Rationalizing Rational Expectations? Tests and Deviations*

Preliminary

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

October 3, 2018

Abstract

In this paper, we build a new test of rational expectations based on the marginal distributions of realizations and subjective beliefs. This test can be used in many different empirical settings, including in the common situation where realizations and subjective 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 inequalities, for which tests have been developed recently in the literature. Next, we define and estimate the minimal deviations from rational expectations than can be rationalized by the data. In the context of structural models, we propose a natural and easy-to-implement way to conduct a sensitivity analysis on the assumed form of expectations. Finally, we use our method to test and quantify deviations from rational expectations about future earnings, and examine the consequences of violations of rational expectations in the context of a life-cycle model of consumption.

Keywords: rational expectations, test, data combination, subjective expectations, sensitivity analysis.

*We thank Peter Arcidiacono, Federico Bugni, Valentina Corradi, Gregory Jolivet, Jia Li, Andrew Patton, Yichong Zhang and seminar participants at Amsterdam, Bocconi, CREST, Duke, Helsinki, Mannheim, National University of Singapore, Singapore Management University, Surrey, Toulouse School of Economics, 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), and the 2018 CEME NSF-NBER conference on "Inference in Nonstandard Problems" (Duke) for useful comments and suggestions.

[†]CREST, xavier.dhaultfoeuille@ensae.fr.

[‡]CREST and TSE, christophe.gaillac@ensae.fr.

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

1 Introduction

Understanding how individuals form their beliefs about uncertain future outcomes is critical to understand decision making. Despite longstanding critiques (see, among many others Tversky and Kahneman, 1992; Manski, 2004), rational expectations remain by far the most popular theory to describe such 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) remain a key building block in many macroeconomic model, but also in most of the dynamic behavioral models that have been estimated over the last two decades (see, e.g., Aguirregabiria and Mira, 2010; Arcidiacono and Ellickson, 2011; Wolpin, 2013; 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. As such, it can be used to test RE in many different empirical settings, including in a data combination context, where individual realizations and subjective beliefs are observed in two different datasets that cannot be matched. Subjective expectations data have been increasingly used in economic research over the last ten years, frequently in situations where actual outcomes and beliefs are observed in two different datasets. The tests of RE implemented so far in this context (see in particular Patton and Timmermann, 2012; Gennaioli et al., 2015) only use specific implications of this hypothesis. In contrast, we develop a test that exploits all possible implications of RE. We show for that purpose that, if one moment equality and infinitely many moment inequalities hold, we can rationalize RE.[1] In other words, if these moment conditions hold, RE cannot be rejected, given the data at our disposal. By exhausting all 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, AS hereafter). By applying their results to our context, we show that the test controls size asymptotically and is consistent over fixed alternatives. We also provide conditions under which the test is not conservative.

We consider several important extensions to our baseline test. First, we show that by observing the same covariates in the two datasets, we can increase our ability to detect violations of the test. Another important issue is the one of unanticipated aggregate shocks. Even if individuals are rational, the mean of observed outcomes may differ from the mean of individual beliefs simply because of such shocks. We show that our test can be easily adapted to account for aggregate shocks. Finally, we show that our test is robust to measurement errors in the following sense. If individuals are rational but both beliefs and outcomes are measured with errors, then our test does not reject provided that the amount of measurement errors

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

on beliefs does not exceed the amount of intervening transitory shocks plus the measurement errors on the outcome variable. In particular, subjective beliefs may 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 the direct test based on a regression of the outcome on beliefs is not robust to any amount of measurement error on beliefs.

Next, we go beyond testing for rational expectations, and introduce the concept of minimal deviations from rational expectations than can be rationalized by the data. We leverage tools recently developed in the optimal transport literature (Villani, 2008) and in particular a recent paper by Gozlan et al. (2018), and provide conditions under which there is a unique transformation of subjective expectations that would make them rational, and which minimizes the transportation cost for all convex and positive loss functions considered. These minimal deviations have an intuitive interpretation as the minimal magnitude of measurement errors that would need to affect the elicited beliefs in order for these to remain compatible with the rational expectations hypothesis. Under some mild regularity conditions, we derive a consistent estimator for this transformation. Importantly for practical purposes, this estimator can be easily implemented, and at a minimal computational cost.

We then extend the concept of minimal deviations from rational expectations to accommodate restrictions on the information set of the agents. We establish existence and uniqueness of the corresponding minimal deviations, for which we derive a consistent and asymptotically normal estimator. In the context of structural models, 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 as a result can be used quite generally. Overall, this method opens up a middle ground between conducting inference on structural choice models based on realized data under the assumption of rational expectations (standard approach a la Rust, 1987; Keane and Wolpin, 1997), and estimating more flexible choice models using subjective beliefs (as in, e.g., Delavande and Zafar, 2018; Wiswall and Zafar, 2018, 2015; Kapor et al., 2017; Arcidiacono et al., 2014; Stinebrickner and Stinebrickner, 2014*b*,*a*; Delavande, 2008).

We apply our method to test and quantify deviations from 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 at any standard levels, our test does reject rational expectations at the 5% level. Taken together, these findings illustrate the practical importance of incorporating the additional restrictions of rational expectations that are embedded in our test.

Finally, we explore the sensitivity of the life-cycle model of earnings and consumption of Kaplan and Violante (2010) to violations of the rational expectations hypothesis. We find that, even though agents are about right on average about their future earnings, some of the findings exhibit significant sensitivity to departures from RE. Notably, we document the existence of substantial changes in the behavioral response of consumers to income shocks. In particular, consumption is less responsive to permanent income shocks if we relax RE. Interestingly, after accounting for these deviations from RE, behavioral responses tend to be more similar to what has been estimated in some of the earlier literature (see, e.g., resp. Blundell et al., 2008; Kaufmann and Pistaferri, 2009).

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 relatively recent literature on data combination (see, e.g., Cross and Manski, 2002, Molinari and Peski, 2006, Fan et al., 2014, Buchinsky et al., 2016, Pacini, 2017, and Ridder and Moffitt, 2007 for a survey), and the literature on testing for (implications of) rational expectations in a micro environment (see, e.g., Gourieroux and Pradel, 1986, and Ivaldi, 1992, for early methodological contributions). This paper also fits into the small but growing literature on the application of optimal transport methods in econometrics (see Galichon, 2016 for a recent overview of optimal transport methods in economics). In the context of our analysis, optimal transport theory offers a very natural and powerful way to quantify deviations from rational expectations.

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, 2014, Van der Klaauw and Wolpin, 2008, Zafar, 2011, Van der Klaauw, 2012, Arcidiacono et al., 2012, 2014, de Paula et al., 2014, Stinebrickner and Stinebrickner, 2014*b*, and Wiswall and Zafar, 2015, 2018). 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.

Our analysis also complements several recent studies which primarily focus on testing for different information sets, while maintaining the rational expectations assumption (see, for a survey, Cunha and Heckman, 2007, and recent articles by Navarro and Zhou, 2017, and Dickstein and Morales, 2018). Unlike these papers, we focus instead on testing for and relaxing rational expectations, while treating the agents' information sets as an infinite dimensional nuisance parameter.

Finally, by developing a new framework allowing to examine the sensitivity of behavioral models to departures from the rational expectations hypothesis, we contribute to a small set of recent papers that estimate structural choice models without imposing rational expectations (see, e.g., Houser et al., 2004; Buchinsky and Leslie, 2010; Stinebrickner and Stinebrickner, 2014*a*; Hoffman and Burks, 2017; Kapor et al., 2017; and Agarwal and Somaini, 2018).

The remainder of the paper is organized as follows. In Section 2, we present the set-up and discuss the main theoretical equivalences that we use to build our testing procedure. In Section 3, we present the statistical tests for rational expectations, and establish their asymptotic properties. Section 4 studies minimal deviations from rational expectations that can be rationalized by the data. Section 5 illustrates the finite sample properties of our tests and estimators through Monte Carlo simulations. Section 6 applies our framework to expectations about future earnings. Finally, Section 7 concludes. The appendix collects various theoretical extensions, additional simulation results, additional material on the application and all the proofs.

2 Set-up and main theoretical equivalences

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, the elicited individual expectation on Y, denoted by ψ. Throughout the paper, 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 use in this paper has been collected in various contexts and used in a number of prior studies (see, among others Delavande, 2008; Zafar, 2011; Arcidiacono et al., 2012, 2014; Armantier et al., 2017; Kuchler and Zafar, 2017; Hoffman and Burks, 2017; Landier et al., 2017).

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). Importantly, we remain agnostic throughout most of our analysis on the information set $\mathcal{I}$.[2] 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. It is easy to see that the RE hypothesis imposes restrictions on the joint distribution of realizations Y and beliefs ψ.

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

where $\sim$ denotes equality in distribution. Rationalizing RE does not mean that the true Y, ψ and $\mathcal{I}$ are such that $\mathbb{E}[Y|\mathcal{I}] = \psi$. Instead, it means that there exists a triplet $(Y', \psi', \mathcal{I}')$

[2] We nevertheless do need to impose stronger restrictions when discussing the role of covariates in Section 2.2.2. For such covariates to be useful, we have to assume that $\sigma(X) \subset \mathcal{I}$ (taking the convention that "$\subset$" includes the case of equality between the two sets). We also accommodate restrictions on the information set of the agents when we discuss in Subsection 4.2 the estimation of the minimal deviations from RE in the context of a behavioral model.

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 Y, subjective beliefs ψ, and information set $\mathcal{I}$ do not satisfy RE ($\mathbb{E}[Y|\mathcal{I}] \neq \psi$) but the converse 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 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, then, 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) ⇒ (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 tests of rational expectations We compare hereafter our test with alternative ones that have been proposed in the literature. In the following discussion and throughout this section, we reason at the population level. Accordingly, we compare the different tests in terms of data generating processes violating the null hypothesis associated with each test.

Our test can clearly detect many more violations of rational expectations than the "naive" test of rational expectations simply based on the equality $\mathbb{E}(Y) = \mathbb{E}(\psi)$. It also detects more violations than a test based on $\mathbb{E}(Y) = \mathbb{E}(\psi)$ and $\mathbb{V}(Y) \geq \mathbb{V}(\psi)$, which has been considered in the macroeconomic literature on the accuracy and rationality of forecasts (see in particular Patton and Timmermann, 2012).[3] On the other hand, and as expected since it is based on the joint distribution of (Y, ψ), the "direct" test of $\mathbb{E}(Y|\psi) = \psi$ can detect more violations of rational expectations than ours.

