Rationalizing Rational Expectations? Tests and Deviations*

Xavier D'Haultfoeuille† Christophe Gaillac‡ Arnaud Maurel§

March 25, 2020

Abstract

In this paper, we build a new test of rational expectations based on the marginal distributions of realizations and subjective beliefs. This test is widely applicable, including in the common situation where realizations and beliefs are observed in two different datasets that cannot be matched. We show that whether one can rationalize rational expectations is equivalent to the distribution of realizations being a mean-preserving spread of the distribution of beliefs. The null hypothesis can then be rewritten as a system of many moment inequality and equality constraints, for which tests have been recently developed in the literature. Next, we go beyond testing by defining and estimating the minimal deviations from rational expectations that can be rationalized by the data in the context of structural models. We build on this concept to propose an easy-to-implement way to conduct a sensitivity analysis on the assumed form of expectations. Finally, we apply our framework to test for rational expectations about future earnings, and examine the consequences of such departures in the context of a life-cycle model of consumption.

Keywords: Rational expectations; Test; Subjective expectations; Data combination; Sensitivity analysis.

*First version: November 2018. We thank Peter Arcidiacono, Levon Barseghyan, Federico Bugni, Pierre Cahuc, Zhuoli Chen, Tim Christensen, Valentina Corradi, Christian Gourieroux, Nathael Gozlan, Gregory Jolivet, Max Kasy, Jia Li, Matt Masten, Magne Mogstad, Andrew Patton, Aureo de Paula, Mirko Wiederholt, Basit Zafar, Yichong Zhang and seminar participants at Amsterdam, Arizona State, Bocconi, CREST, Duke, Helsinki, Mannheim, National University of Singapore, NC State, OFCE, Oxford, Rice, Sciences Po, Singapore Management University, Surrey, Toronto, Toulouse School of Economics, Washington University in St Louis, and attendees of the 2017 Econometric Study Group (Bristol), the Conference on the Intersection of Econometrics and Applied Micro (Toronto, Oct. 17), the 2017 Triangle Econometrics Conference, the 2018 International Association for Applied Econometrics Conference (Montreal), the 2018 CEME NSF-NBER conference on "Inference in Nonstandard Problems" (Duke), and the 2019 Applied Micro Workshop at the University of Essex (ESRC Research Centre on Micro-Social Change) for useful comments and suggestions.

†CREST, xavier.dhaultfoeuille@ensae.fr.

‡CREST and TSE, christophe.gaillac@tse-fr.eu.

§Duke University, NBER and IZA, arnaud.maurel@duke.edu.

1 Introduction

How individuals form their beliefs about uncertain future outcomes is critical to understanding decision making. Despite longstanding critiques (see, among many others, Pesaran, 1987; Manski, 2004), rational expectations remain by far the most popular framework to describe belief formation (Muth, 1961). This theory states that agents have expectations that do not systematically differ from the realized outcomes, and efficiently process all private information to form these expectations. Rational expectations (RE) are a key building block in many macro- and micro-economic models, and in particular in most of the dynamic microeconomic models that have been estimated over the last two decades (see, e.g., Aguirregabiria and Mira, 2010; Blundell, 2017, for recent surveys).

In this paper, we build a new test of RE. Our test only requires having access to the marginal distributions of subjective beliefs and realizations, and, as such, can be applied quite broadly. In particular, this test can be used in a data combination context, where individual realizations and subjective beliefs are observed in two different datasets that cannot be matched. Such situations are common in practice (see, e.g., Delavande, 2008; Arcidiacono, Hotz and Kang, 2012; Arcidiacono, Hotz, Maurel and Romano, 2014; Stinebrickner and Stinebrickner, 2014*a*; Kuchler and Zafar, 2019; Kapor, Neilson and Zimmerman, 2018). Besides, even in surveys for which an explicit aim is to measure subjective expectations, such as the Michigan Survey of Consumers or the Survey of Consumer Expectations of the New York Fed, expectations and realizations can typically only be matched for a subset of the respondents. And of course, regardless of attrition, whenever one seeks to measure long or medium-term outcomes, matching beliefs with realizations does require waiting for a long period of time before the data can be made available to researchers.

The tests of RE implemented so far in this context only use specific implications of the RE hypothesis. In contrast, we develop a test that exploits all possible implications of RE. Using the key insight that we can rationalize RE if and only if the distribution of realizations is a mean-preserving spread of the distribution of beliefs, we show that rationalizing RE is equivalent to satisfying one moment equality and (infinitely) many moment inequalities.[1] As a consequence, if these moment conditions hold, RE cannot be rejected, given the data at our disposal. By exhausting all relevant implications of RE, our test is able to detect much more violations of rational expectations than existing tests.

To develop a statistical test of RE rationalization, we build on the recent literature on inference based on moment inequalities, and more specifically, on Andrews and Shi (2017). By applying their results to our context, we show that our test controls size asymptotically and is consistent

[1] Interestingly, the equivalence on which we rely, which is based on Strassen's theorem (Strassen, 1965), is also used in the microeconomic risk theory literature, see in particular Rothschild and Stiglitz (1970).

over fixed alternatives. We also provide conditions under which the test is not conservative.

We then consider several extensions to our baseline test. First, we show that by using a set of covariates that are common to both datasets, we can increase our ability to detect violations of RE. Another important issue is that of unanticipated aggregate shocks. Even if individuals have rational expectations, the mean of observed outcomes may differ from the mean of individual beliefs simply because of aggregate shocks. We show that our test can be easily adapted to account for such shocks. Finally, we prove that our test is robust to measurement errors in the following sense. If individuals have rational expectations but both beliefs and outcomes are measured with (classical) errors, then our test does not reject RE provided that the amount of measurement errors on beliefs does not exceed the amount of intervening transitory shocks plus the measurement errors on the realized outcomes. In particular, this allows for elicited beliefs to be noisier than realized outcomes. This provides a rationale for our test even in cases where realizations and beliefs are observed in the same dataset, since a direct test based on a regression of the outcome on the beliefs (see, e.g., Lovell, 1986) is, at least at the population level, not robust to any amount of measurement errors on the subjective beliefs.

Next, we go beyond testing and introduce the concept of minimal deviations from rational expectations that can be rationalized by the data, in the context of a structural model that imposes restrictions on the agents' information sets. To do so, we use tools from the optimal transport literature (see Galichon, 2016, for an overview). Importantly, the deviations do not depend on the particular choice of distance between random variables that we consider. The proposed approach yields a natural and easy-to-implement sensitivity check on the assumed form of expectations. This procedure does not require observing the beliefs in the same dataset as the one used to estimate the model, and can thus be used quite generally. Overall, this method offers a middle ground between estimating structural choice models based on realized data only (standard approach a la Rust, 1987; Keane and Wolpin, 1997), and estimating more flexible choice models using subjective beliefs (as in, e.g., Stinebrickner and Stinebrickner, 2014*b*; Delavande and Zafar, 2019). In that sense, our approach shares similarities with Van der Klaauw (2012), who also made use of subjective beliefs observed from an auxiliary dataset to estimate a dynamic structural model. However, in contrast to our work, this paper maintains the RE hypothesis, focusing instead on the efficiency gains from incorporating subjective expectations data in the analysis.

We apply our framework to test for rational expectations about future earnings. To do so, we combine elicited beliefs about future earnings with realized earnings, using data from the Labor Market module of the Survey of Consumer Expectations (SCE, New York Fed), and test whether household heads form rational expectations on their annual labor earnings. While a naive test of equality of means between earnings beliefs and realizations shows that earnings expectations are realistic in the sense of not being significantly biased, thus not rejecting the rational expectations hypothesis, our test does reject rational expectations at the 1% level. Taken together, our findings illustrate the practical importance of incorporating the additional restrictions of rational expectations that are embedded in our test. The results of our test also indicate that the RE hypothesis is more credible for certain subpopulations than others. For instance, we reject RE for individuals without a college degree, who exhibit substantial deviations from RE. On the other hand, we fail to reject the hypothesis that college-educated workers have rational expectations on their future earnings.

Finally, we explore the sensitivity of a standard life-cycle incomplete markets model of consumption to violations of the rational expectations hypothesis. Even though agents are about right on average about their future earnings, we show that minimal deviations from RE entail substantial changes in the predicted responses of consumers to income shocks. In addition to underlining the sensitivity of the model to the RE hypothesis, our results show that departures from RE account for some of the over-insurance to permanent income shocks, as well as the excess sensitivity of consumption to transitory shocks that have been documented in the literature (see, e.g., Hall and Mishkin, 1982; Blundell, Pistaferri and Preston, 2008; Kaplan and Violante, 2010).

By developing a test of rational expectations in a setting where realizations and subjective beliefs are observed in two different datasets, we bring together the literature on data combination (see, e.g., Cross and Manski, 2002, Molinari and Peski, 2006, Fan, Sherman and Shum, 2014, Buchinsky, Li and Liao, 2018, and Ridder and Moffitt, 2007 for a survey), and the literature on testing for rational expectations in a micro environment (see, e.g., Lovell, 1986; Gourieroux and Pradel, 1986; Ivaldi, 1992, for seminal contributions).

On the empirical side, we contribute to a rapidly growing literature on the use of subjective expectations data in economics (see, e.g., Manski, 2004; Delavande, 2008; Van der Klaauw and Wolpin, 2008; Van der Klaauw, 2012; Arcidiacono, Hotz, Maurel and Romano, 2014; de Paula, Shapira and Todd, 2014; Stinebrickner and Stinebrickner, 2014*b*; Wiswall and Zafar, 2015). In this paper, we show how to incorporate all of the relevant information from subjective beliefs combined with realized data to test for, and measure deviations from rational expectations.

By developing a new framework allowing to examine the sensitivity of behavioral models to departures from the rational expectations hypothesis, we also contribute to a small but growing body of research estimating structural choice models without imposing rational expectations (see, e.g., Buchinsky and Leslie, 2010; Stinebrickner and Stinebrickner, 2014*a*; Barseghyan, Molinari and Teitelbaum, 2016; Kapor, Neilson and Zimmerman, 2018; and Agarwal and Somaini, 2018). We add to this literature by showing how a sensitivity analysis of the RE hypothesis can be conducted in frequent situations where the data used to estimate the structural model does not include beliefs, but such beliefs are observed in another dataset. At a broad level, our approach based on minimal deviations from rational expectations shares similarities with earlier work by Hansen and co-authors which attempts to find, in the context of empirical asset pricing, the minimal beliefs distorsion that is required to rationalize Euler equations (Hansen, 2014). It is also related to the cost statistic approach proposed by Barseghyan et al. (2016) to quantify the extent to which choice data violates the restrictions implied by expected utility maximization.

Finally, our approach also shares some similarities with the sensitivity analysis methods recently proposed in the econometrics literature by Andrews, Gentzkow and Shapiro (2017), Armstrong and Kolesár (2018), Bonhomme and Weidner (2018), Christensen and Connault (2019) and Masten and Poirier (2019). Importantly, like Christensen and Connault (2019) and Masten and Poirier (2019), our approach allows for deviations that do not become negligible as the sample size grows.

The remainder of the paper is organized as follows. In Section 2, we present the main theoretical equivalences underlying our RE test, and the corresponding statistical tests. Section 3 illustrates the finite sample properties of our tests through Monte Carlo simulations. In Section 4, we discuss how to use deviations from rational expectations in order to conduct a sensitivity analysis in structural models. Section 5 applies our framework to expectations about future earnings. Finally, Section 6 concludes. The appendix collects various theoretical extensions, additional simulation results, additional material on the application, and all the proofs. The companion R package RationalExp, described in the user guide (D'Haultfœuille, Gaillac and Maurel, 2018), performs the test of RE.

2 Testing rational expectations

2.1 Set-up

We assume that the researcher has access to a first dataset containing the individual outcome variable of interest, which we denote by Y. She also observes, through a second dataset drawn from the same population, the elicited individual expectation on Y, denoted by ψ. The two datasets, however, cannot be matched. We focus on situations where the researcher has access to elicited beliefs about mean outcomes, as opposed to probabilistic expectations about the full distribution of outcomes. The type of subjective expectations data we consider in the paper has been collected in various contexts, and used in a number of prior studies (see, among others, Delavande, 2008; Zafar, 2011*b*; Arcidiacono, Hotz and Kang, 2012; Arcidiacono, Hotz, Maurel and Romano, 2014; Hoffman and Burks, 2020).

Formally, $\psi = \mathcal{E}[Y|\mathcal{I}]$, where $\mathcal{I}$ denotes the σ-algebra corresponding to the agent's information set and $\mathcal{E}[\cdot|\mathcal{I}]$ is the subjective expectation operator (i.e. for any U, $\mathcal{E}[U|\mathcal{I}]$ is a $\mathcal{I}$-measurable random variable). We are interested in testing the rational expectations (RE) hypothesis

$\psi = \mathbb{E}[Y|\mathcal{I}]$, where $\mathbb{E}[\cdot|\mathcal{I}]$ is the conditional expectation operator generated by the true data generating process. Importantly, we remain agnostic throughout most of our analysis on the information set $\mathcal{I}$. Our setting is also compatible with heterogeneity in the information different agents use to form their expectations. To see this, let $(U_1, ..., U_m)$ denote m variables that agents may or may not observe when they form their expectations, and let $D_k = 1$ if U_k is observed, 0 otherwise. Then, if $\mathcal{I}$ is the information set generated by $(D_1U_1, ..., D_mU_m)$, agents will use different subsets of the $(U_k)_{k=1...m}$ (i.e., different pieces of information) depending on the values of the $(D_k)_{k=1...m}$. By remaining agnostic on the information set, our analysis complements several studies which primarily focus on testing for different information sets, while maintaining the rational expectations assumption (see Cunha and Heckman, 2007, for a survey).

It is easy to see that the RE hypothesis imposes restrictions on the joint distribution of realizations Y and beliefs ψ. In this data combination context, the relevant question of interest is then whether one can rationalize RE, in the sense that there exists a triplet $(Y', \psi', \mathcal{I}')$ such that (i) the pair of random variables (Y', ψ') are compatible with the marginal distributions of Y and ψ; and (ii) ψ' correspond to the rational expectations of Y', given the information set $\mathcal{I}'$, i.e., $\mathbb{E}(Y'|\mathcal{I}') = \psi'$. Hence, we consider the test of the following hypothesis:

$$\text{H}_0 : \text{there exists a pair of random variables } (Y', \psi') \text{ and a sigma-algebra } \mathcal{I}' \text{ such that}$$
$$\sigma(\psi') \subset \mathcal{I}', Y' \sim Y,\ \psi' \sim \psi \text{ and } \mathbb{E}\left[Y'\middle|\mathcal{I}'\right] = \psi',$$

where $\sim$ denotes equality in distribution. Rationalizing RE does not mean that the true realizations Y, beliefs ψ and information set $\mathcal{I}$ are such that $\mathbb{E}[Y|\mathcal{I}] = \psi$. Instead, it means that there exists a triplet $(Y', \psi', \mathcal{I}')$ consistent with the data and such that $\mathbb{E}[Y'|\mathcal{I}'] = \psi'$. In other words, rejecting H_0 implies that RE does not hold, in the sense that the true realizations, beliefs, and information set do not satisfy RE ($\mathbb{E}[Y|\mathcal{I}] \neq \psi$). The converse, however, is not true.

2.2 Equivalences

2.2.1 Main equivalence

Let $\delta = \mathbb{E}[Y] - \mathbb{E}[\psi]$, F_ψ and F_Y denote the cumulative distribution functions (cdf) of ψ and Y, $x^+ = \max(0, x)$, and define

$$\Delta(y) = \int_{-\infty}^{y} F_Y(t) - F_\psi(t)dt.$$

Throughout most of our analysis, we impose the following regularity conditions on the distributions of realized outcomes (Y) and subjective beliefs (ψ):

Assumption 1 $\mathbb{E}(|Y|) < +\infty$ *and* $\mathbb{E}(|\psi|) < +\infty$.

The following preliminary result will be useful subsequently.

Lemma 1 *Suppose that Assumption 1 holds. Then H_0 holds if and only if there exists a pair of random variables (Y', ψ') such that $Y' \sim Y$, $\psi' \sim \psi$ and $\mathbb{E}[Y'|\psi'] = \psi'$.*

Lemma 1 states that in order to test for H_0, we can focus on the constraints on the joint distribution of Y and ψ, and ignore those related to the information set. This is intuitive given that we impose no restrictions on this set. Our main result is Theorem 1 below. It states that rationalizing RE (i.e., H_0) is equivalent to a set of many moment inequality and equality constraints.

Theorem 1 *Suppose that Assumption 1 holds. The following statements are equivalent:*

(i) *H_0 holds;*

(ii) *(F_Y is a mean-preserving spread of F_ψ) $\Delta(y) \geq 0$ for all $y \in \mathbb{R}$ and $\delta = 0$;*

(iii) $\mathbb{E}\left[(y - Y)^+ - (y - \psi)^+\right] \geq 0$ *for all* $y \in \mathbb{R}$ *and* $\delta = 0$.

The implication (i) $\Rightarrow$ (iii) and the equivalence between (ii) and (iii) are simple to establish. The key part of the result is to prove that (iii) implies (i). To show this, we first use Lemma 1, which states that H_0 is equivalent to the existence of (Y', ψ') such that $Y' \sim Y$, $\psi' \sim \psi$ and $\mathbb{E}[Y'|\psi'] = \psi'$. Then the result essentially follows from Strassen's theorem (Strassen, 1965, Theorem 8).

It is interesting to note that Theorem 1 is related to the theory of risk in microeconomic theory. In particular, using the terminology of Rothschild and Stiglitz (1970), (ii) states that realizations (Y) are more risky than beliefs (ψ). The main value of Theorem 1, from a statistical point of view, is to transform H_0 into the set of moment inequality (and equality) restrictions given by (iii). We show in Section 2.3 how to build a statistical test of these conditions.

Comparison with alternative approaches We now compare our approach with alternative ones that have been proposed in the literature. In the following discussion, as in this whole section, we reason at the population level and thus ignore statistical uncertainty. Accordingly, the "tests" we consider here are formally deterministic, and we compare them in terms of data generating processes violating the null hypothesis associated with each of them.

Our approach can clearly detect many more violations of rational expectations than the "naive" approach based solely on the equality $\mathbb{E}(Y) = \mathbb{E}(\psi)$. It also detects more violations than the approach based on the restrictions $\mathbb{E}(Y) = \mathbb{E}(\psi)$ and $\mathbb{V}(Y) \geq \mathbb{V}(\psi)$ (approach based on the variance), which has been considered in particular in the macroeconomic literature on the accuracy and rationality of forecasts (see, e.g. Patton and Timmermann, 2012). On the other hand, and as expected since it relies on the joint distribution of (Y, ψ), the "direct" approach for testing RE, based on $\mathbb{E}(Y|\psi) = \psi$, can detect more violations of rational expectations than ours.

To better understand the differences between these four different approaches ("naive", variance, "direct", and ours), it is helpful to consider important particular cases. Of course, if $\psi = \mathbb{E}[Y|\mathcal{I}]$, individuals are rational and none of the four approaches leads to reject RE. Next, consider departures from rational expectations of the form $\psi = \mathbb{E}[Y|\mathcal{I}] + \eta$, with η independent of $\mathbb{E}[Y|\mathcal{I}]$. If $\mathbb{E}(\eta) \neq 0$, subjective beliefs are biased, and individuals are on average either over-pessimistic or over-optimistic. It follows that $\mathbb{E}(Y) \neq \mathbb{E}(\psi)$, implying that all four approaches lead to reject RE.

More interestingly, if $\mathbb{E}(\eta) = 0$, individuals' expectations are right on average, and the naive approach does not lead to reject RE. However, it is easy to show that, as long as deviations from RE are heterogeneous in the population ($\mathbb{V}(\eta) > 0$), the direct approach always leads to a rejection. In this setting, our approach constitutes a middle ground, in which rejection of RE depends on the degree of dispersion of the deviations from RE (η) relative to the uncertainty shocks ($\varepsilon = Y - \mathbb{E}(Y|\mathcal{I})$). In other words and intuitively, we reject RE whenever departures from RE dominate the uncertainty shocks affecting the outcome. Formally, and using similar arguments as in Proposition 4 in Subsection 2.2.4, one can show that if ε is independent of $\mathbb{E}[Y|\mathcal{I}]$, we reject H_0 as long as the distribution of the uncertainty shocks stochastically dominates at the second-order the distribution of the deviations from RE.

Specifically, if $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ and $\eta \sim \mathcal{N}(0, \sigma_\eta^2)$, we reject RE if and only if $\sigma_\eta^2 > \sigma_\varepsilon^2$. In such a case, our approach boils down to the variance approach mentioned above: we reject whenever $\mathbb{V}(\psi) > \mathbb{V}(Y)$. But interestingly, if the discrepancy (η) between beliefs and RE is not normally distributed, we can reject H_0 even if $\mathbb{V}(\psi) \leq \mathbb{V}(Y)$. Suppose for instance that $\varepsilon \sim \mathcal{N}(0, 1)$ and

$$\eta = a \left(-\mathbb{1}\{U \leq 0.1\} + \mathbb{1}\{U \geq 0.9\}\right), \ U \sim \mathcal{U}[0,1] \text{ and } a > 0.$$

In other words, 80% of individuals are rational, 10% are over-pessimistic and form expectations equal to $\mathbb{E}[Y|\mathcal{I}] - a$, whereas 10% are over-optimistic and expect $\mathbb{E}[Y|\mathcal{I}] + a$. Then one can show that our approach leads to reject RE when $a \geq 1.755$, while for $a = 1.755$, $\mathbb{V}(\eta) \simeq 0.616 \leq \mathbb{V}(\varepsilon) = 1$.

Binary outcome Our equivalence result does not require the outcome Y to be continuously distributed. In the particular case where Y is binary, our test reduces to the naive test of $\mathbb{E}(Y) = \mathbb{E}(\psi)$. Indeed, when Y is a binary outcome and $\psi \in [0,1]$, one can easily show that as long as $\mathbb{E}(Y) = \mathbb{E}(\psi)$, the inequalities $\mathbb{E}\left[(y - Y)^+ - (y - \psi)^+\right] \geq 0$ automatically hold for all

$y \in \mathbb{R}$. This applies to expectations about binary events, such as, e.g., being employed or not at a given date.

Interpretation of the boundary condition To shed further light on our test and on the interpretation of H_0, it is instructive to derive the distributions of $Y|\psi$ that correspond to the boundary condition ($\Delta(y) = 0$). The proposition below shows that, in the presence of rational expectations, agents whose beliefs ψ lies at the boundary of H_0 have perfect foresight, i.e. $\psi = \mathbb{E}[Y|\mathcal{I}] = Y$.[2]

Proposition 1 *Suppose that (Y, ψ) satisfies RE, $u \mapsto F^{-1}_{Y|\psi}(\tau|u)$ is continuous for all $\tau \in (0,1)$, and $\Delta(y_0) = 0$ for some y_0 in the interior of the support of ψ. Then $Y|\psi = y_0$ is degenerate.*

2.2.2 Equivalence with covariates

In practice we may observe additional variables $X \in \mathbb{R}^{d_X}$ in both datasets. Assuming that X is in the agent's information set, we modify H_0 as follows:[3]

H_{0X} : there exists a pair of random variables (Y', ψ') and a sigma-algebra $\mathcal{I}'$ such that $\sigma(\psi', X) \subset \mathcal{I}'$, $Y'|X \sim Y|X$, $\psi'|X \sim \psi|X$ and $\mathbb{E}\left[Y'\middle|\mathcal{I}'\right] = \psi'$.

