. *Quarterly Journal of Economics 133*(1), 457–507.

Wolpin, K. and F. Gonul (1985). On the use of expectations data in micro surveys: The case of retirement. Report of the Employment and Training Administration, U.S. Department of Labor, Washington, DC.

Wu, S., R. Stevens, and S. Thorp (2015). Cohort and target age effects on subjective survival probabilities: Implications for models of the retirement phase. *Journal of Economic Dynamics and Control 55*, 39–56.

Zafar, B. (2013). College Major Choice and the Gender Gap. *Journal of Human Resources 48*(3), 545–595.