To better understand the difference between these four different tests, 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 tests reject their null hypothesis. 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 tests are rejected.

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

Specifically, if $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ and $\eta \sim \mathcal{N}(0, \sigma_\eta^2)$, our test rejects if and only if $\sigma_\eta^2 > \sigma_\varepsilon^2$. In such a case, our test boils down to the second test mentioned above: we reject whenever $\mathbb{V}(\psi) > \mathbb{V}(Y)$. But interestingly, if η 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 test rejects when $a \geq 1.755$, while for $a = 1.755$, $\mathbb{V}(\eta) \simeq 0.616 \leq \mathbb{V}(\varepsilon) = 1$.

[3] We also refer the reader to Elliott et al. (2005), Jin et al. (2017) and references therein, for other recent contributions to the literature on the accuracy and rationality of forecasts. The framework in this literature differs however from ours in several aspects, and in particular by focusing on the evolution over time of forecasts.

Finally, in a particular case, our test reduces to the naive test of $\mathbb{E}(Y) = \mathbb{E}(\psi)$. When Y is a binary outcome and $\psi \in [0, 1]$, one can easily show that as long as $\delta = 0$, 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. This also applies to situations where expectations about the distribution of continuous outcomes 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. We refer the reader to Manski (2004) and Delavande (2014) for discussions of papers analyzing this type of probabilistic expectations data. In such cases, one can apply our analysis after replacing, for the different values y at which the subjective beliefs were elicited, Y by $\mathbb{1}\{Y > y\}$, and defining ψ as the subjective survival function evaluated at y.

Interpretation of the boundary condition Finally, to provide a deeper understanding of our test and of 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, i.e. $\psi \in \{y : \Delta(y) = 0\}$, have perfect foresight, i.e. $\psi = \mathbb{E}[Y|\mathcal{I}] = Y$. For any cdf F, we let below F^{-1} denote its quantile function, namely $F^{-1}(\tau) = \inf\{x : F(x) \geq \tau\}$.

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:[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 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;

[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.

(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

There may be cases where the restriction $\mathbb{E}[Y|\psi] = \psi$ (or, in the presence of covariates, $\mathbb{E}[Y|\psi, X] = \psi$) is too strong, in the sense that such a restriction may be violated, even though the rational expectations hypothesis ($\psi = \mathbb{E}[Y|\mathcal{I}]$) holds. This occurs in particular in situations where the outcome Y is affected by unpredictable, aggregate shocks. While these types of shocks arise in a variety of contexts, we consider in the following the particular case of individual income.

Suppose that the logarithm of income of individual i at period t, denoted by y_{it}, satisfies a Restricted Income Profile (MaCurdy, 1982) model:

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

where β_t capture aggregate (macro) shocks, ε_{it} is distributed following 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 $c_t = \beta_t - \mathbb{E}\left[\beta_t|(\beta_{t-k})_{k\geq 1}\right]$. It follows that, under the previous assumptions and although individuals form rational expectations, $\mathbb{E}\left[y_{it}|\psi_{it}\right] = \psi_{it} + c_t \neq \psi_{it}$.[5] In this setting it is therefore natural to test instead for $\mathbb{E}[y|\psi] = c_0 + \psi$, for some $c_0 \in \mathbb{R}$.

A similar reasoning applies to the case of multiplicative aggregate shocks. In those cases, we would like to test for $\mathbb{E}(y|\psi) = c_0\psi$, for some $c_0 > 0$. Note that 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. Generalizing these examples, we consider the following null hypothesis for testing RE in the presence of aggregate shocks:

$$\mathrm{H}_{0S}: \exists\left(Y', \psi', \mathcal{I}'\right):\ \sigma(\psi') \subset \mathcal{I}',\ Y' \sim Y,\ \psi' \sim \psi \text{ and } \mathbb{E}\left[q\left(Y', c_0\right) \middle| \mathcal{I}'\right] = \psi'.$$

where $q(.,.)$ a known function such that $q(y,.)$ is strictly monotonic. In the two previous cases of additive and multiplicative aggregate shocks, we have respectively $q(y,c) = y - c$ and $q(y,c) = y/c$. Theorem 1 then implies directly the following result. Similarly to $\delta = \mathbb{E}(Y) - \mathbb{E}(\psi)$, we define hereafter $\delta_c = \mathbb{E}(q(Y,c)) - \mathbb{E}(\psi)$.

[5] Note that, since c_t is common to all individuals, the expectation we are considering here is implicitly conditional on c_t.

Assumption 2 *$q(y,.)$ is strictly monotonic, and for all c, $\mathbb{E}\left(\left|q\left(Y,c\right)\right|\right)<+\infty$ and $\mathbb{E}\left(\left|\psi\right|\right)<+\infty$.*

Proposition 3 *Suppose that Assumption 2 holds. Then c_0 is identified as the unique c satisfying $E\left[q(Y,c)\right]=E[\psi]$. Moreover, the following statements are equivalent:*

(i) H_{0S} holds;

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

A couple of remarks are in order. First, in the two particular cases above, the condition $\delta_{c_0}=0$ is satisfied by construction of c_0. It follows that, in those cases, testing for H_{0S} is equivalent to testing for the moment inequalities $\mathbb{E}\left[\left(y-q\left(Y,c_0\right)\right)^+-\left(y-\psi\right)^+\right]\geq 0$. Second, 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 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 et al., 2001). We explore in the following the extent to which our test is robust to such 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. 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, so that $\mathbb{E}\left[Y|\psi\right]=\psi$, 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 ψ. Even if individuals are rational, the direct test will reject in the presence of even a small degree of measurement errors on the beliefs ψ. The following proposition shows that our test, on the other hand, is robust to a certain degree of measurement errors on ψ. As above, we let $\varepsilon=Y-\psi$ denote the uncertainty shocks.

Proposition 4 *Suppose that $\mathbb{E}\left[Y|\psi\right]=\psi$ 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 $\widehat{\psi}$ and $\widehat{Y}$ 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 $\xi_Y=0$, the 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 subjective beliefs to be noisier than realized outcomes, a setting which may be relevant in practice. Taken together, 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.[6]

2.2.5 Other extensions

We briefly discuss here other directions in which Theorem 1 can be extended. Another 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 ψ (see, e.g., Manski and Molinari, 2010) We show in Appendix B that our test can be generalized to accommodate this rounding practice.

Finally, we have assumed implicitly so far that the two samples are drawn from the same population. In Appendix C, we relax this assumption and show how to allow for sample selection under unconfoundedness, by using an appropriate reweighting of the observations.

3 Statistical tests

In this section we propose a testing procedure for H_{0X}, which can be easily adapted to the case where no covariate 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). 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:

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

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

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

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 AS, who consider the issue of testing many conditional moment inequalities and equalities.[7] The initial step is to transform the

[6] Clearly, the naive test $\mathbb{E}(Y) = \mathbb{E}(\psi)$ is even more robust to measurement errors. The test will never reject its null hypothesis under any kind of measurement errors, provided that they satisfy $\mathbb{E}(\xi_\psi) = \mathbb{E}(\xi_Y) = 0$. On the other hand, such a test is not robust to aggregate shocks. It is also unable to reject rational expectations in several cases of interest, as discussed above.

[7] Other testing procedures could be used to implement our test, such as those proposed by Linton et al. (2010) and Chernozhukov et al. (2014).

conditional moments into the following unconditional moments conditions:

$$\mathbb{E}\left[W\left(y-\widetilde{Y}\right)^{+}h_1(X)\right]\geq 0,\quad \mathbb{E}\left[\left(Y-\psi\right)h_2(X)\right]=0.$$

for all $y\in\mathbb{R}$, and (h_1,h_2) in a suitable class of functions.

We suppose to observe here a sample $(D_i,X_i,\widetilde{Y}_i)_{i=1\dots n}$ of n i.i.d. copies of $(D,X,\widetilde{Y})$. For notational convenience, we let $\widetilde{X}_i$ denote the nontransformed vector of covariates and redefine X_i as the transformed vector in the following way:

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

where $\Phi_0(x)=(\Phi(x_1),\dots,\Phi\left(x_{d_X}\right))$. 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.

Now that $X_i\in[0,1]^{d_X}$, we consider for h_1 and h_2 indicators of belonging to specific hypercubes within $[0,1]^{d_X}$. Namely, we consider the class of functions $\mathcal{H}_r=\{h_{a,r},\ a\in A_r\}$, with $h_{a,r}(x)=\mathbb{1}\{x\in C_{a,r}\}$ and

$$\mathcal{C}_r=\left\{C_{a,r}:=\prod_{u=1}^{d_X}\left(\frac{a_u-1}{2r},\frac{a_u}{2r}\right]\in[0,1]^{d_X},\, a=(a_1,...,a_{d_X})^{\top}\in A_r\equiv\{1,2,\dots,2r\}^{d_X}\right\}.$$

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

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

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

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

where $\widehat{\Sigma}_n(h,y)$ is the sample covariance matrix of $\sqrt{n}\overline{m}_n\left(h,y\right)$ and $\widehat{\mathbb{V}}\left(\widetilde{Y}\right)$ is the empirical variance of $\widetilde{Y}$.[8] We then denote by $\overline{\Sigma}_{n,jj}(j=1,2)$ the j-th diagonal term of $\overline{\Sigma}_n$.

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

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

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

[8]As in AS, we fix ϵ to 0.05.

Note that to test for rational expectations in the absence of covariates, one can use the formulation (iii) of H_0 given in Proposition 1, and simply omit the transformation of conditional moments to unconditional ones in the derivation above.

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 the Generalized Moment Selection (GMS) method (see Andrews and Soares, 2010; Andrews and Shi, 2017). Specifically, we follow these three steps:

1. Compute the GMS function $\overline{\varphi}_n(y,h) = \left(\overline{\varphi}_{n,1}(y,h), 0\right)^\top$ for (y,h) in $\widehat{\mathcal{Y}} \times \cup_{r=1}^{r_n} \mathcal{H}_r$ with

$$\overline{\varphi}_{n,1}(y,h) = \overline{\Sigma}_{n,11}^{1/2} B_n \mathbb{1}\left\{\frac{n^{1/2}}{\kappa_n} \overline{\Sigma}_{n,11}^{-1/2} \overline{m}_{n,1}(y,h) > 1\right\},$$

where $B_n = b_0 \left(\ln(n)/\ln(\ln(n))\right)^{1/2}$, $b_0 > 0$, and $\kappa_n = (0.3\ln(n))^{1/2}$.

2. Let $(D_i^*, \widetilde{Y}_i^*, X_i^*)_{i=1,\dots,n}$ 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 $\overline{m}_n^*$ and $\overline{\Sigma}_n^*$. Then compute T^* like T, replacing $\overline{\Sigma}_n(y, h_{a,r})$ and $\sqrt{n}\overline{m}_n(y, h_{a,r})$ by $\overline{\Sigma}_n^*(y, h_{a,r})$ and

$$\sqrt{n}\left(\overline{m}_n^* - \overline{m}_n\right)(y, h_{a,r}) + \overline{\varphi}_n(y, h_{a,r}).$$

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

We now turn to the asymptotic properties of the test. For that purpose, it is convenient to introduce additional notation. Let $\mathcal{Y}$ and $\mathcal{X}$ denote the support of Y and X respectively, and

$$\mathcal{L}_F = \left\{(y, h_{a,r}) : y \in \mathcal{Y},\ (a,r) \in A_r \times \mathbb{N} :\ \mathbb{E}_F\left[W\left(y - \widetilde{Y}\right)^+ h_{a,r}(X)\right] = 0\right\},$$

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 $\mathcal{F}_0$ the subset of $\mathcal{F}$ such that H_0 holds. We impose the following conditions on $\mathcal{F}$ and $\mathcal{F}_0$:

Assumption 3

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

(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* $0 < \inf_{F\in\mathcal{F}} \mathbb{E}_F[D] \le \sup_{F\in\mathcal{F}} \mathbb{E}_F[D] < 1$;

(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* $\mathcal{Y} \times \cup_{r\ge1}\mathcal{H}_r$ *with uniform metric* d *defined by*

$$d(K, K') = \sup_{(y,h,y',h') \in \left(\mathcal{Y} \times \cup_{r\ge1} H_r\right)^2} \left\| K(y,h,y',h') - K'(y,h,y',h') \right\|.$$

The main result of this section is Theorem 2, which 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, h_0) in $\{1,2\} \times \mathcal{L}_{F_0}$ such that $K_{F_0,jj}(y_0, h_0, y_0, h_0) > 0$, 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, the main difference being that, in our case, one needs to account for the fact that the proportions $\mathbb{E}[D_i]$ and $\mathbb{E}[1 - D_i]$ are estimated. 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. 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. Finally, Theorem 2 (iii), which is based on Theorem 6.1 in AS, shows that the test is consistent over fixed alternatives.

Extensions 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 $\widetilde{Y}_{\widehat{c}}$, where $\widehat{c}$ denotes a consistent estimator of c_0. The resulting test is given by $\varphi_{n,\alpha,\widehat{c}} = \mathbb{1}\left\{T(\widehat{c}) > c^*_{n,\alpha}\right\}$. 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 example of additive shocks ($q(y, c) = y - c$). We refer the reader to Appendix A for a detailed discussion of this extension.

This testing procedure can also be modified to handle the case where both Y_i and ψ_i are observed for a subset of the population. In such a case, let $\widetilde{D}$ denote the dummy variable equal to one if we observe (Y, ψ), and suppose that conditional on X, $\widetilde{D}$ is independent of (Y, ψ). Then, under rational expectations, the following conditional moment condition also holds:

$$\mathbb{E}\left(\widetilde{D}(Y - \psi)\psi \middle| X\right) = 0. \tag{3}$$

In practice, this only requires augmenting m_2 defined in (2) with this additional moment function. The rest of the procedure remains identical.[9]

[9] By adding (3) to the set of conditional moment equalities, we consider, as is common in the literature (see, e.g., Arcidiacono et al., 2017), the implication of rational expectations that the slope of the (conditional) linear regression of Y on ψ is equal to one. But other implications, such as $\mathbb{E}\left(\widetilde{D}(Y - \psi)q(\psi)\middle| X\right) = 0$ for any $q(.)$ such that $\mathbb{E}[|(Y - \psi)q(\psi)|] < +\infty$, could be easily added as well.

4 Minimal deviations from rational expectations

In this section we introduce the concept of minimal deviations rom rational expectations, and build on optimal transport methods to provide conditions under which these minimal deviations exist and are unique. We first consider in Section 4.1 such deviations while remaining agnostic on the information set of the agents. Then, in Section 4.2, we characterize such deviations when the information set is known, as is typically the case in behavioral models. We show therein how these deviations can be used to assess the sensitivity of structural models to violations of rational expectations.

4.1 Unconstrained information set

4.1.1 Existence and uniqueness

For the cases where H_0 is rejected, we propose a way to quantify the degree to which subjective expectations differ from rational expectations. To do so, we consider the minimal modifications - in a sense to be made precise below - to the distribution of subjective beliefs ψ that are such that the modified distribution of beliefs is compatible with the rational expectations hypothesis. We refer to these as the minimal deviations from rational expectations. In the same spirit as in Section 2, we first consider such deviations without imposing any constraints on the information set of the agents.

Formally, let us define the set:

$$\mathbf{\Psi} = \left\{(Y', \psi', \psi'') : \; Y' \sim Y, \; \psi' \sim \psi \text{ and } \mathbb{E}(Y'|\psi'') = \psi''\right\}. \tag{4}$$

In this set, (Y', ψ') corresponds to a vector that is compatible with the data, whereas ψ'' correspond to possible individual expectations, in a counterfactual situation where people would form rational expectations on their future outcomes. Thus, the subset of $\mathbf{\Psi}$ for which $\psi' = \psi''$ corresponds to the set of random variables (Y', ψ') that are compatible with the data and with the rational expectations hypothesis. However, if H_0 does not hold - which is the relevant situation here - such a subset is, by definition, empty. The idea is then to try and find the vector $(Y', \psi', \psi'') \in \mathbf{\Psi}$ such that ψ' and ψ'' are closest, in the sense of a family of metrics defined below.

The following theorem shows that there exists a solution to this problem. Importantly, this solution is, for a large class of metrics, independent of the specific metric considered. This solution is also unique.

Assumption 4 *$\mathbb{E}(Y^2) < +\infty$, $\mathbb{E}(\psi^2) < +\infty$, and F_ψ has no atom.*

Theorem 3 *Suppose that Assumption 4 holds. Then there exists a unique function g^* such that:*
(i) $g^*(\psi)$ is consistent with RE (namely, there exists Y' such that $(Y', \psi, g^*(\psi)) \in \Psi$);

(ii) for any convex function $\rho : \mathbb{R}^+ \to \mathbb{R}^+$ satisfying $\rho(0) = 0$,

$$\mathbb{E}[\rho(|\psi - g^*(\psi)|)] = \inf_{(Y',\psi',\psi'')\in\mathbf{\Psi}} \mathbb{E}[\rho(|\psi' - \psi''|)]. \tag{5}$$

Moreover, g^* is non-decreasing.

Theorem 3 shows that there exists a (unique) transformation of the subjective beliefs (ψ) that is such that i) the transformed beliefs $g^*(\psi)$ are consistent with RE, and, remarkably, ii) this transformation is minimal *for all* metrics (indexed by ρ) used to measure the distance between the true and modified beliefs distributions. Moreover, under this minimal modification, the modified beliefs are obtained as a monotonically increasing change of the original beliefs. These minimal modifications can be geometrically interpreted as the projections, in the sense of the family of metrics defined in Equation 5, from the set of true beliefs to the set of beliefs that are consistent with RE.[10]

The proof of Theorem 3 can be summarized as follows. We first show, using in particular Theorem 1 in our paper and Proposition 3.1 in Gozlan et al. (2018), that

$$\inf_{(Y',\psi',\psi'')\in\mathbf{\Psi}} \mathbb{E}[\rho(|\psi' - \psi''|)] = \inf_{(Y',\psi'):\, Y'\sim Y,\, \psi'\sim\psi} \mathbb{E}\left[\rho\left(\left|\psi' - \mathbb{E}\left[Y'|\psi'\right]\right|\right)\right].$$

The optimization problem on the right-hand side is an optimal transport problem, in the sense that it corresponds to an optimization over probability measures whose marginals are fixed. Though non-standard, as it involves $\mathbb{E}\left[Y'|\psi'\right]$, this problem has been recently studied by Gozlan et al. (2018). In particular, it follows from their results that there exists a cdf G^* such that

$$\inf_{(Y',\psi',\psi'')\in\mathbf{\Psi}} \mathbb{E}[\rho(|\psi' - \psi''|)] = \inf_{(\psi',\psi''):\, \psi'\sim\psi,\, \psi''\sim G^*} \mathbb{E}\left[\rho\left(\left|\psi' - \psi''\right|\right)\right].$$

Then, by a strict convexity argument based on Theorem 1 again and Pass (2013), we show that such a G^* is unique. Finally, using standard results in optimal transport, we show that $g^* = G^{*-1} \circ F_\psi$ is the unique function satisfying (5).

4.1.2 Consistent estimation

Though g^* does not have a simple form in general, we propose in the following a simple procedure to construct a consistent estimator of it, based on i.i.d. copies $(Y_i)_{i=1\dots L}$ and $(\psi_i)_{i=1\dots L}$ of Y and ψ. For simplicity, we suppose hereafter that the two samples have equal size.[11]

[10] Another interpretation of $g^*(\psi)$ is the following. One may wonder how large measurement errors on the subjective beliefs would need to be in order for the measurement-error free beliefs to remain consistent with RE. Assume that, instead of observing measurement error-free beliefs $\widetilde{\psi}$, we only observe $\psi = \widetilde{\psi} + \nu$, where ν denotes the measurement error. Then Theorem 3 ensures that $\psi - g^*(\psi)$ corresponds to the minimal measurement error ν in the L^p sense (for any $p \geq 1$) that are needed to rationalize RE.

[11] If both samples do not have equal size, one can first apply our analysis after taking a random subsample of the larger one, with the same size as the smaller one. Then we can compute the average of the estimates over a large number of such random subsamples.

To define our estimator, note first that from the proof of Theorem 3, we have

$$g^* = \arg\min_{g\in\mathcal{G}_0} \mathbb{E}\left[(\psi - g(\psi))^2\right], \tag{6}$$

where the set $\mathcal{G}_0$ is defined by

$$\mathcal{G}_0 = \left\{g \text{ non-decreasing} : \mathbb{E}\left[(y-Y)^+ - (y-g(\psi))^+\right] \geq 0\ \forall y \in \mathbb{R},\ E[g(\psi)] = E[Y]\right\}.$$

In other words, g^* is the (increasing) function such that (i) $g^*(\psi)$ is closest to ψ for the L^2 norm; (ii) g^* belongs to $\mathcal{G}_0$, which means by Theorem 1 that we can rationalize $\mathbb{E}(Y|g^*(\psi)) = g^*(\psi)$.

To estimate g^*, we basically replace expectations and cdfs by their empirical counterpart, in (6) and in $\mathcal{G}_0$. Let us denote by $(Y_{(i)})_{i=1...L}$ and $(\psi_{(i)})_{i=1...L}$ the ordered statistics of $(Y_i)_{i=1...L}$ and $(\psi_i)_{i=1...L}$. We first focus on the estimation of $\left(g^*(\psi_{(1)}), ..., g^*(\psi_{(L)})\right)$. The empirical counterpart $\widehat{\mathcal{G}}_0$ of $\mathcal{G}_0$ is

$$\widehat{\mathcal{G}}_0 = \Bigg\{\left(\widetilde{\psi}_{(1)}, ..., \widetilde{\psi}_{(L)}\right) : \widetilde{\psi}_{(1)} < ... < \widetilde{\psi}_{(L)},\ \sum_{i=1}^{L}(y - Y_{(i)})^+ - \left(y - \widetilde{\psi}_{(i)}\right)^+ \geq 0\ \forall y \in \mathbb{R},$$
$$\sum_{i=1}^{L} Y_{(i)} - \widetilde{\psi}_{(i)} = 0\Bigg\}. \tag{7}$$

Here we consider vectors $\left(\widetilde{\psi}_{(1)}, ..., \widetilde{\psi}_{(L)}\right)$ instead of functions g as in $\mathcal{G}$, since g may be assimilated with a vector when ψ has a finite support. On the surface, the set $\widehat{\mathcal{G}}_0$ appears to be complicated because of the infinitely many inequalities. However, one can show (see, e.g., Proposition 2.6 in Gozlan et al., 2018) that $\widehat{\mathcal{G}}_0$ actually boils down to the following set, which only involves a finite number of inequalities:

$$\widehat{\mathcal{G}}_0 = \left\{\left(\widetilde{\psi}_{(1)}, ..., \widetilde{\psi}_{(L)}\right) : \widetilde{\psi}_{(1)} < ... < \widetilde{\psi}_{(L)},\ \sum_{i=j}^{L} Y_{(i)} - \psi_{(i)} \geq 0\ j = 2, ..., L,\ \sum_{i=1}^{L} Y_{(i)} - \widetilde{\psi}_{(i)} = 0\right\}.$$

Our estimator of $\left(g^*(\psi_{(1)}), ..., g^*(\psi_{(L)})\right)$ is the empirical counterpart of (6), which is the solution of a convex quadratic programming problem:

$$\left(\widehat{g}^*(\psi_{(1)}), ..., \widehat{g}^*(\psi_{(L)})\right) = \arg\min_{\widetilde{\psi}_{(1)} < \cdots < \widetilde{\psi}_{(L)}} \sum_{i=1}^{L}\left(\psi_{(i)} - \widetilde{\psi}_{(i)}\right)^2 \text{ s.t.} \sum_{i=j}^{L} Y_{(i)} - \widetilde{\psi}_{(i)} \geq 0,\ j = 2...L,$$
$$\sum_{i=1}^{L} Y_{(i)} - \widetilde{\psi}_{(i)} = 0. \tag{8}$$

Finally, for any $t \in \mathbb{R}$, we let

$$\widehat{g}^*(t) = \widehat{g}^*\left(\min\{(\psi_i)_{i=1...L} : \psi_i \geq \min\{t, \psi_{(L)}\}\right).$$

Theorem 4 shows that $\widehat{g}^*$ is consistent.

Theorem 4 (Convergence of empirical minimal deviations) *Suppose that Assumption 4 holds. Then, for all t that is a continuity point of g^* and such that $F_\psi(t) \in (0,1)$,*

$$\widehat{g}^*(t) \to g^*(t) \quad a.s.$$

Program (8) is a particular convex quadratic programming problem, which turns out to be solvable very efficiently. Indeed, the following algorithm, devised by Suehiro et al. (2012), shows that $\left(\widehat{g}^*(\psi_{(1)}), ..., \widehat{g}^*(\psi_{(L)})\right)$ can be obtained with only $O(L^2)$ elementary operations. This implies that g^* can be estimated simply and at a fairly low computational cost.

Computation of $\left(\widehat{g}^*(\psi_{(1)}), ..., \widehat{g}^*(\psi_{(L)})\right)$.

1. Let $t = 0$ and $i_0 = 0$.
2. While $i_t < L$:
   (a) Let $t = t + 1$.
   (b) Let $C^t(i) = \sum_{k=L+1-i}^{L-i_{t-1}} \left(Y_{(k)} - \psi_{(k)}\right)/(i - i_{t-1})$, for $i = i_{t-1}+1, ..., L$ and let $i_t = \text{argmin}_{i \in \{i_{t-1}+1,...,L\}} C^t(i)$. If there are multiple minimizers, choose the largest one as i_t.
   (c) Set $\widehat{g}^*(\psi_{(k)}) = \psi_{(k)} + C^t(i_t)$, for $k \in \{L+1-i_t, ..., L-i_{t-1}\}$.

The idea of the algorithm is to rely on the first-order conditions of the program, which have a simple form. To provide some intuition, we illustrate this algorithm with $L = 3$, $(Y_{(1)}, Y_{(2)}, Y_{(3)}) = (1, 2.5, 3)$, and $(\psi_{(1)}, \psi_{(2)}, \psi_{(3)}) = (0.75, 2.75, 2.95)$. The black curve in Figure 1 corresponds to $y \mapsto \sum_{i=1}^{L} (y - Y_i)^+ - (y - \psi_i)^+$. In view of (7), any negative value thus corresponds to violations of the constraints. Also, for y large, this function should be equal to 0, in view of the equality constraint on the means. In the first step ($t = 1$), the algorithm picks $i_1 = 2$, corresponding to the negative value $C_2 = (Y_{(2)} - \psi_{(2)})/2 = -0.125$. It then adds this value to $\psi_{(2)}$ and $\psi_{(3)}$. In the second step, it modifies the value of $\psi_{(1)}$, ensuring that the equality constraint on the means holds. The modification on $\psi_{(2)}$ and $\psi_{(3)}$ in the first step also ensures that the green curve is always positive.

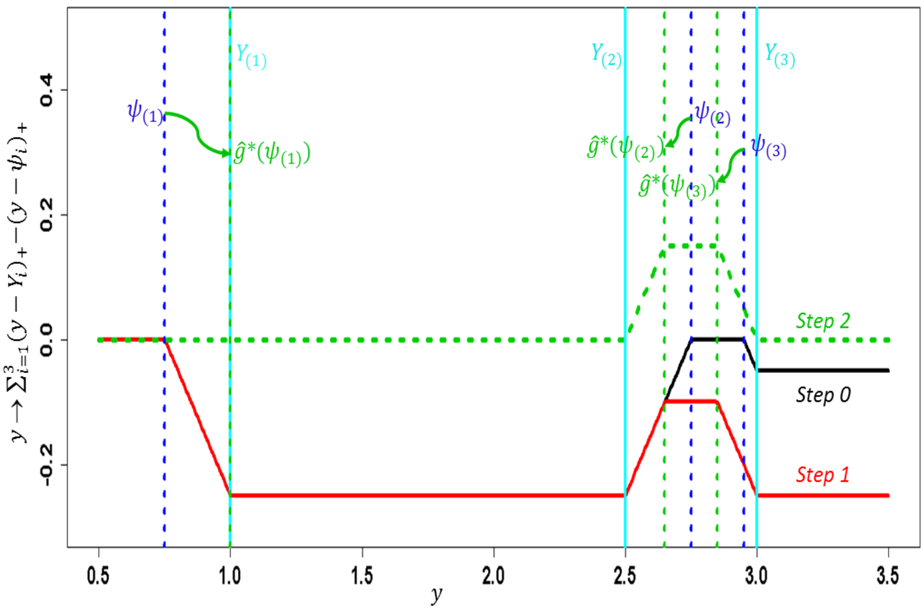


Figure 1: Illustration of the construction of $\left(\widehat{g}^*(\psi_{(1)}), ..., \widehat{g}^*(\psi_{(L)})\right)$.

4.2 Constrained information set and sensitivity analysis in structural models

4.2.1 Existence and uniqueness

We now consider minimal deviations from rational expectations in the presence of constraints on the information set. Such constraints are typically imposed in structural models, along with the rational expectations hypothesis. An important motivation for considering minimal deviations in this setting, then, is to assess the sensitivity of structural models to the RE hypothesis. A more direct way of evaluating how critical the rational expectations hypothesis is for a given model would be to solve it and estimate it, using elicited beliefs about future outcomes both on and off the agent's actual choice path. However, the data requirements are formidable, and, as a consequence, this approach has only been pursued in relatively few studies (see, e.g., Arcidiacono et al., 2014; Stinebrickner and Stinebrickner, 2014*a*,*b*; Wiswall and Zafar, 2015, 2018). We propose here an alternative approach that can be used in a less demanding setting, that is when the data used to estimate the structural model do not include elicited beliefs, but such beliefs are observed in an auxiliary dataset.

Specifically, consider a structural model that imposes both a rational expectation formation process and an information set $\mathcal{I}^M$ of the agents, such that individual rational expectations about the outcome Y are given by $\mathbb{E}\left[Y|\mathcal{I}^M\right]$. Hereafter, 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 ψ, we can test for the restricted RE by simply testing whether $F_\psi = F_{\mathbb{E}[Y|\mathcal{I}^M]}$.

Suppose that such a test is rejected. Then, consider the set

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

As with the set $\mathbf{\Psi}$ in the unconstrained case, if the test above is rejected, there is no couple of the form (ψ', ψ') in $\mathbf{\Psi}^M$. The goal here is then to find a pair $(\psi', \psi'') \in \mathbf{\Psi}^M$ such that ψ' is as close to ψ'' as possible. Such a ψ' corresponds to the minimal deviations from the restricted RE that are consistent with the data on subjective beliefs. Similarly to Theorem 3 in the absence of constraints on the information set, Theorem 5 shows that there exists a solution to this problem, which is moreover independent of the metric. To define this solution, we introduce $h^M = F_\psi^{-1} \circ F_{\mathbb{E}[Y|\mathcal{I}^M]}$.

Theorem 5 *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*

$$\left(h^M\left(\mathbb{E}\left[Y\middle|\mathcal{I}^M\right]\right), \mathbb{E}\left[Y\middle|\mathcal{I}^M\right]\right) \in \arg\min_{(\psi',\psi'')\in\mathbf{\Psi}^M} \mathbb{E}[\rho(|\psi' - \psi''|)]. \tag{9}$$

Moreover, if ρ is strictly convex, $h^M\left(\mathbb{E}\left[Y\middle|\mathcal{I}^M\right]\right)$ is unique in the sense that for any other ψ' such that $(\psi', \mathbb{E}\left[Y\middle|\mathcal{I}^M\right]) \in \mathbf{\Psi}^M$ satisfying (9), $\psi' = h^M\left(\mathbb{E}\left[Y\middle|\mathcal{I}^M\right]\right)$ almost surely.

Theorem 5 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 5 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 here the minimum 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 simply obtained by an equipercentile mapping from the distribution of rational expectations to the distribution of the true subjective beliefs. It follows that the minimal deviations can be easily estimated, as we will discuss in more detail below. Second, $h^M\left(\mathbb{E}\left[Y\middle|\mathcal{I}^M\right]\right)$ is also $\mathcal{I}^M$-measurable, which implies that it is compatible with the information set $\mathcal{I}^M$ imposed by the model. Finally, 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 $\psi = h^M\left(\mathbb{E}\left[Y\middle|\mathcal{I}^M\right]\right)$.[12] 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 $h^M\left(\mathbb{E}\left[Y\middle|\mathcal{I}^M\right]\right)$ does not correspond in general to the true subjective expectations ψ. By construction, the pseudo-beliefs are identifiable.

Having computed the pseudo-beliefs for a given structural model, we can then compare the results obtained with these pseudo-beliefs with those obtained under the baseline RE model. Importantly, this 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. Findings from the baseline model that exhibit significant sensitivity to these minimal deviations should then be interpreted with caution.

[12] On the other hand, it is generally impossible to rationalize the model-free beliefs generated from g^*, namely $(g^*)^{-1}\left(\mathbb{E}\left[Y\middle|\mathcal{I}^M\right]\right)$. Their distribution does not coincide with F_ψ in general.

4.2.2 Consistent estimation

Estimation of h^M is simpler than that of g^*, given its simple, explicit form. We suppose in the following that $F_{\mathbb{E}[Y|\mathcal{I}^M]}$ is known. This includes cases where the parameters of the structural model are known, or calibrated. Alternatively, $F_{\mathbb{E}[Y|\mathcal{I}^M]}$ may depend on unknown parameters θ of the structural model that need to be estimated. In the latter case, $F_{\mathbb{E}[Y|\mathcal{I}^M]}$ remains a known function of θ, and our reasoning holds conditional on θ. We consider the following estimator of h^M:

$$\widehat{h}^M = \widehat{F}_\psi^{-1} \circ F_{\mathbb{E}[Y|\mathcal{I}^M]}. \tag{10}$$

where $\widehat{F}_\psi^{-1}$ denotes the empirical quantile function of the subjective beliefs ψ. Theorem 6 ensures that, under mild regularity conditions, $\widehat{h}^M$ is asymptotically normal, and, importantly for practical purposes, also ensures the validity of the bootstrap.

Theorem 6 *For all t such that $F_{\mathbb{E}[Y|\mathcal{I}^M]}(t) \in (0,1)$ is a continuity point of F_ψ^{-1},*

$$\widehat{h}^M(t) \to h^M(t) \quad a.s.$$

Moreover, for all t such that F_ψ is differentiable at $h^M(t)$ with positive derivative (F'_ψ), we have

$$\sqrt{n}\left(\widehat{h}^M(t) - h^M(t)\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{F_{\mathbb{E}[Y|\mathcal{I}^M]}(t)(1 - F_{\mathbb{E}[Y|\mathcal{I}^M]}(t))}{F'^2_\psi(h^M(t))}\right).$$

Finally, conditional on the sample and with probability tending to one, the bootstrap counterpart of $\sqrt{n}\left(\widehat{h}^M(t) - h^M(t)\right)$ converges to the same limit.[13]

5 Monte-Carlo simulations

In this section 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

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

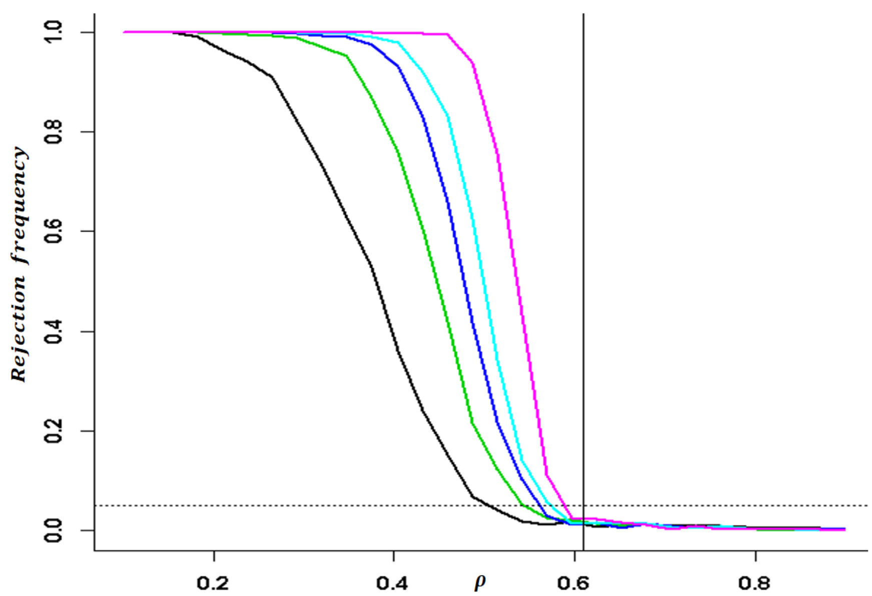
where ζ, U and ψ are mutually independent, $\zeta \sim \mathcal{N}(2, 0.1)$ and $U \sim \mathcal{U}[0,1]$.

In this setup, expectations are rational if and only if $\rho = 1$. But given that we observe Y and ψ in two different datasets, there are values of $\rho \neq 1$ for which we cannot reject our test. More precisely, we can show that as n tends to infinity, we reject our test if and only

[13]For a formal definition of conditional convergence, see e.g. Van der Vaart (2000), Section 23.2.1.

if $\rho \geq \rho^* \simeq 0.616$. In this context also, the naive test $E(Y) = E(\psi)$ always fails to reject RE. The RE test based on variances is only able to detect a subset of violations of RE that corresponds to $\rho < 0.445$.

Results reported in Figure 2 show the power curves of the test φ_α for five sets of sample sizes ($n_\psi = n_Y = n \in \{400; 800; 1,200; 1,600; 3,200\}$), using 800 simulations for each value of ρ. We also rely on 500 bootstrap simulations to compute the critical values of the test. The test statistic T involves the two tuning parameters b_0 (in $B_n = b_0(\ln(n)\ln(\ln(n)))^{1/2}$) and p. As described p.644 of 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 (2018), we choose them as the smallest (resp. highest) value such that the rejection rate under the null is below the nominal size 0.05, and obtain $b_0 = 0.3$ and $p = 0.05$.

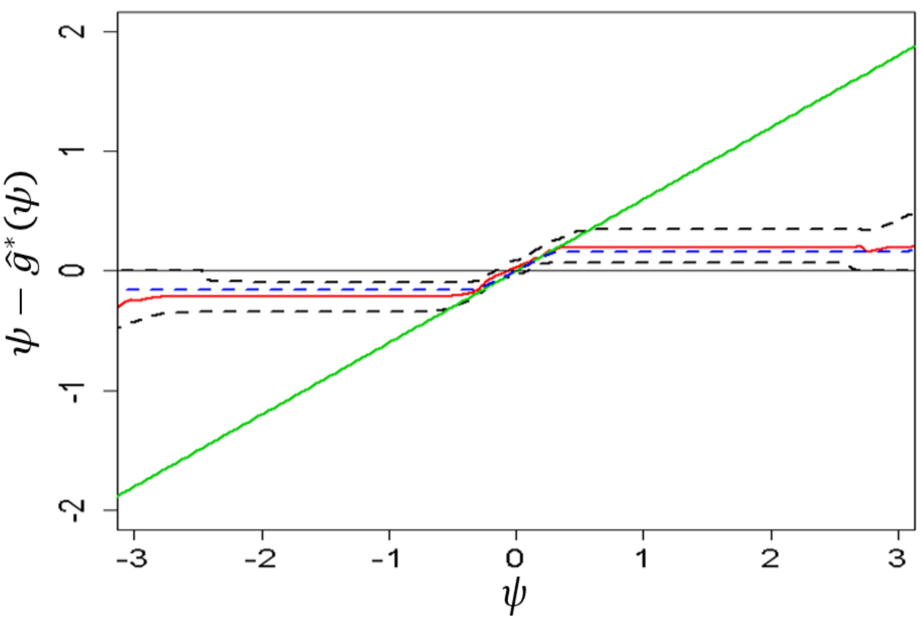


Note: The curves from right to left correspond to $n = 400, 800, 1200, 1600$ and 3200. The vertical line correspond to the theoretical limit for the rejection of our test.

Figure 2: Power curves for the test without covariates.

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 RE test is conservative in the sense that rejection frequencies under the null are smaller than $\alpha = 0.05$, for all sample sizes. Note that 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 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 close to the nominal level.

Next, we report in Figure 3 below the estimated minimal deviations from rational expectations. Specifically, we plot the differences between the beliefs ψ and the modified beliefs $\widehat{g}^*(\psi)$, where the transformation $\widehat{g}^*$ is computed using the estimator of Section 4.1.2, for $\rho = 0.3$ and $n = 800$. In the same figure, we also report the true minimal deviations $\psi - g^*(\psi)$, obtained by solving (8) with a large number of observations ($n = 10,000$) as g^* does not have a closed form representation in this setting. Comparing these two curves shows that the estimator g^* exhibits a small bias over the support of ψ. The 2.5% and 97.5% quantile of $\psi - \widehat{g}^*(\psi)$, in dotted black lines, are also fairly close to each other, showing that the estimator is already accurate with $n = 800$. Besides, the coverage of the bootstrap confidence intervals is generally very close to the nominal rates. For $\rho = 0.3$ and $n = 800$, the mean coverage rates over values of ψ in $[-3, 3]$ are equal to 98.6% and 95.4% for nominal rates of 99% and 95%, respectively. Noteworthy, we obtain very similar patterns on the accuracy of the estimator for alternative values of ρ.



Note: The plain red curve corresponds to the average of $\psi - \widehat{g}^*(\psi)$ over 1, 000 simulations (with $n = 800$), and the dotted black curves are the 2.5% and 97.5% quantiles of $\psi - \widehat{g}^*(\psi)$. The dotted blue curve is the true $g^*(\psi) - \psi$, whereas the green line is the true function $\psi - \mathbb{E}[Y|\psi] = 0.7\psi$.

Figure 3: Estimation and true value of $\psi - g^*(\psi)$.

Finally, the discrepancy between the minimal and the true deviations from rational expectations $\psi - \mathbb{E}[Y|\psi]$, provides a graphical illustration of the loss of information induced by the data combination problem. Whereas, by construction, the minimal deviations never exceed in magnitude the true deviations from RE, it is interesting to note that, in this context, the discrepancy between both deviations is much larger in the tails than in the center of the distribution of beliefs.

6 Application to earnings expectations

6.1 Data

We use data from the Survey of Consumer Expectations (SCE), a monthly household survey conducted by the Federal Reserve Bank of New York since 2012 (see Armantier et al., 2017, for a detailed description of the survey, and Kuchler and Zafar, 2017, and Conlon et al., 2018, for recent articles using the SCE). The SCE is a rotating internet-based panel of about 1,200 household heads, in which respondents participate for up to twelve months.[14] The SCE is conducted with the primary goal of eliciting consumer expectations about inflation, household finance, labor market, as well as housing market. Each survey takes on average about fifteen minutes to complete, and respondents are paid $15 per survey completed.

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 job earnings expectations (ψ) over the next four months: "What do you believe your annual earnings will be in four months?". In this module, respondents are also asked about current job outcomes, and, among them, 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 data on job earnings expectations (ψ), which are available for three cross-sectional samples of household heads who were working either full-time or part-time and responded to the labor market module in March 2015, July 2015, and November 2015, respectively.[15] We combine this data with current earnings (Y) declared in July 2015, November 2015 and March 2016 by the respondents who are working full-time or part-time at the time of the survey. 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, obtained from a total of 1,499 household heads.[16]

6.2 Are earnings expectations rational?

Using the SCE data and the rationality tests discussed in Section 3, we now investigate whether household heads form rational expectations on their future earnings. In Table 1 below, we report the results from the naive test of RE ($\mathbb{E}(Y) = \mathbb{E}(\psi)$), and our preferred generalized test ("Full RE").

Several remarks are in order. First, using our generalized test, we do reject for the whole population the hypothesis that agents form rational expectations over their future earnings.

[14] Each month the panel consists of about 180 entrants, and 1,100 repeated respondents. While SCE entrants are overall fairly similar to SCE repeated respondents, they are slightly older and have slightly lower incomes, reflecting differential attrition (see Table 1 in Armantier et al., 2017).

[15] 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.

[16] 51% (1,536) only of these observations correspond to the sub-sample of respondents who are reinterviewed.

Second, we reject RE when we apply our test separately by college degree attainment, and numeracy test score. We also reject the RE hypothesis for the subpopulation of females, for workers who have spent six months or less in their current job, and for whites (at the 10% level only). Third, using the naive test of equality of means between earnings beliefs and realizations would instead generally result in not rejecting the null at any standard levels. The only exception being the subgroup of workers without a college degree, for whom the naive test yields rejection of RE at the 5% level. While individuals as a whole form expectations over their earnings in the near future - four months out - 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 results highlight the importance of incorporating the additional restrictions of rational expectations that are embedded in our test, using the full distributions of subjective beliefs and realized outcomes to detect violations of RE.

Table 1: Tests of RE on annual earnings

	$\mathbb{E}(Y-\psi)/\mathbb{E}(Y)$	Naive RE (p-val)	Full RE (p-val)	Nb. obs. ψ	Nb. obs. Y	Abs. Relat. Deviation (in %) Top 1%
All	0.034	0.19	0.000^{**}	1,565	1,428	47.0
Women	0.059	0.11	0.000^{**}	730	649	54.1
Men	0.025	0.46	0.513	835	779	35.9
White	0.031	0.27	$0.088^{\dagger}$	1,280	1,182	44.2
Minorities	0.047	0.45	0.268	285	246	71.2
Coll. Degr.	-0.001	0.96	0.000^{**}	1,106	1,053	51.1
No Coll. Degr.	0.090	0.04^{*}	0.036^{*}	459	375	45.7
High Num.	0.033	0.25	0.000^{**}	1,158	1,070	51.4
Low Num.	0.055	0.27	0.026^{*}	407	358	38.4
Exp $\leq$ 6 mth.	0.105	0.21	0.002^{**}	271	180	48.5
Exp $>$ 6 mth.	0.007	0.78	0.547	1,294	1,248	31.0

Notes: significance levels: †: 10%, *: 5%, **: 1%. Annual earnings = An. Earn, Low/high Num. = low/high numerical ability, Exp $\leq$ 6 mth. = less than 6 months of experience in the current main job. "Naive RE" denotes the naive RE test of equality of means between Y and ψ. "Full RE" denotes the Generalized test without covariates, where we consider $Y_{it} = c_t \psi_{it}$ and test H_0' with $E\left[Y_{it}/c_t|\psi_{it}\right] = \psi_{it}$. Distributions of realized earnings (Y) and earnings beliefs (ψ) are both Winsorized at the 95% quantile.

We do not report in this table the results of the direct test of RE. 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 key issues associated with the direct test. First, as already discussed in Appendix B, the direct test is not robust to measurement errors on the subjective beliefs ψ. Second, and importantly, attrition from the survey may be endogenous. We explore the issue of attrition in Table 2 below, where we report the estimation results from a logit model of attrition on earnings beliefs, gender, race/ethnicity, college degree attainment, numeracy test score, and tenure. The main takeaway from this table is that earnings beliefs are significantly associated with attrition, even after controlling for this extensive set of characteristics.

These results suggest that attrition is endogenous, with individuals for whom we observe both earnings expectations and realizations expecting to earn more than those who are not followed across these two waves. Along the same lines, a Kolmogorov-Smirnov test rejects at the 1% level the equality of the distributions of realized and expected earnings between the whole sample, and the subsample used for the direct test (see Figure 4 for the estimated distributions).[17] Taken together, these results indicate that the direct test of rational expectations is unlikely to be valid.

Finally, going beyond testing, Figure 5 offers additional insights regarding the deviations from rational expectations on earnings. We focus here on the whole population, for which, using our test, we strongly reject (at the 1% level) the hypothesis of rational expectations (see Table 1). This figure shows that, for these individuals, the deviations from rational expectations are primarily due to the coexistence of over-pessimistic (i.e., individuals for whom $\psi < E(Y|\psi)$) and over-optimistic (individuals for whom $\psi \geq E(Y|\psi)$) individuals. Both types of deviations from rational expectations partially offset one another when computing the average across all observations, so that the naive test fails to detect this pattern of violations from rational expectations. In contrast, our test, which exploits the full distributions of earnings beliefs and realizations, is able to detect these deviations from rational expectations. While our approach is conservative in the sense that these are the minimal deviations from RE that are consistent with the data, it is interesting to note that some of the deviations are quantitatively large. This is illustrated in Table 1 above, where we report for each subgroup the 99th percentile of the absolute relative minimal deviations from RE.[18]

Table 2: Logit model of no attrition on earnings expectations and demographics

Population	Intercept	ψ	Male	White	Coll. Degree	Low Num.	Exp > 6
All individuals	-1.054**	5.279e-06**	-0.091	0.250	0.070	0.110	0.559**
	(0.226)	(1.341e-06)	(0.1080)	(0.1697)	(0.1191)	(0.1253)	(0.1407)

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

[17] P-values are equal to 0.004 and 0.006 for the distributions of realized and expected earnings, respectively.
[18] This quantity can be interpreted as a maximin type of measure.

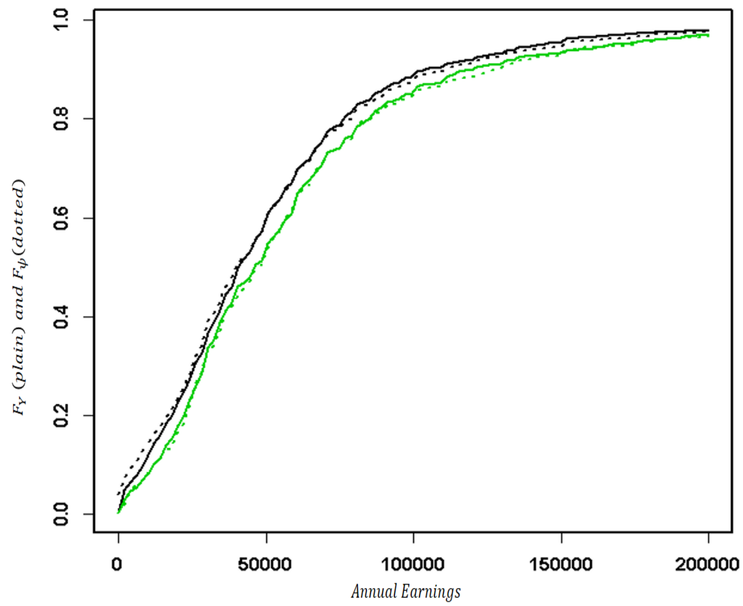


Figure 4: Estimates of F_Y for the subset of individuals where the direct RE test is possible (green) and all the population (black). Estimates of F_ψ for those subsets are represented in dotted.

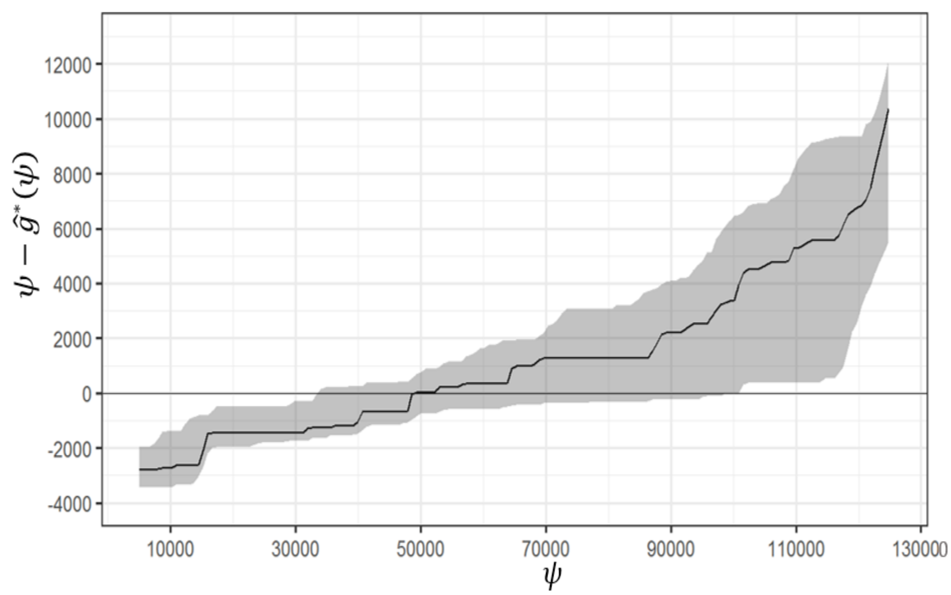


Figure 5: Average estimated minimal deviations from RE $\psi - \widehat{g}^*(\psi)$ (plain black) for annual earnings for the whole population. The shaded grey area corresponds to the 95% bootstrap pointwise confidence interval. All results are displayed in 2015 US dollars.

6.3 Deviation from RE in a life-cycle consumption model

In this section, we revisit the life-cycle consumption model of Kaplan and Violante (2010) (KV) by relaxing the assumption that individuals form rational expectations about their future earnings.
A common feature of benchmark life-cycle standard incomplete market (SIM) consumption models is that the rational expectation hypothesis is maintained, for both analytical tractability and data availability reasons. However, if a substantial fraction of the individuals do not form rational expectations on their future earnings, conclusions that one can draw from this model can be misleading. In the following we address this issue using our methodology developed in Section 4 based on minimal deviations from rational expectations. Specifically, we use the benchmark SIM model used in KV as a starting point, which we modify to account for the fact that individuals may not form rational expectations about their income process. Using this framework, we then illustrate how the type of deviations from rational expectations which are consistent with the SCE data impacts self-insurance mechanisms, and in particular the role played by transitory and permanent income shocks on consumption.

6.3.1 Model

Time is assumed to be discrete. The economy is constituted of agents (household heads) who work for T^{ret} periods, before retiring. For any given $t > T^{ret}$, agents are characterized by an unconditional probability of surviving until age t denoted by ξ_t and are all assumed to die with probability 1 by $t = T$. We assume that income and death are the only sources of uncertainty. Households have an expected lifetime utility given by:

$$\mathcal{E}\left[\sum_{t=1}^{T}\beta^t\xi_t u\left(C_{i,t}\right)\middle|\mathcal{I}_0\right], \tag{11}$$

where $\mathcal{E}\left[\cdot|\mathcal{I}_0\right]$ denotes the *subjective* expectation operator conditional on $\mathcal{I}_0$ the initial information set, $u(.)$ denotes the flow utility of consumption, for which we assume a quadratic specification of the form $u(C) = (C^* - C)^2/2$ (where C^* is a constant), and β is the discount factor. We assume that the subjective expectation operator satisfies the usual mathematical properties of expectations.[19] Importantly though, this framework accommodates a wide range of violations of rational expectations, since subjective expectations $\mathcal{E}\left[\cdot|\mathcal{I}_0\right]$ may not coincide with mathematical expectations $E\left[\cdot|\mathcal{I}_0\right]$. In particular, unlike the benchmark model, this specification allows for biased income expectations formation.

We now turn to the description of the income process. During worklife, realized log income $\log(Y_{i,t})$ is supposed to be given by the sum of a deterministic experience profile, $\kappa_{i,t}$, a

[19] See also, e.g., Brunnermeier and Parker (2005) for a similar assumption.

permanent component, $z_{i,t-1}$, a permanent shock, $\eta_{i,t}$, and a transitory shock, ϵ_{it}:

$$\log(Y_{i,t}) = \kappa_{i,t} + y_{i,t}$$
$$y_{i,t} := z_{i,t} + \epsilon_{i,t}$$
$$z_{i,t} = z_{i,t-1} + \eta_{i,t},$$

where $z_{i,0}$ is drawn from a normal distribution with mean zero and variance σ^2_{z0}. The shocks $\epsilon_{i,t}$ and $\eta_{i,t}$ also have mean zero and are normally distributed with variances σ^2_ϵ and σ^2_η, and are mutually independent and independent over time and across household heads in the economy.[20] 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.

Next, before deriving the budget constraints, we need to introduce a couple of additional notations. We denote by $A_{i,t+1}$ the amount of the tradable risk-free one-period bond which pays rate of return $(1+r)$ detained by households, and assume that they begin their life with $A_{i,0}$ drawn from a specific distribution $H(A_{i,0})$ and that they face a lower bound $\underline{A} \leq 0$ on their assets. We also denote by $Y^S_{i,t}$ the post-retirement social security transfers, which are computed as a function of the lifetime average individual labor income (see KV for additional details).

Agents make their consumption choices by maximizing their present value of subjective expected lifetime utility, under the following budget constraints:

$$C_{i,t} + A_{i,t+1} = (1+r)A_{i,t} + Y_{i,t} \text{ if } t < T^{ret} \tag{12}$$
$$C_{i,t} + \left(\frac{\xi_t}{\xi_{t+1}}\right) A_{i,t+1} = (1+r)A_{i,t} + Y^S_{i,t} \text{ if } t \geq T^{ret}. \tag{13}$$

Given these assumptions, we derive the optimal consumption paths in two situations, namely: i) all individuals form rational expectations on their future outcomes (as in KV), and ii) individual beliefs may deviate from rational expectations. We implement the latter scenario by assuming that subjective expectations deviate minimally from rational expectations satisfying model restrictions, following our discussion in Section 4.2. Specifically, using the notations defined earlier, the (pseudo) subjective income beliefs are computed in this case as a function of the rationally expected income:

$$\mathcal{E}\left[Y_{i,t}|\mathcal{I}_{i,t-1}\right] = h^M\left(E\left[Y_{i,t}|\mathcal{I}_{i,t-1}\right]\right), \tag{14}$$

where h^M is estimated using the estimator of Section 4.2.2 defined in Equation (14).[21] In this estimation we use that $\log\left(E\left[Y_{i,t}|\mathcal{I}_{i,t-1}\right]\right)$ is normally distributed with mean $\kappa_{i,t}$ and

[20] Note that with quadratic preferences, the variance of transitory shocks has no impact on consumption (see Appendix E).

[21] We implicitly restrict the sensitivity analysis to a class of deviations of rational expectations that are such that h^M is constant over time. While it would be conceptually easy to extend our method and relax this assumption by estimating and using instead a sequence of mappings $(h^M_t)_t$, this is arguably the most natural departure from the RE hypothesis, whereby $h^M = Id$ for all time periods. Restricting to time-invariant h^M also yields sizable precision gains, resulting in turn in a more informative sensitivity analysis.

variance $\sigma^2_{z_0} + (t-1)\sigma^2_\eta$ and use a quantile regression estimator for the quantile of income expectations conditional on age $F^{-1}_{\psi|t}$. Since in KV, $Y_{i,t}$ is interpreted as household income after taxes and transfers whereas we only have data on expectations about individual labor earnings, we use an equipercentile mapping to generate from the distribution of either realized or expected individual labor earnings a distribution of realized or expected household income. We estimate this equipercentile mapping using the dataset from Blundell et al. (2008), built from the PSID, that has both realized individual labor earnings and household income from 1989 to 1992. We use the same parameters for the income process as those used in KV and estimated in Blundell et al. (2008), namely $\sigma^2_\epsilon = 0.05$, $\sigma^2_\eta = 0.01$, and $\sigma^2_{z_0} = 0.15$. We consider alternative values on σ^2_η and $\sigma^2_{z_0}$ as robustness checks.

6.3.2 Results

We use the main parameters chosen by KV to reproduce key features of the US economy. Specifically, we assume that in the utility function, $C^* = 200,000$ and that the interest rate r is 3%. As in KV, who based their estimates on the 1989 and 1992 Survey of Consumer Finances data, the discount factor β is set to match an aggregate wealth-income target ratio of 2.5.

Because we reject RE for all individuals, we consider in our second scenario deviations from RE for the whole population. In other words, subjective expectations of individuals are computed using (14). Figure 6 displays $\widehat{h}^M$ used in equation (14).[22] Finally, following KV, we simulate the model both with a borrowing constraint at 0, and without borrowing constraint, for an artificial panel of 10,000 households for 70 periods. A Kolmogorov-Smirnov test rejects at the 1% level the equality of the distributions of simulated expected income and the subjective expected income obtained using (14) (with a p-value lower than 10^{-5} for $\sigma^2_\eta = 0.01$ and $\sigma^2_\eta = 0.02$), thus indicating that, consistent with the earlier findings discussed in Section 6.2, RE does not hold in this context.

Our main object of interest is the 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:

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

where the variance and covariance are taken cross-sectionally over the entire population of households. We also consider below ϕ^x_t, which is the same quantity but computed conditionally on being of age t.

We report the estimates of ϕ^η and ϕ^ϵ on the whole population in Table 3. In the baseline rows, we reproduce the results of KV under rational expectations. We then display the results based on minimal deviations of RE. Interestingly, some coefficients appear to be quite sensitive to these minimal deviations. Shifting from RE to expectations based on (14), an important

[22] To account for the estimation error on h^M and compute standard errors, we use 200 bootstrap replications.

takeaway from this table is that, while ϕ^ϵ (insurance coefficient to transitory shocks) does not generally change much, ϕ^η (insurance coefficient to persistent shocks) is substantially affected. This holds true in the baseline model both with and without borrowing constraints, as well as for the alternative values of $\sigma^2_{z_0}$ and σ^2_η.

These findings also have some implications regarding the link with previous empirical estimates. In the baseline case with borrowing constraints, the estimate of model under RE ($\phi^\eta = 0.22$) is quite different from the Blundell et al. (2008, BPP hereafter) estimates based on US data ($\phi^\eta = 0.34$). Accounting for minimal deviations from RE using our method, the insurance coefficients are then equal respectively to 0.33 and 0.38. Hence, even though we still observe a gap between the estimates from the model and from the data, this gap is much lower when we relax the assumption that agents form rational expectations on their future outcomes. Overall, this suggests that a sizeable share of the discrepancy between the estimates of the insurance coefficient, and the insurance coefficient obtained from the consumption model, is in fact attributable to deviations from rational expectations.

We also observe significant differences between the model with RE, and the version of the model that accounts for minimal deviations from RE in the profile of the insurance coefficients as a function of age under the different scenarii (Figure 8 in Appendix). In particular, households save significantly more in the presence of permanent shocks when we allow for deviations from RE and less in the presence of transitory shocks when they are between 35 and 50 years old.

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

With borrowing constraints	Baseline		$\sigma^2_\eta = 0.02$	
	ϕ^η	ϕ^ϵ	ϕ^η	ϕ^ϵ
RE, Baseline	0.224	0.778	0.218	0.743
Deviation from RE, Baseline	0.330	0.682	0.383	0.473
	(0.080)	(0.032)	(0.015)	(0.024)
Without borrowing constraints	**Baseline**		**$\sigma^2_\eta = 0.02$**	
	ϕ^η	ϕ^ϵ	ϕ^η	ϕ^ϵ
RE, Baseline	0.114	0.938	0.069	0.937
Deviation from RE, Baseline	0.407	0.883	0.576	0.619
	(0.038)	(0.030)	(0.041)	(0.033)

Notes: the baseline case uses $\sigma^2_\eta = 0.01$, $\sigma^2_\epsilon = 0.05$, $\sigma^2_{z_0} = 0.15$, and an aggregate wealth-income ratio of 2.5. "BPP estimates on US data" are the estimates of BPP using US data. "BPP estimates on the model" are estimates based on BPP method using simulated data we obtain from the model.

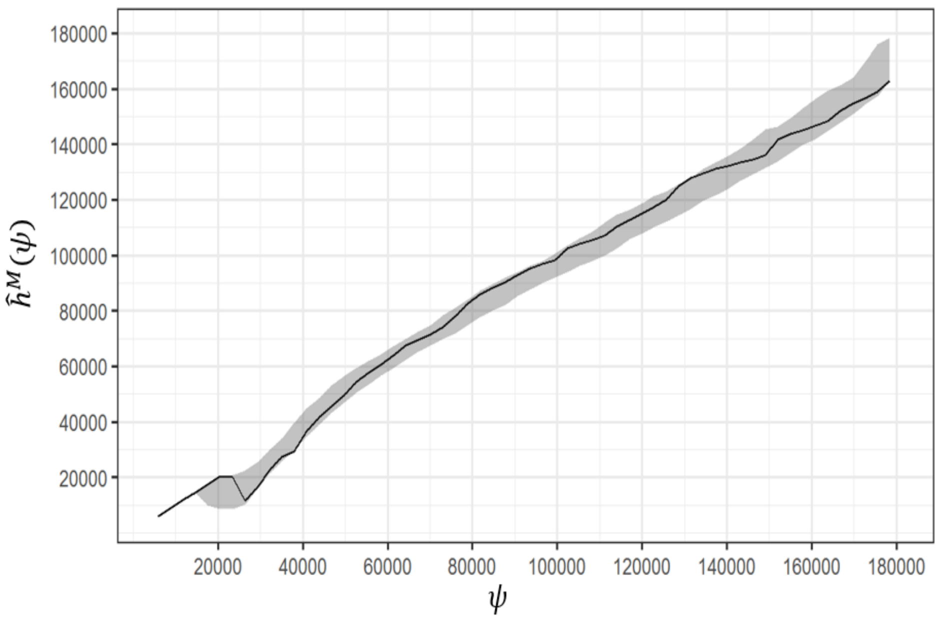


Figure 6: Average estimated function $\widehat{h}^M(\psi)$ (plain dark) for annual income for the whole population. The shaded grey area corresponds to the 95% bootstrap pointwise confidence interval. All results are displayed in 2015 US dollars.

When looking at average lifetime net worth profiles implied by the model with RE or deviations from RE (Figure 9 in Appendix), households heads below 40 appear first to be less indebted than households heads under RE. This comes from the fact that for 60% of them, their income is between 40,000$ and 100,000$ and thus, from Figure 6, they are over-pessimistic. If they are not constrained, they tend to insure more than rationally against permanent shock. After 40, most of the households heads earn on average around or more than 100,000$, which from Figure 6 implies that they are over-optimistic. This translates into an "under-insurance" against transitory shocks (from 0.93 to 0.88 in the case without borrowing constraints), as they have initially accumulated more assets than what would be rationally optimal in order to face retirement. This over-optimism in turn results in a stepper decay of the assets after retirement in Figure 9. Interestingly, our results are in line with the findings of Kaufmann and Pistaferri, 2009 (see also Pistaferri, 2001) who show that using subjective expectations available in the Survey of Household Income and Wealth (SHIW) in Italy lowers the estimated degree of insurance against transitory shocks.

7 Conclusion

In this paper, we develop a new test of rational expectations that can be used in a wide 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 subjective 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 introduce the concept of minimal deviations from rational expectations than can be rationalized by the data. Using recent tools from the optimal transport literature, we show that, under fairly mild regularity conditions, these deviations exist, are unique, and are also easily estimated. In the context of structural models, these deviations offer a new way to conduct a sensitivity analysis on the assumed form of expectations. We apply our method to test and quantify deviations from rational expectations about future earnings. While individuals tend to be right on average about their future earnings, our test rejects rational expectations. Using the deviations from rational expectations within the life-cycle consumption model of Kaplan and Violante (2010), we provide evidence that the behavioral responses of consumers to permanent income shocks are sensitive to departures from rational expectations.

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.

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**, 609–666.

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

Andrews, D. and Soares, G. (2010), 'Inference for parameters defined by moment inequalities using generalized moment selection', *Econometrica* **78**, 119–157.

Arcidiacono, P. and Ellickson, P. (2011), 'Practical methods for estimation of dynamic discrete choice models', *Annual Review of Economics* **3**, 363–394.

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

Arcidiacono, P., Hotz, J., Maurel, A. and Romano, T. (2014), Recovering ex ante returns and preferences for occupations using subjective expectations data, Technical report, National Bureau of Economic Research.

Arcidiacono, P., Hotz, J. V., Maurel, A. and Romano, T. (2017), Ex ante returns and occupational choice, Technical report, Working paper.

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

Beare, B. and Shi, X. (2018), 'An improved bootstrap test of density ratio ordering', *Econometrics and Statistics* **Forthcoming**.

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**, 1887–1921.

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

Brunnermeier, M. K. and Parker, J. A. (2005), 'Optimal expectations', *The American Economic Review* **95**(4), 1092–1118.

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. (2016), '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**, 1591–1608.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2014), 'Testing many moment inequalities', *arXiv preprint arXiv:1312.7614* .

Conlon, J. J., Philossoph, L., Wiswall, M. and Zafar, B. (2018), Labor market search with imperfect information and learning, Technical report, National Bureau of Economic Research.

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**, 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**, 999–1042.

Delavande, A. (2014), 'Probabilistic expectations in developing countries', *Annual Review of Economics* **6**, 1–20.

Delavande, A. and Zafar, B. (2018), 'University choice: the role of expected earnings, non-pecuniary outcomes, and financial constraints', *Journal of Political Economy* **forthcoming**.

Dickstein, M. and Morales, E. (2018), 'What do exporters know?', *Quarterly Journal of Economics* **forthcoming**.

Elliott, G., Komunjer, I. and Timmermann, A. (2005), 'Estimation and testing of forecast rationality under flexible loss', *Review of Economic Studies* **72**, 1107–1125.

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

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

Gennaioli, N., Ma, Y. and Shleifer, A. (2015), Expectations and investment, *in* 'NBER Macroeconomics Annual 2015, Volume 30', University of Chicago 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* **forthcoming**.

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

Hirano, K., Imbens, G. W. and Ridder, G. (2003), 'Efficient estimation of average treatment effects using the estimated propensity score', *Econometrica* **71**, 1161–1189.

Hoffman, M. and Burks, S. V. (2017), Worker overconfidence: Field evidence and implications for employee turnover and returns from training, Technical report, National Bureau of Economic Research.

Houser, D., Keane, M. and McCabe, K. (2004), 'Behavior in a dynamic decision problem: An analysis of experimental evidence using a bayesian type classification algorithm', *Econometrica* **72**, 781–822.

Hsu, Y.-C. (2016), 'Consistent tests for conditional treatment effects', *The econometrics journal* .

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

Jin, S., Corradi, V. and Swanson, N. R. (2017), 'Robust forecast comparison', *Econometric Theory* **33**, 1306–1351.

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

Kapor, A., Neilson, C. A. and Zimmerman, S. (2017), