Estimating Individual Responses When Tomorrow Matters

Stéphane Bonhomme and Angela Denis
DECEMBER 2023
Estimating Individual Responses when Tomorrow Matters*

Stéphane Bonhomme       Angela Denis
University of Chicago    Banco de España
December 15, 2023

Abstract

We propose a regression-based approach to estimate how individuals’ expectations influence their responses to a counterfactual change. We provide conditions under which average partial effects based on regression estimates recover structural effects. We propose a practical three-step estimation method that relies on subjective beliefs data. We illustrate our approach in a model of consumption and saving, focusing on the impact of an income tax that not only changes current income but also affects beliefs about future income. Applying our approach to Italian survey data, we find that individuals’ beliefs matter for evaluating the impact of tax policies on consumption decisions.


Keywords: Dynamics, subjective expectations, beliefs, semi-structural estimation.

*We thank Adeline Delavande, Steven Durlauf, Pamela Giustinelli, Lars Hansen, Jim Heckman, Philip Heiler, Michael Keane, Arnaud Maurel, Pierre-Carl Michaud, Claudio Michelacci, Eduardo Morales, Chris Taber, Arne Uhlendorff, Wilbert Van der Klaauw, Daniel Wilhelm, and participants at various places for comments. Contact: sbonhomme@uchicago.edu; angela.denis@bde.es. The views expressed in this paper are those of the authors and do not necessarily reflect the position of the Banco de España or the Eurosystem.
1 Introduction

Economists often seek to assess how changes in the economic environment affect individual decisions. A leading example is the *ex ante* evaluation of policies that have not yet taken place. However, a key challenge is that, when the environment changes, individual decision rules are generally affected as well. In dynamic settings characterized by uncertainty, it is necessary to consider not only the immediate effect of the change but also its influence on expectations.

A common approach in applied work is to regress outcomes on covariates that one is interested in shifting in the counterfactual (e.g., under a new policy). Average partial effects based on regression estimates can be structurally interpreted as counterfactual policy effects under suitable conditions (Stock, 1989). However, underlying this interpretation is the assumption that the regression function remains invariant in the counterfactual. This invariance assumption can be restrictive in many settings where individuals’ beliefs about the future matter.

Consider the introduction of a permanent income tax in a standard model of consumption and saving (see Deaton, 1992, for a textbook treatment). The effect of the tax can be estimated by regressing consumption on income (in logs), and by then computing an average partial effect associated with the tax change. However, such an effect is likely to be empirically misleading, since both current income and beliefs about future income will be affected by the tax. Not accounting for the change in beliefs will produce biased predictions of the effect of the tax, as emphasized by Lucas (1976) in his influential critique.

As a second example, consider the effect of a change in the weather process in a model of agricultural production. Suppose that farmers choose dynamic inputs (such as irrigation or a fertilizer) based on their forecasts of future weather. In addition to affecting contemporaneous weather conditions, a change in the weather process will affect farmers’ beliefs about future weather, which may lead them to modify their input choices. Not accounting for farmers’ adaptation will bias calculations of the impact of a change in the weather process (Deschênes and Greenstone, 2007, Burke and Emerick, 2016).

In this paper, our aim is to study and estimate average partial effects in an analysis that explicitly accounts for the role of individual expectations. In our setup, individual beliefs are determinants of decisions, and they enter as additional state variables in the agent’s decision problem. In this setting, we show how to assess the total effect of a counterfactual change by means of average partial effects calculations. In addition, we show how to decompose this total effect into a contemporaneous effect where beliefs are held fixed, and a purely dynamic effect that solely reflects the change in beliefs.
To implement this approach we rely on data on subjective expectations. Beliefs data are increasingly available in a variety of settings (Manski, 2004). Given estimates of subjective densities based on such survey responses, we account for those in the definition and estimation of average partial effects. There are many examples of the use of beliefs data on the right-hand side of a regression. Our contribution is to show how to interpret the estimates of such regressions, and to provide conditions under which those can be used for counterfactual prediction.

To interpret regression-based average partial effects, we propose a structural dynamic framework where agents choose actions based on their beliefs about the future. Following a semi-structural approach, we use the framework to justify the use of average partial effects, yet we do not specify or estimate a structural model. As a result, the counterfactuals we focus on are restricted to changes in covariates and beliefs, and our approach cannot answer other counterfactual questions related to changes in preferences or technology, for example.

In the structural framework that we outline, beliefs are time-varying state variables in the agent’s decision problem. Variation in beliefs over time is crucial, since it allows us to control for preference heterogeneity, which we assume to be constant over time, by including individual fixed effects. We assume that current beliefs provide sufficient information to predict future beliefs, an assumption that we refer to as “belief sufficiency”. We show this assumption is compatible with various popular models of belief formation, with and without rational expectations, including various forms of learning.

The structural framework implies that the agent’s decision rule is a function of exogenous state variables such as income or the weather, beliefs about them, and endogenous dynamic state variables such as assets or capital. We assess the effects of a change in the exogenous state variables by computing average partial effects which, unlike in the static case, account for changes in beliefs. Such effects correspond to well-defined structural counterfactuals under the assumption that the dynamic decision rule is invariant to the change. Hence, while we rely on a weaker type of invariance assumption than standard average partial effects that do not allow for belief responses, a certain form of invariance is still needed to structurally interpret average partial effects in our setup.

To estimate average partial effects, we proceed in three easy-to-implement steps. In the first step, we estimate the belief densities. To account for the fact that survey responses on subjective beliefs tend to be coarse, we assume that belief densities depend on a low-dimensional parameter vector. However, we also describe semi- and nonparametric extensions that can be implemented with rich beliefs data. In the second step, we estimate the regression function (i.e., the individual’s decision rule). In the third step, we use these estimates to compute
the impact of a change in the covariates (i.e., of a change in the state variables). Without additional assumptions, nonparametric identification is limited to the empirical support of the conditioning covariates.

As an empirical illustration, we study how consumption decisions depend on current income and beliefs about future income. We rely on Italian data from the Survey on Household Income and Wealth (SHIW), which contain information on respondents’ probabilistic income expectations. We then use our approach to predict the impact of various counterfactual income taxes, involving transitory or permanent increases in marginal tax rates, and a change in the degree of progressivity of the tax. We find that, conditional on current income, income beliefs shape consumption responses, and that they matter for predicting the effects of income taxes.

**Related literature and outline.** Subjective belief data are commonly included on the right-hand side of regressions. For example, Guiso and Parigi (1999) study how a firm’s investment depends on its beliefs about future demand; Hurd, Smith, and Zissimopoulos (2004) study the effects of subjective survival probabilities on decisions about retirement and social security claims; Dominitz and Manski (2007) analyze how beliefs about equity returns affect portfolio choice; Bover (2015) studies how subjective expectations about future home prices affect car and secondary home purchases; and Attanasio, Cunha, and Jervis (2019) study how parental investment in children is influenced by beliefs about the production function. We provide assumptions under which such regressions can be interpreted structurally and used for counterfactuals. A different line of research considers (state-contingent) subjective expectations as dependent variables (Arcidiacono, Hotz, Maurel, and Romano, 2020, Giustinelli and Shapiro, 2019).

Our focus on the estimation of policy effects without a full structural model follows Marschak (1953), Ichimura and Taber (2000, 2002), and Keane and Wolpin (2002a,b), among others; see also Wolpin (2013). In our approach, we rely on subjective belief data and do not assume rational expectations.

There is a growing literature on the combination of structural models and subjective beliefs data, see among others Van der Klaauw and Wolpin (2008), Delavande (2008), Van der Klaauw (2012), Stinebrickner and Stinebrickner (2014), Wiswall and Zafar (2015), and Koşar and O’Dea (2022); see also the recently released handbook on economic expectations (Bachmann, Topa, and van der Klaauw, 2022). Our approach, which is tailored to specific counterfactuals, does not require to specify a full structural model.

Lastly, elicited beliefs about future income are increasingly available. Surveys with this
information include the SHIW in Italy, the Survey of Economic Expectations and the Survey of Consumer Expectations in the US, the Survey of Household Finances in Spain, and the Copenhagen Life Panel in Denmark, among others. Previous contributions using income belief data include, among others, Pistaferri (2001), Guiso, Jappelli, and Pistaferri (2002), and Kaufmann and Pistaferri (2009), who use data on income expectations in the SHIW in combination with models of consumption and saving; Stoltenberg and Uhlenorff (2022), who estimate a structural model with subjective income expectations using the same data; Lee and Sæverud (2023), who use data on subjective expectations and earnings realizations in Denmark to estimate a model where agents have partial information about earnings shocks; and Attanasio, Kovacs, and Molnar (2020), who combine data on subjective expectations with data on actual income and estimate an Euler equation for consumption.

The outline is as follows. In Section 2 we introduce average partial effects for dynamic settings. In Section 3 we describe a structural framework and discuss the interpretation of average partial effects in this context. We present two examples in Section 4. We study identification and estimation in Section 5, and we present our consumption application in Section 6. Finally, in Section 7 we describe some extensions of the approach. Replication files are available online.

2 Average partial effects when tomorrow matters

Consider an individual outcome $y_{it}$ that depends on some covariates $x_{it}$ and $z_{it}$. Suppose that, for some function $g_i$,

$$y_{it} = g_i(x_{it}, z_{it}) + \varepsilon_{it},$$

(1)

where $\varepsilon_{it}$ has zero mean given $x_{it}$ and $z_{it}$. To fix ideas, we will refer to the case where $y_{it}$ denotes consumption, $x_{it}$ is income, and $z_{it}$ includes other determinants such as assets.

Consider an exogenous change in $x_{it}$, from $x_{it} = x$ to some other value $x_{it} = x^{(\delta)}$. For example, if one is interested in a mean shift of (log) income by a $\delta$ amount, corresponding to a proportional tax or subsidy, one will set $x^{(\delta)} = x + \delta$. A standard average partial effect associated with the change in $x_{it}$ is then

$$\Delta_i^{\text{APE}}(\delta, x, z) = g_i(x^{(\delta)}, z) - g_i(x, z),$$

(2)

possibly averaged across individual observations. By estimating quantities such as $\Delta_i^{\text{APE}}$, one can document how individual responses vary across individuals and values of $x$ and $z$. 
However, to interpret \( \Delta_i^\text{APE} \) as the impact on outcomes when \( x_{it} \) changes from \( x \) to \( x^{(\delta)} \), one needs to assume that, as \( x_{it} \) changes while \( z_{it} \) is kept constant, the function \( g_i \) remains constant (Stock, 1989). This invariance assumption is often implausible in applications where dynamics matter. Indeed, in many settings where the current value of \( x_{it} \) changes, beliefs about future \( x_{it} \)'s, which are implicitly contained in the function \( g_i \), are likely to change as well. For example, under a (permanent) income tax, both current income and beliefs about future income change.

Our approach to alleviate this well-known issue is to augment (1) by including beliefs about future \( x_{it} \) values as additional determinants of \( y_{it} \). Letting \( \pi_{it} \) denote the subjective density of \( x_{i,t+1} \) at time \( t \), we postulate that, for some function \( \phi_i \),

\[
y_{it} = \phi_i(x_{it}, \pi_{it}, z_{it}) + \epsilon_{it},
\]

where \( \epsilon_{it} \) has zero mean given \( x_{it} \), \( \pi_{it} \) and \( z_{it} \). In the consumption example, this amounts to including income beliefs as additional determinants of the consumption function.

In model (3), we will be interested in documenting the effects of a change from \( x_{it} = x \) to \( x_{it} = x^{(\delta)} \), associated with a change in beliefs from \( \pi_{it} = \pi \) to \( \pi_{it} = \pi^{(\delta)} \). As an example, consider again the effect of shifting current (log) income \( x_{it} \) by a \( \delta \) amount. In this case, \( \pi^{(\delta)} \) is the belief about future income \( x_{i,t+1} \) under the \( \delta \) mean shift. The tax has two distinct effects on outcomes: a contemporaneous effect associated with the change in \( x_{it} \), and a dynamic effect associated with the change in beliefs \( \pi_{it} \).

In this setup, we define the total average partial effect, or TAPE, as

\[
\Delta_i^\text{TAPE}(\delta, x, \pi, z) = \phi_i(x^{(\delta)}, \pi^{(\delta)}, z) - \phi_i(x, \pi, z).
\]

We then further decompose this total effect as the sum of two terms: a contemporaneous APE (or CAPE), where beliefs are held constant, and a dynamic APE (or DAPE), which solely captures the change in beliefs. Formally, we decompose

\[
\Delta_i^\text{TAPE}(\delta, x, \pi, z) = \Delta_i^\text{CAPE}(\delta, x, \pi, z) + \Delta_i^\text{DAPE}(\delta, x, \pi, z) = \phi_i(x^{(\delta)}, \pi^{(\delta)}, z) - \phi_i(x, \pi, z) - \phi_i(x^{(\delta)}, \pi, z) + \phi_i(x^{(\delta)}, \pi^{(\delta)}, z) - \phi_i(x, \pi^{(\delta)}, z).
\]

To interpret \( \Delta_i^\text{TAPE} \) as the impact on outcomes when \( x_{it} \) changes from \( x \) to \( x^{(\delta)} \) and \( \pi_{it} \) changes from \( \pi \) to \( \pi^{(\delta)} \), one needs to assume that the function \( \phi_i \) remains invariant in the counterfactual. Although this assumption is not without loss of generality (and we will discuss it in the context of a structural framework in the next section), it is weaker than the assumption that \( g_i \) in (1) is invariant to the change. The key difference is that, unlike (1), (3) explicitly accounts for variation in beliefs.
In the next section we will describe a class of structural models under which (3) is the individual’s optimal decision rule in an intertemporal economic model. This will allow us to transparently discuss the assumptions needed to structurally interpret the above average partial effects (TAPE, CAPE and DAPE).

The structural framework has two main features. First, $\pi_{it}$ is sufficient to predict future beliefs $\pi_{i,t+1}$, as formally defined in Assumption 2 in the next section. This implies that $x_{it}$, $\pi_{it}$, and $z_{it}$ are the state variables in the economic model (in addition to some shocks subsumed in $\varepsilon_{it}$). This belief sufficiency assumption imposes restrictions on the belief formation process, however we show it is satisfied in several popular models of beliefs. Second, in the structural model $\phi_i$ depends on preferences, discounting, the law of motion of $z_{it}$, and the law of motion of the beliefs $\pi_{it}$. To guarantee the invariance of $\phi_i$, one will need to assume that none of these quantities varies under the policy change.

Assuming that the law of motion of the beliefs, which we denote as $\rho_i$, is invariant requires that, while agents account for the impact of the change on their beliefs about $x_{i,t+1}$, the way they update their beliefs after period $t+1$ is unaffected. Under this assumption, $\rho_i$ is an individual “type” that is invariant to the change. We will see that this assumption is automatically satisfied in a popular version of the consumption example.¹

Finally, note that, when beliefs matter in (3), an approach based on (1) is incorrect for two reasons. The first one is that beliefs $\pi_{it}$, which are generally correlated with $x_{it}$, are omitted variables in (1). Hence, not controlling for $\pi_{it}$ gives incorrect estimates of the contemporaneous APE. The second one is that relying on (1) makes it impossible to estimate the total APE, and to decompose it into contemporaneous and dynamic APEs. Hence, when (3) holds, $\Delta_i^{APE}$ defined in equation (2) is not economically interpretable in general.

### 3 Structural interpretation

In this section we describe a structural dynamic framework where individual decision rules take the form (3), and we provide a structural interpretation for average partial effects.

¹Relaxing this assumption is conceptually straightforward in our framework, by defining $\pi_{it}$ in (3) as beliefs about a sequence of future $x$’s, $x_{i,t+1}, x_{i,t+2}, ..., x_{i,t+S}$. However, doing so imposes stronger demands on the data. We will return to this point in Section 7.
3.1 Economic environment

Consider an individual $i$’s intertemporal decision making process in discrete time. In the presentation we focus on a stationary infinite-horizon environment. However, in Remark 1 we will show how to apply the framework to finite-horizon environments.

The timing is as follows. At the end of period $t-1$, the individual’s information includes the history of exogenous state variables $x_{i,t-1}, x_{i,t-2}, \ldots$, endogenous state variables $z_{i,t-1}, z_{i,t-2}, \ldots$, actions $y_{i,t-1}, y_{i,t-2}, \ldots$, and shocks $\nu_{i,t-1}, \nu_{i,t-2}, \ldots$. In addition, the individual may have observed other information, such as signals, that are relevant to her beliefs and future actions.

Then, at the beginning of period $t$, $z_{it}, x_{it}$ and $\nu_{it}$ are realized and observed by the individual, and additional signals about future values $x_{i,t+1}$ may be observed as well. We denote the information set at that moment as $\Omega_{it}$. Given this information, the individual forms beliefs about $x_{i,t+1}$. Finally, she chooses the action $y_{it}$ based on the state variables in $\Omega_{it}$.

We make the following assumption regarding beliefs, where we use $A \sim B$ to denote that $A$ and $B$ follow the same distribution, and densities are defined with respect to appropriate measures.

Assumption 1. (beliefs)

$$(x_{i,t+1} | y_{it}, \Omega_{it}) \sim (x_{i,t+1} | \Omega_{it}).$$

We denote the corresponding conditional density as $\pi_{it}(x_{i,t+1})$.

Assumption 1 requires that beliefs about $x_{i,t+1}$, which are relevant to the choice of $y_{it}$, do not depend on $y_{it}$. In other words, beliefs are not contingent on actions. At the same time, Assumption 1 allows past choices $y_{i,t-1}, y_{i,t-2}, \ldots$ to influence current beliefs $\pi_{it}$. We will outline a generalization where agents have state-contingent beliefs in Section 7. The framework is unchanged in this case, except for the fact that $\pi_{it}$ then consists of a set of conditional densities indexed by action values $y$.

In Assumption 1, $\pi_{it}$ is the subjective density of $x_{i,t+1}$, which we will refer to as the belief density, or simply beliefs. $\pi_{it}$ is included in $\Omega_{it}$, and is a random function. Note that we do not impose a rational expectations assumption, so perceived and realized laws of motion may not coincide. This is an important point since we will show that, when subjective expectations data are available, it is not necessary to impose rational expectations in order to estimate decision rules.

We make the following assumption regarding belief updating.

Assumption 2. (belief sufficiency)

$$(\pi_{i,t+1} | x_{i,t+1}, y_{it}, \Omega_{it}) \sim (\pi_{i,t+1} | x_{i,t+1}, y_{it}, \pi_{it}, x_{it}, \nu_{it}).$$
We denote the corresponding conditional density as \( \rho_i(\pi_{i,t+1}; x_{i,t+1}, y_{it}, \pi_{it}, x_{it}, \nu_{it}) \).

We will refer to \( \rho_i \) as the belief updating rule. Belief sufficiency, as stated by Assumption 2, is a key condition in our framework. It requires that current beliefs \( \pi_{it} \), along with \( x_{it} \) and \( \nu_{it} \), be sufficient statistics for \( \Omega_{it} \) when predicting future beliefs. Assumption 2 allows future beliefs \( \pi_{i,t+1} \) to depend on actions \( y_{it} \). However, a special case of the assumption is

\[
\left( \pi_{i,t+1} \mid x_{i,t+1}, y_{it}, \Omega_{it} \right) \sim \left( \pi_{i,t+1} \mid x_{i,t+1}, \pi_{it}, x_{it} \right),
\]

which implies that \( \pi_{it} \) is an exogenous process. We will assume (6) in our illustration to consumption and income.

We make the following assumption regarding the endogenous state variables \( z_{it} \).

**Assumption 3. (endogenous state variables)**

\[
z_{i,t+1} = \gamma_i(z_{it}, x_{it}, y_{it}),
\]

where \( \gamma_i \) is a non-stochastic function.

Assumption 3 can be generalized in various ways without affecting the rest of the framework. For example, \( z_{it} \) might depend on beliefs \( \pi_{i,t-1} \), or on an idiosyncratic i.i.d. shock whose distribution is known to the agent.

Lastly, we make the following assumption regarding the shocks \( \nu_{it} \).

**Assumption 4. (shocks)**

\[
\left( \nu_{i,t+1} \mid x_{i,t+1}, \pi_{i,t+1}, y_{it}, \Omega_{it} \right) \sim \nu_{i,t+1}.
\]

We denote the corresponding density as \( \tau_i(\nu_{i,t+1}) \).

### 3.2 Compatibility with belief formation models

We now illustrate that our belief sufficiency condition, Assumption 2, is consistent with several models of belief formation in economics, see Pesaran and Weale (2006) for references.

As a first example, suppose that agents have rational expectations, and that agent \( i \)'s information set at time \( t \) is \( \Omega_{it} = \{ x_{it}, x_{i,t-1}, ..., \eta_{it}, \eta_{i,t-1}, ... \} \), where \( x_{it} = \eta_{it} + \varepsilon_{it} \), \( \eta_{it} \) is homogeneous first-order Markov, and \( \varepsilon_{it} \) is independent of \( \eta_{it} \) with a stationary distribution.

---

2 The framework is unchanged if, in addition, \( \pi_{i,t+1} \) depends on \( z_{it} \), in which case the belief updating rule is \( \rho_i(\pi_{i,t+1}; x_{i,t+1}, y_{it}, \pi_{it}, x_{it}, z_{it}, \nu_{it}) \).
An example is a permanent-transitory specification of the income process, which we will study in our consumption example. Since \( \pi_{it} \) is the conditional density of \( x_{i,t+1} \) given \( \Omega_{it} \), it coincides with the conditional density of \( x_{i,t+1} \) given \( \eta_{it} \). Given that \( \eta_{it} \) is an exogenous first-order Markov process, Assumption 2 is thus satisfied. However, it generally fails if \( \eta_{it} \) is not first-order Markov.

As a second example, suppose that \( x_{it} = \alpha_i + \varepsilon_{it} \), yet agents do not know \( \alpha_i \) and try to learn it given the observations \( x_{it} \). We show in Appendix A that, when \( \varepsilon_{it} \) is Gaussian and Bayesian agents have Gaussian priors about \( \alpha_i \) and rational expectations, belief sufficiency, as stated by Assumption 2, holds. Note that this example does not allow for learning from past choices, since beliefs are exogenous.

As a third example, consider a case where there are two possible choices \( y_{it} = 1 \) and \( y_{it} = 0 \). Suppose that the agent observes \( x_{it} = \alpha_i + \varepsilon_{it} \) no matter what action she chooses, and that she observes an additional signal \( s_{it} = \alpha_i + v_{it} \) only when choosing \( y_{i,t-1} = 1 \). We show in Appendix A that, when \( (\varepsilon_{it}, v_{it}) \) is Gaussian and agents have rational expectations and a Gaussian prior about \( \alpha_i \), Assumption 2 is satisfied. In this example, beliefs are endogenous, affected by past choices.\(^3\)

Our setup is also compatible with some models of non-rational expectations. As a fourth example, consider a simple model of adaptive expectations, where mean beliefs evolve as

\[
\mathbb{E}_{\pi_{it}}(x_{i,t+1}) = \mathbb{E}_{\pi_{i,t-1}}(x_{it}) + \lambda_i \left( x_{it} - \mathbb{E}_{\pi_{i,t-1}}(x_{it}) \right). \tag{7}
\]

Armona, Fuster, and Zafar (2019) refer to individuals with \( \lambda_i > 0 \) as “extrapolators”, to those with \( \lambda_i = 0 \) as “non-updators”, and to those with \( \lambda_i < 0 \) as “mean reverters”. Assumption 2 is satisfied if (7) holds, and, say, beliefs are normally distributed with constant variance \( \sigma_i^2 \). More generally, Assumption 2 is consistent with models of adaptive expectations where the entire belief density \( \pi_{it} \) depends on \( \pi_{i,t-1} \) and \( x_{it} \).

This discussion provides several examples of belief formation models where belief sufficiency, as stated by Assumption 2, holds. Under this assumption, along with Assumptions 1, 3 and 4, the vector \( (x_{it}, \pi_{it}, z_{it}, \nu_{it}) \) contains all the relevant state variables when making the decision. An advantage of our approach is that, since beliefs \( \pi_{it} \) are state variables, we can perform counterfactual exercises that account for changes in beliefs without the need for a full-fledged structural model.

\(^3\)However, beliefs are not state-contingent in this example, and Assumption 1 holds. We will show in Section 7 that our framework can be extended to allow for state-contingent beliefs, and we will provide a learning model as an illustration.
3.3 Decisions and policy rule

Let $u_i(y_{it}, x_{it}, z_{it}, \nu_{it})$ denote period $t$’s contemporaneous payoffs.$^4$ Here the action may be continuous or discrete, so our framework covers structural dynamic discrete choice models as well as models with continuous choices. It also covers settings with vector-valued actions, including mixed discrete-continuous choices (e.g., Bruneel-Zupanc, 2022). Let $\beta_i$ denote the time discount factor. The individual solves the infinite horizon program

$$(y_{i1}, y_{i2}, \ldots) = \max_{(y_1, y_2, \ldots)} \left[ \sum_{t=1}^{\infty} \beta_i^{t-1} u_i \left(y_t, x_t, z_t, \nu_t \right) \right],$$

where, by Assumptions 1, 2, 3, and 4, the expectation is taken with respect to the process of $x_{it}, \pi_{it}, z_{it}, \nu_{it}$.

Let $V_i(x, \pi, z, \nu)$ denote the value function associated with any given state $(x, \pi, z, \nu)$. Bellman’s principle then implies

$$V_i(x_t, \pi_t, z_t, \nu_t) = \max_{y_t} \left\{ u_i(y_t, x_t, z_t, \nu_t) \right\}$$

$$+ \beta_i \int V_i(x_{t+1}, \pi_{t+1}, \gamma_i(z_t, x_t, y_t), \nu_{t+1}) \pi_i(x_{t+1}) \rho_i(\pi_{t+1}; x_{t+1}, y_t, \pi_t, x_t, \nu_t) \tau_i(\nu_{t+1}) dx_{t+1} d\pi_{t+1} d\nu_{t+1} \right\}.$$

(8)

The implied policy rule for actions is then, under suitable regularity conditions (e.g., Stokey, Lucas, and Prescott, 1989),

$$y_{it} = \phi \left(x_{it}, \pi_{it}, z_{it}, \nu_{it}, \rho_i, u_i, \beta_i, \gamma_i, \tau_i \right),$$

(9)

for some function $\phi$. Then, let

$$\phi_i \left(x_{it}, \pi_{it}, z_{it} \right) = \int \phi \left(x_{it}, \pi_{it}, z_{it}, \nu_{it}, \rho_i, u_i, \beta_i, \gamma_i, \tau_i \right) \tau_i(\nu_{it}) d\nu_{it}$$

denote the average decision rule with respect to the shocks $\nu_{it}$. It follows from Assumption 4 that

$$\phi_i \left(x_{it}, \pi_{it}, z_{it} \right) = \mathbb{E} \left[ \phi \left(x_{it}, \pi_{it}, z_{it}, \nu_{it}, \rho_i, u_i, \beta_i, \gamma_i, \tau_i \right) \mid x_{it}, \pi_{it}, z_{it} \right].$$

$^4$Here $\pi_{it}$ are not payoff-relevant. However, the nonparametric decision rule below will remain the same if payoffs $u_i(y_{it}, x_{it}, \pi_{it}, z_{it}, \nu_{it})$ depend on $\pi_{it}$.

$^5$Here the integral in $(x_{t+1}, \pi_{t+1}, \nu_{t+1})$ is taken relative to an appropriate measure.

$^6$We treat $\rho_i, u_i, \beta_i, \gamma_i, \tau_i$ as non-random quantities.
Hence, (3) holds for $\varepsilon_i = y_i - \phi_i(x_{it}, \pi_{it}, z_{it})$, which has zero mean given $x_{it}, \pi_{it}, z_{it}$. In this framework, $\phi_i$ in (3) can thus be interpreted as the individual’s decision rule averaged over the shocks $\nu_{it}$.7

Remark 1. (Finite horizon)

In a finite horizon environment where $t \in \{1, ..., T\}$, the Bellman equation (8) becomes, for $t < T_i$ and some terminal value $V_{i,T_i},$

$$V_i(x_t, \pi_t, z_t, \nu_t) = \max_{y_t} \left\{ u_i(y_t, x_t, z_t, \nu_t) \right. + \beta_i \int V_{i,t+1}(x_{t+1}, \pi_{t+1}, \gamma_i(z_{t+1}, x_{t+1}, y_{t+1}), \nu_{t+1}) \pi_t(x_{t+1}) \rho_i(\pi_{t+1}; x_{t+1}, y_{t+1}, \pi_t, x_t, \nu_t) \tau_i(\nu_{t+1}) dx_{t+1} d\pi_{t+1} d\nu_{t+1} \left\}.$$

Here, differently from the infinite-horizon setup, the transitions $\rho_i$ between $\pi_i$ and $\pi_{i,t+1}$ are time-specific, and $z_{i,t+1} = \gamma_i(z_{it}, x_{it}, y_{it})$. In this case, actions take the form

$$y_i = \phi_i(x_{it}, \pi_{it}, z_{it}) + \varepsilon_{it},$$

where the dependence of $\phi$ on $(i,t)$ stems from the presence of $u_i$, $\beta_i$, $\tau_i$, the terminal value $V_{i,T_i}$, and the $\rho_{is}$ and $\gamma_{is}$ in all periods $s \geq t$. A difference with the infinite-horizon case is that, since $\phi_{it}$ is time-varying, it is no longer possible to identify individual responses while leaving the individual heterogeneity fully unrestricted.

3.4 Interpreting average partial effects

Structurally interpreting an average partial effect as the effect of a counterfactual change requires $\phi_i$ to remain invariant in the counterfactual. We now discuss this invariance condition.

Keeping $u_i$ and $\beta_i$ constant requires assuming that $u_i$ (such as preferences) and $\beta_i$ (discounting) are invariant to changes in the environment. This is a common assumption in dynamic structural models. Invariance of the density of taste shocks $\tau_i$ is also commonly assumed. In turn, keeping $\gamma_i$ constant requires assuming that the process through which past actions and states feed back onto future $z_{it}$ values is invariant in the counterfactual. When $z_{it}$ is a stock that depreciates over time or an asset with some return, for example, this requires assuming away the presence of general equilibrium effects through which the return or the depreciation rate might change in the counterfactual.

7It is straightforward to include additional state variables in (9), under the assumption that beliefs about them are constant and invariant to counterfactual changes. Accounting for additional state variables can be empirically relevant, and we will include a number of such variables as controls in our application.
In addition, as our framework makes clear, structurally interpreting average partial effects generally requires assuming that the belief updating rule \( \rho_i \) remains constant in the counterfactual. A change in \( \rho_i \) corresponds to a steady-state or “long-run” counterfactual where the entire process of \( x_{it} \), as perceived by the agent, changes. In our setup, we allow for policies or other counterfactuals to affect beliefs \( \pi_{it} \), yet we assume that the belief updating rule \( \rho_i \) is an individual characteristic that remains unaffected. In Section 7 we will describe how to extend the approach to account for beliefs over longer horizons, hence making the invariance assumption about \( \rho_i \) less restrictive. Our focus on counterfactuals involving changes in \( x_{it} \) and \( \pi_{it} \), while \( \rho_i \) is kept constant, can be viewed as an intermediate case between a static counterfactual where only \( x_{it} \) varies, and a long-run, steady-state counterfactual where the entire long-run belief process, including the belief updating rule \( \rho_i \), is allowed to vary.\(^8\)

4 Examples

In this section, we describe two examples of our framework. In the first one, we consider a model of consumption, savings, and income, with the aim to assess the effects on consumption of a change in the income process. In the second example, we outline a model of agricultural production that allows farmers to adapt to new climate, with the goal to estimate the effects of current and expected weather.

4.1 Consumption, saving, and income

In the first example, we consider a standard incomplete markets model of consumption and saving behavior. For simplicity, we focus on infinite-horizon environment, as in Chamberlain and Wilson (2000), although the analysis can easily be adapted to a life-cycle environment.

In the model, \( y_{it} \) is household \( i \)'s log consumption in period \( t \), and household utility over consumption is \( u_i(y_{it}, \nu_{it}) \), where \( u_i \) is an increasing utility function and \( \nu_{it} \) are i.i.d. taste shocks with density \( \tau_i \). Household \( i \)'s discount factor is \( \beta_i \). Log income \( x_{it} \) and beliefs \( \pi_{it} \) about \( x_{i,t+1} \) are exogenous, and Assumptions 1 and 2 hold. Households can self-insure using a risk-free

\(^8\)To identify such long-run counterfactuals in a semi-structural, regression-based approach, one would need to recover the effect of the belief updating rule \( \rho_i \) on decisions. This would require the availability of empirical counterparts for \( \rho_i \), as well as suitable cross-sectional exogeneity assumptions (or a valid instrument for \( \rho_i \)). Both conditions would impose strong demands on the data. In particular, \( \rho_i \) is a subjective process perceived by the agent, which is not directly informed by responses to subjective expectations questions (since \( \rho_i \) need not coincide with the process of realized beliefs \( \pi_{it} \)).
bond with constant interest rate $r_i$, and assets $z_{it}$ follow

$$z_{i,t+1} = (1 + r_i)(z_{it} + w_{it}) - c_{it}, \quad (10)$$

where $w_{it} = \exp(x_{it})$ and $c_{it} = \exp(y_{it})$ denote income and consumption, respectively. As in (9), the (log) consumption rule takes the form

$$y_{it} = \phi(x_{it}, \pi_{it}, z_{it}, \psi_{it}, \rho_i, u_i, \beta_i, r_i, \tau_i).$$

As a specific example for the income process perceived by the agent, consider a permanent-transitory model (e.g., Hall and Mishkin, 1982):

$$x_{it} = \eta_{it} + u_{it}, \quad \eta_{it} = \eta_{i,t-1} + \nu_{it},$$

where $u_{it} \sim N(0, \sigma^2_{iu})$ and $\nu_{it} \sim N(0, \sigma^2_{iv})$ are independent over time and independent of each other at all leads and lags. At time $t$, the agent observes $x_{it}$ and $\eta_{it}$, but neither $x_{i,t+1}$ nor $\eta_{i,t+1}$. In this case, we have

$$\pi_{it}(x) = \frac{1}{\sqrt{\sigma^2_{iu} + \sigma^2_{iv}}} \varphi\left(\frac{x - \eta_{it}}{\sqrt{\sigma^2_{iu} + \sigma^2_{iv}}}\right), \quad (11)$$

where $\varphi$ is the standard Gaussian density, and Assumption 2 holds (in fact, beliefs are exogenous in this case, and the stronger condition (6) holds). In this specific example, only the mean of $\pi_{it}$ varies over time and its variance is constant.

Suppose we wish to assess the impact on consumption of a proportional income tax $T(w) = (1 - \delta)w$ introduced at time $t$, where recall that $w = \exp(x)$ denotes household income. Under the tax, log income is thus $x^{(\delta)} = x + \log \delta$. Suppose households believe the tax change will continue being implemented in the future, and they fully adjust their beliefs to the tax. When $\pi_{it}$ is given by (11) in the absence of the tax, implementing the tax will lead to the new beliefs

$$\pi_{it}^{(\delta)}(x) = \frac{1}{\sqrt{\sigma^2_{iu} + \sigma^2_{iv}}} \varphi\left(\frac{x - \eta_{it} - \log \delta}{\sqrt{\sigma^2_{iu} + \sigma^2_{iv}}}\right).$$

Hence, the tax affects both the mean of log income and the perceived conditional mean of future log income.

In this model, a proportional tax does not affect the belief updating rule $\rho_i$.\(\textsuperscript{10}\) Hence, the total APE fully captures the effect of the tax on consumption. In this case, the contemporaneous

\(\textsuperscript{9}\)Alternatively, in a finite-horizon environment, we obtain a counterpart to this equation, as in Remark 1, which involves a time-varying $\phi_i$.

\(\textsuperscript{10}\)Indeed, the introduction of the tax is isomorphic to a change in the permanent component, from $\eta_{it}$ to $\eta_{it}^{(\delta)} = \eta_{it} + \log \delta$. Moreover, the distribution of $(x_{i,t+1}, \eta_{i,t+1})$ given $(x_{it}, \eta_{it})$ does not change under the tax.
APE corresponds to the effect of a purely transitory tax at $t$ that will disappear at $t + 1$; equivalently, it is the effect of a log $\delta$-shift in the transitory income shock $u_{it}$. In turn, the dynamic APE can be interpreted as the effect of a tax that is announced at $t$ and will be implemented at $t + 1$. Lastly, the total APE, which is the sum of the contemporaneous and dynamic APEs, corresponds to the effect of a log $\delta$-shift in the permanent income shock $v_{it}$.

This model relies on specific assumptions about the income process, information, and beliefs. Those assumptions could be incorrect; for example, agents might have different beliefs about future income. In our approach we do not assume that the consumption model with permanent-transitory income beliefs describes the data. However, interpreting an average partial effect as the structural effect of a counterfactual tax requires that, while beliefs $\pi_{it}$ are affected by the tax, the belief updating rule $\rho_i$ is not.

4.2 Structural and semi-structural tax counterfactuals: a comparison

To illustrate how structural modeling and our approach relate to each other in the context of this example, we simulate a large sample from a life-cycle model of consumption and savings based on Kaplan and Violante (2010), where identical, risk-averse households save to smooth consumption while facing borrowing constraints. We study two versions of the model, with rational and adaptive expectations, respectively. In both cases, income beliefs, which are key state variables in the model, can be summarized by their time-varying means, which follow a first-order Markov process jointly with log income.

Under both versions of the model, we compute the true effect of a 10% permanent proportional income tax, and we decompose it under the model into a contemporaneous effect due to current income and a dynamic effect due to beliefs. Then, we compare these counterfactual predictions with our average partial effects (TAPE, CAPE, and DAPE), which we obtain by estimating consumption regressions in the simulated sample. Since the model has a finite horizon, the consumption function $\phi_t$ is age-dependent, and we proxy for this dependence by controlling for age and its square in the regressions. Note that, as we discussed, the belief updating rule $\rho_i$ is invariant under the counterfactual in the rational expectations version of the model. In the adaptive expectations version we assume that invariance is satisfied as well. We provide details about the model, parameter values, and calculation of counterfactuals in Appendix C.

11 The DAPE in (5) is evaluated at income $x^{(b)}$ after the tax, so that the CAPE and the DAPE add up to the TAPE. It is also possible to compute an alternative DAPE evaluated under income $x$ before the tax, $\Delta_i^{DAPE}(\delta, x, \pi, z) = \phi_t(x, \pi^{(b)}, z) - \phi_t(x, \pi, z)$.  

15
Table 1: Tax counterfactuals under rational and adaptive expectations

<table>
<thead>
<tr>
<th></th>
<th>Rational expectations</th>
<th>Adaptive expectations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Structural</td>
<td>Semi-structural</td>
</tr>
<tr>
<td></td>
<td>Linear</td>
<td>Quadratic</td>
</tr>
<tr>
<td>CAPE</td>
<td>-0.0163</td>
<td>-0.0151</td>
</tr>
<tr>
<td>DAPE</td>
<td>-0.0802</td>
<td>-0.0917</td>
</tr>
<tr>
<td>TAPE</td>
<td>-0.0965</td>
<td>-0.1068</td>
</tr>
</tbody>
</table>

Notes: Effects of a 10% permanent income tax on log consumption in two model economies, where households have rational (in the left panel) or adaptive expectations (in the right panel), respectively. In both economies, log income follows a permanent-transitory process. For the structural counterfactuals we compute the effect of the tax under the model. For semi-structural ones we regress log consumption on log income, income belief and its interaction with log income, age, age squared, and a function of log assets (linear, quadratic, or 20-knot spline depending on the specification). Households with positive assets, age 26–49.

We report the counterfactual calculations in Table 1. Focusing first on the version with rational expectations (in the left panel), the model predicts a decrease in log consumption of $-0.097$, which is almost one-for-one with the tax increase, as is expected in this model, and a large part can be attributed to a change in beliefs. The semi-structural predictions, which do not rely on the knowledge of the structure and the parameter values of the structural model but are computed using regressions, come close to these numbers. We report the results of three specifications, where we control for linear, quadratic, or spline functions of log assets, and all of them give comparable results in this case.

Turning next to the version with adaptive expectations (in the right panel), the model predicts a smaller effect of the tax ($-0.062$), given the expectations process that we assume. When using a structural approach to predict counterfactuals, specifying belief formation correctly is key. However, the semi-structural predictions, which do not rely on correct specification of the model (including the belief formation part of the model), again come close to the tax effects, albeit in this case only when the regression specification is flexible enough (i.e., quadratic or spline).

### 4.3 Weather and agricultural production

In the second example, we consider a model of agricultural production with costly investment. Output $q_{i,t+1} = g_i(x_{i,t+1}, k_{i,t+1})$ depends on the weather $x_{i,t+1}$ and on a dynamic input $k_{i,t+1}$.
(such as capital). The weather $x_{it}$, and farmer $i$’s beliefs $\pi_{it}$ about $x_{i,t+1}$, satisfy Assumptions 1 and 2. The farmer can invest $y_{it}$ in the dynamic input $k_{it}$ at a cost $c_i(y_{it}, \nu_{it})$, for some i.i.d. cost shifters $\nu_{it}$ with density $\tau_i$. The dynamic input follows the law of motion $k_{i,t+1} = (1 - \delta_i)k_{it} + y_{it}$. The farmer decides on $y_{it}$ after observing today’s weather $x_{it}$ and her beliefs $\pi_{it}$ about tomorrow’s weather, but before observing $x_{i,t+1}$. Lastly, the instantaneous profit in period $t$ is $q_{it} - c_i(y_{it}, \nu_{it})$, and the farmer’s discount factor is $\beta_i$.

The state variables of the decision problem are $x_{it}$, $\pi_{it}$, $k_{it}$, and $\nu_{it}$, and, under suitable regularity conditions, the optimal investment rule takes the form

$$y_{it} = \phi(x_{it}, \pi_{it}, k_{it}, \nu_{it}, \rho_i, \beta_i, c_i, \delta_i, g_i, \tau_i),$$

(12)

for some function $\phi$. Substituting (12) into the output equation, output in period $t + 1$ can thus be written as

$$q_{i,t+1} = \tilde{\phi}(x_{i,t+1}, x_{it}, \pi_{it}, k_{it}, \nu_{it}, \rho_i, \beta_i, c_i, \delta_i, g_i, \tau_i),$$

(13)

for some function $\tilde{\phi}$. The presence of $\pi_{it}$ in (12) and (13) reflects that the farmer may adapt to the prospect of harmful weather in the future by investing today.\textsuperscript{12}

The production function in (13) motivates regressing output on current and past weather and on the weather beliefs. Exploiting changes over time in $x_{it}$ and $\pi_{it}$, within farmer, is robust to the presence of individual heterogeneity. As an application, one can estimate our belief-augmented average partial effects to assess the impact of a change in the weather process that affects both weather realizations and weather beliefs. In this case as well, structurally interpreting the total APE as reflecting the total effect of such a change relies on the assumption that $\rho_i$, the belief updating process, is invariant. While this assumption may be tenable in the short or medium run, the total APE will not capture the full impact of long-run changes in the climate under which $\rho_i$ could be affected.

\textsuperscript{12}Farmers’ adaptation has been studied in the literature using various approaches. Burke and Emerick (2016) rely on a long-difference approach to account for farmers’ responses to a changing climate. Shrader (2020) proposes a framework to account for adaptation in a model where, in contrast with our dynamic framework, the firm’s current choice does not affect outcomes (i.e., profit) in later periods. See also Dell, Jones, and Olken (2014) and Keane and Neal (2020). Other approaches in the literature rely on specific aspects of the production model, such as envelope condition arguments (Hsiang, 2016, Lemoine, 2018, Gammans, Mérel, Paroissien, et al., 2020).
5 Estimating average partial effects

In this section we study identification and estimation of $\phi_i$ and average partial effects based on model (3).

5.1 Specification and identification

We impose the following mean independence condition,

$$\mathbb{E}[\varepsilon_{it} | x_{it}, \pi_{it}, z_{it}] = 0. \quad (14)$$

Note that (14) is satisfied in the structural framework of Section 3. To enhance the plausibility of this condition in applications, one can control for additional time-varying regressors (which can be interpreted as additional state variables), as well as for time-invariant fixed-effects. We will account for both factors in our empirical application.\(^{13}\)

Our approach to the measurement of beliefs $\pi_{it}$ relies on data about respondents’ expectations. Eliciting such responses is becoming increasingly common following the work of Dominitz and Manski (1997), see Manski (2004) for a review. Responses to subjective probabilistic expectations questions provide information about some features of $\pi_{it}$. Typically, the responses can be interpreted as some functionals $m_{it} = m(\pi_{it})$, such as the mean, variance, or some other moments of $\pi_{it}$.

We assume that $\pi_{it}$ is parametrically specified; that is, that there exists a finite-dimensional vector $\theta_{it}$ such that

$$\pi_{it} = \pi(\cdot ; \theta_{it}), \quad (15)$$

where $\pi(\cdot ; \theta)$ is known given $\theta$. Parametric specifications are commonly used in the literature on subjective expectations, due to the fact that expectations data are often coarse. For example, in the 1995 and 1998 waves of the SHIW in Italy, respondents are asked about the minimum and maximum earnings that they expect to receive if employed in the following year, together with the probability that their earnings will be below the mid-point between those two values. Using these data, Kaufmann and Pistaferri (2009) assume that income beliefs follow a triangular distribution conditional on employment, so that $\theta_{it}$ in (15) contains three elements. At the end of this section, we will discuss how one could relax the parametric specification on $\pi_{it}$ with rich enough data on beliefs.

\(^{13}\)In certain applications, (14) may not be plausible, but one may have access to instruments $w_{it}$ (e.g., instruments that exploit some policy variation in sample) such that $\mathbb{E}[\varepsilon_{it} | w_{it}] = 0$. Identification of $\phi_i$ will then require suitable relevance conditions (see Newey and Powell, 2003).
Given (3), (14), and (15), we have

$$
\phi_i(x_{it}, \pi_{it}, z_{it}) = \mathbb{E}[y_{it} \mid x_{it}, \theta_{it}, z_{it}],
$$

so \(\phi_i(x, \pi, z)\) is identified for all \(x, \pi, z\) in the empirical support of \(x_{it}, \pi_{it} = \pi(\cdot; \theta_{it})\), and \(z_{it}\). In turn, given \(\phi_i\), average partial effects (TAPE, CAPE and DAPE) are all identified, provided the support of covariates after the change in \(x_{it}\) and \(\pi_{it}\) lies within the support before the change.

The definition of an average partial effect depends on a change in \(x_{it}\) and an associated change in beliefs \(\pi_{it}\). We assume that beliefs remain in the same parametric family after the change, so \(\pi_{it}^{(\delta)}\). As a benchmark, we assume that the individual fully incorporates the effect of the change from \(x\) to \(x^{(\delta)}\) in her beliefs, and set

$$
\theta^{(\delta)} = \arg\max_{\tilde{\theta}} \mathbb{E} \left[ \log \left( \pi \left( x_{t+1}^{(\delta)} ; \tilde{\theta} \right) \right) \right],
$$

where the expectation is taken with respect to \(\pi(x_{t+1}; \theta)\), the belief density before the change. As an example, suppose \(x\) is log income before a counterfactual tax and \(x^{(\delta)} = x + \delta\) is post-tax income. Suppose \(\pi_{it}\) is normal with mean \(\mu_{it}\) and variance \(\sigma^2_{it}\), so \(\theta_{it} = (\mu_{it}, \sigma^2_{it})\). Under (17), \(\pi_{it}^{(\delta)}\) remains normal after the tax, with mean and variance \(\theta_{it}^{(\delta)} = (\mu_{it} + \delta, \sigma^2_{it})\).

**Remark 2.** (Partial adjustment of beliefs)

One can define average partial effects associated with other changes in beliefs. For example, assuming that individuals face a cost of adjusting their beliefs that is proportional to the Kullback-Leibler divergence between the beliefs before and after the change, one can replace (17) by

$$
\theta^{(\delta)} = \arg\max_{\tilde{\theta}} \left\{ \mathbb{E} \left[ \log \left( \pi \left( x_{t+1}^{(\delta)} ; \tilde{\theta} \right) \right) \right] - \xi \text{KL} (\tilde{\theta}, \theta) \right\},
$$

where \(\text{KL} (\tilde{\theta}, \theta) = \mathbb{E} \left[ \log \left( \frac{\pi(x_{t+1}; \theta)}{\pi(x_{t+1}; \tilde{\theta})} \right) \right]\). According to (18), \(\theta^{(\delta)}\) is given by (17) when the adjustment cost \(\xi\) is zero, \(\theta^{(\delta)} = \theta\) unchanged when the cost is infinite, and the individual partially adjusts her beliefs for intermediate values of \(\xi\). For example, consider a change \(x_{it}^{(\delta)} = x_{it} + \delta\). If \(\pi_{it}\) is normal with mean and variance \(\theta_{it} = (\mu_{it}, \sigma^2_{it})\), then \(\pi_{it}^{(\delta)}\) has mean and variance \(\theta_{it}^{(\delta)} = \left( \mu_{it} + \delta, \sigma^2_{it} + \xi \left( \frac{\delta}{1+\xi} \right)^2 \right)\). In our application we will focus on the benchmark case \(\xi = 0\) where individuals fully adjust their beliefs (while also commenting on the case \(\xi = \infty\), which corresponds to the contemporaneous APE where beliefs are held fixed). Alternatively,
when empirical variation in policies and beliefs is available, one could rely on such variation to discipline $\xi$.\footnote{Such empirical variation may take the form of hypothetical questions. For example, Briggs, Caplin, Leth-Petersen, Tonetti, and Violante (2020) and Roth, Wiederholt, and Wohlfart (2023) elicit individual expectations under various policy counterfactual scenarios.}

**Remark 3.** (Identification in short panels)

Here we focus on identifying $\phi_i$ and average partial effects for every individual $i$, which is relevant for applications with a long time dimension. In short panels, it is not possible to allow for unrestricted individual heterogeneity in $\phi_i$. A simple approach is to replace (3) by

$$y_{it} = \phi(x_{it}, \pi_{it}, z_{it}) + \alpha_i + \varepsilon_{it},$$

(19)

where $\phi$ is common across individuals, and $\alpha_i$ is an additive individual fixed effect. Under suitable exogeneity assumptions,\footnote{For example, if $(x_{it}, \pi_{it})$ are strictly exogenous and $z_{it}$ are predetermined, one can replace (14) by

$$\mathbb{E}[\varepsilon_{it} \mid x_{iT}, \pi_{iT}, \ldots, x_{i1}, \pi_{i1}, z_{iT}, z_{i,t-1}, z_{i,1}] = 0.$$}

identification of $\phi$ can be based on moment restrictions (e.g., Arellano and Bond, 1991).

## 5.2 Estimation

For estimation we proceed in three steps. First, we estimate the parameters $\theta_{it}$ that govern the belief density. Assuming that subjective expectations responses $m_{it} = m(\pi_{it})$ are available, a minimum-distance estimator solves

$$\hat{\theta}_{it} = \arg\min_{\theta} d(m_{it}, m(\pi(\cdot; \theta))),$$

where $d$ is some distance function (e.g., Euclidean).

In the second step, we estimate $\phi_i$ as the conditional expectation function in (16). Many approaches are available. For example, Stock (1989) proposes a partially linear semiparametric approach. We will rely on an linear specification of $\phi_i$ in a basis of functions,

$$\phi_i(x, \theta, z; \alpha) = \sum_{r=1}^{R} \alpha_{ir} P_r(x, \theta, z),$$

(21)

where $P_r$ is a family of functions, such as polynomials, and $R$ is the number of terms. In short panels (as in our application), we restrict $\alpha_{ir}$ not to depend on $i$, except the coefficient that corresponds to the intercept in the regression (see Remark 3).
Given observations $y_{it}, x_{it}, z_{it}$ and estimates $\hat{\theta}_{it}$, for $i = 1, \ldots, n$ and $t = 1, \ldots, T$, we estimate $\alpha_{ir}$ using penalized least squares regression

$$\hat{\alpha} = \arg\min_{\alpha} \sum_{i=1}^{n} \sum_{t=1}^{T} \left( y_{it} - \sum_{r=1}^{R} \alpha_{ir} P_r(x_{it}, \hat{\theta}_{it}, z_{it}) \right)^2 + \text{Pen}(\alpha). \quad (22)$$

In our empirical application we will rely on two choices for the penalty term: no penalty (i.e., $\text{Pen}(\alpha) = 0$) so the estimator is simply OLS, and an $\ell^1$ penalty (i.e., $\text{Pen}(\alpha) = \lambda \sum_{i,r} |\alpha_{ir}|$) corresponding to the Lasso estimator.

Lastly, in the third step we estimate counterfactuals by plugging in the estimates $\hat{\theta}_{it}$ and $\hat{\alpha}_{ir}$ in the APE formulas, averaged over individuals and time periods. For example, we estimate the total APE averaged over individuals and time periods as

$$\hat{\Delta}_{\text{TAPe}}(\delta) = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \sum_{r=1}^{R} \hat{\alpha}_{ir} \left( P_r(x_{i(t), \hat{\theta}_{it}, z_{it}}) - P_r(x_{it}, \hat{\theta}_{it}, z_{it}) \right), \quad (23)$$

with analogous expressions for the contemporaneous and dynamic APEs. When including a large number $R$ of terms in the expansion and relying on a penalty for regularization, plug-in estimators such as (23) may be biased. To address this issue, in our application we implement the double Lasso method of Belloni, Chernozhukov, and Hansen (2014).

Remark 4. (Without expectations data)

In certain applications, subjective expectations data are not available. For example, in an agricultural production setting (as in Subsection 4.3) farmers’ subjective beliefs about future weather may not be available. In such cases, our approach is still applicable provided one can recover estimates of the belief density $\pi_{it}$. A strategy to do so is to assume that agents have rational expectations, and to make assumptions about the dynamic process of $x_{it}$.

5.3 Relaxing parametric assumptions on beliefs

The parametric approach we adopt in our application is motivated by the coarse belief information available in the SHIW. In other applications with richer information, a nonparametric treatment of the belief density $\pi_{it}$ may be feasible. Póczos, Singh, Rinaldo, and Wasserman (2013) propose a nonparametric regression estimator that, given a nonparametric estimate $\hat{\pi}_{it}$, can be used to consistently estimate $\phi_i$ and average partial effects. However, their estimator suffers from a slow convergence rate in general. An alternative is to assume that $\phi_i$ in (3) is linear, or more generally polynomial, in beliefs, as in the literature on functional regression.
(see, e.g., Ramsay and Dalzell, 1991, and Yao and Müller, 2010). Under linearity in beliefs, there exists a function \( \varphi_i \) such that

\[
\phi_i(x, \pi, z) = \int \varphi_i(x, \tilde{x}, z) \pi(\tilde{x}) d\tilde{x},
\]

and one can estimate \( \varphi_i \) using functional regression estimators based on principal components analysis or Tikhonov regularization (Hall and Horowitz, 2007). However, all these methods require large samples and the availability of rich information about \( \pi_{it} \).

When subjective data are too coarse, the information in the expectations responses \( m_{it} \) may not be sufficient to point-identify \( \pi_{it} \) nonparametrically. One possibility is to impose parametric assumptions, as we do in our application. An alternative approach is to follow a partial identification strategy. We outline such an approach in Appendix B.3.

6 Income, consumption, and income expectations

In this section we apply our approach to empirically study how consumption depends on current and expected income, and to conduct various tax counterfactuals.

6.1 Data

The Italian Survey on Household Income and Wealth (SHIW) is a cross-sectional survey that collects information on annual consumption, disposable income, and wealth of Italian families. Since 1989, it includes a panel component. We use the 1989–1991 waves and the 1995–1998 waves, which include questions about income expectations asked to a subsample of households.

The expectations questions differ in both sets of waves. However, as we show in Appendix D, the results are qualitatively similar when analyzing the waves separately, so we pool them together to increase power. In 1989 and 1991, individuals are asked about the probability their income growth will fall within a set of predetermined intervals. In 1995 and 1998, individuals are asked the maximum and minimum amounts they expect to earn if employed, and the probability of earning less than the mid-point between the maximum and minimum. We assume beliefs about log income in the following year follow a normal distribution. In Appendix D we describe our approach to estimate the mean \( \mu_{it} \) and standard deviation \( \sigma_{it} \) of the beliefs for each individual and time period, which follows Arellano, Bonhomme, De Vera, Hospido, and Wei (2022). We will also comment on robustness checks obtained under different assumptions and estimation strategies.
We focus on employed household heads, while excluding the self-employed. Our cross-sectional sample with information on beliefs has 7,796 household-year observations, and our panel sample with data on beliefs in two consecutive waves for the same head has 1,646 observations. In Appendix Tables H2 and H3 we report descriptive statistics about income expectations questions. In Appendix Table H4 we provide descriptive statistics about income, consumption, assets, and the estimated means and variances of log income beliefs. Belief questions are about individual income, while consumption, assets, and current income are reported at the household level. We will account for this discrepancy in our construction of average partial effects, and we will also report estimates that control for spousal beliefs when available. Another issue with the belief data in the SHIW is that expectations questions about income in the next 12 months are asked a few months after the end of the calendar year. We will return to this issue in the next subsection. As a preliminary validation check for the expectations questions, in Appendix Table H5 we document that beliefs have explanatory power for future log income, even conditional on current log income and other controls, in line with what Kaufmann and Pistaferri (2009) found for the 1995-1998 waves.

6.2 Estimates of the consumption function

We estimate several versions of the following regression of log consumption:

\[ y_{it} = \phi_i(x_{it}, \pi_{it}, z_{it}) + \varepsilon_{it} \]

\[ = \beta_x x_{it} + \beta_y \theta_{it} + \beta_{x} \theta_{it} x_{it} + \beta_{z} \theta_{it} z_{it} + \alpha_i + \varepsilon_{it}, \]  

(25)

where \( y_{it} \) is log consumption, \( x_{it} \) is log income, \( \theta_{it} \) contains the mean and variance of income beliefs, and \( z_{it} \) include log assets as well as a variety of controls (including age, household composition, and a wave indicator).

6.2.1 Main estimates

We show our main estimates in Table 2, where we estimate equation (25) by OLS in first differences in both sets of waves. In the table we show standard errors clustered at the household level.\(^{16}\) The results in columns (2) and (3) show that the mean of log income beliefs influences consumption decisions significantly over and beyond current income, while the variance of the beliefs has an insignificant effect.

\(^{16}\)Standard errors in Table 2 do not account for the estimation of the means and variances of beliefs. We will return to the impact of belief elicitation on our estimates at the end of this subsection.
It is also interesting to compare the estimates in column (2) with those in column (1) that do not account for beliefs. When including beliefs, the coefficient of family income decreases from 0.58 to 0.44. This finding is consistent with the presence of an upward omitted variable bias in column (1).

In column (4) of Table 2, we interact the mean income beliefs with current income. While the estimates suggest the effect of the mean belief tends to be larger for higher-income households, the interaction effect is only marginally significant. Lastly, in column (5) we add the variance of beliefs and its interaction with income. We find small differences compared to column (4), with insignificant coefficients associated with the variance of beliefs.

In addition to these specifications we also estimated flexible models using the Lasso, and used them to produce average partial effects (see the next subsection).

6.2.2 Robustness checks

In Appendix E we report a series of robustness checks. Our main estimates are obtained using a particular approach to construct the mean and variance of log income beliefs. We first probe the robustness of our estimates to different assumptions about the distribution of beliefs, and to different construction methods for the mean and variance of beliefs. The results reported in Appendix Table H6 show only minor differences compared to our baseline estimates.

While consumption and income correspond to households, the income beliefs questions correspond to individual income. In the baseline results we only use the beliefs of household heads (and adjust our counterfactual calculations). In a robustness check we control for spouses’ beliefs about their own income in the consumption regression. The results, also reported in Appendix Table H6, are again very similar to our main estimates.

The estimates in Table 2 are obtained by pooling two sets of waves, 1989–1991 and 1995–1998. Economic conditions, as well as the belief elicitation strategies, differ between these two periods. As a robustness check, we report estimates for the two sets of waves separately. The results, reported in Appendix Table H7, show general qualitative agreement and some quantitative differences between the two periods, with a stronger effect of beliefs in 1995–1998.

Lastly, although assets are important determinants of consumption, their measurement in the SHIW is imperfect. Indeed, respondents are asked about end-of-year assets, while the state variable in the consumption function is beginning-of-period assets. We assess the robustness of our results in this dimension in two ways. First, following Stoltenberg and Uhlendorff (2022) we construct an alternative measure of assets by subtracting yearly savings from end-of-year assets. A concern with this specification in our context is that savings in the SHIW are con-
Table 2: Estimates of the log consumption function

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean expected log income</td>
<td>0.235</td>
<td>0.238</td>
<td>0.229</td>
<td>0.231</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.094)</td>
<td>(0.095)</td>
<td>(0.093)</td>
<td>(0.093)</td>
<td></td>
</tr>
<tr>
<td>(Mean expect. log income)-(Log family income)</td>
<td>0.104</td>
<td>0.104</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.061)</td>
<td>(0.061)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var expected log income</td>
<td>-2.590</td>
<td>-2.613</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.876)</td>
<td>(1.941)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Var expect. log income)-(Log family income)</td>
<td>-1.144</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.499)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log family income</td>
<td>0.584</td>
<td>0.439</td>
<td>0.439</td>
<td>0.439</td>
<td>0.440</td>
</tr>
<tr>
<td></td>
<td>(0.070)</td>
<td>(0.089)</td>
<td>(0.089)</td>
<td>(0.089)</td>
<td>(0.089)</td>
</tr>
<tr>
<td>Log family assets</td>
<td>0.010</td>
<td>0.018</td>
<td>0.018</td>
<td>0.019</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td>(0.023)</td>
<td>(0.023)</td>
<td>(0.023)</td>
<td>(0.023)</td>
</tr>
<tr>
<td>Household fixed effect</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Controls</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>N observations</td>
<td>1,536</td>
<td>1,536</td>
<td>1,536</td>
<td>1,536</td>
<td>1,536</td>
</tr>
<tr>
<td>N households</td>
<td>768</td>
<td>768</td>
<td>768</td>
<td>768</td>
<td>768</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.24</td>
<td>0.26</td>
<td>0.26</td>
<td>0.26</td>
<td>0.26</td>
</tr>
<tr>
<td>Pvalue F beliefs</td>
<td>0.01</td>
<td>0.03</td>
<td>0.02</td>
<td>0.05</td>
<td></td>
</tr>
</tbody>
</table>

Notes: SHIW, 1989–1991 and 1995–1998. Regression for household heads. The expectations variables (mean and variance) and log family income are centered around the weighted average in the sample. Controls include age and age squared, existence of a spouse, marital status, family size, number of children 0-5, 6-13, 14-17 years old in the household, number of children outside the household, number of income earners in the household, and a wave indicator. Regression results are weighted using survey weights. Standard errors (shown in parenthesis) are clustered at the household level.
structed by netting out consumption expenditures from total income, so measurement error in consumption might bias our regression coefficients. Given this, we also report the results of a second specification where we do not include any control for assets. In addition to these checks, we also report results based on an IV strategy that relies on first-period assets and income as instruments for current assets. All the results for current income and income beliefs that we report in Appendix Table H8 are overall quite similar to our main estimates.

6.2.3 Measurement error in beliefs

A possible concern with the estimates in Table 2 is measurement error in beliefs data. To explore this issue, we focus on the 1989–1991 waves. In those two waves, individuals are asked to distribute 100 balls into 12 bins, corresponding to different intervals of beliefs about log income growth. Assuming log income growth beliefs to be normally distributed, a simple model of the responses is that individuals draw 100 i.i.d. values from their normal belief distributions, and put those in the bins.

However, this simple model does not provide a good approximation to individuals’ responses in the SHIW. Indeed, by simulating income beliefs responses from the model, we document that, if they were indeed drawing 100 values, respondents would be reporting a larger number of bins than they do in the data (specifically, 3.61 bins on average according to the model compared to 1.75 in the data). The results of this comparison are presented in Appendix Table H9.

As an alternative model, we postulate that individuals only draw $M < 100$ values. We interpret these values as $M$ income growth “scenarios” that the respondent contemplates before giving her answer. The simulations reported in Appendix Table H9 show that, when $M$ is of the order of 5 or 10 draws, instead of 100, the predicted number of bins reported by the individuals is much closer to the data.

Given this model of measurement error, for any given $M$ we implement a “small-$\sigma$” approximation (e.g., Evdokimov and Zeleneev, 2022), and use it to bias-correct our regression estimates. While different $M$ values can imply very different belief responses, we find that the resulting coefficient estimates vary little across values of $M$. We provide details about this approach in Appendix F and report the main results in Appendix Figure H3. At the same time, we acknowledge that, while this sensitivity analysis exercise is reassuring, it relies on a specific model of measurement error and our ability to entertain other models is limited by the short panel dimension available in the SHIW.

Lastly, a possible source of measurement error specific to the SHIW, and not captured by the model we have just outlined, relates to the timing of the expectations questions. As pointed
out by Pistaferri (2001), since income and consumption refer to the previous calendar year, yet expectations are asked a few months after the end of the year, one needs to assume that individuals do not update their information sets during these few months.\footnote{Alternatively, one could instead follow a structural approach and specify a complete structural model of consumption choices and belief formation. Stoltenberg and Uhlendorff (2022) propose such an approach and find that income beliefs, corrected for the timing discrepancy within the structure of their model (which assumes rational expectations), have larger effects on consumption than the original beliefs.}

### 6.3 Counterfactual taxes

We now use our framework, and our estimates of the consumption function, to assess the effects of a counterfactual income tax on consumption. We assume that the tax schedule takes the parametric form $T(w_g) = w_g - \lambda w_g^{1-\tau}$, where $w_g$ denotes gross income (e.g., Benabou, 2002). To define a baseline level of the tax, we rely on the estimates obtained by Holter, Krueger, and Stepanchuk (2019) for Italy, averaged over family composition characteristics in our sample.

We consider three counterfactuals, corresponding to changes in the $\lambda$ and $\tau$ parameters that index the tax schedule. In the transitory tax and permanent tax counterfactuals, we increase the average tax by 10 percentage points by decreasing $\lambda$, only for one period in the former case and in all subsequent periods in the latter. In the regressivity counterfactual, we set the parameter $\tau$ to its value in the French tax system (which is somewhat less progressive than the Italian one) while at the same time decreasing $\lambda$ such that the tax change is neutral in terms of total tax revenue.

To estimate the effects of the counterfactuals we compute average partial effects. We report estimates of TAPE, CAPE, and DAPE obtained using linear regression (see Table 2), as well as estimates obtained using the Lasso. For the latter, we rely on the double/debiased Lasso method introduced by Belloni, Chernozhukov, and Hansen (2014), based on interactions and powers of the covariates up to the third order. In the calculations for the permanent tax and regressivity counterfactuals, we assume that individuals fully adjust their beliefs to the new tax; i.e., we implement the formula in (17). We report point estimates and standard errors based on the bootstrap in Appendix Table H10.

The top panel in Figure 1 shows average partial effects based on the estimates from column (5) in Table 2, while the bottom panel corresponds to estimates based on the Lasso. On the left graphs we show the effects on log consumption of a 10\% transitory tax. The overall effect based on OLS is $-0.049$, and it is very similar according to the Lasso. In addition, in both specifications there is only moderate variation along income quantiles (indicated on the x axis).
Figure 1: Average partial effects for various tax counterfactuals

A. Average partial effects based on OLS estimates

(a) Transitory tax  (b) Permanent tax  (c) Regressivity

B. Average partial effects based on the Lasso

(d) Transitory tax  (e) Permanent tax  (f) Regressivity

Notes: SHIW, 1989–1991 and 1995–1998, cross-sectional sample. Black bars correspond to contemporaneous APE and grey bars correspond to dynamic APE. Total APE are the sums of CAPE and DAPE. In the top panel we report results based on OLS estimates, see column (5) in Table 2. In the bottom panel we report estimates based on the double/debiased Lasso, for a dictionary including interactions and powers of the covariates up to the third order. See Appendix Figure H5 for results corresponding to second and fourth order interactions and powers.

On the middle graphs we show the effect of a 10% permanent tax. Note that the contemporaneous average partial effect (CAPE) coincides with the effect of a transitory tax (compare with the left graphs). Beyond this contemporaneous effect, we find sizable dynamic effects. The dynamic APE (DAPE), which reflects the impact of a changes in beliefs, contributes an additional -0.024 according to OLS, and -0.028 according to the Lasso. The total change in consumption, which is approximately -0.073 in both specifications, is less than the 10% decrease in income, as is expected if households are only partially insured against income changes (Blundell, Pistaferri, and Preston, 2008). Moreover, the estimates from both specifications indicate that dynamic effects are larger for higher-income households.

Lastly, on the right graphs we show the effect of a revenue-neutral decrease in the progressivity of the tax. While the total effects averaged over all households are relatively small
(around −0.011), they show substantial heterogeneity along the income distribution: reducing progressivity tends to favor the rich, and it hurts the log consumption of the poor proportionally more. The estimates of OLS and the Lasso are very similar. However, in this case estimates are less precise, see Appendix Table H10. As in the other two counterfactuals, we observe that the contemporaneous and dynamic effects of the tax have the same sign.

It is interesting to compare these estimates to average partial effects calculations that do not account for the role of beliefs. In that case, the average consumption effect over all households of a 10% permanent income tax is −0.065. This is larger than the contemporaneous effect (−0.049), consistently with beliefs being an omitted yet relevant regressor in the specification without beliefs. However, this is lower than the total effect that accounts for both contemporaneous and dynamic margins (−0.073). These differences underscore the need to account for beliefs when computing average partial effects. In addition, note that an estimation method that does not include beliefs cannot account for the difference in impact between a permanent tax and a transitory one.

Lastly, it is worth emphasizing that two conditions are needed in order to interpret the average partial effects in Figure 1 as structural tax counterfactuals. The first one is that individual beliefs respond one-to-one to the tax. By varying the parameter $\xi$ in (18), we can predict tax effects under different assumptions about belief responses, in the spirit of sensitivity analysis. The second condition is that the belief updating rule $\rho_i$ is invariant under the tax. When tax changes have a long-lasting effect, changes in $\rho_i$ may occur and induce a third margin of response, beyond contemporaneous and dynamic effects (i.e., beyond CAPE and DAPE). While this third margin may be small or zero in certain cases (as in the permanent-transitory model with a proportional tax, see Subsection 4.1), accounting for it may be important in other cases. The extension to beliefs over longer horizons that we outline in the next section provides a possible way forward.

7 Extensions

In this paper we provide a method to account for the role of individual expectations in assessing the impact of policies and other counterfactuals. Our approach is semi-structural, in the sense that it is justified under dynamic structural assumptions, yet implementing the method does not require full specification and estimation of a structural model.

Among possible extensions of the method, it is interesting to allow for endogenous and exogenous state variables in $x_{it}$, and for state-contingent beliefs about them. For example, in
a model of occupational choice, individual income beliefs contingent on occupational choice may be available (e.g., Patnaik, Venator, Wiswall, and Zafar, 2020, Arcidiacono, Hotz, Maurel, and Romano, 2020). In that case, our framework is unchanged except for the fact that the state-contingent beliefs enter as arguments in the decision rule. We describe this extension in Appendix B.1, and provide a learning model as an illustration.

A second extension is to introduce beliefs over longer horizons. If one had access to data on the sequence of beliefs about $x_{i,t+1}, x_{i,t+2}, ...$ into the far future, accounting for those as determinants of the decision, and shifting them in the counterfactual, would provide valid predictions without the need for an invariance assumption about some $\rho_i$ process. To go one step in this direction, one can elicit beliefs over multiple horizons $x_{i,t+1}, x_{i,t+2}, ..., x_{i,t+S}$ (Koşar and van der Klaauw, 2022), and account for variation in those beliefs in estimation and counterfactuals. We describe this second extension in Appendix B.2.
References


APPENDIX

A Belief formation models with learning

In this section of the appendix we describe two models of belief formation with learning that we mentioned in Subsection 3.2.

A.1 Exogenous beliefs

We start with the model where beliefs are not affected by past actions. Suppose that

\[ x_{it} = \alpha_i + \varepsilon_{it}, \]

where \( \varepsilon_{it} \) are i.i.d. \( \mathcal{N}(0, \sigma^2_{\varepsilon_{it}}) \). Suppose agents have rational expectations, with information set \( \Omega_{it} = \{x_{it}, x_{i,t-1}, \ldots\} \), which does not include \( \alpha_i \). Furthermore, assume agents are Bayesian learners with prior beliefs about \( \alpha_i \) that are normally distributed. Then, by Bayes rule, posterior beliefs about \( \alpha_i \) over time are also normally distributed with mean \( \mu_{it} \) and variance \( \sigma^2_{it} \) satisfying

\[
\mu_{it} = \mu_{i,t-1} + \frac{\sigma^2_{it}}{\sigma^2_{\varepsilon_{it}}} \left( x_{it} - \mu_{i,t-1} \right), \\
(\sigma^2_{it})^{-1} = (\sigma^2_{i,t-1})^{-1} + (\sigma^2_{\varepsilon_{it}})^{-1}.
\]  

(A1)

(A2)

Then, \( \pi_{it} \) is a normal density with mean \( \mathbb{E}_{\pi_{it}}(x_{i,t+1}) = \mu_{it} \) and variance \( \text{Var}_{\pi_{it}}(x_{i,t+1}) = \sigma^2_{it} + \sigma^2_{\varepsilon_{it}} \). Hence, by (A1)-(A2) the belief process satisfies Assumption 2. Note that the mean beliefs in (A1) are as in the adaptive expectations case, see (7), but with a parameter \( \lambda_{it} = \frac{\sigma^2_{\varepsilon_{it}}}{\sigma^2_{\varepsilon_{it}}} \) that is time-varying and converges to zero over time.

A.2 Endogenous beliefs

We now describe a variation of the previous model, where actions \( y_{it} \in \{0, 1\} \) are binary, and the agent observes an additional signal about \( \alpha_i \),

\[ s_{it} = \alpha_i + v_{it}, \]

only when \( y_{i,t-1} = 1 \). We assume that \( v_{it} \) are i.i.d. \( \mathcal{N}(0, \sigma^2_v) \), independent of \( \varepsilon_{it} \) at all leads and lags. The posterior mean of \( \alpha_i \) is \( \mathcal{N}(\mu_{it}, \sigma^2_{it}) \), where now \( \mu_{it} \) and \( \sigma^2_{it} \) depend on \( y_{i,t-1} \). When \( y_{i,t-1} = 0 \), \( \mu_{it} \) and \( \sigma^2_{it} \) are given by (A1)-(A2), while when \( y_{i,t-1} = 1 \) they are given by

\[
\mu_{it} = \mu_{i,t-1} + \frac{\sigma^2_{it}}{\sigma^2_{\varepsilon_{it}}} \left( x_{it} - \mu_{i,t-1} \right) + \frac{\sigma^2_{it}}{\sigma^2_{v_i}} \left( s_{it} - \mu_{i,t-1} \right), \\
(\sigma^2_{it})^{-1} = (\sigma^2_{i,t-1})^{-1} + (\sigma^2_{\varepsilon_{it}})^{-1} + (\sigma^2_{v_i})^{-1}.
\]  

(A3)

(A4)
Now, denoting $\tilde{\sigma}_{it}^2 = \left[ (\sigma_{i,t-1}^2)^{-1} + (\sigma_{\xi}^2)^{-1} \right]^{-1}$, we have
\[
(s_{it} \mid x_{it}, y_{it-1} = 1, \Omega_{i,t-1}) \sim \mathcal{N} \left( \mu_{i,t-1} + \tilde{\sigma}_{it}^2 \left( x_{it} - \mu_{i,t-1} \right), \sigma_{it}^2 + \sigma_{v_i}^2 \right).
\]
Hence, by (A3),
\[
(\mu_{it} \mid x_{it}, y_{it-1} = 1, \Omega_{i,t-1}) \sim \mathcal{N} \left( \mu_{i,t-1} + \left( \frac{\sigma_{it}^2}{\sigma_{\xi_i}^2} + \frac{\sigma_{it}^2 \tilde{\sigma}_{it}^2}{\sigma_{v_i}^2} \frac{1}{\sigma_{\xi_i}^2} \right) \left( x_{it} - \mu_{i,t-1} \right), \frac{\sigma_{it}^4}{\sigma_{v_i}^2} \left( \tilde{\sigma}_{it}^2 + \sigma_{v_i}^2 \right) \right).
\]
\[(A5)\]
It thus follows from (A4)-(A5) in the case $y_{i,t-1} = 1$, and from (A1)-(A2) in the case $y_{i,t-1} = 0$, that $\pi_{it}$, which is the normal density with mean $\mu_{it}$ and variance $\sigma_{it}^2 + \sigma_{\xi_i}^2$, satisfies Assumption 2. Note that, in this case, beliefs $\pi_{it}$ depend on past actions $y_{i,t-1}$, so (6) does not hold.

**B  Extensions**

In this section we outline three extensions of our approach that we mentioned in the main text.

**B.1 State-contingent beliefs**

Our framework can easily be generalized to allow for endogenous and exogenous states in $x_{it}$, and for contingent beliefs about them. To see this, suppose for simplicity that actions $y_{it}$ belong to a finite set $\mathcal{Y}$ with $n$ elements. In this case, one can define $\pi_{it} = \{ \pi_{it}(\cdot; y) : y \in \mathcal{Y} \}$ to be a set of $n$ conditional densities where, for all $y \in \mathcal{Y}$, $\pi(\cdot; y)$ is the conditional density of $(x_{i,t+1} \mid y_{it} = y, \Omega_{it})$. With this new definition of $\pi_{it}$, and the associated change in the definition of $\rho_i$ in Assumption 2, the framework is unchanged relative to Section 3. In particular, the decision rule is still given by (9), so actions depend on the $n$ belief densities $\pi_{it}(\cdot; y)$.

As an example of a model with state-contingent beliefs, suppose $x_{it} = \alpha_i + \varepsilon_{it}(k)$ when $y_{i,t-1} = k$, for $k \in \{0,1\}$.\footnote{This is equivalent to assuming the individual only observes $x_{it}(k) = \alpha_i + \varepsilon_{it}(k)$ when $y_{i,t-1} = k$. As an extension, $\alpha_i$ may also depend on $k$ (for example, $\alpha_i$ may represent a vector of occupation-specific abilities), and $x_{it}(k) = \alpha_i(k) + \varepsilon_{it}(k)$. In that case, the updating formulas (A6)-(A7) need to be adjusted to vector-valued $\mu_{it}$ and matrix-valued $\sigma_{it}^2$. See Arcidiacono, Aucejo, Maurel, and Ransom (2016) for a recent example.} Suppose in addition that $\varepsilon_{it}(k) \sim \mathcal{N}(0, \sigma_{\varepsilon_i(k)}^2)$, independent across $i$ and $t$, and that agents are Bayesian with a normal prior on $\alpha_i$. The posterior distribution of $\alpha_i$ when $y_{i,t-1} = k$ is then $\mathcal{N}(\mu_{it}, \sigma_{it}^2)$, where $\mu_{it}$ and $\sigma_{it}^2$ are functions of $k$ satisfying
\[
\begin{align*}
\mu_{it} &= \mu_{i,t-1} + \frac{\sigma_{it}^2}{\sigma_{\varepsilon_i(k)}^2} \left( x_{it} - \mu_{i,t-1} \right), \\
(\sigma_{it}^2)^{-1} &= \left( (\sigma_{it-1}^2)^{-1} + (\sigma_{\varepsilon_i(k)}^2)^{-1} \right)^{-1}. \quad \text{(A7)}
\end{align*}
\]
\[(A6)\]
We then define beliefs as \( \pi_{it} = (\pi_{it}(0), \pi_{it}(1)) \), where \( \pi_{it}(k) \) is the normal density with mean \( \mu_{it} \) and variance \( \sigma^2_{it} + \sigma^2_{\varepsilon_i(k)} \). It follows from (A6)-(A7) that Assumption 2, for these beliefs \( \pi_{it} \), is satisfied.

**B.2 Beliefs over longer horizons**

A key feature of the framework is that, while beliefs about next period’s state variables change in the data and counterfactual, the belief updating rule \( \rho_i \) is constant in sample and invariant to the counterfactual change. This assumption can be relaxed by introducing beliefs over multiple horizons.

To describe such an approach, let us replace Assumption 1 by the following, for some \( S \geq 1 \):

\[
(x_{i,t+S}, \ldots, x_{i,t+1} \mid y_{it}, \Omega_{it}) \sim (x_{i,t+S}, \ldots, x_{i,t+1} \mid \Omega_{it}), \tag{A8}
\]

and denote the corresponding conditional density as \( \pi_{it}(x_{i,t+S}, \ldots, x_{i,t+1}) \).

In this case, (8) becomes

\[
V_i(x_t, \pi_t, z_t, \nu_t) = \max_{\mu_t} \left\{ u_i(y_t, x_t, z_t, \nu_t) + \beta \int V_i(x_{t+1}, \pi_{t+1}; y_t, \nu_t) \pi^{(1)}_i (x_{t+1}) \rho_i(\pi_{t+1}; x_{t+1}, y_t, \pi_t, x_t, \nu_t) d\pi_t \right\},
\]

where \( \pi^{(1)}_i \) denotes the marginal of \( \pi_t \) corresponding to period-\( t+1 \) outcomes. Hence, equation (9) is satisfied for the \( \pi_{it} \) corresponding to (A8).

**B.3 Partial identification**

Lastly, we outline an approach to partially identify the function \( \phi_i \) in cases where we do not impose a parametric specification on the belief density (i.e., when we do not impose (15)).

To proceed, let us omit the reference to \( x \) and \( z \) for conciseness. In this case, the conditional mean \( \phi_i(\pi_{it}) = \mathbb{E}[y_{it} \mid \pi_{it}] \) is bounded as follows:

\[
\inf_{\pi \in \Pi(m_{it})} \phi_i(\pi) \leq \mathbb{E}[y_{it} \mid \pi_{it}] \leq \sup_{\pi \in \Pi(m_{it})} \phi_i(\pi),
\]

where \( \Pi(m_{it}) = \{ \pi : m(\pi) = m_{it} \} \).

These bounds imply the following moment inequalities on \( \phi_i \):

\[
\mathbb{E} \left[ y_{it} - B^L_i(m_{it}; \phi_i) \mid m_{it} \right] \geq 0, \quad \mathbb{E} \left[ y_{it} - B^U_i(m_{it}; \phi_i) \mid m_{it} \right] \leq 0.
\]
C  Structural and semi-structural counterfactuals

In this section of the appendix we present the details of the calibration that we used to produce Table 1, and report additional output from the simulation.

C.1 Model

The model closely follows Kaplan and Violante (2010), with some differences. Agents live for \( T \) periods, and work until age \( T_{ret} \), where both \( T \) and \( T_{ret} \) are exogenous and fixed. \textit{Ex ante} identical households maximize expected life-time utility

\[
E_0 \left[ \sum_{t=1}^{T} \beta^{t-1} u(c_{it}) \right].
\]

During working years \( 1 \leq t \leq T_{ret} \), agents receive after-tax labor income \( w_{it} = \exp(x_{it}) \), the log of which is the sum of a deterministic experience profile \( \kappa_t \), a permanent component \( \eta_{it} \), and a transitory component \( \varepsilon_{it} \),

\[
x_{it} = \kappa_t + \eta_{it} + \varepsilon_{it},
\]

\[
\eta_{it} = \eta_{it-1} + v_{it},
\]

where \( \eta_{i1} \) is drawn from an initial normal distribution with mean zero and variance \( \sigma^2_{\eta_{i1}} \). The shocks \( \varepsilon_{it} \) and \( v_{it} \) have zero mean, are independent at all leads and lags, and are normally distributed with variances \( \sigma^2_{\varepsilon} \) and \( \sigma^2_{v} \), respectively.

We define gross labor income as \( \tilde{w}_{it} = G(w_{it}) \), where \( G \) is the inverse of the tax function

\[
\tau(\tilde{w}_{it}) = \tilde{w}_{it} - \tilde{\lambda} \tilde{w}_{it}^{1-r}.
\]

After retirement, agents receive after-tax social security transfers \( w_{it}^{ss} \), which are a function of average individual gross income over the last few years of their working life,

\[
w_{it}^{ss} = P \left( \frac{1}{T_{ret} - T_{cont}} \sum_{t=T_{cont}}^{T_{ret}-1} \tilde{w}_{it} \right).
\]

Lastly, throughout their lifetime households can save (but not borrow) through a single risk-free, one-period bond whose constant return is \( 1 + r \), and they face a period-to-period...
budget constraint

\[ z_{i,t+1} = (1 + r)z_{it} + w_{it} - c_{it} \quad \text{if } t < T_{ret} \]
\[ z_{i,t+1} = (1 + r)z_{it} + w_{it} - c_{it} \quad \text{if } t \geq T_{ret} \]

We consider two cases:

- A case with rational expectations, where individuals observe \( \eta_{it} \) each period, and beliefs about after-tax log income next period are normally distributed with

\[
\mathbb{E}_t(x_{i,t+1}) = \kappa_{t+1} + \eta_{it}, \\
\text{Var}_t(x_{i,t+1}) = \sigma_v^2 + \sigma_\varepsilon^2.
\]

- A case with adaptive expectations, where beliefs about after-tax log income next period are normally distributed with

\[
\mathbb{E}_t(x_{i,t+1}) = \kappa_{t+1} + (\mathbb{E}_{t-1}(x_{it}) - \kappa_t) + \Gamma \cdot (x_{it} - \mathbb{E}_{t-1}(x_{it})) + u_{it}, \quad u_{it} \sim \mathcal{N}(0, V_u), \\
\text{Var}_t(x_{i,t+1}) = \sigma_v^2 + \sigma_\varepsilon^2,
\]

where \( \Gamma \) is a constant, \( u_{it} \) are independent of all other shocks in the model, and initial mean beliefs are given by \( \mathbb{E}_1(x_{i2}) = \kappa_2 + \eta_{i1} \).

C.2 Calibration

We closely follow the calibration strategy in Kaplan and Violante (2010).

Demographics. The model period is one year. Agents enter the labor market at age 25, retire at age 60, and die with certainty at age 95. So we set \( T_{ret} = 35 \), and \( T = 70 \).

Preferences. The utility function is CRRA, \( u(c) = c^{1-\gamma}/(1-\gamma) \), where the risk aversion parameter is set to \( \gamma = 2 \).

Discount factor and interest rate. The interest rate is \( r = 0.03 \), and \( \beta = 1/(1 + r) \).

Income process. We use the deterministic age profile \( \kappa_t \) from Kaplan and Violante (2010). For the stochastic components of the income process, we set \( \sigma_{\eta_t}^2 = 0.15, \sigma_v^2 = 0.01, \) and \( \sigma_\varepsilon^2 = 0.05 \).
Initial wealth and borrowing limit. Households’ initial assets are set to 0 and there is no borrowing possible.

Tax system. We use parameters derived from Holter, Krueger, and Stepanchuk (2019), $\tilde{\lambda} = 3.826$, $\tau = 0.137$.

Social security benefits. Social security benefits are a function of average individual gross earnings between the ages of 50 and 60, $w_{it}^{ss} = P\left(\frac{1}{T_{ret}-T_{cont}} \sum_{t=T_{cont}}^{T_{ret}-1} \bar{w}_{it}\right)$, where $T_{cont} = 25$. Pre-tax benefits are equal to 90% of average past earnings up to a given bend point, 32% from this first bend point to a second bend point, and 15% beyond that. The two bend points are set at, respectively, 0.18 and 1.10 times cross-sectional average gross earnings. Benefits are then scaled proportionately so that a worker earning average labor between ages 50 and 60 is entitled to a pre-tax replacement rate of 45%. There is also a cap on pre-tax earnings contributing to pensions (cap of 2.2) and only 85% of pre-tax pensions are taxed.

Adaptive beliefs. We take $\Gamma = 0.5$ and $V_u = 0.2$.

There are two main differences between our calibration and the one from Kaplan and Violante (2010), besides including the adaptive expectations case and using a different tax function. First, pensions depend on contributions made between ages 50 and 60, so the history of past income is not a relevant state variable before age 50. Second, we do not consider random mortality during retirement years.

C.3 Additional simulation results

In this subsection we report results based on the calibrated structural model that we introduced in Subsection 4.1.

In Table H1 we report structural and semi-structural counterfactual effects of a permanent 10% income tax, as in Table 1, for three different ages: 26, 35, and 45. We see that, under rational expectations (left panel), the contemporaneous effect of the tax is higher for the young than for older households, while the dynamic impact is lower. This reflects the fact that households start their working life without assets, and that they cannot borrow. The semi-structural average partial effects reproduce the structural policy effects well. In the case of adaptive expectations (right panel) there is less variation by age, and while a linear specification
tends to produce too high a contemporaneous effect for the old, the quadratic and spline specifications agree well with the structural predictions. For completeness, in Figures H1 and H2 we plot the policy rules and the mean and variance profiles of consumption, assets and income under the model.

D Beliefs data

In this section of the appendix we describe the income belief questions in the SHIW, and explain how we estimate the parameters of the belief densities.

D.1 Expectations questions in the SHIW

The SHIW includes questions about income expectations in waves 1989–1991 and 1995–1998; however the expectations questions differ in the two sets of waves.

The 1989–1991 waves include a question about expected income growth:

Thinking now of your total income from work or retirement and its evolution [for the next 12 months]... Which categories would you exclude? Suppose you have 100 points to distribute among the remaining categories, how many would you give to each?

The possible categories are more than 25%, between 20% and 25%, between 15% and 20%, between 13% and 15%, between 10% and 13%, between 8% and 10%, between 7% and 8%, between 6% and 7%, between 5% and 6%, between 3% and 5%, between 0% and 3%, or less than 0%, and in that case, by how much. In Table H2 we report descriptive statistics corresponding to this question.

The 1995–1998 waves include three questions about expected income level:

Minimum amount expected to earn: Assuming that you remain in or find employment in the next 12 months, can you say what is the minimum overall annual amount you expect to earn, net of taxes, including overtime, bonuses, fringe benefits, etc?

Maximum amount expected to earn: Assuming again that you remain in or find employment in the next 12 months, can you say what is the maximum overall annual amount you expect to earn, net of taxes, including overtime, bonuses, fringe benefits, etc?

Probability of earning less than half: What is the probability that you will earn less than X (the amount obtained for (maximum + minimum)/2)? If you had to give a score of between 0 and 100 to the chances of earning less than X, what would it be? ("0" if certain of earning more than X, "100" if certain of earning less than X).
In Table H3 we report descriptive statistics corresponding to these questions. In these two waves, the survey also includes a question about the probability of being employed next year that we use in a robustness check specific to those waves.

D.2 Estimation of income beliefs

We assume log income beliefs are normally distributed, with mean $\mu_{it}$ and variance $\sigma^2_{it}$, and use the expectations questions to estimate these two parameters for each individual and wave. In this subsection, we omit the reference to $i$ and $t$ for ease of notation.

**First two waves.** For the 1989–1991 waves, we use the survey expectations questions to estimate the mean and variance of the beliefs of log income growth, which are normally distributed under our assumptions, with mean $\mu_g = \mu - x$ (where $x$ is the current log income), and variance $\sigma^2_g = \sigma^2$. Given estimates of $\mu_g$ and $\sigma^2_g$, we then recover estimates of $\mu$ and $\sigma^2$.

Let $\hat{p}_j$ denote the fraction of points the respondent assigns to bin $j$ (out of 100 points), for $j = 1, \ldots, J$, where $J = 12$. For each bin, one could interpret $\hat{p}_j$ as the probability that a $N(\mu_g, \sigma^2_g)$ draw takes values within the interval corresponding to that bin. Under this interpretation, one could estimate $\mu_g$ and $\sigma_g$ using maximum likelihood or minimum distance given the fractions $\hat{p}_j$. However, this approach does not work well in practice since many of the $\hat{p}_j$’s are exactly 0 or 1.

Instead of assuming that respondents report exact, normality-based probabilities, we follow Arellano, Bonhomme, De Vera, Hospido, and Wei (2022) and assume that, when answering the survey expectations questions, individuals sample $M$ draws from their underlying $N(\mu_g, \sigma^2_g)$ distribution, and use those draws to provide their answers $\hat{p}_j$. Given that, in the survey, individuals are asked to distribute 100 points among the 12 bins, we take $M = 100$ as our baseline. Hence, the answers $\hat{p}_j$ are obtained from $M = 100$ trials from a multinomial distribution with true probabilities $p_j$.

To estimate the $p_j$, we assume an uninformative (Jeffreys) prior on $(p_1, \ldots, p_J)$. It then follows that the posterior means of the $p_j$ are

$$\tilde{p}_j = \frac{\hat{p}_j + \frac{1}{2M}}{1 + \frac{J}{2M}}, \quad j = 1, \ldots, J.$$  

(A9)

The estimates $\tilde{p}_j$ are regularized counterparts to the $\hat{p}_j$. An advantage is that they take values in the open interval $(0, 1)$, which allows one to implement minimum distance or maximum likelihood estimation strategies based on them. We have performed robustness checks using
other regularization devices, including different $M$ values, and found only minor impacts on
the results (see Section E of this appendix).

Given the regularized responses $\tilde{p}_j$ in (A9), we then construct the cumulative probabilities,
$\tilde{c}_j = \sum_{k=1}^{j} \tilde{p}_k$, and estimate $\mu_g$ and $\sigma_g$ based on the following system of linear equations:
\begin{equation}
\Phi^{-1}(\tilde{c}_j) \cdot \sigma_g + \mu_g = v_j, \quad j = 1, \ldots, J - 1, \tag{A10}
\end{equation}
where $v_j$ correspond to the right endpoint of the $j$-th bin, and $\Phi$ denotes the standard normal
cdf. Since the first and last bins in the survey question are unbounded, we add bounds to those
(-10% for the bin below 0%, and 35% for the bin above 25%).\footnote{We verified that our estimates of the log consumption function remain similar when using different bounds,
and when excluding observations that assign all points to the first or last bin.} This amounts to working with
14 bins in total. We then estimate $\mu_g$ and $\sigma_g$ using OLS based on a subset of the equalities in
(A10). Specifically, we use all the bins $j$ for which $\tilde{p}_j > 0$, and use in addition one unbounded
bin to the left and one unbounded bin to the right of those. The reason for only using a subset
of the restrictions in (A10) is to reduce the influence of the regularization for bins with $\tilde{p}_j = 0$.

As an example, consider an individual who assigns 60 points to the 5–6% bin, and 40 points
to the 6–7% bin. In this case we use the intervals $(0.05,0.06)$ and $(0.06,0.07)$, both of which
have positive $\tilde{p}_j$, and we add the intervals $(-\infty,0.05)$ and $(0.07,+\infty)$, to the left and to the
right, respectively. We then compute the sums of the $\tilde{p}_j$ in (A9), in each of these four intervals.
Lastly, we use these cumulative probabilities to estimate $\mu_g$ and $\sigma_g$ by OLS. Since, in the
fourth interval, the cumulative probability is equal to 1, in this example we only rely on three
independent linear restrictions to estimate $\mu_g$ and $\sigma_g$.

\textbf{Last two waves.} For the 1995–1998 waves, we use the survey expectations questions to
estimate the mean $\mu$ and variance $\sigma^2$ of log income beliefs directly (since in these waves the
questions are about income, not income growth). We interpret the answers as probabilities
assigned to two bins (between the minimum and the mid-point, and between the mid-point and
the maximum). As in the 1989–1991 waves, we add two additional bins, one below the reported
minimum and another one above the reported maximum, which amounts to be working with 4
bins in total. These additional bins have a positive but low probability $\tilde{p}_j = \frac{1}{2M+4}$, which might
reflect that respondents interpret the minimum and maximum questions as asking them to
report quantiles of their distributions (see Delavande, Giné, and McKenzie, 2011). In the 1995–
1998 waves, the locations and widths of the bins come from individuals’ responses, providing
\footnote{We found that using all bins with $\tilde{p}_j = 0$ tended to artificially increase the variance of estimated beliefs.}
more information to capture beliefs, in particular beliefs with very small variance. For example, when the reported minimum and maximum coincide, the implied estimate of $\sigma$ is equal to zero.

**Descriptives and predictive power.** In Table H4 we provide descriptive statistics about the beliefs that we estimate and the main variables in the consumption equation.

In Table H5 we assess the predictive power of these beliefs: we regress $\log(w_{i,t+1})$ in columns (1) to (4), and $\log(w_{i,t+1}) - \log(w_{it})$ in columns (5) to (8), as functions of the estimated mean beliefs $\mu_{it}$ and other controls. In this table, we use log individual income as our dependent variable. The estimates suggest that individual beliefs predict future income, even conditional on current income.

### E Robustness checks

In this section of the appendix we provide several robustness checks for the estimation of the consumption function.

In columns (1) and (2) in Table H6 we show the estimates are robust to relying on different distributional assumptions for beliefs: a discrete distribution for waves 1989–1991 (as in Pistaferri, 2001), and a triangular distribution for waves 1995-1998 (as in Kaufmann and Pistaferri, 2009). In columns (3) to (6) we show that estimates are robust to the value of $M$ used for estimation (see (A9), where the baseline corresponds to $M = 100$). In columns (7) and (8) we also control for the spouse’s beliefs about their own income, when available. Results remain virtually unchanged, and spousal beliefs don’t appear to play a major role in household consumption for this sample.

In Table H7 we estimate the consumption function, focusing on the specification with mean beliefs interacted with log current income, separately for waves 1989–1991 and 1995–1998. The point estimates are different in the two samples, with a larger effect of beliefs in the 1995-1998 waves. However, in both cases beliefs play a significant role in household consumption.

---

4When spousal beliefs are not available, we set the variable to zero and add binary indicators for missingness, distinguishing between spouses that are homemakers, employed, or other labor status. Note that only 32% and 17% of the 768 households in columns (7) and (8), respectively, are households where data on spousal beliefs are available in at least one or in both waves.

5In each pair of waves, we also control for other expectations questions available: inflation expectations in 1989–1991, and expectations about future employment in 1995–1998.

6Using the 1995–1998 waves, we also estimated the consumption function including unemployed household heads in the sample and controlling for beliefs about future employment, and found similar results. In the 1989–1991 waves expectations questions were not asked to the unemployed.
Lastly, in Table H8 we present estimates obtained under different approaches for dealing with assets. As mentioned in the main text, the estimates of current income and income beliefs are quite similar across specifications, although we see some quantitative differences, especially in the case of the IV specification in columns (3) and (4).

F Measurement error

In this section of the appendix we describe how we correct for measurement error in the beliefs responses, by relying on the 1989–1991 waves. In our baseline specification, we estimate the mean and variance of beliefs by relying on a model that assumes individuals draw $M = 100$ different scenarios from their underlying beliefs to answer the expectations questions (see Subsection D.2 of this appendix). This choice is motivated by the format of the questions, where respondents are asked to distribute 100 points among the bins.

However, this model may not provide a good approximation to the response process of individuals when answering the questions in the SHIW. In fact, it is possible that respondents are only able to imagine a smaller number $M < 100$ of “income growth scenarios”, corresponding to events that they expect might happen in the next year, such as a promotion or a demotion, a job change, etc. To provide empirical support for this possibility, we predict, for each respondent, the number of non-empty bins reported by the respondent under the model, for various values of $M$. The estimates in Table H9 show that taking $M = 100$ implies that, on average, respondents should report 3.6 non-empty bins, while in the data this number is only 1.7. Besides, the table shows that taking smaller values of $M$ provides a better approximation to the distribution of the number of non-empty bins across individuals.

With this motivation, here we entertain an alternative parametric model for the responses, where individuals draw $M < 100$ values from a $\mathcal{N}(\mu_g, \sigma_g^2)$, and distribute those among the bins.\(^7\) Given this model, we propose a correction for measurement error and apply it to revisit our baseline estimates of the consumption function (see Table 2). Our approach is based on a “small-$\sigma$” approximation (e.g., Evdokimov and Zeleneev, 2022). Given that, for a given $M$ value, the model of measurement error is parametric, the correction can be implemented using a simple parametric bootstrap method, which we now describe.\(^8\)

\(^7\)In the model of measurement error that we propose, $M$ is constant across individuals. An alternative model would let $M_i$ vary across individuals. Manski and Molinari (2010) exploit repeated responses by the same individual to infer individual types of measurement error in responses.

\(^8\)Since the measurement error model is parametric, one could alternatively rely on an exact approach for deconvolving the measurement error, without the need for an approximation. An advantage of the specific
We consider the specification of the consumption function in column (3) of Table H7, which only accounts for mean beliefs. We draw $S = 1,000$ samples where, for each respondent, we draw $M$ observations from a $\mathcal{N}(\hat{\mu}_g; \hat{\sigma}_g^2)$, for $\hat{\mu}_g$ and $\hat{\sigma}_g^2$ our original estimates of $\mu_g$ and $\sigma_g^2$, respectively. This gives us simulated responses $\hat{p}_j^{(s)}$, for each sample $s$, from which we estimate $\mu_g$ and $\sigma_g$ and, based on those, the coefficients of the consumption function, exactly in the same way as we did to obtain the estimates in Table H7. Let $\hat{\beta}_{BC}$ denote the estimated coefficients in this last regression. We then construct the bootstrapped bias-corrected counterpart to the original coefficients $\hat{\beta}_{OLS}$ as

$$\hat{\beta}_{BC} = 2\hat{\beta}_{OLS} - \frac{1}{S} \sum_{s=1}^{S} \hat{\beta}^{(s)}.$$ 

We repeat this exercise for values of $M$ between 1 and 100.

In Figure H3 we report the bias-corrected estimator $\hat{\beta}_{BC}$ for two of the regression parameters: the coefficient of the mean income beliefs, and the coefficient of current log income. We report the results for different values of $M$. The figure shows that the results are fairly robust to this form of measurement error, with $\hat{\beta}_{BC}$ and $\hat{\beta}_{OLS}$ being close to each other irrespective of $M$. In addition, the variability induced by this form of measurement error, as captured by the dashed lines in the figure, appears moderate.

**G Tax counterfactuals: details about estimation**

In this section of the appendix we detail the calculations of tax counterfactuals and present additional empirical estimates.

**G.1 Tax schedule**

We assume the tax schedule takes the parametric form $T(\tilde{w}_r) = \tilde{w}_r - \lambda \tilde{w}_r^{1-\tau}$, where $\tilde{w}_r$ denotes gross income in multiples of its population average, as in Benabou (2002). This parametric form can be re-written as a similar function that depends on gross income $\tilde{w}$, with the same parameter $\tau$ but a different parameter $\tilde{\lambda}$. For the baseline level of the tax, we rely on the estimates obtained by Holter, Krueger, and Stepanchuk (2019) for Italy, averaged over family composition characteristics in our sample: $\lambda_0 = 0.94$ and $\tau_0 = 0.196$. The approach that we implement here is its simplicity.

9In particular, we still consider a likelihood model with 100 trials and an uninformative prior.

10Specifically, $\tilde{\lambda} = \lambda K^\tau$, for $K$ the average gross income in the population.
Let $\lambda_1$ and $\tau_1$ denote the parameters defining the tax schedule under a counterfactual scenario. We assume the tax schedule applies to gross family income, and that each individual pays taxes proportionally to their contribution in the family, $r_{it}$, a proportion we assume does not change in counterfactual scenarios. Let $x_{1it}$ denote log family income and $(\mu_{0it}, \sigma^2_{0it})$ denote the parameters of income beliefs under a counterfactual scenario. Let $(\mu_{1it}, \sigma^2_{1it})$ denote their baseline values, observed in sample. In this case, 

$$\mu_{1it} - \mu_{0it} = \left[ \log(\tilde{\lambda}_1) - \frac{1 - \tau_1}{1 - \tau_0} \log(\tilde{\lambda}_0) \right] + \frac{(\tau_0 - \tau_1)}{(1 - \tau_0)} \mu_{0it} + \log(r_{it}) \frac{\tau_1 - \tau_0}{1 - \tau_0},$$

$$\sigma^2_{1it} - \sigma^2_{0it} = \sigma^2_{0it} \left[ \frac{(1 - \tau_1)^2}{(1 - \tau_0)^2} - 1 \right],$$

$$x_{1it} - x_{0it} = \log(\tilde{\lambda}_1) - \log(\tilde{\lambda}_0) + \left[ \frac{x_{0it} - \log(\tilde{\lambda}_0)}{1 - \tau_0} \right] (\tau_0 - \tau_1).$$

Given a counterfactual tax schedule $(\lambda_1, \tau_1)$, we can use these values to compute average partial effects.

We consider three counterfactual scenarios. In the transitory tax increase and permanent tax increase counterfactuals, we set $\lambda_1 = \lambda_0 - 0.1$ and $\tau_1 = \tau_0$. In the regressivity counterfactual, we set $\tau_1 = 0.142$, the progressivity parameter of the tax system in France according to Holter, Krueger, and Stepanchuk (2019), and set $\lambda_1$ such that the tax change is revenue neutral.\footnote{Assuming that family gross income is log-normally distributed with parameters $\mu_{r \omega}$ and $\sigma^2_{r \omega}$, a change in the parameters of the tax system is revenue neutral if

$$\log(\tilde{\lambda}_1) - \log(\tilde{\lambda}_0) = \frac{1}{2} \sigma^2_{r \omega} \left[ (1 - \tau_0)^2 - (1 - \tau_1)^2 \right] + \mu_{r \omega} (\tau_1 - \tau_0).$$

Furthermore, $\mu_{r \tilde{\omega}} = (\mu_{x \omega} - \log(\tilde{\lambda}_0))/(1 - \tau_0)$ and $\sigma^2_{r \tilde{\omega}} = \sigma^2_{x \omega}/(1 - \tau_0)$, where $\mu_{x \omega}$ and $\sigma^2_{x \omega}$ are the mean and variance of the log of disposable family income, which we estimate from the SHIW.}

### G.2 Double Lasso estimation

In this subsection we describe how we estimate the consumption function using the double Lasso method introduced by Belloni, Chernozhukov, and Hansen (2014). Consider the equation,

$$y_{it} = d' \Psi(s_{it}) + \beta_k k_{it} + \alpha_i + \epsilon_{it}, \quad (A11)$$

where $\Psi(s_{it})$ includes polynomial functions of the main covariates (age, log income, log assets, and the income beliefs’ means and variances), and $k_{it}$ includes the other demographic controls. Under this specification, an average partial effect corresponding to a counterfactual of interest
is given by
\[ d' \left( \frac{1}{nT} \sum_{i,t} (\Psi(\tilde{s}_{it}) - \Psi(s_{it})) \right) \]
where \( s_{it} \) are the main covariates under the baseline, and \( \tilde{s}_{it} \) are the main covariates under the counterfactual.

Letting
\[ v = \frac{1}{nT} \sum_{i,t} (\Psi(\tilde{s}_{it}) - \Psi(s_{it})) , \]
we first reparameterize the polynomials so that the average partial effect of interest coincides with the coefficient of the first regressor. To that end, we construct an invertible matrix \( A \) whose first column is equal to \( v \).

Then, we rewrite (A11) using the reparameterized polynomials \( \tilde{\Psi}(s_{it}) = A^{-1} \Psi(s_{it}) \), and obtain
\[ y_{it} = (A'a)' \tilde{\Psi}(s_{it}) + \beta_k k_{it} + \alpha_i + \varepsilon_{it}. \] (A12)

Note that the coefficient of the first covariate in (A12) is equal to \( d'v \), which is the average partial effect of interest.

To estimate \( d'v \), we apply the double Lasso estimator to (A12). To account for household fixed effects, we take first differences. We always include (i.e., we do not penalize) the following regressors: the first order polynomials (age, log income, log assets, and the beliefs’ means and variances), as well as the variables in \( k_{it} \) (existence of a spouse, marital status, family size, number of children 0-5, 6-13, 14-17 years old in the household, number of children outside the household, number of income earners in the household, and a wave indicator).

The double Lasso method is implemented in two steps. In a first step, we apply the Lasso to regress the first element in \( \tilde{\Psi}(s_{it}) \) on its second to last elements and \( k_{it} \), in first differences. In the second step, we again apply the Lasso to regress \( y_{it} \) on the second to last elements of \( \tilde{\Psi}(s_{it}) \) and \( k_{it} \), in first differences. In both steps, we choose the penalty parameters by 10-fold cross-validation (Chetverikov, Liao, and Chernozhukov, 2021). Lastly, we regress \( y_{it} \) on the first element in \( \tilde{\Psi}(s_{it}) \) and all the controls selected in the two Lasso steps, again in first differences. We account for estimation uncertainty (in particular, for the fact that \( v \) is estimated) by computing bootstrapped standard errors.

\[ ^{12}\text{For example, we set } A = [v \ t_2 \ldots \ t_L], \text{ where } t_\ell \text{ are the canonical vectors in } \mathbb{R}^L \text{ and } L = \dim \Psi, \text{ provided such a matrix } A \text{ is invertible.} \]
G.3 Empirical estimates

In Table H10 we report average partial effects based on OLS estimates of the consumption function, and average partial effects based on the double Lasso. We show these in graphical form in Figures H4 and H5, respectively. Overall, the results are quite consistent across specifications.

H Appendix tables and figures

Table H1: Simulated tax counterfactuals under rational and adaptive expectations by age

<table>
<thead>
<tr>
<th>Age</th>
<th>Rational expectations</th>
<th>Adaptive expectations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Structural</td>
<td>Semi-structural</td>
</tr>
<tr>
<td></td>
<td>Linear</td>
<td>Quadratic</td>
</tr>
<tr>
<td>CAPE</td>
<td>-0.0663</td>
<td>-0.0599</td>
</tr>
<tr>
<td>DAPE</td>
<td>-0.0471</td>
<td>-0.0550</td>
</tr>
<tr>
<td>TAPE</td>
<td>-0.1134</td>
<td>-0.1149</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age</th>
<th>Rational expectations</th>
<th>Adaptive expectations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Structural</td>
<td>Semi-structural</td>
</tr>
<tr>
<td></td>
<td>Linear</td>
<td>Quadratic</td>
</tr>
<tr>
<td>CAPE</td>
<td>-0.0110</td>
<td>-0.0097</td>
</tr>
<tr>
<td>DAPE</td>
<td>-0.0921</td>
<td>-0.0982</td>
</tr>
<tr>
<td>TAPE</td>
<td>-0.1031</td>
<td>-0.1079</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age</th>
<th>Rational expectations</th>
<th>Adaptive expectations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Structural</td>
<td>Semi-structural</td>
</tr>
<tr>
<td></td>
<td>Linear</td>
<td>Quadratic</td>
</tr>
<tr>
<td>CAPE</td>
<td>-0.0058</td>
<td>-0.0062</td>
</tr>
<tr>
<td>DAPE</td>
<td>-0.0794</td>
<td>-0.0877</td>
</tr>
<tr>
<td>TAPE</td>
<td>-0.0852</td>
<td>-0.0939</td>
</tr>
</tbody>
</table>

Notes: See the notes to Table 1. Results by age.
Table H2: Descriptive statistics on income expectations questions 1989–1991

<table>
<thead>
<tr>
<th>Income growth</th>
<th>Cross-sectional sample</th>
<th>Panel sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Obs</td>
<td>P25</td>
</tr>
<tr>
<td>Income growth &gt; 25%</td>
<td>5,486</td>
<td>0</td>
</tr>
<tr>
<td>Income growth 20 – 25%</td>
<td>5,486</td>
<td>0</td>
</tr>
<tr>
<td>Income growth 15 – 20%</td>
<td>5,486</td>
<td>0</td>
</tr>
<tr>
<td>Income growth 13 – 15%</td>
<td>5,486</td>
<td>0</td>
</tr>
<tr>
<td>Income growth 10 – 13%</td>
<td>5,486</td>
<td>0</td>
</tr>
<tr>
<td>Income growth 8 – 10%</td>
<td>5,486</td>
<td>0</td>
</tr>
<tr>
<td>Income growth 7 – 8%</td>
<td>5,486</td>
<td>0</td>
</tr>
<tr>
<td>Income growth 6 – 7%</td>
<td>5,486</td>
<td>0</td>
</tr>
<tr>
<td>Income growth 5 – 6%</td>
<td>5,486</td>
<td>0</td>
</tr>
<tr>
<td>Income growth 3 – 5%</td>
<td>5,486</td>
<td>0</td>
</tr>
<tr>
<td>Income growth 0 – 3%</td>
<td>5,486</td>
<td>0</td>
</tr>
<tr>
<td>Income growth &lt; 0%</td>
<td>5,486</td>
<td>0</td>
</tr>
<tr>
<td>Income growth - by how much if &lt; 0%</td>
<td>163</td>
<td>3</td>
</tr>
</tbody>
</table>

Notes: Descriptive statistics are weighted using the survey's weights.

Table H3: Descriptive statistics on income expectations questions 1995–1998

<table>
<thead>
<tr>
<th></th>
<th>Cross-sectional sample</th>
<th>Panel sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Obs</td>
<td>P25</td>
</tr>
<tr>
<td>Minimum amount expected to earn</td>
<td>2,310</td>
<td>13,515.1</td>
</tr>
<tr>
<td>Maximum amount expected to earn</td>
<td>2,310</td>
<td>16,109.9</td>
</tr>
<tr>
<td>Prob. of earning less than half</td>
<td>2,302</td>
<td>40.00</td>
</tr>
</tbody>
</table>

Notes: Amounts are in 2010 euros. Descriptive statistics are weighted using the survey's weights.
Table H4: Descriptive statistics

<table>
<thead>
<tr>
<th></th>
<th>Cross-sectional sample</th>
<th>Panel sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Obs</td>
<td>P25</td>
</tr>
<tr>
<td>Log family consumption</td>
<td>7,796</td>
<td>9.78</td>
</tr>
<tr>
<td>Log family assets</td>
<td>7,496</td>
<td>10.03</td>
</tr>
<tr>
<td>Log family income</td>
<td>7,795</td>
<td>10.03</td>
</tr>
<tr>
<td>Log individual income</td>
<td>7,791</td>
<td>9.69</td>
</tr>
<tr>
<td>Mean expected log income</td>
<td>7,796</td>
<td>9.72</td>
</tr>
<tr>
<td>SD expected log income</td>
<td>7,796</td>
<td>0.005</td>
</tr>
</tbody>
</table>

Notes: Amounts are in 2010 euros. Descriptive statistics are weighted using the survey’s weights. Individual income excludes property income and income from transfers. Individual-level variables (i.e., income and income expectations) corresponds to the household head.

Table H5: Predictive power of income beliefs

<table>
<thead>
<tr>
<th></th>
<th>log(\text{w}_{i,t+1})</th>
<th>log(\text{w}_{i,t+1}) - log(\text{w}_t)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Mean expected log income</td>
<td>0.596</td>
<td>0.367</td>
</tr>
<tr>
<td></td>
<td>(0.036)</td>
<td>(0.082)</td>
</tr>
<tr>
<td>Mean expected change in log income</td>
<td>0.659</td>
<td>0.367</td>
</tr>
<tr>
<td></td>
<td>(0.116)</td>
<td>(0.082)</td>
</tr>
<tr>
<td>Log individual income</td>
<td>0.566</td>
<td>0.239</td>
</tr>
<tr>
<td></td>
<td>(0.041)</td>
<td>(0.083)</td>
</tr>
</tbody>
</table>

Controls: Yes Yes Yes Yes Yes Yes Yes Yes
N observations: 2,994 2,994 2,994 2,994 2,994 2,994 2,994 2,994
R-squared: 0.290 0.466 0.400 0.470 0.047 0.098 0.196 0.211

Notes: SHIW, 1989–1991 and 1995–1998. Regression for household heads. Controls include age and age squared, gender, education, indicator of spouse, marital status, family size, number of children 0-5, 6-13, 14-17 years old in the household, number of children outside the household, area, number of income earners in the household, and a wave indicator. Regression estimates are weighted using survey weights. Standard errors (shown in parenthesis) are clustered at the household level.
Table H6: Estimates of the log consumption function: robustness checks

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean expected log income head</td>
<td>0.235</td>
<td>0.229</td>
<td>0.237</td>
<td>0.230</td>
<td>0.235</td>
<td>0.229</td>
<td>0.245</td>
<td>0.242</td>
</tr>
<tr>
<td></td>
<td>(0.095)</td>
<td>(0.095)</td>
<td>(0.095)</td>
<td>(0.094)</td>
<td>(0.095)</td>
<td>(0.093)</td>
<td>(0.095)</td>
<td>(0.093)</td>
</tr>
<tr>
<td>Mean expected log income spouse</td>
<td>0.106</td>
<td>-0.022</td>
<td>0.104</td>
<td>0.103</td>
<td>0.106</td>
<td>0.062</td>
<td>0.111</td>
<td>0.099</td>
</tr>
<tr>
<td></td>
<td>(0.054)</td>
<td>(0.064)</td>
<td>(0.054)</td>
<td>(0.091)</td>
<td>(0.091)</td>
<td>(0.091)</td>
<td>(0.023)</td>
<td>(0.023)</td>
</tr>
<tr>
<td>Log family income</td>
<td>0.438</td>
<td>0.438</td>
<td>0.438</td>
<td>0.438</td>
<td>0.439</td>
<td>0.439</td>
<td>0.428</td>
<td>0.439</td>
</tr>
<tr>
<td></td>
<td>(0.091)</td>
<td>(0.090)</td>
<td>(0.090)</td>
<td>(0.090)</td>
<td>(0.089)</td>
<td>(0.089)</td>
<td>(0.091)</td>
<td>(0.091)</td>
</tr>
<tr>
<td>Log family assets</td>
<td>0.016</td>
<td>0.017</td>
<td>0.018</td>
<td>0.019</td>
<td>0.018</td>
<td>0.018</td>
<td>0.018</td>
<td>0.020</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.024)</td>
<td>(0.023)</td>
<td>(0.023)</td>
<td>(0.023)</td>
<td>(0.023)</td>
<td>(0.023)</td>
<td>(0.023)</td>
</tr>
<tr>
<td>Household fixed effect</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Controls</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Distribution assumption</td>
<td>Disc - Triang</td>
<td>Disc - Triang</td>
<td>Log-normal</td>
<td>Log-normal</td>
<td>Log-normal</td>
<td>Log-normal</td>
<td>Log-normal</td>
<td>Log-normal</td>
</tr>
<tr>
<td>M draws</td>
<td>10</td>
<td>10</td>
<td>50</td>
<td>50</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>N observations</td>
<td>1,514</td>
<td>1,514</td>
<td>1,536</td>
<td>1,536</td>
<td>1,536</td>
<td>1,536</td>
<td>1,536</td>
<td>1,536</td>
</tr>
<tr>
<td>N households</td>
<td>757</td>
<td>757</td>
<td>768</td>
<td>768</td>
<td>768</td>
<td>768</td>
<td>768</td>
<td>768</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.26</td>
<td>0.26</td>
<td>0.26</td>
<td>0.26</td>
<td>0.26</td>
<td>0.26</td>
<td>0.26</td>
<td>0.26</td>
</tr>
<tr>
<td>Mean F beliefs head</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Mean F beliefs spouse</td>
<td>0.74</td>
<td>0.45</td>
<td>0.26</td>
<td>0.26</td>
<td>0.26</td>
<td>0.26</td>
<td>0.26</td>
<td>0.26</td>
</tr>
<tr>
<td>Mean F beliefs head and spouse</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Notes: SHIW, regression for household heads. In columns (1) and (2) we assume a different distribution of beliefs (discrete distribution in waves 1989–1991 and triangular distribution in waves 1995–1998). In columns (3) to (6) we vary the number M of draws used in estimation. In columns (7) and (8), we add spouse’s beliefs (for spouses that are employees and have beliefs questions, and 0 for everyone else). The expectations variables and log family income are centered around the weighted average in the sample. Controls include age and age squared, existence of a spouse, marital status, family size, number of children 0-5, 6-13, 14-17 years old in the household, number of children outside the household, number of income earners in the household, and a wave indicator. In columns (7) and (8), we also control for a categorical variable indicating spousal situation (no spouse, spouse is homemaker, spouse is employee with beliefs questions, spouse is employee without beliefs questions, other). Regression estimates are weighted using survey weights. Standard errors (shown in parenthesis) are clustered at the household level.
Table H7: Estimates of the log consumption function by wave

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean expected log income</td>
<td>0.235</td>
<td>0.229</td>
<td>0.212</td>
<td>0.242</td>
<td>0.323</td>
<td>0.342</td>
</tr>
<tr>
<td></td>
<td>(0.094)</td>
<td>(0.093)</td>
<td>(0.110)</td>
<td>(0.108)</td>
<td>(0.171)</td>
<td>(0.172)</td>
</tr>
<tr>
<td>(Mean expect. log income)-(Log family income)</td>
<td>0.104</td>
<td>0.113</td>
<td>-0.125</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.061)</td>
<td>(0.060)</td>
<td></td>
<td></td>
<td>(0.177)</td>
<td></td>
</tr>
<tr>
<td>Log family income</td>
<td>0.439</td>
<td>0.439</td>
<td>0.461</td>
<td>0.442</td>
<td>0.277</td>
<td>0.264</td>
</tr>
<tr>
<td></td>
<td>(0.089)</td>
<td>(0.089)</td>
<td>(0.101)</td>
<td>(0.100)</td>
<td>(0.169)</td>
<td>(0.168)</td>
</tr>
<tr>
<td>Log family assets</td>
<td>0.018</td>
<td>0.019</td>
<td>0.046</td>
<td>0.048</td>
<td>-0.063</td>
<td>-0.060</td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td>(0.023)</td>
<td>(0.027)</td>
<td>(0.026)</td>
<td>(0.039)</td>
<td>(0.039)</td>
</tr>
<tr>
<td>Household fixed effect</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Controls</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>N observations</td>
<td>1,536</td>
<td>1,536</td>
<td>962</td>
<td>962</td>
<td>512</td>
<td>512</td>
</tr>
<tr>
<td>N households</td>
<td>768</td>
<td>768</td>
<td>481</td>
<td>481</td>
<td>256</td>
<td>256</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.26</td>
<td>0.26</td>
<td>0.35</td>
<td>0.37</td>
<td>0.16</td>
<td>0.17</td>
</tr>
<tr>
<td>Pvalue F beliefs</td>
<td>0.01</td>
<td>0.02</td>
<td>0.05</td>
<td>0.03</td>
<td>0.06</td>
<td>0.14</td>
</tr>
</tbody>
</table>

Notes: SHIW, regression for household heads. The expectations variables and log family income are centered around the weighted average in the sample. Controls include age and age squared, existence of a spouse, marital status, family size, number of children 0-5, 6-13, 14-17 years old in the household, number of children outside the household, number of income earners in the household, and a wave indicator. When available, we also control for other expectations variables: columns (3) and (4) also control for mean expected inflation, and columns (5) and (6) also control for the beliefs about the probability of being employed next year. Regression estimates are weighted using survey weights. Standard errors (shown in parenthesis) are clustered at the household level.
Table H8: Estimates of the log consumption function: robustness to assets

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean expected log income</td>
<td>0.245</td>
<td>0.238</td>
<td>0.167</td>
<td>0.159</td>
<td>0.191</td>
<td>0.186</td>
<td>0.223</td>
<td>0.216</td>
</tr>
<tr>
<td></td>
<td>(0.097)</td>
<td>(0.095)</td>
<td>(0.107)</td>
<td>(0.106)</td>
<td>(0.091)</td>
<td>(0.089)</td>
<td>(0.096)</td>
<td>(0.095)</td>
</tr>
<tr>
<td>(Mean expect. log income) - (Log family income)</td>
<td>0.095</td>
<td>0.093</td>
<td>0.038</td>
<td>0.102</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.061)</td>
<td>(0.062)</td>
<td>(0.068)</td>
<td>(0.060)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log family income</td>
<td>0.410</td>
<td>0.413</td>
<td>0.642</td>
<td>0.648</td>
<td>0.494</td>
<td>0.499</td>
<td>0.475</td>
<td>0.476</td>
</tr>
<tr>
<td></td>
<td>(0.097)</td>
<td>(0.097)</td>
<td>(0.144)</td>
<td>(0.144)</td>
<td>(0.096)</td>
<td>(0.095)</td>
<td>(0.097)</td>
<td>(0.096)</td>
</tr>
<tr>
<td>Log family assets</td>
<td>0.033</td>
<td>0.032</td>
<td>-0.084</td>
<td>-0.087</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td>(0.032)</td>
<td>(0.055)</td>
<td>(0.054)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Log family assets)^2</td>
<td>0.007</td>
<td>0.006</td>
<td></td>
<td></td>
<td>0.051</td>
<td>0.050</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.006)</td>
<td></td>
<td></td>
<td>(0.022)</td>
<td>(0.022)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log (family assets - savings)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.051</td>
<td>0.050</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.022)</td>
<td>(0.022)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: SHIW, regression for household heads. In columns (1) and (2) we control for log assets squared. In columns (3) and (4) we instrument the difference of log family assets by first-period assets and income. In columns (5) and (6) we replace end-of-year family assets by end-of-year family assets minus savings during the year. Lastly, in columns (7) and (8) we do not include any controls for assets. The expectations variables and log family income are centered around the weighted average in the sample. Controls include age and age squared, existence of a spouse, marital status, family size, number of children 0-5, 6-13, 14-17 years old in the household, number of children outside the household, number of income earners in the household, and a wave indicator. Regression estimates are weighted using survey weights. Standard errors (shown in parenthesis) are clustered at the household level.
Table H9: Predicted distribution of number of bins by number of draws $M$

<table>
<thead>
<tr>
<th>Number of bins with non-zero frequencies</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>0.59</td>
<td>0.24</td>
<td>0.09</td>
<td>0.05</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.75</td>
</tr>
<tr>
<td>$M = 1$</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>$M = 2$</td>
<td>0.68</td>
<td>0.32</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.32</td>
</tr>
<tr>
<td>$M = 3$</td>
<td>0.57</td>
<td>0.35</td>
<td>0.08</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.51</td>
</tr>
<tr>
<td>$M = 4$</td>
<td>0.50</td>
<td>0.36</td>
<td>0.12</td>
<td>0.02</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.66</td>
</tr>
<tr>
<td>$M = 5$</td>
<td>0.45</td>
<td>0.37</td>
<td>0.14</td>
<td>0.04</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.78</td>
</tr>
<tr>
<td>$M = 6$</td>
<td>0.42</td>
<td>0.37</td>
<td>0.15</td>
<td>0.05</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.88</td>
</tr>
<tr>
<td>$M = 7$</td>
<td>0.39</td>
<td>0.37</td>
<td>0.16</td>
<td>0.06</td>
<td>0.02</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.96</td>
</tr>
<tr>
<td>$M = 8$</td>
<td>0.36</td>
<td>0.38</td>
<td>0.16</td>
<td>0.06</td>
<td>0.02</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>2.03</td>
</tr>
<tr>
<td>$M = 9$</td>
<td>0.34</td>
<td>0.38</td>
<td>0.17</td>
<td>0.07</td>
<td>0.03</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>2.10</td>
</tr>
<tr>
<td>$M = 10$</td>
<td>0.32</td>
<td>0.39</td>
<td>0.18</td>
<td>0.07</td>
<td>0.03</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>2.16</td>
</tr>
<tr>
<td>$M = 20$</td>
<td>0.17</td>
<td>0.41</td>
<td>0.24</td>
<td>0.09</td>
<td>0.05</td>
<td>0.02</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>2.59</td>
</tr>
<tr>
<td>$M = 30$</td>
<td>0.09</td>
<td>0.39</td>
<td>0.30</td>
<td>0.11</td>
<td>0.06</td>
<td>0.03</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>2.87</td>
</tr>
<tr>
<td>$M = 40$</td>
<td>0.05</td>
<td>0.36</td>
<td>0.34</td>
<td>0.13</td>
<td>0.06</td>
<td>0.03</td>
<td>0.02</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>3.07</td>
</tr>
<tr>
<td>$M = 50$</td>
<td>0.03</td>
<td>0.31</td>
<td>0.38</td>
<td>0.14</td>
<td>0.07</td>
<td>0.04</td>
<td>0.02</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>3.22</td>
</tr>
<tr>
<td>$M = 60$</td>
<td>0.01</td>
<td>0.28</td>
<td>0.41</td>
<td>0.15</td>
<td>0.07</td>
<td>0.04</td>
<td>0.02</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>3.33</td>
</tr>
<tr>
<td>$M = 70$</td>
<td>0.01</td>
<td>0.24</td>
<td>0.43</td>
<td>0.16</td>
<td>0.08</td>
<td>0.04</td>
<td>0.02</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>3.42</td>
</tr>
<tr>
<td>$M = 80$</td>
<td>0.00</td>
<td>0.21</td>
<td>0.45</td>
<td>0.16</td>
<td>0.08</td>
<td>0.04</td>
<td>0.02</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>3.49</td>
</tr>
<tr>
<td>$M = 90$</td>
<td>0.00</td>
<td>0.19</td>
<td>0.46</td>
<td>0.16</td>
<td>0.09</td>
<td>0.04</td>
<td>0.03</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>3.55</td>
</tr>
<tr>
<td>$M = 100$</td>
<td>0.00</td>
<td>0.16</td>
<td>0.48</td>
<td>0.17</td>
<td>0.09</td>
<td>0.05</td>
<td>0.03</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>3.61</td>
</tr>
</tbody>
</table>

Notes: SHIW, 1989–1991, sample from column (3) in Table H7. Each row reports the simulated distribution of the number of non-empty bins in data simulated from a measurement error model with $M$ draws, averaged across observations and $S = 1,000$ simulations.
Table H10: Average partial effects estimates

<table>
<thead>
<tr>
<th>Quintile</th>
<th>Transitory tax counterfactual</th>
<th>Permanent tax counterfactual</th>
<th>Regressivity counterfactual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CAPE</td>
<td>DAPE</td>
<td>TAPE</td>
</tr>
<tr>
<td>A. OLS estimates</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-0.0449</td>
<td>0.0000</td>
<td>-0.0449</td>
</tr>
<tr>
<td></td>
<td>(0.0105)</td>
<td>(0.0000)</td>
<td>(0.0105)</td>
</tr>
<tr>
<td>2</td>
<td>-0.0482</td>
<td>0.0000</td>
<td>-0.0482</td>
</tr>
<tr>
<td></td>
<td>(0.0102)</td>
<td>(0.0000)</td>
<td>(0.0102)</td>
</tr>
<tr>
<td>3</td>
<td>-0.0489</td>
<td>0.0000</td>
<td>-0.0489</td>
</tr>
<tr>
<td></td>
<td>(0.0102)</td>
<td>(0.0000)</td>
<td>(0.0102)</td>
</tr>
<tr>
<td>4</td>
<td>-0.0498</td>
<td>0.0000</td>
<td>-0.0498</td>
</tr>
<tr>
<td></td>
<td>(0.0103)</td>
<td>(0.0000)</td>
<td>(0.0103)</td>
</tr>
<tr>
<td>5</td>
<td>-0.0528</td>
<td>0.0000</td>
<td>-0.0528</td>
</tr>
<tr>
<td></td>
<td>(0.0105)</td>
<td>(0.0000)</td>
<td>(0.0105)</td>
</tr>
<tr>
<td>Total</td>
<td>-0.0489</td>
<td>0.0000</td>
<td>-0.0489</td>
</tr>
<tr>
<td></td>
<td>(0.0102)</td>
<td>(0.0000)</td>
<td>(0.0102)</td>
</tr>
<tr>
<td>B. Double Lasso estimates</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-0.0371</td>
<td>0.0000</td>
<td>-0.0371</td>
</tr>
<tr>
<td></td>
<td>(0.0264)</td>
<td>(0.0000)</td>
<td>(0.0264)</td>
</tr>
<tr>
<td>2</td>
<td>-0.0438</td>
<td>0.0000</td>
<td>-0.0438</td>
</tr>
<tr>
<td></td>
<td>(0.0153)</td>
<td>(0.0000)</td>
<td>(0.0153)</td>
</tr>
<tr>
<td>3</td>
<td>-0.0455</td>
<td>0.0000</td>
<td>-0.0455</td>
</tr>
<tr>
<td></td>
<td>(0.0127)</td>
<td>(0.0000)</td>
<td>(0.0127)</td>
</tr>
<tr>
<td>4</td>
<td>-0.0452</td>
<td>0.0000</td>
<td>-0.0452</td>
</tr>
<tr>
<td></td>
<td>(0.0146)</td>
<td>(0.0000)</td>
<td>(0.0146)</td>
</tr>
<tr>
<td>5</td>
<td>-0.0491</td>
<td>0.0000</td>
<td>-0.0494</td>
</tr>
<tr>
<td></td>
<td>(0.0179)</td>
<td>(0.0000)</td>
<td>(0.0179)</td>
</tr>
<tr>
<td>Total</td>
<td>-0.0452</td>
<td>0.0000</td>
<td>-0.0452</td>
</tr>
<tr>
<td></td>
<td>(0.0129)</td>
<td>(0.0000)</td>
<td>(0.0129)</td>
</tr>
</tbody>
</table>

Notes: SHIW, 1989–1991 and 1995–1998, cross-sectional sample. In the top panel we report results based on OLS estimates, see column (5) in Table 2. In the bottom panel we report estimates based on the double/debiased Lasso, for a dictionary including interactions and power of the covariates up to the third order. Standard errors are based on 1,000 bootstrap replications.
Figure H1: Policy rules by type of expectations and age

A. Rational expectations

(a) 26 years old  (b) 35 years old  (c) 45 years old

B. Adaptive expectations

(a) 26 years old  (b) 35 years old  (c) 45 years old

Notes: The top panel plots policy rules under rational expectations and the bottom panel plots policy rules under adaptive expectations. The horizontal axes show log income and mean beliefs, and the vertical axis shows log consumption. In each figure, assets are fixed at the median value among simulated cases with positive assets. The colors represent the number of observations in the corresponding simulated data set.
Figure H2: Simulation results, rational versus adaptive expectations

A. Consumption

(a) Mean

(b) Variance

B. Assets

(a) Mean

(b) Variance

B. Income

(a) Mean

(b) Variance

Notes: Simulations results based on the structural model. Black lines are results under rational expectations, blue lines are results under adaptive expectations.
Figure H3: Bias-corrected coefficients of mean beliefs and log income

(a) $\beta$ for mean income beliefs

(b) $\beta$ for current log income

Notes: SHIW, 1989–1991, sample from column (3) in Table H7. The horizontal dotted lines show the corresponding elements of $\hat{\beta}^{OLS}$ from column (3) in Table H7. The solid lines show $\hat{\beta}^{BC}$, and the dashed lines add a band of plus or minus twice the standard deviation of $\hat{\beta}^{(s)}$ across simulations. 1,000 simulations.
Figure H4: Average partial effects estimates (OLS)

A. Mean beliefs only

(b) Transitory tax  
Transitory tax

(b) Permanent tax  
Permanent tax

(c) Regressivity
Regressivity

B. Mean beliefs interacted with current log income

(d) Transitory tax  
Transitory tax

(e) Permanent tax  
Permanent tax

(f) Regressivity
Regressivity

C. Mean and variance of beliefs interacted with current log income

(g) Transitory tax  
Transitory tax

(h) Permanent tax  
Permanent tax

(i) Regressivity
Regressivity

Notes: SHIW, 1989–1991 and 1995–1998, cross-sectional sample. Black bars correspond to contemporaneous APE and grey bars correspond to dynamic APE. Total APE are the sums of CAPE and DAPE. The top panel is based on column (2) in Table 2, the middle panel on column (4), and the bottom panel on column (5).
Figure H5: Average partial effects estimates (Lasso)

A. Double Lasso estimates, degree 2

(a) Transitory tax  (b) Permanent tax  (c) Regressivity

B. Double Lasso estimates, degree 3

(d) Transitory tax  (e) Permanent tax  (f) Regressivity

C. Double Lasso estimates, degree 4

(g) Transitory tax  (h) Permanent tax  (i) Regressivity

Notes: SHIW, 1989–1991 and 1995–1998, cross-sectional sample. Black bars correspond to contemporaneous APE and grey bars correspond to dynamic APE. Total APE are the sums of CAPE and DAPE. Double Lasso estimates. The top panel is based on polynomials of degree 2, the middle panel on polynomials of degree 3, and the bottom panel on polynomials of degree 4.