Rationalizing Rational Expectations: Characterizations and Tests*

Xavier D'Haultfoeuille[†] Christophe Gaillac[‡] Arnaud Maurel[§]

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. The test is robust to measurement errors under some restrictions and can be extended to account for aggregate shocks. Finally, we apply our methodology to test for rational expectations about future earnings. While individuals tend to be right on average about their future earnings, our test strongly rejects rational expectations.

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

*This paper is based on portions of our previous working paper D'Haultfœuille et al. (2018*b*). We thank the co-editor Andres Santos, three anonymous referees, 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 participants of various seminars and conferences for useful comments and suggestions.

[†]CREST-ENSAE, xavier.dhaultfoeuille@ensae.fr. Xavier D'Haultfoeuille thanks the hospitality of PSE where part of this research was conducted.

[‡]CREST-ENSAE 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*; Gennaioli, Ma and Shleifer, 2016; Kuchler and Zafar, 2019; Boneva and Rauh, 2018; Biroli, Boneva, Raja and Rauh, 2020). 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.[1]

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

[1]Situations where realizations can be perfectly predicted beforehand, such as in school choice settings where assignments are a known function of observed inputs, are notable exceptions.

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.[2] As a consequence, if these moment conditions hold, RE cannot be **refuted,** 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 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 **we can still rationalize RE with such data** 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 that specific sense, imperfect data quality does not jeopardize the validity of our test. 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.

We apply our framework to test for rational expectations about future earnings. To

[2]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).

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.

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, 2019, 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 rational expectations.

The remainder of the paper is organized as follows. In Section 2, we present the general set-up and the main theoretical equivalences underlying our RE test. In Section 3, we introduce the corresponding statistical tests and study their asymptotic properties. Section 4 illustrates the finite sample properties of our tests through Monte Carlo simulations. Section 5 applies our framework to expectations about future earnings. Finally, Section 6 concludes. The appendix gathers the proofs of the equivalence results. We consider in the Web Appendix various theoretical extensions, additional simulation results, additional material on the application, and all the remaining proofs. Finally, the companion R package RationalExp, described in the user guide (D'Haultfœuille, Gaillac and Maurel, 2018*a*), performs the test of RE.

2 Set-up and characterizations

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 $(D_k)_{k=1...m}$. Our setup encompasses a wide variety of situations, where individuals have private information and form their beliefs based on their information set. This includes various contexts where individuals form their expectations about future outcomes, including education, labor market as well as health outcomes. 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:

H_0 : there exists a pair of random variables (Y', ψ') and a sigma-algebra $\mathcal{I}'$ such that $\sigma(\psi') \subset \mathcal{I}', Y' \sim Y,\ \psi' \sim \psi$ and $\mathbb{E}\left[Y'|\mathcal{I}'\right] = \psi'$,

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 $\mathbb{E}\left[Y|\mathcal{I}\right] = \psi$. Instead, it means that there exists a triplet $(Y', \psi', \mathcal{I}')$ consistent with the data and such that $\mathbb{E}\left[Y'|\mathcal{I}'\right] = \psi'$. In other words, **a violation of** H_0 implies that RE does not hold, in the sense that the true realizations, beliefs, and information set do not satisfy RE ($\mathbb{E}\left[Y|\mathcal{I}\right] \neq \psi$). The converse, however, is not true.

2.2 Equivalences

2.2.1 Main equivalence

Let 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 continuum of moment inequalities, and one moment equality.

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

(i) *H_0 holds;*

(ii) *(F_Y mean-preserving spread of F_ψ) $\Delta(y) \geq 0$ for all $y \in \mathbb{R}$ and $\mathbb{E}[Y] = \mathbb{E}[\psi]$;*

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

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 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 $\mathbb{E}\left[Y|\mathcal{I}\right] - a$, whereas 10% are over-optimistic and expect $\mathbb{E}\left[Y|\mathcal{I}\right] + a$. Then one can show that our approach leads to reject RE when $a \geq 1.755$, while for $a = 1.755$, $\mathbb{V}(\eta) \simeq 0.616 < \mathbb{V}(\varepsilon) = 1$.

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. $\psi = \mathbb{E}[Y|\mathcal{I}] = Y$.[3]

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 the distribution of Y conditional on $\psi = y_0$ is degenerate: $P(Y = y_0|\psi = y_0) = 1$.*

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

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:[4]

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

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 continuum of conditional moment inequalities, and one conditional moment equality.

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, say), along with the outcome 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 $\mathbb{E}\left[Y|C=c,\psi\right] = \psi$ is generally violated, even though the rational expectations hypothesis holds. It follows that one cannot directly apply our previous

[4]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.

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 $\mathcal{I}_{it-1} = \sigma\left(\alpha_i, (\beta_{t-k})_{k\geq 1}, (\varepsilon_{it-k})_{k\geq 1}\right)$. 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

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

Thus, $Y_{it} = \psi_{it} + C_t + \varepsilon_{it} - \varepsilon_{it-1}$, with $C_t = \beta_t - \mathbb{E}\left[\beta_t|(\beta_{t-k})_{k\geq 1}\right]$. 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 $\mathbb{E}\left[Y|\mathcal{I}\right] = \psi$, but instead to $\mathbb{E}\left[Y|\mathcal{I}\right] = c_0 + \psi$ 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 $\mathbb{E}\left[Y|\mathcal{I}\right] = c_0\psi$, 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, by $c_0 = \mathbb{E}(Y)/\mathbb{E}(\psi)$ in the multiplicative case. Moreover, 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 (Y', ψ'), 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 few remarks on this proposition 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 1.1. 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. Third, in Appendix 1.2, we examine whether one can extend the results above to test for RE when aggregate shocks affect the outcomes in a more general way. Proposition S2 establishes a negative result in this respect: as long as one allows for a sufficiently flexible dependence between the outcome and the aggregate shocks, any given distribution of subjective expectations is arbitrarily close to a distribution for which RE can be rationalized. This implies that, within this more general class of outcome models, there does not exist any almost-surely continuous RE test that has non-trivial power.

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. By robust, we mean that **we still rationalize RE,** when they in fact hold. Specifically, assume that the true variables (ψ and Y) are unobserved. Instead, we only observe $\widehat{\psi}$ and $\widehat{Y}$, which are affected by classical measurement errors.[5] Namely:

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

The following proposition shows that our test is robust to a certain degree of measurement errors on the beliefs.

Proposition 4 *Suppose that Y and ψ satisfy H_0, and let $\varepsilon = Y - \psi$ and $\left(\widehat{\psi}, \widehat{Y}\right)$ 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 is likely to be relevant in practice. One should not infer, however, that measurement errors are innocuous in our set-up. Indeed, the converse of Proposition 4 does not hold: $\widehat{Y}$ **and** $\widehat{\psi}$ **may satisfy** H_0**, though** Y **and** ψ **do not.** As a simple example, suppose that $Y \sim \mathcal{N}(0, \sigma_Y^2)$, $\psi \sim \mathcal{N}(0, \sigma_\psi^2)$, $\xi_Y \sim \mathcal{N}(0, \sigma_3^2)$, $\xi_\psi = 0$ and $\sigma_\psi^2 \in (\sigma_Y^2, \sigma_Y^2 + \sigma_3^2]$. Then, $\widehat{Y}$ and $\widehat{\psi}$ satisfy H_0, since $\sigma_\psi^2 \leq \sigma_Y^2 + \sigma_3^2$, whereas Y and

[5] 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.

ψ do not, since $\sigma_\psi^2 > \sigma_Y^2$. Importantly though, Proposition 4 does show that our test is conservative in the sense that measurement errors cannot result in incorrectly concluding that the RE hypothesis does not hold.

In situations where $(\widehat{Y}, \widehat{\psi})$ are jointly observed, one could in principle alternatively implement the direct test. However, in contrast to our test, the direct test is not robust to any measurement errors on the subjective beliefs ψ. Indeed, if RE holds, 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 $\text{Cov}(\xi_Y, \widehat{\psi}) = \text{Cov}(\xi_\psi, Y) = 0$ and $\mathbb{V}(\xi_\psi) > 0$. In other words, even if individuals have rational expectations, the direct test will reject the null hypothesis in the presence of even an arbitrarily small degree of measurement errors on the elicited beliefs.

Also, it is unclear whether, in the presence of measurement errors on the elicited beliefs and beyond the restrictions on the marginal distributions, there are restrictions on the copula of $(\widehat{Y}, \widehat{\psi})$ that are implied by RE. For instance, we show in Proposition S3 in Appendix 2 that under RE, and without imposing restrictions on the dependence between $\xi_Y + \varepsilon$ and ξ_ψ, the coefficient of the (theoretical) linear regression of $\widehat{Y}$ on $\widehat{\psi}$ remains unrestricted.[6] On the other hand, if one assumes that $\text{Cov}(\xi_Y + \varepsilon, \xi_\psi) \geq 0$ and $\mathbb{V}(\psi)/\mathbb{V}(\xi_\psi) \geq \underline{\lambda}$ for some $\underline{\lambda} \geq 0$, Proposition S3 also shows that the coefficient of the linear regression of $\widehat{Y}$ on $\widehat{\psi}$ is bounded from below under RE. Such a restriction, which does require to take a stand on the signal-to-noise ratio $\mathbb{V}(\psi)/\mathbb{V}(\xi_\psi)$, can be easily added to the moment inequalities of our test if $(\widehat{Y}, \widehat{\psi})$ is observed.

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 3 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 4, we

[6] There might of course possibly be additional relevant information in the higher-order moments, although we have not been able to find any.

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.

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). 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.[7] The initial step is to transform the conditional moments into the following unconditional moments conditions:

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

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

We suppose to observe a sample $(D_i, X_i, \widetilde{Y}_i)_{i=1\dots n}$ 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 $\widetilde{X}_i$ 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 $x = (x_1, \dots, x_{d_X})$, we let $\Phi_0(x) = \left(\Phi(x_1), \dots, \Phi\left(x_{d_X}\right)\right)^{\top}$. Here Φ denotes the standard normal cdf, $\widehat{\Sigma}_{\widetilde{X},n}$ is the sample covariance matrix of $\left(\widetilde{X}_i\right)_{i=1\dots n}$ 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 $A_r = \{1, 2, \dots, 2r\}^{d_X}$ $(r \geq 1)$, $g_{a,r}(x) = \mathbb{1}\{x \in C_{a,r}\}$ and, for any $a = (a_1, ..., a_{d_X})^{\top} \in A_r$,

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

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

Finally, to define the test statistic T, we need to introduce additional notations. First, let $w_i = nD_i/\sum_{j=1}^n D_j - n(1-D_i)/\sum_{j=1}^n(1-D_j)$ and define, for any $y \in \mathbb{R}$,

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

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

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

where $\widehat{\mathcal{Y}} = \left[\min_{i=1,\dots,n} \widetilde{Y}_i, \max_{i=1,\dots,n} \widetilde{Y}_i\right]$, $p \in (0,1)$ is a parameter weighting 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:

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

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 $\varphi_{n,\alpha} = \mathbb{1}\left\{T > c^*_{n,\alpha}\right\}$ where the estimated critical value $c^*_{n,\alpha}$ is obtained by bootstrap using as in AS the Generalized Moment Selection method. Specifically, we follow 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.[8]

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

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

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

(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 1.1 for a detailed discussion of this extension.

4 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 5.

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

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

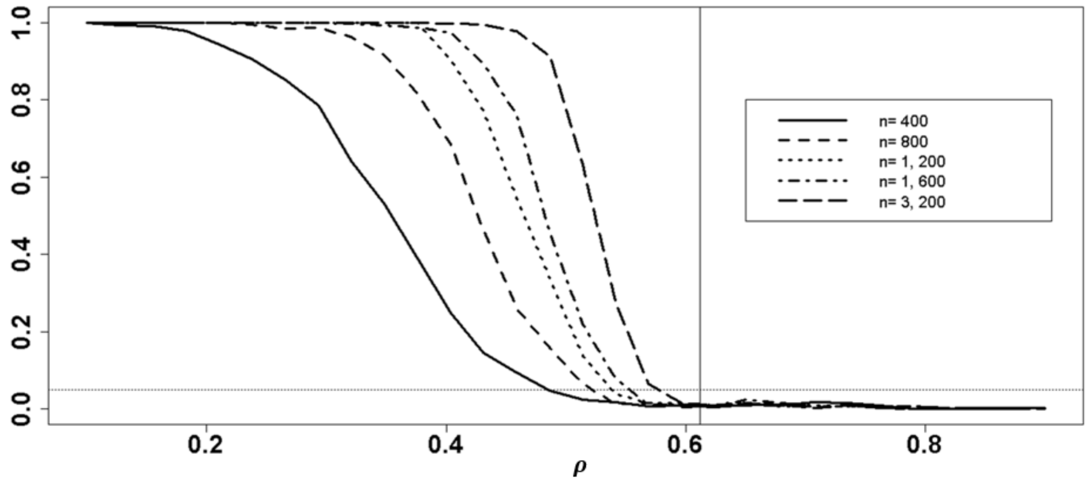
where ζ, U and ψ are mutually independent, $\zeta \sim \mathcal{N}(2, 0.1)$ and $U \sim \mathcal{U}[0, 1]$. In this setup, $\mathbb{E}(Y|\psi) = \rho\psi$ and expectations are rational if and only if $\rho = 1$. But since we observe Y and ψ in two different datasets, there are values of $\rho \neq 1$ for which our test **is not consistent.** More precisely, we can show that **the test is consistent if and only if** $\rho \leq \rho^* \simeq 0.616$. Besides, given this data generating process, the naive test $\mathbb{E}(Y) = \mathbb{E}(\psi)$ **is not consistent for any** ρ**,** while the RE test based on variances is only able to detect a subset of violations of RE that correspond to $\rho < 0.445$.

To compute our test, we need to choose the tuning parameters b_0, κ, ϵ and η (see Section 3 for definitions). As mentioned in Section 3, we set $\epsilon = 0.05$ and $\eta = 10^{-6}$, following Andrews and Shi (2017). Andrews and Shi (2013) show that there exists in practice a large range of admissible values for the other tuning parameters

parameters. Regarding b_0 and κ, we follow Beare and Shi (2019, Section 4.2) and compute, for a grid of candidate parameters, the rejection rate under the null and under one alternative (namely, $\rho = 0.5$), through Monte Carlo simulations. Then, we set (b_0, κ) so as to maximize the power subject to the constraint that the rejection rate under the null is below the nominal size 0.05. That way, we obtain $b_0 = 0.3$ and $\kappa = 0.001$. The parameter p has a distinct effect, in that its choice does not affect size, at least asymptotically. Rather, this parameter selects to what extent the test aims power at the equality constraint $E(Y - \psi) = 0$ versus the inequalities $E[(y-Y)^+ - (y-\psi)^+] \geq 0$ $(y \in \mathbb{R})$. Setting p to 0.05 leads to slightly higher power in our DGP, but values of$p$ in $[0, 0.31]$ provide similar finite sample performances, with power always greater than 90% of the maximal power.

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.

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 Section 5 of the Web Appendix, rejection frequencies under the null are very close to the nominal level.



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.

5 Application to earnings expectations

5.1 Data

Using the tests developed in Section 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, 2020 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.[9] Each month, the panel consists of about 180 entrants, and 1,100 repeated respondents. While entrants are overall fairly similar to the repeated respondents,

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

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.[10] 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?".

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.[11] 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. 51% (1,536) of these observations correspond to the sub-sample of respondents who are reinterviewed at least once. We refer to Table 1 for additional details on our sample.

[10]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 2 below). This provides empirical support for this assumption.

[11]Throughout our analysis (with the exception of the number of observations reported in Table 2) 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.

Table 1: 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

5.2 Implementation of the test

We summarize how we implemented the test in practice, either on the overall sample or on each subsample corresponding to the binary covariates in Table 1. For each case, we start by winsorizing the distribution of realized earnings (Y) and earnings beliefs (ψ) at the 95% level.[12] Then, we perform the test without covariates, where we allow for multiplicative aggregate shock and thus test H_{0S}, with $q(y;c) = y/c$.[13] Then, we use the function `test` of our companion R package RationalExp.[14] We choose the same values for the tuning parameters $b_0 = 0.3$ and $\kappa = 0.001$ as in the Monte-Carlo simulations in Section 4. We also set $p = 0.05$, $\epsilon = 0.05$, and $\eta = 10^{-6}$. Following Andrews and Shi (2017), the interval $\widehat{\mathcal{Y}}$ is approximated by a grid of length 100 from $\min_{i=1,\dots,n} \widetilde{Y}_i$ to $\max_{i=1,\dots,n} \widetilde{Y}_i$. Finally, we use 5,000 bootstrap simulations to compute the critical values of the test.

5.3 Are earnings expectations rational?

In Table 2 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.

[12] We show in Table S1 of the Web Appendix that our results are robust to other levels of Winsorization.

[13] In our application, the parameter c is estimated using survey weights from the SCE.

[14] See Section 3 in our user's guide (D'Haultfœuille et al., 2018*a*) for details on this function.

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.

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.[15]

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

[15] 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).

Table 2: 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	1,294	1,248

Notes: "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.

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. These results also suggest that, in order to rationalize the realized and expected earnings data, one should consider alternative models of expectation formation that primarily differ from RE in their third, or higher-order moments.

The results of the direct test of RE on the subsample of individuals who are followed over four months are reported in Table 3 below. While these results generally paint a similar picture to the results of our test, there are some differences. In particular, the direct test rejects RE at the 5% level for men and at 1% for individuals with tenure greater than 6 months, whereas we do not reject RE for the former group and only marginally so, at the 10% level, for the latter. The direct test also rejects with less power than our test for certain groups (low numeracy, tenure lower than 6 months, and minorities). This lower power may seem surprising given that the direct test can exploit the joint distribution of (Y, ψ), but is simply due to the important reduction in sample size when focusing on the subsample of individuals who are followed over four months results.

There are also important issues associated with the direct test, which generally warrant caution when interpreting the results from this test. Most importantly, as already discussed in Section 2.2.4, the direct test is not robust to measurement errors on the subjective beliefs ψ. As shown in Proposition SS3 in the Web Appendix, it is however possible to derive a restriction on β under RE. Specifically, if ξ_ψ is positively correlated with $\varepsilon + \xi_Y$, we have, under RE,

$$\beta \geq 1 - \frac{1}{1 + \underline{\lambda}}, \tag{3}$$

where $\underline{\lambda}$ is a lower bound on the signal-to-noise ratio $\mathbb{V}(\psi)/\mathbb{V}(\xi_\psi)$. Table 3 also reports the results of tests combining (3) with the restrictions on the marginal distributions used in our full RE test. Adding the restriction (3) does not change the results for values of signal-to-noise ratio between 5 and 20 (i.e., for noise-to-signal ratios between 5% and 20%). Overall, using the subsample of linked data (Y, ψ) through this additional restriction does not add much to our test, at least once we account for possible measurement errors on the elicited beliefs. Another significant concern with the direct test, and, more generally, the use of linked data on (Y, ψ), is that attrition may be endogenous. We discuss this issue in more details in Appendix 6.2.

Table 3: Direct test, our test, and combined test of RE on annual earnings

	β	Direct test	Full RE	Combined test		Number of obs.		
Bound on signal/noise $\underline{\lambda}$				5	20			
Implied bound on β				0.833	0.952			
		(p-val)	(p-val)	(p-value)	(p-value)	ψ	Y	(ψ, Y)
All	0.954	0.001	< 0.001	< 0.001	< 0.001	1,565	1,428	768
Women	0.956	0.002	< 0.001	< 0.001	< 0.001	730	649	356
Men	0.960	0.021	0.210	0.276	0.276	835	779	412
White	0.963	0.004	0.021	0.019	0.010	1,200	1,097	596
Minorities	0.928	0.010	0.006	0.007	0.005	365	331	172
College degree	0.974	0.060	0.130	0.182	0.182	1,106	1,053	560
No college degree	0.954	0.044	0.013	0.017	0.017	459	375	208
High numeracy	0.959	0.001	0.012	0.016	0.016	1,158	1,070	573
Low numeracy	0.954	0.094	0.022	0.030	0.030	407	358	195
Tenure $\leq$ 6 months	0.942	0.015	0.001	0.002	0.001	271	180	98
Tenure $>$ 6 months	0.956	0.001	0.091	0.094	0.094	1,294	1,248	670

Notes: "Direct test" denotes the direct test of RE when (ψ, Y) is observed. β 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. "Combined RE test" denotes the test without covariates, where we test H_{0S} with $q(y, c) = y/c$, which is the "Full RE" test, combined with the additional restriction $\beta \geq 1 - 1/(1 + \underline{\lambda})$, where $\underline{\lambda}$ is an a priori bound on the signal-to-noise ratio. Distributions of realized earnings (Y) and earnings beliefs (ψ) are both Winsorized at the 95% quantile.

Coming back to our test, the rejection of RE for the overall population but also for most of the subpopulations are, in view of Proposition 4, unlikely to be due to data quality issues. In that sense, these results may be seen as robust evidence against the RE hypothesis for individual earnings, at least in this context. As a result, conclusions of behavioral models based on the assumption that agents form rational expectations about their future earnings may be misleading. Exploring this important question requires one to go beyond testing though, by quantifying the extent to which model predictions are actually sensitive to the violations from rational expectations that have been detected with our test. We investigate this issue in D'Haultfœuille et al. (2018b) in the context of a life-cycle consumption model.

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. As such, it can be applied in frequent cases where realizations and beliefs are observed in two separate datasets, or only observed for a selected sub-population. By bypassing the need to link beliefs to future realizations, our approach also enables to test for rational expectations without having to wait until the outcomes of interest are realized and made available to researchers. 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. 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.

Beyond testing, in this application as in any other situations where rational expectations are violated, a natural next step is to evaluate the deviations from rational expectations that one can rationalize from the available data. In the context of structural analysis, a central question then becomes to which extent the main predictions of the model are sensitive to those departures from rational expectations. We explore this important issue and propose in D'Haultfœuille et al. (2018*b*) a tractable sensitivity analysis framework on the assumed form of expectations.

References

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

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.

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.

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

Biroli, P., Boneva, T., Raja, A. and Rauh, C. (2020), 'Parental beliefs about returns to child health investments', *Journal of Econometrics* **Forthcoming**.

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

Boneva, T. and Rauh, C. (2018), 'Parental beliefs about returns to educational investments - the later the better?', *Journal of the European Economic Association* **16**(6), 1669–1711.

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

Buchinsky, M., Li, F. and Liao, Z. (2019), 'Estimation and inference of semiparametric models using data from several sources', *Journal of Econometrics* **Forthcoming**.

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.

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.

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

D'Haultfœuille, X., Gaillac, C. and Maurel, A. (2018*b*), Rationalizing rational expectations? Tests and deviations. NBER Working Paper No. 25274.

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. (2020), 'What would you do with $500? Spending responses to gains, losses, news and loans', *Review of Economic Studies* **Forthcoming**.

Gennaioli, N., Ma, Y. and Shleifer, A. (2016), 'Expectations and investment', *NBER Macroeconomics Annual* **30**(1), 379–431.

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.

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.

Kuchler, T. and Zafar, B. (2019), 'Personal experiences and expectations about aggregate outcomes', *Journal of Finance* **74**, 2491–2542.

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.

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.

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.

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.

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.

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 Proofs of the equivalence results

A.1 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,

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

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.

A.2 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.

A.3 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 $\text{Supp}(Y|\psi = u) \subset [y_0, \infty)$ or $\text{Supp}(Y|\psi = u) \subset (-\infty, y_0]$. Because $\mathbb{E}[Y|\psi] = \psi$, $\text{Supp}(Y|\psi = u) \subset [y_0, \infty)$ for almost all $u > y_0$ and $\text{Supp}(Y|\psi = u) \subset (-\infty, y_0]$ 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.

A.4 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 $\mathcal{I}' = \sigma\left(X', \psi'\right)$. Then $\psi' = \mathbb{E}(Y'|\psi', X) = \mathbb{E}(Y'|\mathcal{I}')$ and H_{0X} holds. The proposition then follows as Theorem 1.

A.5 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.