Adding covariates increases the number of restrictions that are implied by the rational expectation hypothesis, thus improving our ability to detect violations of rational expectations. Proposition 2 below formalizes this idea and shows that H_{0X} can be expressed as a system of many conditional moment inequalities and equalities.

Proposition 2 *Suppose that Assumption 1 holds. The following two statements are equivalent:*

(i) H_{0X} holds;

(ii) Almost surely, $\mathbb{E}\left[(y-Y)^+ - (y-\psi)^+\middle|X\right] \geq 0$ for all $y \in \mathbb{R}$ and $\mathbb{E}\left[Y - \psi|X\right] = 0$.

Moreover, if H_{0X} holds, H_0 holds as well.

2.2.3 Equivalence with unpredictable aggregate shocks

Oftentimes, the outcome variable is affected not only by individual-specific shocks, but also by aggregate shocks. We denote by C the random variable corresponding to the aggregate shocks. The issue, in this case, is that we observe a single realization of C (c), along with the outcome

[2] For any cdf F, we let F^{-1} denote its quantile function, namely $F^{-1}(\tau) = \inf\{x : F(x) \geq \tau\}$.

[3] See complementary work by Gutknecht et al. (2018), who use subjective expectations data to relax the rational expectations assumption, and propose a method allowing to test whether specific covariates are included in the agents' information sets.

variable conditional on that realization $C = c$. In other words, we only identify $F_{Y|C=c}$ rather than F_Y, as the latter would require to integrate over the distribution of all possible aggregate shocks. Moreover, the restriction $\mathbb{E}[Y|C = c, \psi] = \psi$ is generally violated, even though the rational expectations hypothesis holds. It follows that one cannot directly apply our previous results by simply replacing F_Y by $F_{Y|C=c}$. In such a case, one has to make additional assumptions on how the aggregate shocks affect the outcome.

To illustrate our approach, let us consider the example of individual income. Suppose that the logarithm of income of individual i at period t, denoted by Y_{it}, satisfies a Restricted Income Profile model:

$$Y_{it} = \alpha_i + \beta_t + \varepsilon_{it},$$

where β_t capture aggregate (macroeconomic) shocks, ε_{it} follows a zero-mean random walk, and α_i, $(\beta_t)_t$ and $(\varepsilon_{it})_t$ are assumed to be mutually independent. Let $\mathcal{I}_{it-1}$ denote individual i's information set at time $t-1$, and suppose that $\mathcal{I}_{it-1} = \sigma\left(\alpha_i, (\beta_{t-k})_{k\geq1}, (\varepsilon_{it-k})_{k\geq1}\right)$. If individuals form rational expectations on their future outcomes, their beliefs in period $t-1$ about their future log-income in period t are given by

$$\psi_{it} = \mathbb{E}\left[Y_{it}|\mathcal{I}_{it-1}\right] = \alpha_i + \mathbb{E}\left[\beta_t|(\beta_{t-k})_{k\geq1}\right] + \varepsilon_{it-1}.$$

Thus, $Y_{it} = \psi_{it} + C_t + \varepsilon_{it} - \varepsilon_{it-1}$, with $C_t = \beta_t - \mathbb{E}\left[\beta_t|(\beta_{t-k})_{k\geq1}\right]$. The corresponding conditional expectation is given by:

$$\mathbb{E}\left[Y_{it}|\mathcal{I}_{it-1}, C_t = c_t\right] = \psi_{it} + c_t \neq \psi_{it}.$$

To get closer to our initial set-up, we now drop indexes i and t and maintain the conditioning on the aggregate shocks $C = c$ implicit. Under these conventions, rationalizing RE does not correspond to $\mathbb{E}[Y|\mathcal{I}] = \psi$, but instead to $\mathbb{E}[Y|\mathcal{I}] = c_0 + \psi$ for some $c_0 \in \mathbb{R}$. A similar reasoning applies to multiplicative instead of additive aggregate shocks. In such a case, the null takes the form $\mathbb{E}[Y|\mathcal{I}] = c_0\psi$, for some $c_0 > 0$. In these two examples, c_0 is identifiable: by $c_0 = \mathbb{E}(Y) - \mathbb{E}(\psi)$ in the additive case, and by $c_0 = \mathbb{E}(Y)/\mathbb{E}(\psi)$ in the multiplicative case. Moreover, note that there exists in both cases a known function $q(y, c)$ such that $\mathbb{E}(q(Y, c_0)) = \mathbb{E}(\psi)$, namely $q(y, c) = y - c$ and $q(y, c) = y/c$ for additive and multiplicative shocks, respectively.

More generally, we consider the following null hypothesis for testing RE in the presence of aggregate shocks:

> H_{0S} : there exist random variables $\left(Y', \psi'\right)$, a sigma-algebra $\mathcal{I}'$ and $c_0 \in \mathbb{R}$ such that $\sigma(\psi') \subset \mathcal{I}'$, $Y' \sim Y$, $\psi' \sim \psi$ and $\mathbb{E}\left[q\left(Y', c_0\right)|\mathcal{I}'\right] = \psi'$.

where $q(.,.)$ is a known function supposed to satisfy the following restrictions.

Assumption 2 *$\mathbb{E}(|\psi|) < +\infty$ and for all c, $\mathbb{E}(|q(Y,c)|) < +\infty$. Moreover, $\mathbb{E}[q(Y,c)] = \mathbb{E}[\psi]$ admits a unique solution, c_0.*

By applying our main equivalence result (Theorem 1) to $q(Y, c_0)$ and ψ, we obtain the following result.

Proposition 3 *Suppose that Assumption 2 holds. Then the following statements are equivalent:*

(i) H_{0S} *holds;*

(ii) $\mathbb{E}\left[(y - q(Y, c_0))^+ - (y - \psi)^+\right] \geq 0$ *for all* $y \in \mathbb{R}$.

A couple of remarks are in order. First, this result can be extended in a straightforward way to a setting with covariates. This is important not only to increase the ability of our test to detect violations of RE, but also because this allows for aggregate shocks that differ across observable groups. We discuss further this extension, and the corresponding statistical test, in Appendix A. Second, in the presence of aggregate shocks, the null hypothesis does not involve a moment equality restriction anymore; the corresponding moment is used instead to identify c_0. Related, a clear limitation of the naive test ($\mathbb{E}(Y) = \mathbb{E}(\psi)$) is that, unlike our test, it is not robust to aggregate shocks. In this case, rejecting the null could either stem from violations of the rational expectation hypothesis, or simply from the presence of aggregate shocks.

2.2.4 Robustness to measurement errors

We have assumed so far that Y and ψ were perfectly observed; yet measurement errors in survey data are pervasive (see, e.g. Bound, Brown and Mathiowetz, 2001). We explore in the following the extent to which our test is robust to measurement errors. Specifically, assume that the true variables (ψ, Y) are unobserved. Instead, we only observe $\widehat{\psi}$ and $\widehat{Y}$, which are affected by classical measurement errors.[4]

Namely:

$$\begin{aligned} \widehat{\psi} &= \psi + \xi_\psi \quad \text{with} \quad \xi_\psi \perp\!\!\!\perp \psi,\ \mathbb{E}[\xi_\psi] = 0 \\ \widehat{Y} &= Y + \xi_Y \quad \text{with} \quad \xi_Y \perp\!\!\!\perp Y,\ \mathbb{E}[\xi_Y] = 0. \end{aligned} \tag{1}$$

Then one can show that if RE holds (and assuming away aggregate shocks, for simplicity), so that $\mathbb{E}[Y|\psi] = \psi$, it is nevertheless the case that $\mathbb{E}\left[\widehat{Y}\middle|\widehat{\psi}\right] \neq \widehat{\psi}$, as long as $\mathrm{Cov}(\xi_Y, \widehat{\psi}) = \mathrm{Cov}(\xi_\psi, Y) = 0$ and $\mathbb{V}(\xi_\psi) > 0$. In other words, the direct test is not robust to any measurement errors on the

[4] See Zafar (2011*a*) who does not find evidence of non-classical measurement errors on subjective beliefs elicited from a sample of Northwestern undergraduate students. We conjecture that our test is robust to some forms of non-classical measurement errors. However, it seems difficult in this case to obtain a general result similar to the one in Proposition 4.

subjective beliefs ψ. Even if individuals have rational expectations, the direct test will reject the null in the presence of even a small degree of measurement errors on the elicited beliefs. The following proposition shows that our test, on the other hand, is robust to a certain degree of measurement errors on the beliefs. As above, we let $\varepsilon = Y - \psi$ denote the uncertainty shocks.

Proposition 4 *Suppose that Y and ψ satisfy H_0 and let $\left(\widehat{\psi}, \widehat{Y}\right)$ be defined as in (1). Suppose also that $\varepsilon + \xi_Y \perp\!\!\!\perp \psi$ and F_{ξ_ψ} dominates at the second order $F_{\xi_Y+\varepsilon}$. Then $\widehat{Y}$ and $\widehat{\psi}$ satisfy H_0.*

The key condition is that F_{ξ_ψ} dominates at the second order $F_{\xi_Y+\varepsilon}$, or, equivalently here, that $F_{\xi_Y+\varepsilon}$ is a mean-preserving spread of F_{ξ_ψ}. Recall that in the case of normal variables, $\xi_\psi \sim \mathcal{N}(0, \sigma_1^2)$ and $\xi_Y + \varepsilon \sim \mathcal{N}(0, \sigma_2^2)$, this is in turn equivalent to imposing $\sigma_1^2 \leq \sigma_2^2$. Thus, even if there is no measurement error on Y, so that $\xi_Y = 0$, this condition may hold provided that the variance of measurement errors on ψ is smaller than the variance of the uncertainty shocks on Y. More generally, this allows elicited beliefs to be - potentially much - noisier than realized outcomes, a setting which may be relevant in practice. Overall, these results support the use of our test rather than the direct test even in cases where realizations and beliefs are observed in the same dataset.

2.2.5 Other extensions

We now briefly discuss other relevant directions in which Theorem 1 can be extended. First, another potential source of uncertainty on ψ is rounding. Rounding practices by interviewees are common in the case of subjective beliefs. Under additional restrictions, it is possible in such a case to construct bounds on the true beliefs ψ (see, e.g., Manski and Molinari, 2010). We show in Appendix B that our test can be generalized to accommodate this rounding practice.

Second, we have implicitly maintained the assumption so far that subjective beliefs and realized outcomes are drawn from the same population. In Appendix C, we relax this assumption and show that our test can be easily extended to allow for sample selection under unconfoundedness, through an appropriate reweighting of the observations.

Third, our equivalence result and our test can be extended to accommodate situations with multiple outcomes $(Y_k)_{k=1,..,K}$ and multiple subjective beliefs $(\psi_k)_{k=1,..,K}$ associated with each of these outcomes. Specifically, whether one can rationalize rational expectations in this environment can be written as:

$$\mathbb{E}(Y_k|\psi_1, ..., \psi_K) = \psi_k, \text{ for all } k \in \{1, ..., K\}$$

which, in turn, is equivalent to the distribution of the outcomes Y_k being a mean-preserving spread of the distribution of the beliefs ψ_k. This situation arises in various contexts, including cases where respondents declare their subjective probabilities of making particular choices among $K+1$ possible alternatives. This also arises in situations where expectations about the distribution of a continuous outcome Y are elicited through questions of the form "what do you think is the percent chance that [Y] will be greater than [y]?", for different values $(y_k)_{k=1,..,K}$. In such cases, it is natural to build a RE test based on the multiple outcomes $(\mathbb{1}\{Y > y_k\})_{k=1,..,K}$ and subjective beliefs $(\psi_k)_{k=1,..,K}$, where ψ_k is the subjective survival function of Y evaluated at y_k.

2.3 Statistical tests

We now propose a testing procedure for H_{0X}, which can be easily adapted to the case where no covariate common to both datasets is available to the analyst. To simplify notation, we use a potential outcome framework to describe our data combination problem. Specifically, instead of observing (Y,ψ), we suppose to observe only, in addition to the covariates X, $\widetilde{Y} = DY + (1-D)\psi$ and D, where $D = 1$ (resp. $D = 0$) if the unit belongs to the dataset of Y (resp. ψ). As in Subsection 2.1, we assume that the two samples are drawn from the same population, which amounts to supposing that $D \perp\!\!\!\perp (X, Y, \psi)$ (see Assumption 3-(i) below).[5] In order to build our test, we use the characterization (ii) of Proposition 2:

$$\mathbb{E}\left[(y-Y)^+ - (y-\psi)^+ \middle| X\right] \geq 0 \quad \forall y \in \mathbb{R} \;\text{ and }\; \mathbb{E}\left[Y - \psi | X\right] = 0.$$

Equivalently but written more compactly with $\widetilde{Y}$ only,

$$\mathbb{E}\left[W\left(y-\widetilde{Y}\right)^+ \middle| X\right] \geq 0 \quad \forall y \in \mathbb{R} \;\text{ and }\; \mathbb{E}\left[W\widetilde{Y} \middle| X\right] = 0,$$

where $W = D/\mathbb{E}(D) - (1-D)/\mathbb{E}(1-D)$. This formulation of the null hypothesis allows us to apply the instrumental functions approach of Andrews and Shi (2017, AS), who consider the issue of testing many conditional moment inequalities and equalities. We then build on their results to establish that our test controls size asymptotically and is consistent over fixed alternatives.[6] The initial step is to transform the conditional moments into the following unconditional moments conditions:

$$\mathbb{E}\left[W\left(y-\widetilde{Y}\right)^+ g(X)\right] \geq 0, \quad \mathbb{E}\left[(Y-\psi)\, g(X)\right] = 0.$$

for all $y \in \mathbb{R}$ and g belonging to a suitable class of non-negative functions.

We suppose to observe a sample $(D_i, X_i, \widetilde{Y}_i)_{i=1...n}$ of n i.i.d. copies of $(D, X, \widetilde{Y})$. We consider instrumental functions g that are indicators of belonging to specific hypercubes within $[0,1]^{d_X}$, hence we tranform the variables X_i to lie in $[0,1]^{d_X}$. For notational convenience, we let $\widetilde{X}_i$

[5]See Appendix C for a discussion of how to extend our test to allow for sample selection under unconfoundedness.

[6]Other testing procedures could be used to implement our test, such as that proposed by Linton et al. (2010).

denote the nontransformed vector of covariates, and redefine X_i as:

$$X_i = \Phi_0\left(\widehat{\Sigma}_{\widetilde{X},n}^{-1/2}\left(\widetilde{X}_i - \overline{\widetilde{X}}_i\right)\right),$$

where, for any $x = (x_1, \ldots, x_{d_X})$, we let $\Phi_0(x) = (\Phi(x_1), \ldots, \Phi(x_{d_X}))^\top$. Here Φ denotes the standard normal cdf, $\widehat{\Sigma}_{\widetilde{X},n}$ is the sample covariance matrix of $\left(\widetilde{X}_i\right)_{i=1\ldots n}$ and $\overline{\widetilde{X}}_n$ its sample mean.

Specifically, we consider instrumental functions g belonging to the class of functions $\mathcal{G}_r = \{g_{a,r},\ a \in A_r\}$, with $A_r = \{1, 2, \ldots, 2r\}^{d_X}$ $(r \geq 1)$, $g_{a,r}(x) = \mathbb{1}\{x \in C_{a,r}\}$ and, for any $a = (a_1, ..., a_{d_X})^\top \in A_r$,

$$C_{a,r} = \prod_{u=1}^{d_X}\left(\frac{a_u - 1}{2r}, \frac{a_u}{2r}\right].$$

Finally, to define the test statistic T, we need to introduce additional notations. First, we define, for any given y,

$$m\left(D_i, \widetilde{Y}_i, X_i, g, y\right) = \begin{pmatrix} m_1\left(D_i, \widetilde{Y}_i, X_i, g, y\right) \\ m_2\left(D_i, \widetilde{Y}_i, X_i, g, y\right)\end{pmatrix} = \begin{pmatrix} w_i\left(y - \widetilde{Y}_i\right)^+ g\left(X_i\right) \\ w_i \widetilde{Y}_i g\left(X_i\right)\end{pmatrix}, \tag{2}$$

where $w_i = nD_i / \sum_{j=1}^n D_j - n(1-D_i)/\sum_{j=1}^n (1-D_j)$. Let $\overline{m}_n(g,y) = \sum_{i=1}^n m\left(D_i, \widetilde{Y}_i, X_i, g, y\right)/n$ and define similarly $\overline{m}_{n,j}$ for $j = 1, 2$. For any function g and any $y \in \mathbb{R}$, we also define, for some $\epsilon > 0$,

$$\overline{\Sigma}_n(g,y) = \widehat{\Sigma}_n(g,y) + \epsilon \text{Diag}\left(\widehat{\mathbb{V}}\left(\widetilde{Y}\right), \widehat{\mathbb{V}}\left(\widetilde{Y}\right)\right),$$

where $\widehat{\Sigma}_n(g,y)$ is the sample covariance matrix of $\sqrt{n}\overline{m}_n\left(g,y\right)$ and $\widehat{\mathbb{V}}\left(\widetilde{Y}\right)$ is the empirical variance of $\widetilde{Y}$. We then denote by $\overline{\Sigma}_{n,jj}(g,y)$ $(j = 1,2)$ the j-th diagonal term of $\overline{\Sigma}_n(g,y)$.

Then the (Cramér-von-Mises) test statistic T is defined by

$$T = \sup_{y \in \widehat{\mathcal{Y}}} \sum_{r=1}^{r_n} \frac{(2r)^{-d_X}}{(r^2+100)} \sum_{a \in A_r}\left((1-p)\left(\left(-\frac{\sqrt{n}\overline{m}_{n,1}\left(g_{a,r}, y\right)}{\overline{\Sigma}_{n,11}(g_{a,r}, y)^{1/2}}\right)^+\right)^2 + p\left(\frac{\sqrt{n}\overline{m}_{n,2}\left(g_{a,r}, y\right)}{\overline{\Sigma}_{n,22}(g_{a,r}, y)^{1/2}}\right)^2\right),$$

where $\widehat{\mathcal{Y}} = \left[\min_{i=1,\ldots,n} \widetilde{Y}_i, \max_{i=1,\ldots,n} \widetilde{Y}_i\right]$, $p \in (0,1)$ is a parameter that weights the moments inequalities versus equalities and $(r_n)_{n \in \mathbb{N}}$ is a deterministic sequence tending to infinity.

To test for rational expectations in the absence of covariates, we set the instrumental function equal to the constant function $g(X) = 1$, and the test statistic is simply written as:

$$T = \sup_{y \in \widehat{\mathcal{Y}}}\left((1-p)\left(\left(-\frac{\sqrt{n}\overline{m}_{n,1}(y)}{\overline{\Sigma}_{n,11}(y)^{1/2}}\right)^+\right)^2 + p\left(\frac{\sqrt{n}\overline{m}_{n,2}(y)}{\overline{\Sigma}_{n,22}(y)^{1/2}}\right)^2\right),$$

where, using the notations introduced above, $\overline{m}_{n,j}(y) = \overline{m}_{n,j}(1,y)$ and $\overline{\Sigma}_{n,jj}(y) = \overline{\Sigma}_{n,jj}(1,y)$ $(j = 1,2)$.

Whether or not covariates are included, the resulting test is of the form $\varphi_{n,\alpha} = \mathbb{1}\{T > c^*_{n,\alpha}\}$ where the estimated critical value $c^*_{n,\alpha}$ is obtained by bootstrap using as in AS the Generalized Moment Selection method. Specifically, we follow these three steps:

1. Compute the function $\overline{\varphi}_n(y,g) = \left(\overline{\varphi}_{n,1}(y,g), 0\right)^\top$ for (y,g) in $\widehat{\mathcal{Y}} \times \cup_{r=1}^{r_n}\mathcal{G}_r$, with
$$\overline{\varphi}_{n,1}(y,g) = \overline{\Sigma}_{n,11}^{1/2} B_n \mathbb{1}\left\{\frac{n^{1/2}}{\kappa_n}\overline{\Sigma}_{n,11}^{-1/2}\overline{m}_{n,1}(y,g) > 1\right\},$$
and where $B_n = (b_0 \ln(n)/\ln(\ln(n)))^{1/2}$, $b_0 > 0$, $\kappa_n = (\kappa \ln(n))^{1/2}$, and $\kappa > 0$. To compute $\overline{\Sigma}_{n,11}$, we fix ϵ to 0.05, as in AS.

2. Let $\left(D_i^*, \widetilde{Y}_i^*, X_i^*\right)_{i=1,\dots,n}$ denote a bootstrap sample, i.e., an i.i.d. sample from the empirical cdf of $\left(D, \widetilde{Y}, X\right)$, and compute from this sample the bootstrap counterparts of $\overline{m}_n$ and $\overline{\Sigma}_n$, $\overline{m}_n^*$ and $\overline{\Sigma}_n^*$. Then compute the bootstrap counterpart of T, T^*, replacing $\overline{\Sigma}_n(y, g_{a,r})$ and $\sqrt{n}\overline{m}_n(y, g_{a,r})$ by $\overline{\Sigma}_n^*(y, g_{a,r})$ and $\sqrt{n}\left(\overline{m}_n^* - \overline{m}_n\right)(y, g_{a,r}) + \overline{\varphi}_n(y, g_{a,r})$, respectively.

3. The threshold $c^*_{n,\alpha}$ is the quantile (conditional on the data) of order $1 - \alpha + \eta$ of $T^* + \eta$ for some $\eta > 0$. Following AS, we set η to 10^{-6}.

Note that, despite the multiple steps involved, the testing procedure remains computationally easily tractable. In particular, for the baseline sample we use in our application (see Section 5.1), the RE test only takes 2 minutes.[7]

We now turn to the asymptotic properties of the test. For that purpose, it is convenient to introduce additional notations. Let $\mathcal{Y}$ and $\mathcal{X}$ denote the support of Y and X respectively, and
$$\mathcal{L}_F = \left\{(y, g_{a,r}) : y \in \mathcal{Y},\ (a,r) \in A_r \times \mathbb{N} :\ \mathbb{E}_F\left[W\left(y - \widetilde{Y}\right)^+ g_{a,r}(X)\right] = 0\right\},$$
where, to make the dependence on the underlying probability measure explicit, $\mathbb{E}_F$ denotes the expectation with respect to the distribution F of $\left(D, \widetilde{Y}, X\right)$. Finally, let $\mathcal{F}$ denote a subset of all possible cumulative distribution functions of $\left(D, \widetilde{Y}, X\right)$ and $\mathcal{F}_0$ be the subset of $\mathcal{F}$ such that H_{0X} holds. We impose the following conditions on $\mathcal{F}$ and $\mathcal{F}_0$.

Assumption 3

(i) For all $F \in \mathcal{F}$, $D \perp\!\!\!\perp (X, Y, \psi)$;

[7] This CPU time is obtained using our companion R package, on an Intel Xeon CPU E5-2643, 3.30GHz with 256Gb of RAM.

(ii) *There exists* $M > 0$ *such that* $\widetilde{Y} \in [-M, M]$ *for all* $F \in \mathcal{F}$*. Also,* $\inf_{F\in\mathcal{F}} \mathbb{V}_F\left(\widetilde{Y}\right) > 0$ *and* $0 < \inf_{F\in\mathcal{F}} \mathbb{E}_F[D] \leq \sup_{F\in\mathcal{F}} \mathbb{E}_F[D] < 1$*;*

(iii) *For all* $F \in \mathcal{F}_0$*,* K_F*, the asymptotic covariance kernel of* $n^{-1/2}\text{Diag}\left(\mathbb{V}_F\left(\widetilde{Y}\right)\right)^{-1/2}\overline{m}_n$ *is in a compact set* $\mathcal{K}_2$ *of the set of all* 2×2 *matrix valued covariance kernels on* $\mathcal{Y} \times \cup_{r\geq 1}\mathcal{G}_r$ *with uniform metric* d *defined by*

$$d(K, K') = \sup_{(y,g,y',g')\in\left(\mathcal{Y}\times\cup_{r\geq 1}H_r\right)^2} \left\|K(y, g, y', g') - K'(y, g, y', g')\right\|.$$

The main result of this section is Theorem 2. It shows that, under Assumption 3, the test $\varphi_{n,\alpha}$ controls the asymptotic size and is consistent over fixed alternatives.

Theorem 2 *Suppose that* $r_n \to \infty$ *and Assumption 3 holds. Then:*

(i) $\limsup_{n\to\infty} \sup_{F\in\mathcal{F}_0} \mathbb{E}_F[\varphi_{n,\alpha}] \leq \alpha$*;*

(ii) *If there exists* $F_0 \in \mathcal{F}_0$ *such that* $\mathcal{L}_{F_0}$ *is nonempty and there exists* (j, y_0, g_0) *in* $\{1, 2\} \times \mathcal{L}_{F_0}$ *such that* $K_{F_0,jj}(y_0, g_0, y_0, g_0) > 0$*, then, for any* $\alpha \in [0, 1/2)$*,*

$$\lim_{\eta\to 0} \limsup_{n\to\infty} \sup_{F\in\mathcal{F}_0} \mathbb{E}_F[\varphi_{n,\alpha}] = \alpha.$$

(iii) *If* $F \in \mathcal{F}\backslash\mathcal{F}_0$*, then* $\lim_{n\to\infty} \mathbb{E}_F(\varphi_{n,\alpha}) = 1$*.*

Theorem 2 (i) is closely related to Theorem 5.1 and Lemma 2 in AS. It shows that the test $\varphi_{n,\alpha}$ controls the asymptotic size, in the sense that the supremum over $\mathcal{F}_0$ of its level is asymptotically lower or equal to α. To prove this result, the key is to establish that, under Assumption 3, the class of transformed unconditional moment restrictions that characterize the null hypothesis satisfies a manageability condition (see Pollard, 1990). Using arguments from Hsu (2016), we then exhibit cases of equality in Theorem 2 (ii), showing that, under mild additional regularity conditions, the test has asymptotically exact size (when letting η tend to zero). Finally, Theorem 2 (iii), which is based on Theorem 6.1 in AS, shows that the test is consistent over fixed alternatives.

Extension to account for aggregate shocks This testing procedure can be easily modified to accommodate unanticipated aggregate shocks. Specifically, using the notation defined in Section 2.2.3, we consider the same test as above after replacing $\widetilde{Y}$ by $\widetilde{Y}_{\widehat{c}} = Dq(Y, \widehat{c}) + (1 - D)\psi$, where $\widehat{c}$ denotes a consistent estimator of c_0. The resulting test is given by $\varphi_{n,\alpha,\widehat{c}} = \mathbb{1}\left\{T(\widehat{c}) > c^*_{n,\alpha}\right\}$ (where $T(\widehat{c})$ is obtained by replacing $\widetilde{Y}$ by $\widetilde{Y}_{\widehat{c}}$ in the original test statistic). Such tests have the same properties as those above under some mild regularity conditions on $q(\cdot, \cdot)$, which hold in particular for the leading examples of additive and multiplicative shocks ($q(y, c) = y - c$ and $q(y, c) = y/c$). We refer the reader to Appendix A for a detailed discussion of this extension.

3 Monte Carlo simulations

In the following we study the finite sample performances of the test without covariates through Monte Carlo simulations. The finite sample performances of the version of our test that accounts for covariates are reported and discussed in Appendix D.

We suppose that the outcome Y is given by

$$Y = \rho\psi + \varepsilon,$$

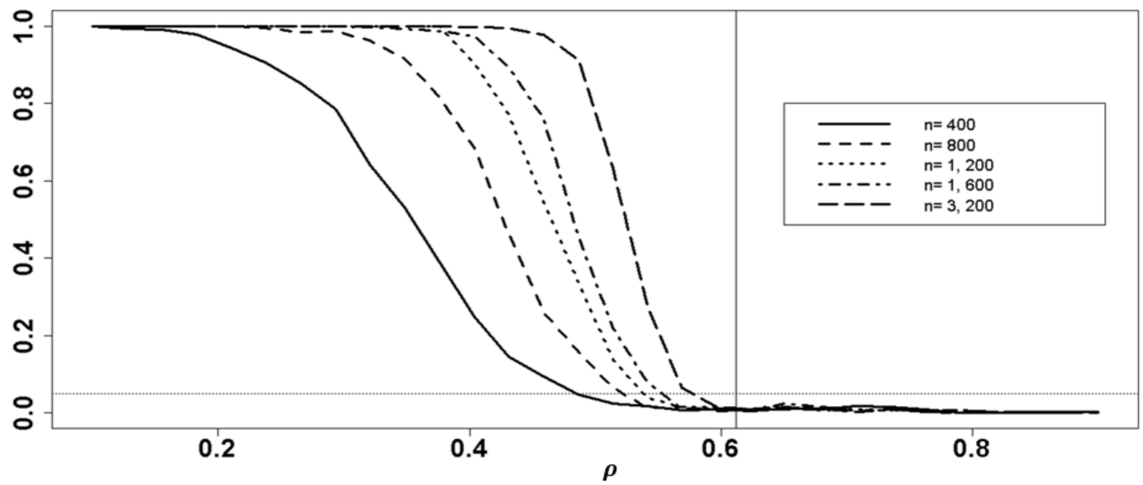
with $\rho \in [0,1]$, $\psi \sim \mathcal{N}(0,1)$ and

$$\varepsilon = \zeta \left(-\mathbb{1}\{U \leq 0.1\} + \mathbb{1}\{U \geq 0.9\} \right),$$

where ζ, U and ψ are mutually independent, $\zeta \sim \mathcal{N}(2, 0.1)$ and $U \sim \mathcal{U}[0,1]$.

In this setup, $\mathbb{E}(Y|\psi) = \rho\psi$ and expectations are rational if and only if $\rho = 1$. But since we observe Y and ψ in two different datasets, there are values of $\rho \neq 1$ for which our test cannot reject the null hypothesis. More precisely, we can show that as the sample size n grows to infinity, we reject the null if and only if $\rho \leq \rho^* \simeq 0.616$. Besides, given this data generating process, the naive test $\mathbb{E}(Y) = \mathbb{E}(\psi)$ always fails to reject RE, while the RE test based on variances is only able to detect a subset of violations of RE that correspond to $\rho < 0.445$.

Results reported in Figure 1 show the power curves of the test φ_α for five different sample sizes ($n_Y = n_\psi = n \in \{400; 800; 1,200; 1,600; 3,200\}$) as a function of the parameter ρ, using 800 simulations for each value of ρ. We use 500 bootstrap simulations to compute the critical values of the test. The test statistic T involves the three tuning parameters b_0, κ, and p (see Section 2.3 for definitions). As described p.643 in Andrews and Shi (2013), there exists in practice a large range of admissible values for these parameters. Following Section 4.2 of Beare and Shi (2019), we set them equal to the smallest (resp. highest) value such that the rejection rate under the null is below the nominal size 0.05, and obtain $b_0 = 0.3$, $\kappa = 0.001$, and $p = 0.05$.



Notes: The vertical line at $\rho \simeq 0.616$ corresponds to the theoretical limit for the rejection of the null hypothesis using our test. The dotted horizontal line corresponds to the 5% level.

Figure 1: Power curves.

Several remarks are in order. First, as expected, under the alternative (i.e. for values of $\rho \leq \rho^* = 0.616$), rejection frequencies increase with the sample size n. In particular, for the largest sample size $n = 3,200$, our test always results in rejection of the RE hypothesis for values of ρ as large as .45. Second, in this setting, our test is conservative in the sense that rejection frequencies under the null are smaller than $\alpha = 0.05$, for all sample sizes. This should not necessarily come as a surprise since the test proposed by AS has been shown to be conservative in alternative finite-sample settings (see, *e.g.* Table 1 p.22 in AS for the case of first-order stochastic dominance tests). However, for the version of our test that accounts for covariates and for the data generating process considered in Appendix D, rejection frequencies under the null are very close to the nominal level.

4 Sensitivity to departures from rational expectations in structural models

In this section, we go beyond testing and show how one can combine subjective beliefs and realizations to perform sensitivity analyses on the assumed form of expectations in structural models. Specifically, we show below that if an auxiliary dataset with subjective beliefs is available to the researcher, one can modify the model-implied rational expectations in a minimal and tractable way so as to make them compatible with the distribution of subjective beliefs. We then recommend to evaluate the sensitivity of parameters or predictions of interest to these departures from rational expectations.[8]

An alternative way of evaluating how critical the rational expectations hypothesis is for a given model would be to relax it and estimate it using instead elicited beliefs about future outcomes both on and off the agents' actual choice paths. However, the data requirements are formidable, and, as a consequence, this approach has only been pursued in a handful of studies (see, e.g., Arcidiacono et al., 2014; Stinebrickner and Stinebrickner, 2014*a*,*b*; Wiswall and Zafar, 2015, 2018). Our approach can be used much more broadly, as it applies to the frequent cases where the model cannot be estimated using available subjective beliefs data.

4.1 Minimal deviations from RE and pseudo-beliefs

We consider a structural model that imposes both a rational expectations formation process and an information set $\mathcal{I}^M$ of the agents, such that individual expectations about the outcome Y are given by $\mathbb{E}\left[Y|\mathcal{I}^M\right]$. In the following, we refer to this assumption ($\psi = \mathbb{E}\left[Y\middle|\mathcal{I}^M\right]$) as the restricted RE hypothesis. Note that with auxiliary data on the subjective beliefs, we can test for the restricted RE hypothesis by simply testing whether $F_\psi = F_{\mathbb{E}[Y|\mathcal{I}^M]}$.

Suppose that the restricted RE hypothesis is rejected. Then, consider the set

$$\boldsymbol{\Psi}^M = \left\{(\psi', \psi'') : \ \psi' \sim \psi, \ \psi'' \sim \mathbb{E}\left[Y\middle|\mathcal{I}^M\right]\right\}.$$

If we reject RE, there is no pair of the form (ψ', ψ') in $\boldsymbol{\Psi}^M$.[9] The goal here is then to find a pair $(\psi', \psi'') \in \boldsymbol{\Psi}^M$ such that ψ' is as close to ψ'' as possible, in the sense of a family of metrics defined below. The discrepancy between the restricted model-based RE and the beliefs ψ' corresponds to the minimal deviations from RE that are consistent with the data on subjective beliefs.[10] Theorem 3 below shows that there exists a solution to this problem, which turns out to be independent of the metric. To define this solution, we introduce $h^M = F_\psi^{-1} \circ F_{\mathbb{E}[Y|\mathcal{I}^M]}$.

Theorem 3 *Suppose that $F_{E[Y|\mathcal{I}^M]}$ has no atom. Then, for any convex function $\rho : \mathbb{R}^+ \to \mathbb{R}^+$ satisfying $\rho(0) = 0$, we have*

$$\left(h^M\left(\mathbb{E}\left[Y\middle|\mathcal{I}^M\right]\right), \mathbb{E}\left[Y\middle|\mathcal{I}^M\right]\right) \in \arg \min_{(\psi', \psi'') \in \boldsymbol{\Psi}^M} \mathbb{E}[\rho(|\psi' - \psi''|)]. \quad (3)$$

Moreover, if ρ is strictly convex, $h^M\left(\mathbb{E}\left[Y\middle|\mathcal{I}^M\right]\right)$ is unique: for any other ψ' such that

$$(\psi', \mathbb{E}\left[Y\middle|\mathcal{I}^M\right]) \in \arg \min_{(\psi', \psi'') \in \boldsymbol{\Psi}^M} \mathbb{E}[\rho(|\psi' - \psi''|)],$$

[8] Of course, as is the case with sensitivity analysis in general, such an exercise is especially informative if the other assumptions of the structural model are satisfied.

[9] Note that the distribution of rational expectations $\mathbb{E}\left[Y\middle|\mathcal{I}^M\right]$ is identified. It follows that the set $\boldsymbol{\Psi}^M$ only involves the distribution of expectations and does not depend on the distribution of realizations.

[10] At a high level, it is interesting to note that our approach to measuring the deviations from RE is similar in spirit to the approach proposed by Hansen and Jagannathan (1997) to quantify specification error when estimating stochastic discount factors in the context of GMM asset pricing models.

we have $\psi' = h^M\left(\mathbb{E}\left[Y\middle|\mathcal{I}^M\right]\right)$ *almost surely.*

Theorem 3 implies that among all random variables that are consistent with the true subjective beliefs, $h^M\left(\mathbb{E}\left[Y\middle|\mathcal{I}^M\right]\right)$ is closest to the rational expectations $\mathbb{E}\left[Y\middle|\mathcal{I}^M\right]$, for any metric indexed by ρ. Theorem 3 relies on results on optimal transport on the real line. In such a case, the optimal map has been shown to be independent of the cost function (see, e.g., Rachev and Rüschendorf, 1998, Chapter 3), which is why the minimal deviations from RE do not depend on the specific metric considered.

A couple of remarks are in order. First, $h^M\left(\mathbb{E}\left[Y\middle|\mathcal{I}^M\right]\right)$ is $\mathcal{I}^M$-measurable, which implies that it is compatible with the information set $\mathcal{I}^M$ imposed by the model. Second, by construction, $h^M\left(\mathbb{E}\left[Y\middle|\mathcal{I}^M\right]\right)$ is consistent with the observed subjective beliefs, since their marginal distributions coincide. Hence, given the data and the constraints imposed by the model on the information set, we can rationalize that $\psi = h^M\left(\mathbb{E}\left[Y\middle|\mathcal{I}^M\right]\right)$. For this reason, we refer to $h^M\left(\mathbb{E}\left[Y\middle|\mathcal{I}^M\right]\right)$ as *pseudo-beliefs*. We use the term pseudo-beliefs here to emphasize that, even though both sets of beliefs are observationally equivalent, $h^M\left(\mathbb{E}\left[Y\middle|\mathcal{I}^M\right]\right)$ does in general not coincide with the true subjective expectations ψ. By construction, the pseudo-beliefs are identifiable. Finally, it directly follows from their definition that the pseudo-beliefs are a monotone transformation of rational expectations. Many of the beliefs distortions that have been considered in the behavioral literature are assumed to follow a similar type of monotonicity property (see, e.g., Section 4.2 in Barseghyan et al., 2016, and references therein).

4.2 Implementation

We describe below how pseudo-beliefs can be used in the context of structural models to conduct a sensitivity analysis to violations of RE. Our proposed approach consists in the following four steps:

1. **Estimation of the baseline model, under RE.** We denote the estimated parameter vector by $\widehat{\theta}_0$, whereas the true parameter vector is denoted by θ_0.

2. **Estimation of the pseudo-beliefs**. First, note that the minimal deviations h^M can be estimated using

$$\widehat{h}^M = \widehat{F}_{\psi}^{-1} \circ F_{\mathbb{E}[Y|\mathcal{I}^M],\widehat{\theta}_0}, \tag{4}$$

   where $F_{\mathbb{E}[Y|\mathcal{I}^M],\widehat{\theta}_0}$ denotes the distribution of RE $\mathbb{E}\left[Y\middle|\mathcal{I}^M\right]$ obtained when the model parameter vector θ is set equal to its estimate $\widehat{\theta}_0$ (computed in Step 1), and $\widehat{F}_{\psi}^{-1}$ is the empirical quantile of the subjective beliefs, which is obtained from the auxiliary subjective data. It then follows that the pseudo-beliefs $h^M\left(\mathbb{E}\left[Y\middle|\mathcal{I}^M\right]\right)$ are simply estimated by an equipercentile mapping from the distribution of rational expectations $F_{\mathbb{E}[Y|\mathcal{I}^M],\widehat{\theta}_0}$ to the distribution of subjective beliefs ψ.

3. **Estimation of the alternative (non-RE) model**. Same as Step 1, after replacing $\mathbb{E}\left[Y\middle|\mathcal{I}^M\right]$ by the pseudo-beliefs $\widehat{h}^M\left(\mathbb{E}\left[Y\middle|\mathcal{I}^M\right]\right)$ estimated in Step 2. We denote the estimated parameter vector by $\widehat{\theta}_1$.

4. **Sensitivity of parameters (or predictions) of interest to departures from RE**. Parameters of interest are a function of the model parameters, as well as, potentially, of the beliefs about future outcomes. For instance, in our application to a life-cycle consumption model discussed in Section 5, we focus on the partial insurance coefficients to transitory and permanent income shocks. These coefficients are computed as a function of the estimated model parameters ($\widehat{\theta}_0$ and $\widehat{\theta}_1$ for the baseline and alternative model) and of the individual beliefs about future earnings ($\mathbb{E}\left[Y\middle|\mathcal{I}^M\right]$ and $\widehat{h}^M(\mathbb{E}\left[Y\middle|\mathcal{I}^M\right])$ in the baseline and alternative case).

This approach provides a way to assess the sensitivity of the findings to violations of RE, holding fixed the restrictions on the information set implied by the model. In particular, findings from the baseline model that exhibit significant sensitivity to these minimal deviations should be interpreted with caution.

We conclude this section by noting that, in certain models, the parameters or predictions of interest may also involve subjective beliefs about additional features of the distribution of future outcomes, such as, e.g., subjective variances. In these cases, one can still use our approach to evaluate the sensitivity of the model predictions to minimal deviations from rational expectations, after replacing rational expectations by pseudo-beliefs about future outcomes, while leaving the higher-order moments unchanged.[11]

4.3 Asymptotic properties

We now turn to the asymptotic properties of the estimated parameters in this sensitivity analysis. As detailed above, the parameter of interest, which we denote by ϕ_1, depends in general on the vector of parameters of the model and on the distribution of individual beliefs. Therefore, for the alternative (non-RE) model we consider in the sensitivity analysis, ϕ_1 is estimated by:

$$\widehat{\phi}_1 = w\left(\widehat{\theta}_1, F_{\widehat{h}^M(\mathbb{E}[Y|\mathcal{I}^M]),\widehat{\theta}_0}\right),$$

where $\widehat{h}^M$ is defined in Equation (4) and $\widehat{\theta}_1$ is obtained by re-estimating the structural model using the estimated pseudo-beliefs $\widehat{h}^M(\mathbb{E}\left[Y\middle|\mathcal{I}^M\right])$ instead of $\mathbb{E}\left[Y\middle|\mathcal{I}^M\right]$ (Step 3 of the implementation discussed above). Because $\widehat{h}^M$ depend both on $\widehat{F}_\psi$ and $\widehat{\theta}_0$, $\widehat{\phi}_1$ can be rewritten as

$$\widehat{\phi}_1 = f\left(\widehat{\theta}_0, \widehat{\theta}_1, \widehat{F}_\psi\right),$$

[11] Of course, in a richer data environment where elicited beliefs about higher moments of the outcome distribution are available to the analyst, one can also use our method to compute the pseudo-beliefs associated with each of these moments, and then incorporate the corresponding departures from RE in the model.

for some known function f. Assuming that θ_0 is estimated by maximum likelihood or GMM, $\widehat{\theta}_1$ can be represented as a two-step GMM estimator, where $(\widehat{\theta}_0, \widehat{F}_\psi)$ are the (root-n consistent) first-step estimators. Thus, under standard regularity conditions, $\widehat{\theta}_1$, along with $\widehat{\theta}_0$ and $\widehat{F}_\psi$, admit a linear expansion (see, e.g. Chen, Linton and Van Keilegom, 2003). Root-n consistency and asymptotic normality of $\widehat{\phi}_1$ follows, as long as f is Hadamard differentiable. Importantly for practical purposes, bootstrap will be valid under standard regularity conditions (Chen et al., 2003).

5 Application to earnings expectations

5.1 Data

Using the tests developed in Section 2.3, we now investigate whether household heads form rational expectations on their future earnings. We use for this purpose data from the Survey of Consumer Expectations (SCE), a monthly household survey that has been conducted by the Federal Reserve Bank of New York since 2012 (see Armantier, Topa, Van der Klaauw and Zafar, 2017, for a detailed description of the survey, and Kuchler and Zafar, 2019; Conlon, Philossoph, Wiswall and Zafar, 2018; Fuster, Kaplan and Zafar, 2018 for recent articles using the SCE). The SCE is conducted with the primary goal of eliciting consumer expectations about inflation, household finance, labor market, as well as housing market. It is a rotating internet-based panel of about 1,200 household heads, in which respondents participate for up to twelve months.[12] Each month, the panel consists of about 180 entrants, and 1,100 repeated respondents. While entrants are overall fairly similar to the repeated respondents, they are slightly older and also have slightly lower incomes (see Table 1 in Armantier et al., 2017).

Of particular interest for this paper is the supplementary module on labor market expectations. This module is repeated every four months since March 2014. Since March 2015, respondents are asked the following question about labor market earnings expectations (ψ) over the next four months: "What do you believe your annual earnings will be in four months?". Implicit throughout the rest of our analysis is the assumption that these elicited beliefs correspond to the mean of the subjective beliefs distribution.[13] In this module, respondents are also asked about current job outcomes, including their current annual earnings (Y), through the following question: "How much do you make before taxes and other deductions at your [main/current] job, on an annual basis?".

[12]Each survey takes on average about fifteen minutes to complete, and respondents are paid $15 per survey completed.

[13]This assumption, while often made in the subjective expectations literature, is *a priori* restrictive. In this application, for the vast majority of the sub-groups of the population, the mean of ψ cannot be statistically distinguished from the one of Y (see Table 1 below). This provides empirical support for this assumption.

Specifically, we use for our baseline test the elicited earnings expectations (ψ), which are available for two cross-sectional samples of household heads who were working either full-time or part-time at the time of the survey, and responded to the labor market module in March 2015 and July 2015 respectively. We combine this data with current earnings (Y) declared in July 2015 and November 2015 by the respondents who are working full-time or part-time at the time of the survey.[14] This leaves us with a final sample of 2,993 observations, which is composed of 1,565 earnings expectation observations, and 1,428 realized earnings observations (see Table 4 in Appendix E.1 for descriptive statistics).[15]

5.2 Are earnings expectations rational?

In Table 1 below, we report the results from the naive test of RE ($\mathbb{E}(Y) = \mathbb{E}(\psi)$), and our preferred test ("Full RE"), where we allow for multiplicative aggregate shocks. We implement the tests both on the overall population and on separate subgroups. The latter approach allows us not only to identify which groups fail to rationalize RE, but also, and importantly, to account for the possibility that aggregate shocks may in fact differ across subgroups.[16]

[14] Throughout our analysis (with the exception of the number of observations reported in Table 1) we use the monthly survey weights of the SCE in order to obtain an estimation sample that is representative of the population of U.S. household heads. See Armantier et al. (2017) for more details on the construction of these weights. We also Winsorize the top 5 percentile of the distributions of realized earnings and earnings beliefs.

[15] 51% (1,536) of these observations correspond to the sub-sample of respondents who are reinterviewed at least once.

[16] In practice we Winsorize the distribution of realized earnings (Y) and earnings beliefs (ψ) at the 95% level. We show in Table 5 in Appendix E.2 that our results are robust to other levels of Winsorization.

Table 1: Tests of RE on annual earnings

	$\mathbb{E}(Y-\psi)/\mathbb{E}(Y)$	Naive RE (p-val)	Variance RE (p-val)	Full RE (p-val)	Number of obs. ψ	Y
All	0.034	0.23	0.71	$< 0.001^{**}$	1,565	1,428
Women	0.059	0.13	0.62	$< 0.001^{**}$	730	649
Men	0.025	0.48	0.58	0.210	835	779
White	0.032	0.31	0.67	0.021^{*}	1,200	1,097
Minorities	0.046	0.43	0.60	$< 0.006^{**}$	365	331
College degree	-0.001	0.96	0.50	0.130	1,106	1,053
No college degree	0.093	0.04^{*}	0.57	0.013^{*}	459	375
High numeracy	0.033	0.28	0.62	0.012^{*}	1,158	1,070
Low numeracy	0.055	0.27	0.58	0.022^{*}	407	358
Tenure $\leq$ 6 months	0.105	0.24	0.63	$< 0.001^{**}$	271	180
Tenure $>$ 6 months	0.007	0.81	0.65	$0.091^{\dagger}$	1,294	1,248

Notes: significance levels: †: 10%, *: 5%, **: 1%. "Naive RE" denotes the naive RE test of equality of means between Y and ψ. "Variance RE" denotes the variance RE test where the null hypothesis is the variance of Y being greater or equal than the variance of ψ, once we account for aggregate, multiplicative shocks. "Full RE" denotes the test without covariates, where we test H_{0S} with $q(y,c) = y/c$. We use 5,000 bootstrap simulations to compute the critical values of the Full RE test. Distributions of realized earnings (Y) and earnings beliefs (ψ) are both Winsorized at the 95% quantile.

Several remarks are in order. First, using our test, we reject for the whole population, at any standard level, the hypothesis that agents form rational expectations over their future earnings. Second, we also reject RE (at the 5% level) when we apply our test separately for whites (non-Hispanics) and minorities, as well as low vs. high numeracy test scores.[17] Third, the results from our test point to beliefs formation being heterogeneous across schooling (college degree vs. no college degree) and tenure (more or less than 6 months spent in current job) levels. In particular, we cannot rule out that the beliefs about future earnings of individuals with more schooling experience correspond to rational expectations with respect to some information set. Similarly, while we reject RE at any standard level for the subgroup of workers who have accumulated less than 6 months of experience in their current job, we can only marginally reject at the 10% level RE for those who have been in their current job for a longer period of time. As such, these findings complement some of the recent evidence from the economics of education and labor economics literatures that individuals have more accurate beliefs about their ability as they progress through their schooling and work careers (see, e.g., Stinebrickner

[17] Respondents' numeracy is evaluated in the SCE through five questions involving computation of sales, interests on savings, chance of winning lottery, of getting a disease and being affected by a viral infection. Respondents are then partitioned into two categories: "High numeracy" (4 or 5 correct answers), and "low numeracy" (3 or fewer correct answers).

and Stinebrickner, 2012; Arcidiacono, Aucejo, Maurel and Ransom, 2016).

Fourth, using the naive test of equality of means between earnings beliefs and realizations, one would instead generally not reject the null at any standard levels. The one exception is the subgroup of workers without a college degree, for whom the naive test yields rejection of RE at the 5% level. But, as discussed before, one cannot rule out that such a rejection is due to aggregate shocks.

Even though individuals in the overall sample form expectations over their earnings in the near future that are realistic, in the sense of not being significantly biased, the result from our preferred test shows that earnings expectations are nonetheless not rational. Taken together, these findings highlight the importance of incorporating the additional restrictions of rational expectations that are embedded in our test, using the distributions of subjective beliefs and realized outcomes to detect violations of rational expectations. That the variance test of RE never rejects the null at any standard levels indicates that it is important in practice to go beyond the first moments, and exploit instead the full distributions of beliefs and outcomes to detect departures from rational expectations.

We do not report in this table the results of the direct test of RE.[18] Beyond the obvious implication that restricting to the subsample of individuals who are followed over four months results in a loss of statistical power, there are a couple of important issues associated with the direct test. First and foremost, as already discussed in Section 2.2.4, the direct test is not robust to measurement errors on the subjective beliefs ψ. To the extent that subjective beliefs are likely measured with some error, this is an important limitation of this test. Second, attrition from the survey may be endogenous. To explore this possibility, we report in Table 2 the estimation results from a logit model of attrition on earnings beliefs, gender, race/ethnicity, college degree attainment, numeracy test score, tenure and a (linear) time trend.

Table 2: Logit model of attrition

Population	Intercept	ψ	Male	White	Coll. Degree	Low Num.	Tenure > 6	Trend
All	1.327**	-6.206e-06**	0.046	-0.311	-0.137	-0.141	-0.786**	-0.040
	(0.293)	(1.621e-06)	(0.138)	(0.222)	(0.139)	(0.162)	(0.164)	(0.033)

Notes: 1,565 observations. Significance levels: †: 10%, *: 5%, **: 1%.

The main takeaway from this table is that earnings beliefs ψ are significantly associated with attrition, even after controlling for this extensive set of characteristics. This result suggests that individuals for whom we observe both earnings expectations and realizations are likely to earn

[18] The results of the direct test are reported in Table 6 in Appendix E.2.

more than those who are not followed across the two waves. Along the same lines, a Kolmogorov-Smirnov test rejects at the 1% level the equality of the distributions of realized earnings between the whole sample and the subsample that would be used for the direct test. Similarly, we reject the equality of the distributions of expected earnings between these two samples. These results indicate that, in this context, the direct RE test is likely to be misleading. Conversely, attrition is unlikely to be an issue with our test, since we use in each wave the observations of all respondents.[19]

5.3 Deviation from RE in a life-cycle consumption model

In this section, we examine the sensitivity of a standard life-cycle incomplete markets (SIM) model of consumption to the relaxation of the assumption that individuals form rational expectations about their future earnings. In the baseline SIM model, as in the vast majority of life-cycle consumption models, the rational expectations hypothesis is maintained. However, if a substantial fraction of the individuals do not have rational expectations on their future earnings, some of the conclusions that are drawn from this model may be misleading. In the following, we address this issue by conducting a sensitivity analysis along the lines of Section 4. Specifically, we use a benchmark SIM model as a starting point, which we modify to account for the fact that individuals may not have rational expectations about their income process. Using this framework, we then illustrate how the minimal deviations from rational expectations that are consistent with the SCE data impact partial insurance mechanisms, and in particular the predicted effects of transitory and permanent income shocks on consumption.

5.3.1 Benchmark model and deviation

The SIM model we consider closely resembles that of Kaplan and Violante (2010, KV hereafter), who also focus on the responsiveness of consumption to income shocks. Time is assumed to be discrete, $t \in \{1, ..., T\}$. The economy is constituted of agents (household heads) who work for $T^{ret} - 1$ periods, before retiring. Their unconditional probability of surviving until period t is denoted by ξ_t, and we assume that $\xi_t = 1$ for all $t < T^{ret}$ (and $\xi_{T+1} = 0$). Agents are assumed not to be altruistic. At each period t, consumption, income and assets of agent i are denoted respectively by $C_{i,t}$, $Y_{i,t}$ and $A_{i,t}$, with $A_{i,1} = 0$. The assets are made of a tradable risk-free one-period bond with a rate of return r. Assuming perfect annuity markets, the budget constraint

[19] The one assumption we need to make is that respondents in the surveys used to measure ψ (i.e., those of March and July 2015) are drawn from the same population as those from the surveys used to measure Y (i.e., those of July and November 2015). That there is no significant time trend in the attrition model (Table 2) suggests that this assumption is reasonable in this context.

can be written as follows:

$$C_{i,t} + \left(\frac{\xi_t}{\xi_{t+1}}\right) A_{i,t+1} = (1+r)A_{i,t} + Y_{i,t}. \tag{5}$$

We also assume that agent i faces at each period a constraint on the level of her assets,

$$A_{i,t} \geq \underline{A}. \tag{6}$$

Agents are forward-looking, and choose at the beginning of period t, if still alive and before observing their income, their sequence of consumption. They do so by sequentially maximizing the present value of subjective expected lifetime utility given their information set, denoted by $\mathcal{I}_{i,t-1}$ for agent i, and given the constraints in (5)-(6). This present value at period t is equal to

$$\mathcal{E}\left[\sum_{t'=t}^{T} \beta^{t'-t} \frac{\xi_{t'}}{\xi_t} u\left(C_{i,t'}\right) \middle| \mathcal{I}_{i,t-1}\right], \tag{7}$$

where β denotes the discount factor and $u(.)$ is the flow utility of consumption. $\mathcal{E}\left[\cdot|\mathcal{I}_{i,t-1}\right]$ denotes the (conditional) subjective expectation operator. In order to make the problem tractable, we assume that this operator shares the same properties as the conditional expectation operator $\mathbb{E}\left[\cdot|\mathcal{I}_{i,t-1}\right]$, the key difference being that it integrates over the subjective - rather than the true - conditional distribution of the data.

During worklife ($t < T^{ret}$), the log income $\ln(Y_{i,t})$ is supposed to be the sum of a deterministic experience profile, $\kappa_{i,t}$, a permanent component, $z_{i,t-1}$, a permanent shock, $\eta_{i,t}$, and a transitory shock, $\epsilon_{i,t}$:

$$\ln\left(Y_{i,t}\right) = \kappa_{i,t} + z_{i,t} + \epsilon_{i,t},$$
$$z_{i,t} = z_{i,t-1} + \eta_{i,t}.$$

We also assume that $\epsilon_{i,t}$ and $\eta_{i,t}$ are normally distributed with mean zero and variances σ_ϵ^2 and σ_η^2 respectively. They are mutually independent and independent over time and across agents. The initial permanent shock $z_{i,0}$ is also normally distributed with mean zero and variance $\sigma_{z_0}^2$. As in KV, we assume that the information set at date t, $\mathcal{I}_t$, is composed of the permanent component $z_{i,t-1}$, as well as past transitory shocks. Finally, when $t \geq T^{ret}$, the post-retirement social security transfers $Y_{i,t}$ are computed as a piecewise constant function of the lifetime individual income, following the procedure described pp.64-65 in KV.

We adopt the following specification for the model. As in KV, we suppose that agents start working in the first period ($t = 1$), at age 25; we then set $T^{ret} = 35$ and $T = 70$ (years). We fix the interest rate r at 3% and consider two extreme cases for $\underline{A}$: a natural borrowing constraint (NBC) economy, with $\underline{A} = -10^8$, and a zero borrowing constraint (ZBC) economy, with $\underline{A} = 0$. Following, e.g., Hall and Mishkin (1982), we use a quadratic specification for the flow utility of consumption, namely $u(C) = -(C^* - C)^2/2$, with $C^* = 200,000$. Finally, as in KV and given the model in hand, the discount factor β is set to match an aggregate wealth-income target ratio of 2.5.

Finally, we consider two alternative specifications regarding the subjective expectations. First, in the benchmark model, all individuals form rational expectations on their future income. Second, we consider an alternative specification in which individual beliefs deviate from rational expectations, and replace the rational expectations on Y_{it} by the pseudo-beliefs, following the approach described in Section 4.[20] Key to the computation of the pseudo-beliefs is the availability of elicited earnings beliefs from an auxiliary dataset, in this case the SCE. Using our previous notation, the pseudo-beliefs on income are computed as a function of the rationally expected income as follows:

$$\mathcal{E}\left[Y_{i,t}|\mathcal{I}_{i,t-1}\right] = h^M\left(\mathbb{E}\left[Y_{i,t}|\mathcal{I}_{i,t-1}\right]\right), \tag{8}$$

where the function h^M is estimated using the empirical counterpart of (denoting by ψ_t the subjective beliefs at period $t < T^{ret}$):

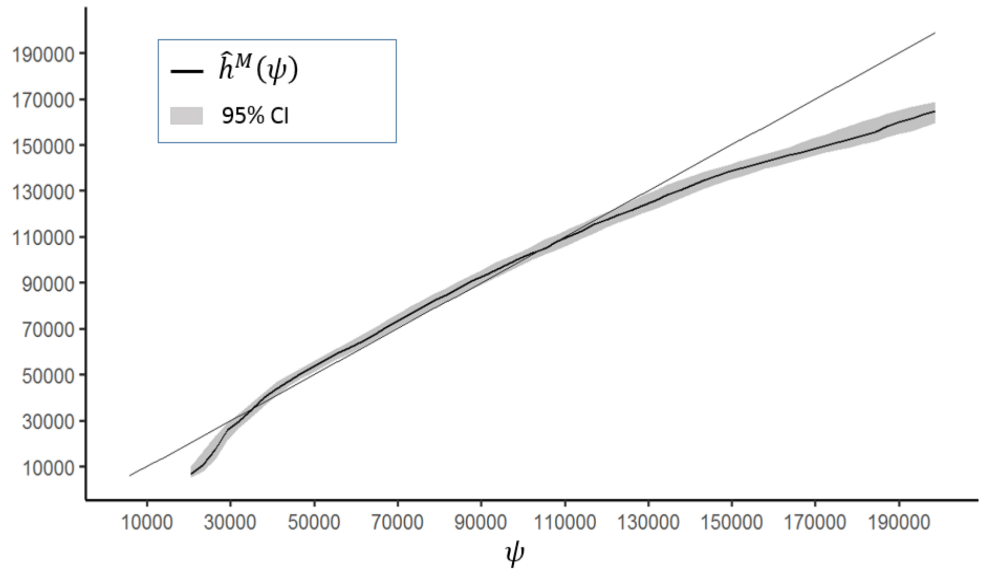
$$h^M(y) = \frac{1}{T^{ret}-1}\sum_{t=1}^{T^{ret}-1} F_{\psi_t}^{-1} \circ F_{\mathbb{E}[Y_t|\mathcal{I}_{t-1}]}(y). \tag{9}$$

We provide additional details regarding the specification of the model, in particular the income process, as well as the estimation of the pseudo-beliefs in Appendix E.3.

5.3.2 Results

A Kolmogorov-Smirnov test rejects at the 1% level the equality of the distributions of rationally expected income and the pseudo income beliefs obtained using (8), with a p-value lower than 10^{-5}. This indicates that, consistent with the earlier findings discussed in Section 6.2, RE does not hold in this context. Figure 2 displays $\widehat{h}^M$ used in equation (8) to compute the pseudo-beliefs from the rational expectations. We return to this graph below when we discuss the sources and consequences of departures from RE in the context of our model.

[20] Given the specification of the model, and in particular the quadratic specification of the utility of consumption, one can show that the optimal consumption path depends on the subjective expectations on Y_{it} only.



Notes: pointwise confidence intervals are obtained by percentile bootstrap, with 200 bootstrap samples. All results are in 2015 US dollars.

Figure 2: Estimate of h^M

Next, and following KV, we simulate the model both in the zero borrowing constraint case and in the natural borrowing constraint case, for an artificial panel of 10,000 households for 70 periods. Our main object of interest is the partial insurance coefficient, namely the share of the variance of the income shock $x_{i,t}$ (with $x \in \{\eta, \epsilon\}$) that does not translate into consumption growth:

$$\phi^x = 1 - \frac{\mathbb{C}ov(\Delta \ln(C_{i,t}), x_{i,t})}{\mathbb{V}(x_{i,t})},$$

where the covariance and variance are computed cross-sectionally over the entire population of agents between ages 25 and 60. We also consider and discuss below ϕ_t^x, which is the same quantity but computed conditionally on being of age $24 + t$.

We report the partial insurance coefficients to permanent income shocks (ϕ^η) and transitory income shocks (ϕ^ϵ) in Table 3 below. We first display the coefficients under rational expectations, and then the estimates obtained under our minimal deviations from RE. The changes in the insurance coefficients across both scenarios reflect the changes in the income expectations (RE vs. pseudo beliefs), combined with the change in the discount factor β which, in both cases, is set to match an aggregate wealth-income target ratio of 2.5. Specifically, β decreases from around 97% to 94-95%, depending on the borrowing constraints assumption (ZBC or NBC).[21]

[21] The direction of the change is consistent with prior evidence from lab and field experiments (see, e.g. Andersen, Harrison, Morten and Rutstrom, 2008; Andreoni and Sprenger, 2012; Belzil, Maurel and Sidibé, 2017), which generally points to discount factors lower than 97%.

Table 3: Insurance coefficients under RE or deviations from RE.

	Zero borrowing constraint			Natural borrowing constraint		
	ϕ^η	ϕ^ϵ	β	ϕ^η	ϕ^ϵ	β
Model with RE	0.223	0.757	0.971	0.100	0.938	0.973
Model with deviations	0.425	0.539	0.938	0.568	0.728	0.947
from RE	(0.019)	(0.019)	(0.003)	(0.068)	(0.038)	(0.003)

Notes: we use $\sigma^2_\eta = 0.02$, $\sigma^2_\epsilon = 0.05$, $\sigma^2_{z_0} = 0.15$, and an aggregate wealth-income ratio of 2.5. Standard errors in parentheses.

Turning to our main parameters of interest, we find that consumption responses to both transitory and permanent income shocks are significantly affected by the minimal deviations from RE. In particular, consumption is found to be significantly less responsive to permanent income shocks when we relax RE, with substantial increases in the partial insurance coefficient ϕ^η for both ZBC and NBC specifications. In that sense, accounting for deviations from RE with our pseudo-beliefs takes the model predictions further away from those obtained with a canonical permanent income hypothesis model (in which $\phi^\eta = 0$). Conversely, accounting for deviations from RE also results in consumption being more responsive to transitory income shocks (i.e., lower insurance coefficient ϕ^ϵ). The age profile of the insurance coefficients is also sensitive to the type of beliefs that are used to simulate the consumption paths. Figure 4 in Appendix E.3 shows that, in particular, household heads between the ages of 35 and 50 tend to smooth permanent (transitory) income shocks significantly more (less) when we allow for deviations from RE.

It is interesting to discuss the findings from Table 3 in light of previous empirical estimates that have been obtained in the consumption literature. In particular, in the presence of borrowing constraints (ZBC), the partial insurance coefficient to permanent shocks implied by the model under RE ($\phi^\eta = 0.22$) is lower than the estimated coefficient obtained by Blundell, Pistaferri and Preston (2008, BPP) ($\phi^\eta = 0.36$) using US data on income and consumption.[22] Accounting for minimal deviations from RE using our method, the insurance coefficient increases substantially, to about $\phi^\eta = 0.43$. Hence, relaxing the assumption that agents form rational expectations about their future incomes reduces the gap between the partial insurance coefficient to permanent shocks implied by the SIM model, and the empirical estimates obtained by BPP. This suggests that part of the over-insurance phenomenon to permanent income shocks that has been documented in the literature (see, e.g., Blundell et al., 2008) may in fact be attributable to

[22] While BPP provide a range of estimates for the insurance coefficient, the estimate 0.36 is obtained when labor income is defined as household earnings after tax and transfers, and, as such, is arguably the relevant benchmark here.

departures from the RE hypothesis that is typically maintained in consumption models. Note that in the natural borrowing constraint (NBC) economy, the model with deviations from RE also results in significantly larger insurance coefficients to permanent shocks than in the baseline RE model (0.57 vs. 0.10). While the estimated coefficient that accounts for deviations from RE is larger than the BPP estimate, the discrepancy remains smaller than in the benchmark RE model.

Turning to the transitory income shocks, we find that relaxing RE results in a significant decrease, of about 20 pp. for both ZBC and NBC cases, in the insurance coefficients ϕ^{ϵ}. This suggests that departures from RE also play a role in accounting for the excess sensitivity of consumption to transitory shocks that has been documented in some of the literature using standard realized data on income and consumption (see, e.g., Hall and Mishkin, 1982), and, more recently, in Kaufmann and Pistaferri (2009), using subjective expectations data from the Survey of Household Income and Wealth in Italy.[23]

In order to shed light on the underlying mechanisms, it is instructive to examine the lifetime net worth profiles that are implied by the model with RE, versus the model where we relax RE using the pseudo-beliefs. These profiles are plotted in Figure 5 in Appendix E.3. A couple of comments are in order. First, household heads between 25 and 35 are more indebted in the model with deviations from RE. This is due to the fact that their average expected income is between 45,000$ and 100,000$ and thus, from Figure 2, they tend to be over-optimistic (i.e., pseudo-beliefs are greater than RE). It follows that they tend to borrow more than in the RE model, as shown with natural borrowing constraints in Figure 5. Second, later in the life-cycle and before retirement, household heads tend to be over-pessimistic. With quadratic preferences, this translates into more savings compared with the RE environment, which results in a steeper increase of the assets before retirement in Figure 5.[24]

Taken together, the findings from this analysis show that accounting for minimal deviations from rational expectations results in significant and sizable changes in the predicted consumption responses to both transitory and permanent income shocks. As such, they highlight the importance of collecting subjective expectations data in order to analyze consumption dynamics while allowing for departures from rational expectations.

[23] Kaufmann and Pistaferri (2009) combine realizations and subjective expectations about future income to tell apart anticipated and unanticipated income changes, while maintaining the RE hypothesis.

[24] As a consequence, the model with deviations from RE fits the data substantially better than the benchmark model, comparing the worth profiles for both models to the local linear regression obtained from the 1992 Survey of Consumer Finance (SCF) data (the dotted curves in Figure 5). This is true in particular around and after retirement age. Over the life cycle, the average prediction error decreases by about 16.5% in both ZBC and NBC cases when we allow for deviations from RE.

6 Conclusion

In this paper, we develop a new test of rational expectations that can be used in a broad range of empirical settings. In particular, our test only requires having access to the marginal distributions of realizations and subjective beliefs, and, as such, can be applied in frequent cases where realizations and beliefs are observed in two separate datasets. We establish that whether one can rationalize rational expectations is equivalent to the distribution of realizations being a mean-preserving spread of the distribution of beliefs, a condition which can be tested using recent tools from the moment inequalities literature. We show that our test can easily accommodate covariates and aggregate shocks, and, importantly for practical purpose, is robust to some degree of measurement errors on the elicited beliefs.

Going beyond testing, we also introduce the concept of minimal deviations from rational expectations that can be rationalized by the data in the context of structural models, which impose constraints on the agents' information sets. Using tools from the optimal transport literature, we show that, under mild regularity conditions, these deviations exist, are unique, and are also easily estimated. These deviations offer a novel and tractable way to conduct a sensitivity analysis on the assumed form of expectations.

We apply our method to test for rational expectations about future earnings, using data from the Survey of Consumer Expectations. While individuals tend to be right on average about their future earnings, our test strongly rejects rational expectations. Using the deviations from rational expectations within a standard life-cycle incomplete markets, we then provide evidence that the behavioral responses of consumers to income shocks are sensitive to departures from rational expectations. Notably, our results suggest that part of the over-insurance to permanent income shocks that has been documented in the literature is attributable to departures from the rational expectations hypothesis.

References

Agarwal, N. and Somaini, P. (2018), 'Demand analysis using strategic reports: an application to a school choice mechanism', *Econometrica* **86**(2), 391–444.

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

Andersen, S., Harrison, G. W., Morten, L. and Rutstrom, E. E. (2008), 'Eliciting risk and time preferences', *Econometrica* **76**(3), 583–618.

Andreoni, J. and Sprenger, C. (2012), 'Estimating time preferences from convex budgets', *American Economic Review* **102**(7), 3333–56.

Andrews, D. (1994), 'Empirical process methods in econometrics', *Handbook of econometrics* **4**, 2247–2294.

Andrews, D. and Shi, X. (2013), 'Inference based on conditional moment inequalities', *Econometrica* **81**(2), 609–666.

Andrews, D. and Shi, X. (2017), 'Inference based on many conditional moment inequalities', *Journal of Econometrics* **196**(2), 275–287.

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

Arcidiacono, P., Aucejo, E., Maurel, A. and Ransom, T. (2016), College attrition and the dynamics of information revelation. NBER Working Paper No. 22325.

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

Arcidiacono, P., Hotz, J. V., Maurel, A. and Romano, T. (2014), Recovering ex ante returns and preferences for occupations using subjective expectations data. NBER Working Paper No. 20626.

Armantier, O., Topa, G., Van der Klaauw, W. and Zafar, B. (2017), 'An overview of the survey of consumer expectations', *Economic Policy Review* **23**(2), 51–72.

Armstrong, T. B. and Kolesár, M. (2018), 'Sensitivity analysis using approximate moment condition models'. Working paper.

Barseghyan, L., Molinari, F. and Teitelbaum, J. C. (2016), 'Inference under stability of risk preferences', *Quantitative Economics* **7**(2), 367–409.

Beare, B. and Shi, X. (2019), 'An improved bootstrap test of density ratio ordering', *Econometrics and Statistics* **10**, 9–26.

Belzil, C., Maurel, A. and Sidibé, M. (2017), Estimating the value of higher education financial aid: Evidence from a field experiment. NBER Working Paper No. 23641.

Blundell, R. (2017), 'What have we learned from structural models?', *American Economic Review: Papers and Proceedings* **107**(5), 287–292.

Blundell, R., Pistaferri, L. and Preston, I. (2008), 'Consumption inequality and partial insurance', *The American Economic Review* **98**(5), 1887–1921.

Bonhomme, S. and Weidner, M. (2018), 'Minimizing sensitivity to model misspecification', *arXiv preprint arXiv:1807.02161* .

Bound, J., Brown, C. and Mathiowetz, N. (2001), 'Measurement error in survey data', *Handbook of Econometrics* **5**, 3705–3843.

Buchinsky, M. and Leslie, P. (2010), 'Educational attainment and the changing u.s. wage structure: Dynamic implications on young individuals' choices', *Journal of Labor Economics* **28**(3), 541–594.

Buchinsky, M., Li, F. and Liao, Z. (2018), 'Estimation and inference of semiparametric models using data from several sources'. Working paper.

Chen, X., Linton, O. and Van Keilegom, I. (2003), 'Estimation of semiparametric models when the criterion function is not smooth', *Econometrica* **71**(5), 1591–1608.

Christensen, T. and Connault, B. (2019), Counterfactual sensitivity and robustness. Working paper.

Conlon, J. J., Philossoph, L., Wiswall, M. and Zafar, B. (2018), Labor market search with imperfect information and learning. NBER Working Paper No. 24988.

Cross, P. J. and Manski, C. F. (2002), 'Regressions, short and long', *Econometrica* **70**(1), 357–368.

Cunha, F. and Heckman, J. J. (2007), 'Identifying and estimating the distributions of ex post and ex ante returns to schooling', *Labour Economics* **14**(6), 870–93.

Davydov, Y. A., Lifshits, M. A. and Smorodina, N. V. (1998), *Local properties of distributions of stochastic functionals*, American Mathematical Society.

de Paula, A., Shapira, G. and Todd, P. E. (2014), 'How beliefs about hiv status affect risky behaviors: Evidence from malawi', *Journal of Applied Econometrics* **29**(6), 944–964.

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

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

D'Haultfœuille, X., Gaillac, C. and Maurel, A. (2018), Rationalexp: Tests of and deviations from rational expectations. Working paper.

Fan, Y., Sherman, R. and Shum, M. (2014), 'Identifying treatment effects under data combination', *Econometrica* **82**(2), 811–822.

Fuster, A., Kaplan, G. and Zafar, B. (2018), What would you do with $500? Spending responses to gains, losses, news and loans. NBER Working Paper No. 24386.

Galichon, A. (2016), *Optimal transport methods in economics*, Princeton University Press.

Gourieroux, C. and Pradel, J. (1986), 'Direct test of the rational expectation hypothesis', *European Economic Review* **30**(2), 265–284.

Gozlan, N., Roberto, C., Samson, P.-M., Shu, Y. and Tetali, P. (2018), 'Characterization of a class of weak transport-entropy inequalities on the line', *Annales de l'IHP* **54**(3), 1667–1693.

Gutknecht, D., Hoderlein, S. and Peters, M. (2018), 'Constrained information processing and individual income expectations'. Working paper.

Hall, R. E. and Mishkin, F. S. (1982), 'The sensitivity of consumption to transitory income: Estimates from panel data on households', *Econometrica* **50**(2), 461–81.

Hansen, L. P. (2014), 'Nobel lecture: Uncertainty outside and inside economic models', *Journal of Political Economy* **122**(5), 945–987.

Hansen, L. P. and Jagannathan, R. (1997), 'Assessing specification errors in stochastic discount factor models', *Journal of Finance* **52**(2), 557–590.

Hoffman, M. and Burks, S. V. (2020), 'Worker overconfidence: Field evidence and implications for employee turnover and returns from training', *Quantitative Economics* **11**, 315–348.

Hsu, Y.-C. (2016), 'Consistent tests for conditional treatment effects', *The Econometrics Journal* **20**(1), 1–22.

Ivaldi, M. (1992), 'Survey evidence on the rationality of expectations', *Journal of Applied Econometrics* **7**(3), 225–241.

Kaplan, G. and Violante, G. L. (2010), 'How much consumption insurance beyond self-insurance?', *American Economic Journal: Macroeconomics* **2**(4), 53–87.

Kapor, A., Neilson, C. A. and Zimmerman, S. (2018), 'Heterogeneous beliefs and school choice mechanisms', *American Economic Review* **Forthcoming**.

Kaufmann, K. and Pistaferri, L. (2009), 'Disentangling insurance and information in intertemporal consumption choices', *American Economic Review* **99**(2), 387–92.

Keane, M. and Wolpin, K. (1997), 'The career decisions of young men', *Journal of Political Economy* **105**(3), 473–522.

Kuchler, T. and Zafar, B. (2019), 'Personal experiences and expectations about aggregate outcomes', *Journal of Finance* **Forthcoming**.

Linton, O., Song, K. and Whang, Y.-J. (2010), 'An improved bootstrap test of stochastic dominance', *Journal of Econometrics* **154**, 186–202.

Lovell, M. C. (1986), 'Tests of the rational expectations hypothesis', *American Economic Review* **76**(1), 110–124.

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

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

Masten, M. A. and Poirier, A. (2019), Salvaging falsified instrumental variable models. Working paper.

Molinari, F. and Peski, M. (2006), 'Generalization of a result on "regressions, short and long"', *Econometric Theory* **22**(1), 159–163.

Muth, J. F. (1961), 'Rational expectations and the theory of price movements', *Econometrica* **29**(3), 315–335.

Patton, A. J. and Timmermann, A. (2012), 'Forecast rationality tests based on multi-horizon bounds', *Journal of Business & Economic Statistics* **30**(1), 1–17.

Pesaran, M. H. (1987), *The limits to rational expectations*, Basil Blackwell.

Pollard, D. (1990), Empirical processes: theory and applications, *in* 'NSF-CBMS regional conference series in probability and statistics', Institute of Mathematical Statistics and the American Statistical Association, pp. i–86.

Rachev, S. T. and Rüschendorf, L. (1998), *Mass Transportation Problems: Volume I: Theory*, Springer Science & Business Media.

Ridder, G. and Moffitt, R. (2007), 'The econometrics of data combination', *Handbook of Econometrics* **6**, 5469–5547.

Rothschild, M. and Stiglitz, J. (1970), 'Increasing risk: I. a definition', *Journal of Economic Theory* **2**(3), 225–243.

Rust, J. (1987), 'Optimal replacement of gmc bus engines: An empirical model of harold zurcher', *Econometrica* **55**(5), 999–1033.

Santambrogio, F. (2015), 'Optimal transport for applied mathematicians', *Birkäuser* .

Stinebrickner, R. and Stinebrickner, T. (2012), 'Learning about academic ability and the college dropout decision', *Journal of Labor Economics* **30**(4), 707–748.

Stinebrickner, R. and Stinebrickner, T. (2014*a*), 'Academic performance and college dropout: Using longitudinal expectations data to estimate a learning model', *Journal of Labor Economics* **32**(3), 601–644.

Stinebrickner, R. and Stinebrickner, T. (2014*b*), 'A major in science? initial beliefs and final outcomes for college major and dropout', *The Review of Economic Studies* **81**(1), 426–472.

Strassen, V. (1965), 'The existence of probability measures with given marginals', *The Annals of Mathematical Statistics* **36**(2), 423–439.

Van der Klaauw, W. (2012), 'On the use of expectations data in estimating structural dynamic choice models', *Journal of Labor Economics* **30**(3), 521–554.

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

Van der Vaart, A. (2000), *Asymptotic statistics*, Cambridge University Press.

Van der Vaart, A. and Wellner, J. (1996), Weak convergence, *in* 'Weak Convergence and Empirical Processes', Springer, pp. 16–28.

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

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

Zafar, B. (2011*a*), 'Can subjective expectations data be used in choice models? evidence on cognitive biases', *Journal of Applied Econometrics* **26**(3), 520–544.

Zafar, B. (2011*b*), 'How do college students form expectations?', *Journal of Labor Economics* **29**(2), 301–348.

A Statistical tests in the presence of aggregate shocks

In this appendix, we show how to adapt the construction of the test statistic and obtain similar results as in Theorem 2 in the presence of aggregate shocks. As explained in Section 2.2.3, we mostly have to replace $\widetilde{Y}$ by $\widetilde{Y}_c = Dq\left(\widetilde{Y}, c\right) + (1-D)\psi$. Because we include covariates here, as in Section 2.3, c is actually a function of X. Also, the true function c_0 has to be estimated. We let $\widehat{c}$ denote such a nonparametric estimator, which is based on $\mathbb{E}[q(Y, c_0(X))|X] = \mathbb{E}[\psi|X]$. When $q(y,c) = y - c$ or $q(y,c) = y/c$, we get respectively $c_0(X) = \mathbb{E}(Y|X) - \mathbb{E}(\psi|X)$ and $c_0(X) = \mathbb{E}(Y|X)/\mathbb{E}(\psi|X)$, and $\widehat{c}$ is easy to compute using nonparametric estimators of $\mathbb{E}(Y|X)$ and $\mathbb{E}(\psi|X)$.

Because in Proposition 3 (ii) we do not test for a moment equality anymore, $m\left(D_i, \widetilde{Y}_i, X_i, g, y\right)$ reduces to $m_1\left(D_i, \widetilde{Y}_{c,i}, X_i, g, y\right)$. We let hereafter $\overline{m}_n(g,y) = \sum_{i=1}^n m_1\left(D_i, \widetilde{Y}_{c,i}, X_i, g, y\right)/n$. In the test statistic T, we replace, for $(y,g) \in \mathcal{Y} \times \cup_{r\geq 1}\mathcal{G}_r$, $\overline{\Sigma}_n(g,y)$ by $\overline{\Sigma}_n(g,y) = \widehat{\Sigma}_n(g,y) + \epsilon \text{Diag}\left(\widehat{\mathbb{V}}\left(\widetilde{Y}_{\hat{c}}\right), \widehat{\mathbb{V}}\left(\widetilde{Y}_{\hat{c}}\right)\right)$, where $\widehat{\Sigma}_n(g,y)$ and $\widehat{\mathbb{V}}\left(\widetilde{Y}_{\hat{c}}\right)$ are respectively the sample covariance matrix of $\sqrt{n}\overline{m}_n(g,y)$ and the empirical variance of $\widetilde{Y}_{\hat{c}}$. The last difference with the test considered in Section 2.3 is that when using the bootstrap to compute the critical value, we also have to re-estimate c_0 in the bootstrap sample.

We obtain in this context a result similar to Theorem 2 above, under the regularity conditions stated in Assumption 4. We let hereafter $\mathcal{C}_s\left([0,1]^{d_X}\right)$ denote the space of continuously differentiable functions of order s on $[0,1]^{d_X}$ that have a finite norm $\|c\|_{s,\infty} = \max_{|\boldsymbol{k}|\leq s} \sup_{x\in[0,1]^{d_X}} \left|c^{(\boldsymbol{k})}(x)\right|$. We also let, for any function f on a set $\mathcal{G}$, $\|f\|_{\mathcal{G}} = \sup_{x\in\mathcal{G}} |f(x)|$. Finally, when the distribution of $\left(D, \widetilde{Y}, X\right)$ is F, K_F denotes the asymptotic covariance kernel of $n^{-1/2}\text{Diag}\left(\mathbb{V}\left(\widetilde{Y}_{c_0}\right)\right)^{-1/2}\overline{m}$.

Assumption 4 *(i)* $\widehat{c}$ *and* c_0 *belong to* $\mathcal{C}_s\left([0,1]^{d_X}\right)$*, with* $s \geq d_X$*. Moreover,* $\|\widehat{c} - c_0\|_{[0,1]^{d_X}} = o_P(1)$.

(ii) For all $y \in \mathcal{Y}$*,* q *is Lipschitz on* $\mathcal{Y} \times [-C, C]$ *for some* $C > \|c_0\|_{[0,1]^{d_X}}$*. Moreover,* $\sup_{(y,c)\in\mathcal{Y}\times[-C,C]} |q(y,c)| \leq M_0$*;*

(iii) For all $c \in \mathbb{R}$*, the function* $q(\cdot, c) : \mathcal{Y} \to \mathcal{Y}$ *is bijective and its inverse* $q^I(\cdot, c)$ *is Lipschitz on* $\mathcal{Y}$*;*

(iv) $F_{\psi|X}(\cdot|x)$*,* $F_{Y|X}(\cdot|x)$ *are Lipschitz on* $\mathcal{Y}$ *uniformly in* $x \in [0,1]^{d_X}$ *with constants* $Q_{F,1}$ *satisfying* $\sup_{F\in\mathcal{F}_0} Q_{F,1} \leq \overline{Q}_1 < +\infty$*. Also,* $F_{q(\psi,c(X))}$*,* $F_{q(Y,c(X))}$ *are Lipschitz on* $[-M_0, M_0]$ *with constants* $Q_{F,2}$ *satisfying* $\sup_{F\in\mathcal{F}_0} Q_{F,2} \leq \overline{Q}_2 < +\infty$*;*

(iv) $\inf_{F\in\mathcal{F}} \mathbb{V}_F\left[\widetilde{Y}_c^2\right] > 0$ *and* $\epsilon_0 \leq \inf_{F\in\mathcal{F}} \mathbb{E}_F[D] \leq \sup_{F\in\mathcal{F}} \mathbb{E}_F[D] \leq 1 - \epsilon_0$ *for some* $\varepsilon_0 \in (0, 1/2)$*. Also,* $\widehat{\mathbb{V}}_F\left[\widetilde{Y}_{\widehat{c}}^2\right]$ *is a consistent estimator of* $\mathbb{V}_F\left[\widetilde{Y}_c^2\right]$.

Part (i) imposes some regularity conditions on c_0 and its nonparametric estimator $\widehat{c}$. It is possible to check such regularity conditions on $\widehat{c}$ with kernel or series estimators of $\mathbb{E}(Y|X)$ and $\mathbb{E}(\psi|X)$. Parts (ii) and (iii) also hold when $q(y,c) = y - c$ and $q(y,c) = q(y)/c$, by imposing in the second case that c belongs to a compact subset of $(0,\infty)$. Proposition 5 shows that under these conditions, the test has asymptotically correct size.

Proposition 5 *Suppose that $r_n \to \infty$ and that Assumptions 3 and 4 hold. Then (i) in Proposition 2 holds, replacing $\varphi_{n,\alpha}$ by $\varphi_{n,\alpha,\widehat{c}}$.*

Results like (ii) and (iii) in Proposition 2 could also be obtained under the conditions of Proposition 5, modifying directly the proof of Proposition 2.

B Tests with rounding practices

We have considered in Section 2.2.4 the possibility of measurement errors on ψ. Another source of uncertainty on ψ is rounding. Rounding practices by interviewees are common. A way to interpret these practices is that in situations of ambiguity, individuals may only be able to bound the distribution of their future outcome Y (Manski, 2004). If individuals round at 5% levels, for instance, an answer $\psi = 0.05$ for the beliefs about percent increase of income should then only be interpreted as $\psi \in [0.025, 0.075]$. Another case where only bounds on ψ are observed is when questions to elicit subjective expectations take the following form: "What do you think is the percent chance that your own [Y] will be below [y]?", for a certain grid of y. If 0 and 100 are always observed, or if we assume that the support of subjective distributions is included in $[\underline{y}, \overline{y}]$, we can still compute bounds on ψ.[25] In such cases, we only observe (ψ_L, ψ_U), with $\psi_L \leq \psi \leq \psi_U$. For a thorough discussion of this issue, and especially of how to infer rounding practices, see Manski and Molinari (2010).

In this setting, rationalizing rational expectations is less stringent than in our baseline set-up since the constraints on the distribution of ψ are weaker. Formally, the null hypothesis takes the following form:

$$\mathrm{H}_{0B}: \ \exists (Y', \psi', \mathcal{I}') : \ \sigma(\psi') \subset \mathcal{I}', \ Y' \sim Y, \ F_{\psi_U} \leq F_{\psi'} \leq F_{\psi_L} \text{ and } \mathbb{E}(Y'|\mathcal{I}') = \psi'.$$

To obtain an equivalent formulation to H_{0B}, a natural idea would be to fix a candidate cdf $F \in [F_{\psi_U}, F_{\psi_L}]$ for F_ψ and apply Theorem 1 with this F. Then, letting $\Delta_F(y) = \int_{-\infty}^{y} F_Y(t) - F(t)dt$ and $\delta_F = \mathbb{E}(Y) - \int u dF(u)$, H_{0B} would hold as long as for some $F \in [F_{\psi_U}, F_{\psi_L}]$, $\Delta_F(y) \geq 0$ for all $y \in \mathbb{R}$ and $\delta_F = 0$. In practice though, directly checking whether such a distribution exists would be very difficult. Fortunately, we show in the following proposition that it is in fact

[25] Note however that in this case, our approach does not take into account all the information on the subjective distribution.

sufficient to check that these conditions hold for a specific candidate distribution. To define the cdf of this distribution, we introduce, for all $b \in \mathbb{R}$, the random variables

$$\psi^b = \psi_U \mathbb{1}\{\psi_U < b\} + \max(b, \psi_L)\mathbb{1}\{\psi_U \geq b\}.$$

We also let $\psi^{-\infty} = \psi_L$ and $\psi^{+\infty} = \psi_U$. The cdf of ψ^b is then $F^b(t) = F_{\psi_U}(t)\mathbb{1}\{t < b\} + F_{\psi_L}(t)\mathbb{1}\{t \geq b\}$, for all $b \in \overline{\mathbb{R}}$. We let $\mathcal{F}_B = \{F^b, b \in \overline{\mathbb{R}}\}$ denote the set of all such cdfs.

Assumption 5 $\mathbb{E}(|Y|) < +\infty$, $\mathbb{E}(|\psi_L|) < +\infty$ *and* $\mathbb{E}(|\psi_U|) < +\infty$.

Proposition 6 *Suppose that Assumption 5 holds. First, if $\mathbb{E}[\psi_L] \leq \mathbb{E}[Y] \leq \mathbb{E}[\psi_U]$, there exists a unique $F^* \in \mathcal{F}_B$ such that $\delta_{F^*} = 0$. Second, the following statements are equivalent:*

(i) H_{0B} holds.

(ii) $\mathbb{E}[\psi_L] \leq \mathbb{E}[Y] \leq \mathbb{E}[\psi_U]$ and $\Delta_{F^*}(y) \geq 0$ for all $y \in \mathbb{R}$.

This test shares some similarities with the test in the presence of aggregate shocks. Specifically, if $\mathbb{E}[\psi_L] \leq \mathbb{E}[Y] \leq \mathbb{E}[\psi_U]$, we first identify $b_0 \in \overline{\mathbb{R}}$ such that the candidate belief ψ^{b_0}, which plays a similar role as the modified outcome $q(Y, c_0)$ in the test with aggregate shocks, satisfies the equality constraint $\mathbb{E}[\psi^{b_0}] = \mathbb{E}[Y]$. Noting that the inequality $\Delta_{F^*}(y) \geq 0$ can be rewritten as $\mathbb{E}\left[(y - Y)^+ - \left(y - \psi^{b_0}\right)^+\right] \geq 0$, it follows from (ii) that rationalizing RE in this context (i.e., H_{0B}) is then equivalent to a set of many moment inequality constraints involving the distributions of realizations Y and candidate belief ψ^{b_0}.

C Tests with sample selection in the datasets

We consider here cases where the two samples are not representative of the same population, or formally, D is not independent of (Y, ψ). This may arise for instance because of oversampling of some subpopulations or differences in nonresponse between the two surveys that are used. We assume instead that selection is conditionally exogenous, that is to say:

$$D \perp\!\!\!\perp (Y, \psi)|X. \tag{10}$$

We show how to use a propensity score weighting to handle such a selection. Denote by $p(x) = P(D = 1|X = x) = \mathbb{E}[D|X = x]$ the propensity score and by

$$W(X) = \frac{D}{p(X)} - \frac{1 - D}{1 - p(X)}.$$

The law of iterated expectations combined with Proposition 2 directly yields the following proposition:

Proposition 7 *Suppose that* (10) *and Assumption 1 hold. Then* H_{0X} *is equivalent to*

$$\mathbb{E}\left[W(X)\left(y-\widetilde{Y}\right)^{+}\middle|X\right]\geq 0$$

for all $y \in \mathbb{R}$ *and* $\mathbb{E}\left[W(X)\widetilde{Y}\middle|X\right]=0$.

This proposition shows that under sample selection, we can build a statistical test of H_{0X} akin to that developed in Section 2.3, by merely estimating nonparametrically $p(X)$. We could consider for that purpose a series logit estimator, for instance. Validity of such a test would follow using very similar arguments as for the test with aggregate shocks considered above.

D Simulations with covariates

We consider here simulations including covariates. The DGP is similar to that considered in Section 3. Specifically, we assume that

$$Y = \rho\psi + \sqrt{X}\varepsilon,$$

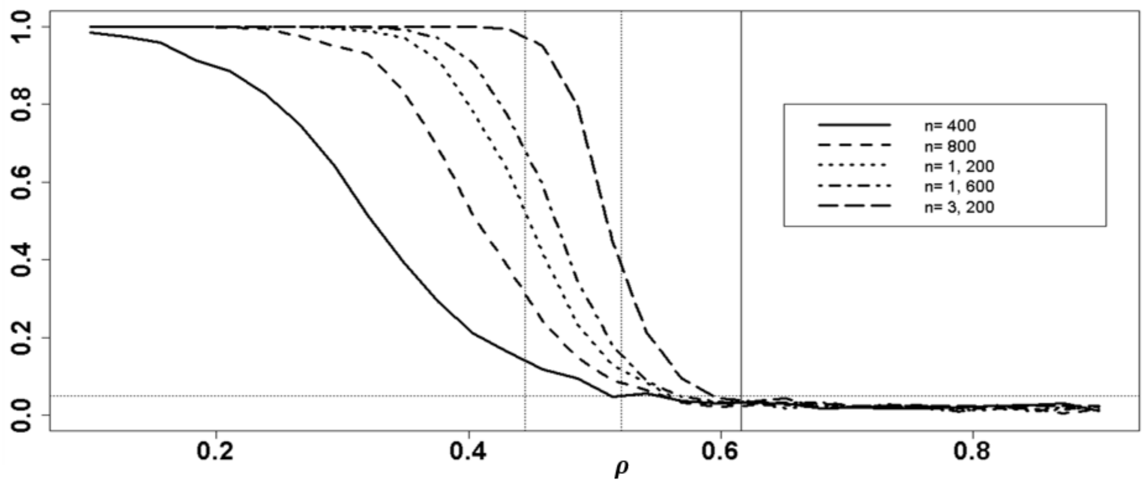
with $\rho \in [0,1]$, $\psi \sim \mathcal{N}(0,1)$, $X \sim \text{Beta}(0.1, 10)$ and

$$\varepsilon = \zeta\left(-\mathbb{1}\{U \leq 0.1\} + \mathbb{1}\{U \geq 0.9\}\right),$$

where $\zeta \sim \mathcal{N}(2, 0.1)$ and $U \sim \mathcal{U}[0,1]$. (ψ, ζ, U, X) are supposed to be mutually independent.

Like in the test without covariates, we can show that the test with covariates is able to reject RE if and only if $\rho < 0.616$. On the other hand, by construction $\mathbb{E}[Y|X] = \mathbb{E}[\psi|X]$, so the naive conditional test has no power. The test based on conditional variances rejects only if $\rho < 0.445$. Finally, we can show that without using X, our test has power only for $\rho < 0.52$. Hence, relying on covariates allows us to gain power for $\rho \in [0.521, 0.616)$.

Again, we consider hereafter $n_\psi = n_Y = n \in \{400; 800; 1,200; 1,600; 3,200\}$, use 500 bootstrap simulations to compute the critical value, and rely on 800 Monte-Carlo replications for each value of ρ and n. We use the same parameters $p = 0.05$ and $b_0 = 0.3$ as above. Figure 3 shows that the RE test with covariates asymptotically outperforms the RE test without covariates. The test exhibits a similar behavior as that without covariates, though, as we could expect, the power converges less quickly to one as n tends to infinity.



Notes: the dotted vertical lines correspond to the theoretical limit for the rejection of the null hypothesis for test based on variance ($\rho \simeq 0.445$), our test without covariates ($\rho \simeq 0.521$) and our tests with covariates ($\rho = 0.616$). The dotted horizontal line corresponds to the 5% level.

Figure 3: Power curves for the test with covariates.

E Additional material on the application

E.1 Descriptive statistics and minimal deviations for non-college graduates

Table 4: Descriptive statistics of the SCE sample

	Mean	Std. dev.
Male	0.53	0.50
White	0.74	0.43
College degree	0.49	0.46
Low numeracy	0.33	0.47
Tenure $\leq$ 6 months	0.17	0.38
Age	45.8	13.0
ψ (Earnings beliefs)	\$50,592	\$40,889
Y (Realized earnings)	\$52,354	\$38,634

E.2 Additional results on the tests of RE

Table 5: Full test of RE with different levels of Winsorization

Winsorization level Full RE	0.95 (p-val)	0.97 (p-val)	0.99 (p-val)
All	$< 0.001^{**}$	0.001^{**}	0.002^{**}
Women	$< 0.001^{**}$	$< 0.001^{**}$	0.001^{**}
Men	0.210	0.254	0.342
White	0.021^{*}	0.030^{*}	0.049^{*}
Minorities	0.006^{**}	0.007^{**}	0.018^{*}
College degree	0.130	0.146	0.196
No college degree	0.013^{*}	0.012^{*}	0.009^{**}
High numeracy	0.012^{*}	0.017^{*}	0.034^{*}
Low numeracy	0.022^{*}	0.026^{*}	0.029^{*}
Tenure $\leq$ 6 months	0.001^{**}	0.005^{**}	0.009^{**}
Tenure $>$ 6 months	$0.091^{\dagger}$	0.118	0.304

Notes: significance levels: †: 10%, *: 5%, **: 1%. "Full RE" denotes the test without covariates, where we test H_{0S} with $q(y,c) = y/c$. We use 5,000 bootstrap simulations to compute the critical values of the Full RE test. Distributions of realized earnings (Y) and earnings beliefs (ψ) are both Winsorized either at the 0.95, 0.97, or 0.99 quantile.

Table 6: Test of RE on annual earnings

	ρ	Direct test (p-val)	Full RE (p-val)	Number of obs. ψ	Y	(ψ, Y)
All	0.954	0.001**	< 0.001**	1,565	1,428	768
Women	0.956	0.002**	< 0.001**	730	649	356
Men	0.960	0.021*	0.210	835	779	412
White	0.963	0.004**	0.021*	1,200	1,097	596
Minorities	0.928	0.010*	0.006**	365	331	172
College degree	0.974	0.060†	0.130	1,106	1,053	560
No college degree	0.954	0.044*	0.013*	459	375	208
High numeracy	0.959	0.001**	0.012*	1,158	1,070	573
Low numeracy	0.954	0.094†	0.022*	407	358	195
Tenure $\leq$ 6 months	0.942	0.015*	0.001**	271	180	98
Tenure > 6 months	0.956	0.001**	0.091†	1,294	1,248	670

Notes: significance levels: †: 10%, *: 5%, **: 1%. "Direct test" denotes the direct test of RE when (ψ, Y) is observed. ρ is the coefficient of the regression of Y on ψ in that case. "Full RE" denotes the test without covariates, where we test H_{0S} with $q(y, c) = y/c$. We use 5,000 bootstrap simulations to compute the critical values of the Full RE test. Distributions of realized earnings (Y) and earnings beliefs (ψ) are both Winsorized at the 95% quantile.

E.3 Additional details and results on the life-cycle consumption model

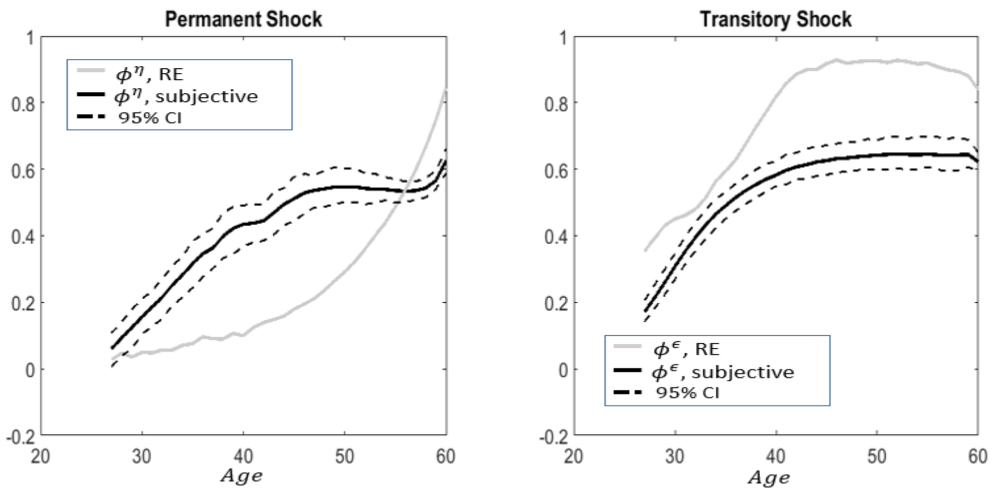
The income process is specified as follows. First, we estimate the deterministic trend $\kappa_{i,t}$ as a smooth function of age (second-order polynomial) using the dataset of Blundell et al. (2008) built from the PSID.[26] We use the same value as in the baseline specification of KV for σ_ϵ^2 and $\sigma_{z_0}^2$, namely $\sigma_\epsilon^2 = 0.05$ and $\sigma_{z_0}^2 = 0.15$. We choose $\sigma_\eta^2 = 0.02$ as it is in the range of Blundell et al. (2008) and appears to fit the 1989 and 1992 Survey of Consumer Finances data better than the baseline value used by KV. Our main results remain qualitatively unchanged when we use the same values as in Blundell et al. (2008) for both variances σ_ϵ^2 (0.037) and σ_η^2 (0.019).

To estimate h^M, we rely on (9). Given the specification of our model, $\mathbb{E}\left[Y_{i,t}|\mathcal{I}_{i,t-1}\right]$, when $t < T^{ret}$, is lognormally distributed with parameters $\kappa_{i,t} + (\sigma_\eta^2 + \sigma_\epsilon^2)/2$ and $\sigma_{z_0}^2 + (t-1)\sigma_\eta^2$. To estimate $F_{\psi_t}^{-1}$, we use the subjective beliefs of individuals between 25 and 60 measured in the SCE survey. Since in KV, $Y_{i,t}$ is interpreted as household income after taxes and transfers, whereas we only observe subjective expectations on individual labor earnings, we use an equipercentile mapping based on the two distributions of realized (expected) individual labor earnings and realized (expected) household income. We estimate this equipercentile mapping using the dataset from Blundell et al. (2008), built from the PSID, where both realized individual labor earnings and household income are observed from 1989 to 1992. Finally, we assume that the quantile of

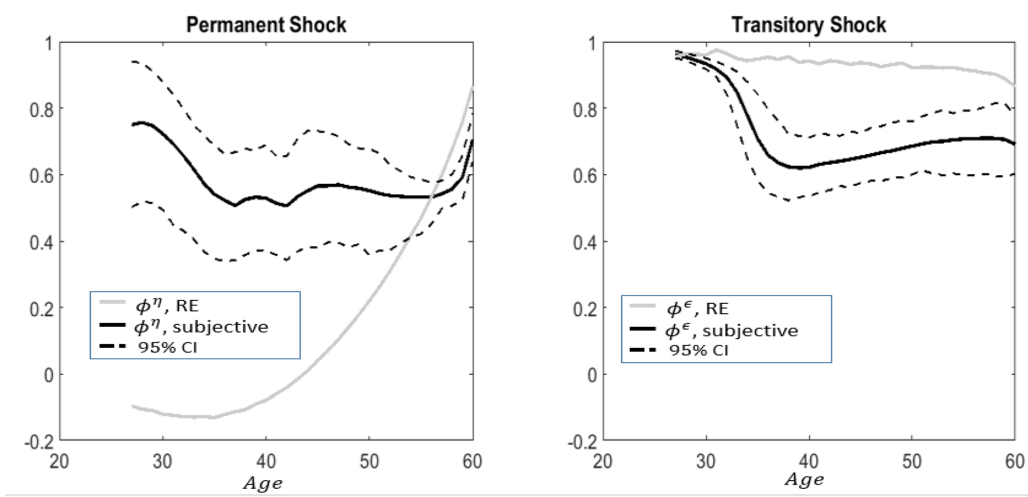
[26] Results are robust to the use of a more flexible fourth-order polynomial for $\kappa_{i,t}$.

income expectations is linear in age, and thus estimate $F_{\psi_t}^{-1}$ by a quantile regression of subjective expectations on age. We finally estimate h^M using (9), replacing $F_{\psi_t}^{-1}$ by the quantile regression estimator.

Figure 4: Age profiles of insurance coefficients.



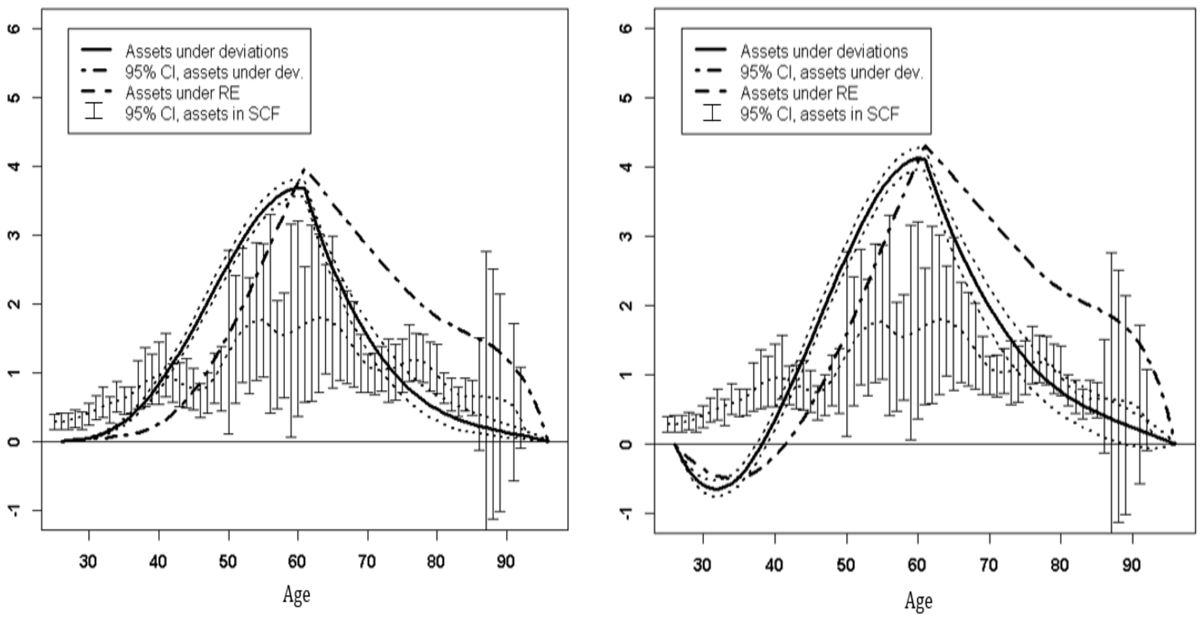
(a) With borrowing constraints



(b) Without borrowing constraints

Notes: the curves in gray (resp. in black) correspond to insurance coefficients under RE (resp. minimal deviations from RE). The dotted black curves are the 2.5 and 97.5 quantiles of bootstrap simulations, taking into account the randomness of $\widehat{h}^M$. They are obtained using 200 bootstrap samples.

Figure 5: Average lifetime net worth (in \$00,000) profiles.



(a) Zero borrowing constraint

(b) Natural borrowing constraint

Notes: SCF stands for the 1992 Survey of Consumer Finance. The dotted black curve is the estimated non-parametric regression function of net worth of households in this dataset on age using local polynomials and a bandwidth selected via cross-validation. The confidence intervals on $\widehat{h}^M$ are obtained with 200 bootstrap samples.

F Proofs

F.1 Notation and preliminaries

For any set $\mathcal{G}$, let us denote by $l^{\infty}(\mathcal{G})$ the collection of all uniformly bounded real functions on $\mathcal{G}$ equipped with the supremum norm $\|f\|_{\mathcal{G}} = \sup_{x\in\mathcal{G}} |f(x)|$. Denote by $L^2(F)$ the square integrable space with respect to the measure associated with F, and let $\|\cdot\|_{F,2}$ be the corresponding norm. We let $N(\epsilon, \mathcal{T}, L_2(F))$ denote the minimal number of ϵ-balls with respect to $\|\cdot\|_{F,2}$ needed to cover $\mathcal{T}$. An ϵ-bracket (with respect to F) is a pair of real functions (l, u) such that $l \leq u$ and $\|u - l\|_{F,2} \leq \epsilon$. Then, for any set of real functions $\mathcal{M}$, we let $N_{[]}(\epsilon, \mathcal{M}, L_2(F))$ denote the minimum number of ϵ-brackets needed to cover $\mathcal{M}$. We denote by $\mathcal{G} = (\cup_{r\geq 1}\mathcal{G}_r)$. For $x \in \mathbb{R}^d$, $d > 1$, we denote by $\|x\|_{\infty} = \max_{j=1,\ldots,d} |x|$.

For a sequence of random variable $(U_n)_{n\in\mathbb{N}}$ and a set $\mathcal{F}_0$, we say that $U_n = O_P(1)$ uniformly in $F \in \mathcal{F}_0$ if for any $\epsilon > 0$ there exist $M > 0$ and $n_0 > 0$ such that $\sup_{F\in\mathcal{F}_0} \mathbb{P}_F(|U_n| > M) < \epsilon$ for all $n > n_0$. Similarly we say that $U_n = o_P(1)$ uniformly in $F \in \mathcal{F}_0$ if for any $\epsilon > 0$, $\sup_{F\in\mathcal{F}_0} \mathbb{P}_F(|U_n| > \epsilon) \to 0$.

Finally, we add stars to random variables whenever we consider their bootstrap versions, as with T^* versus T. We define o_{P^*} and O_{P^*} as above, but conditional on $\left(\widetilde{Y}_i, D_i, X_i\right)_{i=1\dots n}$. Convergence in distribution conditional on $\left(\widetilde{Y}_i, D_i, X_i\right)_{i=1\dots n}$ is denoted by $\to_{d^*}$.

F.2 Proof of Lemma 1

Under H_0, there exist Y', ψ' and $\mathcal{I}'$ such that $Y' \sim Y$, $\psi' \sim \psi$, $\sigma(\psi') \subset \mathcal{I}'$ and $\mathbb{E}(Y'|\mathcal{I}') = \psi'$. Then, by the law of iterated expectations,

$$\mathbb{E}[Y'|\psi'] = \mathbb{E}\left[\mathbb{E}\left[Y'\middle|\mathcal{I}'\right]\middle|\psi'\right] = \mathbb{E}\left[\psi'|\psi'\right] = \psi'.$$

Conversely, if there exists (Y', ψ') such that $Y' \sim Y$, $\psi' \sim \psi$ and $\mathbb{E}[Y'|\psi'] = \psi'$, let $\mathcal{I}' = \sigma(\psi')$. Then $\psi' = \mathbb{E}[Y'|\psi'] = \mathbb{E}[Y'|\mathcal{I}']$ and H_0 holds.

F.3 Proof of Theorem 1.

(i) $\Leftrightarrow$ (iii). By Strassen's theorem (Strassen, 1965, Theorem 8), the existence of (Y, ψ) with margins equal to F_Y and F_ψ and such that $\mathbb{E}[Y|\psi] = \psi$ is equivalent to $\int f dF_\psi \leq \int f dF_Y$ for every convex function f. By, e.g., Proposition 2.3 in Gozlan et al. (2018), this is, in turn, equivalent to (iii).

(ii) $\Leftrightarrow$ (iii). By Fubini-Tonelli's theorem, $\int_{-\infty}^{y} F_Y(t)dt = \mathbb{E}\left[\int_{-\infty}^{y} \mathbb{1}\{t \geq Y\}dt\right] = \mathbb{E}\left[(y - Y)^+\right]$. The same holds for ψ. Hence, $\Delta(y) \geq 0$ for all $y \in \mathbb{R}$ is equivalent to $\mathbb{E}\left[(y - Y)^+\right] \geq \mathbb{E}\left[(y - \psi)^+\right]$ for all $y \in \mathbb{R}$. The result follows.

F.4 Proof of Proposition 1.

First, by Jensen's inequality, we obtain

$$\mathbb{E}[(y_0 - Y)^+|\psi] \geq (y_0 - \mathbb{E}(Y|\psi))^+ = (y_0 - \psi)^+.$$

Moreover, $\Delta(y_0) = 0$ implies that $\mathbb{E}((y_0 - Y)^+) = \mathbb{E}((y_0 - \psi)^+)$. Hence, almost surely, we have

$$\mathbb{E}[(y_0 - Y)^+|\psi] = (y_0 - \psi)^+.$$

Equality in the Jensen's inequality implies that the function is affine on the support of the random variable. Therefore, for almost all u, we either have $\mathcal{S}(Y|\psi = u) \subset [y_0, +\infty)$ or $\mathcal{S}(Y|\psi = u) \subset (-\infty, y_0]$. Because $\mathbb{E}[Y|\psi] = \psi$, $\mathcal{S}(Y|\psi = u) \subset [y_0, +\infty)$ for almost all $u > y_0$ and $\mathcal{S}(Y|\psi = u) \subset (-\infty, y_0]$ for almost all $u < y_0$. Then, for all $\tau \in (0, 1)$, $F^{-1}_{Y|\psi}(\tau|u) \geq y_0$ for almost all $u \geq y_0$ and $F^{-1}_{Y|\psi}(\tau|u) \leq y_0$ for almost all $u \leq y_0$. Thus, for all $\tau \in (0, 1)$, by continuity of $F^{-1}_{Y|\psi}(\tau|\cdot)$, $F^{-1}_{Y|\psi}(\tau|y_0) = y_0$. This implies that $Y|\psi = y_0$ is degenerate.

F.5 Proof of Proposition 2.

We first prove that H_{0X} is equivalent to the existence of (Y', ψ') such that $DY' + (1-D)\psi' = \widetilde{Y}$, $D \perp\!\!\!\perp (Y', \psi')|X$ and $\mathbb{E}((Y'|\psi', X) = \psi'$. First, under H_{0X}, there exists $(Y', \psi', \mathcal{I}')$ such that $DY' + (1-D)\psi' = \widetilde{Y}$, $D \perp\!\!\!\perp (Y', \psi')|X$, $\sigma(\psi', X) \subset \mathcal{I}'$ and $\mathbb{E}(Y'|\mathcal{I}') = \psi'$. Then

$$\mathbb{E}[Y'|\psi', X] = \mathbb{E}\left[\mathbb{E}\left[Y'|\mathcal{I}'\right]|\psi', X\right] = \mathbb{E}\left[\psi'|\psi', X\right] = \psi'.$$

Conversely, if there exists (Y', ψ') such that $DY' + (1-D)\psi' = \widetilde{Y}$, $D \perp\!\!\!\perp (Y', \psi')|X$ and $\mathbb{E}(Y'|\psi', X) = \psi'$, let $\mathcal{I}' = \sigma(X', \psi')$. Then $\psi' = \mathbb{E}(Y'|\psi', X) = \mathbb{E}(Y'|\mathcal{I}')$ and H_{0X} holds. The proposition then follows as Theorem 1.

F.6 Proof of Proposition 4

For all y, $\xi \mapsto \mathbb{E}[(y - \psi - \xi)^+]$ is decreasing and convex. Then, because F_{ξ_ψ} dominates at the second order $F_{\xi_Y + \varepsilon}$, we have

$$\int \mathbb{E}\left[(y - \psi - \xi)^+\right] dF_{\varepsilon + \xi_Y}(\xi) \geq \int \mathbb{E}\left[(y - \psi - \xi)^+\right] dF_{\xi_\psi}(\xi).$$

As a result, for all y, we obtain

$$\begin{aligned}
\mathbb{E}\left[\left(y - \widehat{Y}\right)^+\right] &= \int \mathbb{E}\left[(y - \psi - \varepsilon - \xi_Y)^+ | \varepsilon + \xi_Y = \xi\right] dF_{\varepsilon + \xi_Y}(\xi) \\
&= \int \mathbb{E}\left[(y - \psi - \xi)^+\right] dF_{\varepsilon + \xi_Y}(\xi) \\
&\geq \int \mathbb{E}\left[(y - \psi - \xi)^+\right] dF_{\xi_\psi}(\xi) \\
&= \mathbb{E}\left[(y - \widehat{\psi})^+\right].
\end{aligned}$$

Moreover, $\mathbb{E}\left(\widehat{Y}\right) = \mathbb{E}\left(\widehat{\psi}\right)$. By Theorem 1, $\widehat{Y}$ and $\widehat{\psi}$ satisfy H_0.

F.7 Proof of Theorem 2.

(i) This is a particular case of Proposition 5 below, with $q(Y, c_0) = Y$. The proof is therefore omitted.

(ii) We show that equality holds for $F_0 \in \mathcal{F}_0$ satisfying the conditions stated in (ii). The proof is divided in three steps. We first prove convergence in distribution of T to S defined below, and conditional convergence of T^* towards the same limit. Then we show that the cdf H of S is continuous and strictly increasing in the neighborhood of its quantile of order $1 - \alpha$, for any $\alpha \in (0, 1/2)$. The third step concludes.

1. Convergence in distribution of T and T^*.

Let us introduce some notation. Let $K_{j,j}$ ($j \in \{1,2\}$) be the j-th diagonal element of the covariance kernel K, $\mathcal{S} : (\nu, K) \mapsto (1-p)\left(-\nu_1/K_{1,1}^{1/2}\right)^{+2} + p\left(\nu_2/K_{2,2}^{1/2}\right)^2$, $q(r) = \left(r^2+100\right)^{-1}(2r)^{-d_X}$, and

$$\nu_{n,F_0}(y,g) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \text{Diag}\left(\mathbb{V}_{F_0}\left(\widetilde{Y}\right)\right)^{-1/2}\left(m\left(D_i, \widetilde{Y}_i, X_i, g, y\right) - \mathbb{E}_{F_0}\left[m\left(D_i, \widetilde{Y}_i, X_i, g, y\right)\right]\right).$$

Finally, we define $k_{n,F_0}(y,g) = \sqrt{n}\text{Diag}\left(\mathbb{V}_{F_0}\left(\widetilde{Y}\right)\right)^{-1/2}\mathbb{E}_{F_0}\left[m\left(D_i, \widetilde{Y}_i, X_i, g, y\right)\right]$,

$$K_{n,F_0}(y,g,y',g') = \text{Diag}\left(\mathbb{V}_{F_0}\left(\widetilde{Y}\right)\right)^{-1/2}\widehat{\text{Cov}}\left(\sqrt{n}\overline{m}_n(y,g), \sqrt{n}\overline{m}_n(y',g')\right)\text{Diag}\left(\mathbb{V}_{F_0}\left(\widetilde{Y}\right)\right)^{-1/2},$$

$$\overline{K}_{n,F_0}(y,g,y',g') = K_{n,F_0}(y,g,y',g') + \epsilon\text{Diag}\left(\mathbb{V}_{F_0}\left(\widetilde{Y}\right)\right)^{-1/2}\text{Diag}\left(\widehat{\mathbb{V}}\left(\widetilde{Y}\right)\right)\text{Diag}\left(\mathbb{V}_{F_0}\left(\widetilde{Y}\right)\right)^{-1/2},$$

and use the notations $K_{n,F_0}(y,g) = K_{n,F_0}(y,g,y,g)$ and $\overline{K}_{n,F_0}(y,g) = \overline{K}_{n,F_0}(y,g,y,g)$.

We have, by definition of T,

$$T = \sup_{y\in\mathcal{Y}} \sum_{(a,r):r\in\{1,\dots,r_n\}, a\in A_r} q(r)\mathcal{S}\left(\nu_{n,F_0}(y,g_{a,r}) + k_{n,F_0}(y,g_{a,r}), \overline{K}_{n,F_0}(y,g_{a,r})\right).$$

To characterize the distribution of T (resp. T^*), we first prove the convergence of ν_{n,F_0} and $K_{n,F_0}(y,g_{a,r})$ (resp. ν^*_{n,F_0} and $K^*_{n,F_0}(y,g_{a,r})$). For those purposes, we use a class of functions which is a general form taken by m_1 defined in (2), namely, for any $0 < N_1 < M_1$,

$$\begin{aligned}\mathcal{M}_0 = \{f_{y,\phi_1,\phi_2,g}(\widetilde{y},x,d) = \left(d\phi_1(y-\widetilde{y})^+ - (1-d)\phi_2(y-\widetilde{y})^+\right)g(x),\\ (y,\phi_1,\phi_2,g) \in \mathcal{Y}\times[N_1,M_1]^2\times\mathcal{G}\}.\end{aligned}$$

Remark that $\mathcal{M}_0$ is a particular case of classes $\mathcal{M}$ defined in (15) below. Then, by the proof of Proposition 5 below, Assumptions PS1 and PS2 in AS are satisfied. Thus, the assumptions of Lemma D.2 in AS hold as well. This entails that Assumptions PS4 and PS5 in AS hold. Namely, there exists a Gaussian process ν_{F_0} such that

- $\nu_{n,F_0} \to_d \nu_{F_0}$ and $\nu^*_{n,F_0} \to_{d^*} \nu_{F_0}$;
- For all $r \in \mathbb{N}$ and $(y,g) \in \mathcal{Y}\times\mathcal{G}_r$, $\overline{K}_{n,F_0}(y,g) \to_P K_{F_0}(y,g) + \epsilon I_2$ and $K^*_{n,F_0}(y,g) \to_{P^*} K_{F_0}(y,g) + \epsilon I_2$, where I_2 is the 2×2 identity matrix.

Moreover, letting $k_{F_0}(y,g)$ denote the limit in probability of $k_{n,F_0}(y,g)$, we have $k_{F_0}(y,g) = 0$ if $(y,g) \in \mathcal{L}_{F_0}$ and $+\infty$ otherwise. Note that by assumption, the set $\mathcal{L}_{F_0}$ is nonempty.

Thus, using (D.11) in the proof of Theorem D.3. in AS, which is based on the uniform continuity of the function $\mathcal{S}$ in the sense of Assumption S2 therein, we have, under F_0,

$$\begin{aligned}T &\to_d \sup_{y\in\mathcal{Y}} \sum_{(a,r)\in A_r\times\mathbb{N}} \mathcal{S}\left(\nu_{F_0}(y,g_{a,r}) + k_{F_0}(y,g_{a,r}), K_{F_0}(y,g_{a,r}) + \epsilon I_2\right)\\ &= S := \sup_{y\in\mathcal{Y}} \sum_{(a,r):(y,g_{a,r})\in\mathcal{L}_{F_0}} q(r)\mathcal{S}\left(\nu_{F_0}(y,g_{a,r}), K_{F_0}(y,g_{a,r}) + \epsilon I_2\right),\end{aligned}$$

where the equality follows by definition of $\mathcal{S}$ and $k_{F_0}(y,g)$. Similarly, using Assumption PS5 and (D.11) in AS, replacing T by T^* and quantities $\nu_{n,F_0}(y,g_{a,r})$ and $K_{n,F_0}(y,g_{a,r})$ by their bootstrap counterparts (see the proof of Lemma D.4 in AS) we have $T^* \to_{d^*} S$.

2. The cdf H of S is continuous and strictly increasing in the neighborhood of any of its quantile of order $1-\alpha > 1/2$.

First, the cdf H of S is a convex functional of the Gaussian process ν_{F_0}. Then, as in the proof of Lemma B3 in Andrews and Shi (2013), we can use Theorem 11.1 of Davydov et al. (1998) p.75 to show that H is continuous and strictly increasing at every point of its support except $\underline{r} = \inf\{r \in \mathbb{R} : H(r) > 0\}$. Moreover, for any $r > 0$,

$$\begin{aligned} H(r) &\geq \mathbb{P}\left(\sup_{y\in\mathcal{Y}} \sum_{(a,r):(y,g_{a,r})\in\mathcal{L}_{F_0}} q(r)\mathcal{S}\left(\nu_{F_0}(y,g_{a,r}), K_{F_0}(y,g_{a,r}) + \epsilon I_2\right) < r\right) \\ &\geq \mathbb{P}\left(\sup_{j\in\{1,2\},(y,a,r):(y,g_{a,r})\in\mathcal{L}_{F_0}} \left|(K_{2,F_0,j,j}(y,g_{a,r}) + \epsilon)^{-1/2}\nu_{F_0,j}(y,g_{a,r})\right| < \frac{\sqrt{r/2}}{Q}\right) \\ &> 0, \end{aligned}$$

where $Q = \sum_{(a,r):(y,g_{a,r})\in\mathcal{L}_{F_0}} q(r) < \infty$ and we use Problem 11.3 of Davydov et al. (1998) p.79 for the last inequality. This yields $r > \underline{r}$ and H is continuous and strictly increasing on $(0,\infty)$.

Then, we show that for any $\alpha \in (0,1/2)$, the quantile of order $1-\alpha$ of the distribution of S is positive. By assumption, there exists $(y_0,g_0) \in \mathcal{L}_{F_0}$ such that either $K_{F_0,11}(y_0,g_0) > 0$ or $K_{F_0,2}(y_0,g_0) > 0$. This yields

$$\begin{aligned} \mathbb{P}(S>0) &= 1 - \mathbb{P}\left(\sup_{y\in\mathcal{Y}} \sum_{(a,r):(y,g_{a,r})\in\mathcal{L}_{F_0}} q(r)\mathcal{S}\left(\nu_{F_0}(y,g_{a,r}), K_{F_0}(y,g_{a,r}) + \epsilon I_2\right) = 0\right) \\ &\geq 1 - \mathbb{P}\left(\nu_{F_0,1}(y_0,g_0) \leq 0,\ \nu_{F_0,2}(y,g_0) = 0\right) \\ &\geq 1 - \min\left\{\mathbb{P}\left(\nu_{F_0,1}(y_0,g_0) \leq 0\right), \mathbb{P}\left(\nu_{F_0,2}(y_0,g_0) = 0\right)\right\} \\ &\geq 1/2. \end{aligned} \tag{11}$$

The first inequality holds by definition of the supremum and because $\mathcal{S}$ is nonnegative. To obtain the last inequality, note that either $\nu_{F_0,1}(y_0,g_0)$ is non-degenerate, in which case the first probability is $1/2$ (since $\nu_{F_0,1}(y_0,g_0)$ is normal with zero mean), or $\nu_{F_0,2}(y_0,g_0)$ is non-degenerate, in which case the second probability is 0.

Finally, using that H is strictly increasing on $(0,\infty)$, (11) ensures that any quantile of S of order $1-\alpha$ with $\alpha \in [0,1/2)$ is positive. Hence, H is continuous and strictly increasing in the neighborhood of any such quantiles.

3. Conclusion.

Using $T^* \to_{d^*} S$ in distribution, Step 2 and Lemma 21.2 in Van der Vaart (2000), we have that for $\eta > 0$, $c^*_{n,\alpha} \to_{d^*} c(1-\alpha+\eta)+\eta$, where $c(1-\alpha+\eta)$ is the $(1-\alpha+\eta)$-th quantile of the distribution of S. Because $T \to_d S$ and H is continuous at $c(1-\alpha+\eta)+\eta > 0$, we obtain that

$$\lim_{\eta\to 0}\limsup_{n\to\infty} \mathbb{P}_{F_0}\left(T > c^*_{n,\alpha}\right) = \alpha.$$

Combined with the inequality of Part (i) above, this yields the result.

(iii) This results follows from Theorem E.1 in AS. First, Assumption SIG2 in AS holds for $\sigma_F^2 = \mathbb{V}_F\left(\widetilde{Y}\right)$, following the proof of Lemma 7.2 (b) under Assumption 3-(ii). Second, Assumptions PS4 and PS5 are satisfied using the point (ii) above. Third, Assumptions CI, MQ, S1, S3, S4 in AS are also satisfied by construction of the statistic T. Thus, Theorem E.1 in AS yields the result. □

F.8 Proof of Theorem 3.

Note first that because $F_{\mathbb{E}[Y|\mathcal{I}]}$ is continuous, $F_{\mathbb{E}[Y|\mathcal{I}]}(\mathbb{E}\left[Y|\mathcal{I}\right])$ is uniformly distributed (see, e.g. Van der Vaart, 2000, p.305). In turn, this implies that the cdf of $h^M\left(\mathbb{E}\left[Y|\mathcal{I}\right]\right)$ is F_ψ. Hence, $\left(h^M\left(\mathbb{E}\left[Y|\mathcal{I}\right]\right), \mathbb{E}\left[Y|\mathcal{I}\right]\right) \in \Psi^M$. If for all (ψ', ψ''), $\mathbb{E}(\rho(|\psi'-\psi''|)] = +\infty$, Equality (3) holds. If not, let $(\psi',\psi'') \in \Psi^M$ be such that $\mathbb{E}(\rho(|\psi'-\psi''|)] < +\infty$. Because ρ is convex, we have, for all $x' \geq x$ and $y' \geq y$,

$$\rho(|x'-y'|) - \rho(|x-y'|) - \rho(|x'-y|) + \rho(|x-y|) \leq 0.$$

Then, by Theorem 3.1.2 in Rachev and Rüschendorf (1998),

$$\begin{aligned} E[\rho(|\psi'-\psi''|)] &\geq \int_0^1 \rho\left(\left|F_\psi^{-1}(u) - F^{-1}_{\mathbb{E}(Y|\mathcal{I}^M)}(u)\right|\right) du. \\ &= \int \rho\left(\left|F_\psi^{-1}\circ F_{\mathbb{E}[Y|\mathcal{I}]}(v) - F^{-1}_{\mathbb{E}[Y|\mathcal{I}]}\circ F_{\mathbb{E}[Y|\mathcal{I}]}(v)\right|\right) dF_{\mathbb{E}[Y|\mathcal{I}]}(v) \\ &= \mathbb{E}\left[\rho\left(\left|h^M\left(\mathbb{E}\left[Y|\mathcal{I}\right]\right) - F^{-1}_{\mathbb{E}[Y|\mathcal{I}]}\circ F_{\mathbb{E}[Y|\mathcal{I}]}(\mathbb{E}\left[Y|\mathcal{I}\right])\right|\right)\right]. \end{aligned} \tag{12}$$

Finally, note that $F^{-1}_{\mathbb{E}[Y|\mathcal{I}]}\circ F_{\mathbb{E}[Y|\mathcal{I}]}(v) < v$ only if v is in the interior or at the right end of a "flat" of $F_{\mathbb{E}[Y|\mathcal{I}]}$ (see, e.g., Lemma 21.1 in Van der Vaart, 2000). Because the set of such right end points is countable and $F_{\mathbb{E}[Y|\mathcal{I}]}$ has no atom, $F^{-1}_{\mathbb{E}[Y|\mathcal{I}]}\circ F_{\mathbb{E}[Y|\mathcal{I}]}(\mathbb{E}\left[Y|\mathcal{I}\right]) = \mathbb{E}\left[Y|\mathcal{I}\right]$ almost surely. Combined with Equation (12), this implies (3).

Now, let us suppose that ρ is strictly convex and let $(\psi', \mathbb{E}\left[Y|\mathcal{I}\right]) \in \mathbf{\Psi}^M$ satisfy (3). We can apply the first part of the proof of Theorem 2.2.1 in Santambrogio (2015), remarking that it does not rely on the asssumption of compact supports. This implies that the distribution of $(\psi', \mathbb{E}\left[Y|\mathcal{I}\right])$ is equal to that of $(h^M(\mathbb{E}\left[Y|\mathcal{I}\right]), \mathbb{E}\left[Y|\mathcal{I}\right])$. Hence, conditional on $\mathbb{E}\left[Y|\mathcal{I}\right]$, ψ' is degenerate and equal to $h^M(\mathbb{E}\left[Y|\mathcal{I}\right])$. The result follows.

F.9 Proof of Proposition 5

We introduce $\mathbb{E}_{F,c} = \mathbb{E}_F\left[m\left(D_i, \widetilde{Y}_{c,i}, X_i, g, y\right)\right]$ and

$$\nu_{n,F}(y,g) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \text{Diag}\left(\widehat{\mathbb{V}}_F\left(\widetilde{Y}_{\widehat{c}}\right)\right)^{-1/2}\left(m\left(D_i, \widetilde{Y}_{\widehat{c},i}, X_i, g, y\right) - E_{F,\widehat{c}}\right),$$
$$\overline{\nu}_{n,F}(y,g) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \text{Diag}\left(\mathbb{V}_F\left(\widetilde{Y}_{c_0}\right)\right)^{-1/2}\left(m\left(D_i, \widetilde{Y}_{c_0,i}, X_i, g, y\right) - E_{F,c_0}\right).$$

The proof is based on Theorem 5.1 in AS, hence we have to check that the corresponding assumptions PS1, PS2, and SIG1 hold. Namely, we have to ensure that

- **PS1**: for all sequence $F \in \mathcal{F}$ and all $(d, y', x, g, y, c) \in \{0,1\} \times \mathcal{Y} \times [0,1]^{d_X} \times \mathcal{G}_r \times \mathcal{Y} \times \mathcal{C}_s\left([0,1]^{d_X}\right)$

$$\left|\frac{m(d,y',x,g,y)}{\mathbb{V}_F\left(\widetilde{Y}_{c,i}\right)}\right| \leq M(d,y',x,g,y) \text{ and } \mathbb{E}_F\left[M\left(D_i, \widetilde{Y}_{c,i}, X_i, g, y\right)^{2+\delta}\right] \leq C < \infty,$$

  where $\delta > 0$ and for some function M;

- **PS2**: for all sequence $F_n \in \mathcal{F}$, the i.i.d triangular array of processes

$$\mathcal{T}_n^0 = \left\{\frac{m\left(D_i, \widetilde{Y}_{n,c(X_{n,i})}, X_{n,i}, g, y\right)}{\mathbb{V}_{F_n}\left(\widetilde{Y}_{n,c(X_{n,i})}\right)},\ (c,y,g) \in \mathcal{C}_s\left([0,1]^{d_X}\right) \times \mathcal{Y} \times \mathcal{G},\ i \leq n,\ n \geq 1\right\}$$

  is manageable with respect to some envelope function U_1 (see Pollard, 1990, p.38 for the definition of a manageable class);

- **SIG1**: for all $\zeta > 0$, $\sup_{F\in\mathcal{F}, c\in\mathcal{C}_s\left([0,1]^{d_X}\right)} \mathbb{P}\left(\left|\widehat{\mathbb{V}}_F\left(\widetilde{Y}_{i,c}\right)/\mathbb{V}_F\left(\widetilde{Y}_{i,c}\right) - 1\right| > \zeta\right) \to 0.$

We proceed in two steps, to handle the fact that c_0 and $\text{Diag}\left(\mathbb{V}_F\left(\widetilde{Y}_{c_0}\right)\right)^{-1/2}$ are estimated:

1. We first show that

$$\sup_{F\in\mathcal{F}_0}\ \sup_{g\in\cup_{r\geq 1}\mathcal{G}_r, y\in\mathcal{Y}} \left\|\nu_{n,F}(y,g) - \overline{\nu}_{n,F}(y,g)\right\|_\infty = o_P(1), \tag{13}$$
$$\sup_{F\in\mathcal{F}_0}\ \sup_{g\in\cup_{r\geq 1}\mathcal{G}_r, y\in\mathcal{Y}} \left\|\nu^*_{n,F}(y,g) - \overline{\nu}^*_{n,F}(y,g)\right\|_\infty = o_{P^*}(1). \tag{14}$$

2. Next, we show that m satisfies assumptions PS1, PS2, and that SIG1 in AS also holds for $\sigma_F^2 = \mathbb{V}_F\left(\widetilde{Y}_{c_0}\right)$, where $F \in \mathcal{F}$ and $\widehat{\sigma}_n^2 = n^{-1}\sum_{i=1}^n\left(\widetilde{Y}_{\widehat{c},i} - n^{-1}\sum_{j=1}^n \widetilde{Y}_{\widehat{c},j}\right)^2$.

1. Proof of (13)-(14).

We apply the uniform version over $F \in \mathcal{F}_0$ of Theorem 3 in Chen et al. (2003) to a general class of functions to which pertain the moment condition m (see (2), with $\widetilde{Y}$ replaced here by $\widetilde{Y}_c = Dq\left(\widetilde{Y}, c\right) + (1-D)\psi$ and without the moment equality m_2). Hence, it suffices to verify that Assumptions (3.2) and (3.3) of Theorem 3 in Chen et al. (2003) are satisfied. Let us introduce, for any $0 < N_1 < M_1$, the classes of functions

$$\mathcal{M}_1 = \left\{ f_{c,y,\phi,g}\left(\widetilde{y}, x\right) = \phi\left(y - q\left(\widetilde{y}, c(x)\right)\right)^+ g(x), (c, y, \phi, g) \in \mathcal{C}_s\left([0,1]^{d_X}\right) \times \mathcal{Y} \times [N_1, M_1] \times \mathcal{G} \right\}, \tag{15}$$

$$\begin{aligned}\mathcal{M}_2 =& \left\{ f_{c,y,\phi,g}\left(\widetilde{y}, x\right) = \phi\left(y - \widetilde{y}\right)^+ g(x),\ (c, y, \phi, g) \in \mathcal{C}_s\left([0,1]^{d_X}\right) \times \mathcal{Y} \times [N_1, M_1] \times \mathcal{G} \right\}, \\ \mathcal{M} =& \{ f_{c,y,\phi_1,\phi_2,g}\left(\widetilde{y}, x, d\right) = \left(d g_{c,y,\phi_1,g} - (1-d) q_{c,y,\phi_2,g}\right)\left(\widetilde{y}, x\right),\ g \in \mathcal{M}_1,\ q \in \mathcal{M}_2, \\ & \qquad (c, y, \phi_1, \phi_2, g) \in \mathcal{C}_s\left([0,1]^{d_X}\right) \times \mathcal{Y} \times [N_1, M_1]^2 \times \mathcal{G} \}.\end{aligned}$$

Note that ϕ_1, ϕ_2, and c in the class $\mathcal{M}$ denote components of m that are estimated.

Consider the space $\mathcal{C}_s\left([0,1]^{d_X}\right) \times \mathcal{Y} \times [N_1, M_1]^2 \times \mathcal{G}$ equipped with the norm

$$\|(c, y, \phi_1, \phi_2, g)\| = \max\left\{ \|c\|_{[0,1]^{d_X}}, |y|, |\phi_1|, |\phi_2|, \|g\|_{[0,1]^{d_X}} \right\}.$$

For $v = (c, y, \phi_1, \phi_2, g), v' = (c', y', \phi_1', \phi_2', g') \in \mathcal{C}_s\left([0,1]^{d_X}\right) \times \mathcal{Y} \times [N_1, M_1]^2 \times \mathcal{G}$ and $(\widetilde{y}, x, d) \in \mathcal{Y} \times [0,1]^{d_X} \times \{0,1\}$, we have, by the triangular inequality and Assumptions 4-(i) and 4-(v),

$$\begin{aligned}\left|f_v\left(\widetilde{y}, x, d\right) - f_{v'}\left(\widetilde{y}, x, d\right)\right| \leq& \left| g_{c,y,\phi_1,g}\left(\widetilde{y}, x\right) - g_{c',y',\phi_1',g'}\left(\widetilde{y}, x\right) \right| \\ &+ \left| q_{c,y,\phi_2,g}\left(\widetilde{y}, x\right) - q_{c',y',\phi_2',g'}\left(\widetilde{y}, x\right) \right| \\ \leq& (M + M_0)\left(\left|\phi_1 - \phi_2'\right| + \left|\phi_2 - \phi_2'\right|\right) \\ &+ 2M_1\left[\left|y - y'\right| + \left|q\left(\widetilde{y}, c(x)\right) - q\left(\widetilde{y}, c'(x)\right)\right|\right] \\ &+ 2M_0M_1\Big[\left|\mathbb{1}\left\{q(\widetilde{y}, c(x)) \leq y\right\} - \mathbb{1}\left\{q(\widetilde{y}, c(x)) \leq y'\right\}\right| \\ &\qquad + \left|\mathbb{1}\left\{q\left(\widetilde{y}, c(x)\right) \leq y'\right\} - \mathbb{1}\left\{q\left(\widetilde{y}, c'(x)\right) \leq y'\right\}\right| \\ &\qquad + \left|g(x) - g'(x)\right|\Big].\end{aligned}$$

Denote by $K_q > 0$ the Lipschitz constant of $q(\widetilde{y}, .)$. Then, by convexity of $x \mapsto x^2$, we obtain

$$\begin{aligned}\frac{1}{7}\left|f_v\left(\widetilde{y}, x, d\right) - f_{v'}\left(\widetilde{y}, x, d\right)\right|^2 \leq& (M + M_0)^2\left(\left|\phi_1 - \phi_1'\right|^2 + \left|\phi_2 - \phi_2'\right|^2\right) \\ &+ 4M_1^2\left[\left|y - y'\right|^2 + K_q\left\|c - c'\right\|^2_{[0,1]^{d_X}}\right] \\ &+ 4(M_0M_1)^2\Big[\left|\mathbb{1}\left\{q(\widetilde{y}, c(x)) \leq y\right\} - \mathbb{1}\left\{q(\widetilde{y}, c(x)) \leq y'\right\}\right| \\ &\qquad + \left|\mathbb{1}\left\{q\left(\widetilde{y}, c(x)\right) \leq y'\right\} - \mathbb{1}\left\{q\left(\widetilde{y}, c'(x)\right) \leq y'\right\}\right| \\ &\qquad + \left\|g - g'\right\|^2_{[0,1]^{d_X}}\Big].\end{aligned}$$

Fix $\delta > 0$. If $\|v - v'\| \leq \delta$, this yields

$$\begin{aligned}\frac{1}{7}\left|f_v\left(\widetilde{y}, x, d\right) - f_{v'}\left(\widetilde{y}, x, d\right)\right|^2 \leq & \delta^2\left(2(M+M_0)^2 + 4M_1^2(1+K_q) + 4(M_0M_1)^2\right) \\ & + 4(M_0M_1)^2\big[\mathbb{1}\left\{q(\widetilde{y}, c(x)) \leq y + \delta\right\} - \mathbb{1}\left\{q(\widetilde{y}, c(x)) \leq y - \delta\right\} \\ & + \left|\mathbb{1}\left\{\widetilde{y} \leq q^I\left(y', c(x)\right)\right\} - \mathbb{1}\left\{\widetilde{y} \leq q^I\left(y', c'(x)\right)\right\}\right|\big].\end{aligned}$$

Next, by Assumption 4-(iv), we obtain

$$\begin{aligned}\mathbb{E}\left[\mathbb{1}\left\{q\left(\widetilde{Y}, c(X)\right) \leq y + \delta\right\} - \mathbb{1}\left\{q\left(\widetilde{Y}, c(X)\right) \leq y - \delta\right\}\right] &= F_{q\left(\widetilde{Y}, c(X)\right)}(y+\delta) - F_{q\left(\widetilde{Y}, c(X)\right)}(y-\delta) \\ &\leq 2\overline{Q}_2\delta.\end{aligned}$$

Finally, we have

$$\begin{aligned}&\mathbb{E}\left[\left|\mathbb{1}\left\{Y \leq q^I\left(y', c(X)\right)\right\} - \mathbb{1}\left\{\widetilde{y} \leq q^I\left(y', c'(X)\right)\right\}\right|\right] \\ \leq & \mathbb{E}\left[\mathbb{1}\left\{Y \leq q^I\left(y', c(X)\right) - Q_{F,2}\delta\right\} - \mathbb{1}\left\{\widetilde{y} \leq q^I\left(y', c(X)\right) + Q_{F,2}\delta\right\}\right] \\ \leq & \mathbb{E}\left[F_{Y|X}\left(q^I\left(y', c(X)\right) - Q_{q^I}\delta \middle| X\right) - F_{Y|X}\left(q^I\left(y', c(X)\right) + Q_{q^I}\delta \middle| X\right)\right] \\ \leq & 2Q_{F,1}Q_{q^I}\delta,\end{aligned}$$

where Q_{q^I} is the Lipschitz constant of q^I. Thus, by Assumption 4, there exists $Q > 0$ such that

$$\sup_{F \in \mathcal{F}_0} \mathbb{E}\left[\sup_{\|v - v'\| \leq \delta}\left|f_v\left(\widetilde{Y}, X, D\right) - f_{v'}\left(\widetilde{Y}, X, D\right)\right|^2\right] \leq Q\delta. \tag{16}$$

Therefore the class $\mathcal{M}$ satisfies Condition (3.2) of Theorem 3 in Chen et al. (2003) uniformly in $F \in \mathcal{F}_0$. Moreover, the class $\mathcal{G}$ is manageable and thus Donsker (see Lemma 3 in Andrews and Shi, 2013). Finally, by Remark 3 (ii) in Chen et al. (2003), $\mathcal{C}_s\left([0,1]^{d_X}\right)$ is also Donsker. Then, $\mathcal{C}_s\left([0,1]^{d_X}\right)$, $\mathcal{Y}$, $[N_1, M_1]$, and $\mathcal{G}$ satisfy Condition (3.3) of Theorem 3 in Chen et al. (2003). The result follows by Theorem 3 in Chen et al. (2003).

2. m satisfies PS1 and PS2 of AS and SIG1 of AS also holds for σ_F^2 and $\widehat{\sigma}_n^2$.

From Assumption 4 (iii) and the proof of Lemma 7.2 (a) in AS, PS1 is satisfied replacing B by $\max(M, M_0)$ in the proof of Lemma 7.2-(a) in AS.

We now show that PS2 in AS also holds. As the result is uniform over $\mathcal{F}_0$, we have to consider sequences for the cdfs F_n of $(D_{n,i}, Y_{n,i}, X_{n,i})_{i=1\dots n}$ (with $F_n \in \mathcal{F}_0$). We also define

$$\begin{aligned}\widetilde{Y}_{n,c(X_{n,i})} &= D_{n,i}q\left(Y_{n,i}, c(X_{n,i})\right) + (1 - D_{n,i})\psi_{n,i}, \\ W_{n,i} &= \frac{D_{n,i}}{\mathbb{E}_{F_n}\left[D_{n,i}\right]} - \frac{1 - D_{n,i}}{\mathbb{E}_{F_n}\left[1 - D_{n,i}\right]}, \\ \sigma_{F_n}^2 &= \mathbb{V}_{F_n}\left(\widetilde{Y}_{n,c(X_{n,i})}\right).\end{aligned}$$

Note that by Assumption 3 (iii), $\sigma^2_{F_n} \geq \overline{\sigma} > 0$ for all $F_n \in \mathcal{F}$. Let $(\Omega, \mathbb{F}, F_n)$ be a probability space and let ω denote a generic element in Ω. Showing Assumption PS2 in AS then boils down to prove that for any $0 < N_1 < M_1 := 1/\inf_F \sigma^2_F$, the i.i.d triangular array of processes

$$\mathcal{T}_{1,n,\omega} = \Big\{ W_{n,i}\phi\left(y - \widetilde{Y}_{n,c(X_{n,i})}\right)^+ g(X_{n,i}),\ (c, y, \phi, g) \in \mathcal{C}_s\left([0,1]^{d_X}\right) \times \mathcal{Y} \times [N_1, M_1] \times \mathcal{G},$$
$$i \leq n, n \geq 1 \Big\}$$

is manageable with respect to some envelope function U_1. Lemma 3 in Andrews and Shi (2013) shows that the processes $\{g(X_{n,i}),\ g \in \mathcal{G},\ i \leq n, n \geq 1\}$ are manageable with respect to the constant function 1. Then, using Lemma D.5 in AS, it remains to show that

$$\mathcal{T}'_{1,n,\omega} = \left\{ W_{n,i}\phi\left(y - \widetilde{Y}_{n,c(X_{n,i})}\right)^+,\ (c, y, \phi) \in \mathcal{C}_s\left([0,1]^{d_X}\right) \times \mathcal{Y} \times [N_1, M_1], i \leq n, n \geq 1 \right\},$$

is manageable with respect to some envelope. For such an envelope, we can consider $U'_1(\omega) = (M_0 + M)/(\overline{\sigma}\epsilon_0)$. We now prove the manageability of $\mathcal{T}'_{1,n,\omega}$. Let us define

$$\mathcal{M}' = \Big\{ f_{c,y,\phi_1,\phi_2}\left(\widetilde{y}, x, d\right) = d\phi_1\left(y - q\left(\widetilde{y}, c(x)\right)\right)^+ - (1-d)\phi_2\left(y - \widetilde{y}\right)^+,$$
$$(c, y, \phi_1, \phi_2) \in \mathcal{C}_s\left([0,1]^{d_X}\right) \times \mathcal{Y} \times [N_1, M_1]^2 \Big\}.$$

Reasoning as for the class $\mathcal{M}$ defined in (15), and using the last equation of the proof of Theorem 3 in Chen et al. (2003), p.1607, we have that for $\epsilon > 0$,

$$N_{[\cdot]}\left(\epsilon, \mathcal{M}', \|\cdot\|_2\right) \leq N\left(\epsilon', [N_1, M_1]^2, |\cdot|\right) \times N\left(\epsilon', \mathcal{Y}, |\cdot|\right) \times N\left(\epsilon', \mathcal{C}_s\left([0,1]^{d_X}\right), \|\cdot\|_{[0,1]^{d_X}}\right),$$

with $\epsilon' = (\epsilon/(2Q))^2$ and Q defined in (16). Using Theorem 2.7.1 page 155 in Van der Vaart and Wellner (1996), there exists a constant Q_2 depending only on s, d_X, and $[0,1]^{d_X}$ such that

$$\ln\left(N\left(\epsilon', \mathcal{C}_s([0,1]^{d_X}), \|\cdot\|_{[0,1]^{d_X}}\right)\right) \leq Q_2 \epsilon'^{-d_X/s}.$$

Moreover, because $\mathcal{Y}$ and $[N_1, M_1]$ are compact subsets of two Euclidean spaces, there exist Q_3, Q_4 such that

$$N\left(\epsilon', [N_1, M_1]^2, |\cdot|\right) \leq Q_3 \epsilon'^{-4} \text{ and } N\left(\epsilon', \mathcal{Y}, |\cdot|\right) \leq Q_4 \epsilon'^{-2}. \tag{17}$$

This yields

$$\ln\left(N_{[\cdot]}\left(\epsilon, \mathcal{M}', \|\cdot\|_2\right)\right) \leq (6 + Q_2)\max\left(-\ln(\epsilon'), \epsilon'^{-d_X/s}\right) + \ln(Q_3 Q_4). \tag{18}$$

Let $\odot$ denote element-by-element product and $\mathcal{D}\left(\epsilon\left|\alpha \odot U'_1(\omega)\right|, \alpha \odot \mathcal{T}'_{1,n,\omega}\right)$ denote random packing numbers. By (A.1) in Andrews (1994, p.2284), we have

$$\begin{aligned}\sup_{\omega\in\Omega, n\geq 1,\ \alpha\in\mathbb{R}^n_+} \mathcal{D}\left(\epsilon\left|\alpha \odot U'_1(\omega)\right|, \alpha \odot \mathcal{T}'_{1,n,\omega}\right) &\leq \sup_{F\in\mathcal{F}_0} N\left(\frac{\epsilon}{2}, \mathcal{M}', \|\cdot\|_2\right) \\ &\leq \sup_{F\in\mathcal{F}_0} N_{[\cdot]}\left(\epsilon, \mathcal{M}', \|\cdot\|_2\right),\end{aligned} \tag{19}$$

where the second inequality follows as in e.g., Van der Vaart and Wellner (1996, p.84). Then, (18) ensures (see Definition 7.9 in Pollard (1990), p.38) that

$$\sup_{\omega\in\Omega, n\geq 1,\ \alpha\in\mathbb{R}_+^n} \mathcal{D}\left(\epsilon \left|\alpha\odot U_1'(\omega)\right|, \alpha\odot \mathcal{T}_{1,n,\omega}'\right) \leq \lambda(\epsilon),$$

where $\lambda(\epsilon) = \exp\left((6+Q_2)\max\left(-2\ln\left(\epsilon/(2Q)\right), (\epsilon/(2Q))^{-2d_X/s}\right) + \ln(Q_3Q_4)\right)$. Moreover, by using $\sqrt{a+b}\leq\sqrt{a}+\sqrt{b}$ for all $a,b\geq 0$,

$$\begin{aligned}\int_0^1 \sqrt{\ln(\lambda(\epsilon))}d\epsilon &\leq \sqrt{6+Q_2}\int_0^1 \left[\max\left(-2\ln\left(\epsilon/(2Q)\right), (\epsilon/(2Q))^{-2d_X/s}\right)\right]^{1/2} d\epsilon + \sqrt{\ln(Q_3Q_4)} \\ &< \infty.\end{aligned}$$

Thus, $\mathcal{T}_{1,n,\omega}'$ hence $\mathcal{T}_{1,n,\omega}$ are manageable. Therefore, m satisfies PS2 in AS.

Finally, in order to show that SIG1 in AS is satisfied, we use Assumption 4 (iii) and follow the proof of Lemma 7.2 (b) in AS where we replace Y by $q(Y, c(X))$ and B by $\max(M, M_0)$. The result follows.

F.10 Proof of Proposition 6

We first prove that if $\mathbb{E}[\psi_L]\leq\mathbb{E}[Y]\leq\mathbb{E}[\psi_U]$, there exists a unique $F^*\in\mathcal{F}_B$ such that $\delta_{F^*}=0$. First, suppose that $F^b\neq F^{b'}$ and, without loss of generality, $b>b'$. Then $\psi^b\leq\psi^{b'}$, implying that $F^b(y)\leq F^{b'}(y)$ for all y. Moreover, the inequality is strict for at least one y. As a result, $\mathbb{E}(\psi^b)>\mathbb{E}(\psi^{b'})$. In other words, there is at most one $F^*\in\mathcal{F}_B$ such that $\delta_{F^*}=0$. If $\mathbb{E}[\psi_L]=\mathbb{E}[Y]$ or $\mathbb{E}[\psi_U]=\mathbb{E}[Y]$, such a solution also exists by taking $b=-\infty$ and $b=+\infty$, respectively. Now, suppose that $\mathbb{E}[\psi_L]<\mathbb{E}[Y]<\mathbb{E}[\psi_U]$. For all $+\infty>b>b'>-\infty$,

$$\begin{aligned}\psi^b-\psi^{b'} &= \left(\psi_U-\max(\psi_L,b')\right)\mathbb{1}\{\psi_U\in[b',b)\} + (b-b')\mathbb{1}\{\psi_L<b', \psi_U\geq b\} \\ &\quad + (b-\psi_L)\mathbb{1}\{\psi_L\in[b',b), \psi_U\geq b\}.\end{aligned}$$

As a result, $|\psi^b-\psi^{b'}|\leq|b-b'|$. This implies that $\widetilde{\delta}: b\mapsto\mathbb{E}[\psi^b]$ is continuous. Moreover, $\lim_{b\to-\infty}\widetilde{\delta}(b)=\mathbb{E}[\psi_L]<\mathbb{E}(Y)$ and $\lim_{b\to+\infty}\widetilde{\delta}(b)=\mathbb{E}[\psi_U]>\mathbb{E}(Y)$. By the intermediate value theorem, there exists b^* such that $\widetilde{\delta}(b^*)=\mathbb{E}(Y)$. Hence, there exists $F^*\in\mathcal{F}_B$ such that $\delta_{F^*}=0$. The first part of Proposition 6 follows.

Let us turn to the second part of the proposition. First, if (ii) holds, there exists $b_0\in\overline{\mathbb{R}}$ such that $F^*=F^{b_0}$. Then, by construction and Theorem 1, Y and ψ^{b_0} satisfy H_0. Moreover, $F^{b_0}\in[F_{\psi_U}, F_{\psi_L}]$. Therefore, H_{0B} holds as well.

Now, let us prove that (i) implies (ii). Let us denote by $\mathcal{D}$ the set of all the cdfs for ψ such that H_{0B} holds. By Theorem 1, these are cdfs F satisfying $F_{\psi_U}\leq F\leq F_{\psi_L}$, $\delta_F=0$ and dominating at the second order F_Y. We show below that all $F\in\mathcal{D}$ are dominated at the second order by

F^*. Then, because $F_{\psi_U} \leq F^* \leq F_{\psi_L}$ and $\int ydF^*(y) = \int ydF_Y(y)$, $\mathcal{D}$ is not empty only if F^* dominates at the second order F_Y. The result then follows by Theorem 1.

Thus, we have to show that for all $t \in \mathbb{R}$,

$$F^* = \text{argmin}_{F_\psi \in \mathcal{D}} \int_{-\infty}^{t} F_\psi(y)dy. \tag{20}$$

First, if $F^* = F^{-\infty}$, we have for all $F \neq F^*$, $F(y) \leq F_{\psi_L}(y) = F^*(y)$ for all y, with strict inequality for some y. Then $\delta_F > \delta_{F^*} = 0$ and $\mathcal{D} = \{F^*\}$, implying that (20) holds. Similarly, (20) holds if $F^* = F^{+\infty}$.

Suppose now that $F^* = F^{b_0}$ for some $b_0 \in \mathbb{R}$. Because $F_{\psi_U}(y) \leq F_\psi(y)$ for all $y < b_0$ and all $F_\psi \in \mathcal{D}$, (20) holds for all $t < b_0$. We now prove that (20) holds also for $t \geq b_0$. First suppose that $t \geq \max(b_0, 0)$. For all $F_\psi \in \mathcal{D}$, $\int ydF_Y(y) = \int ydF_\psi(y)dy$. As a result, by Fubini's theorem,

$$\begin{aligned} & -\int_{-\infty}^{0} F^*(y)dy + \int_0^t (1 - F^*(y))\, dy + \int_t^\infty (1 - F^*(y))\, dy \\ & = -\int_{-\infty}^{0} F_\psi(y)dy + \int_0^t (1 - F_\psi(y))\, dy + \int_t^\infty (1 - F_\psi(y))\, dy. \end{aligned}$$

Because $F_\psi \leq F_{\psi_L} = F^*$ on $[b_0, +\infty]$, this implies that

$$-\int_{-\infty}^{0} F^*(y)dy + \int_0^t (1 - F^*(y))\, dy \geq -\int_{-\infty}^{0} F_\psi(y)dy + \int_0^t (1 - F_\psi(y))\, dy$$

and thus (20) holds for $t \geq \max(b_0, 0)$. Now, if $b_0 < 0$ and $t \in (b_0, 0)$, we have

$$\begin{aligned} & -\left(\int_{-\infty}^{t} F^*(y)dy + \int_t^0 F^*(y)dy\right) + \int_0^\infty (1 - F^*(y))\, dy \\ & = -\left(\int_{-\infty}^{t} F_\psi(y)dy + \int_t^0 F_\psi(y)dy\right) + \int_0^\infty (1 - F_\psi(y))\, dy. \end{aligned}$$

Using again $F_\psi \leq F_{\psi_L} = F^*$ on $[t, +\infty)$ yields

$$-\int_t^0 F^*(y)dy + \int_0^\infty (1 - F^*(y))\, dy \leq -\int_t^0 F_\psi(y)dy + \int_0^\infty (1 - F_\psi(y))\, dy.$$

Therefore, the result also follows in this case.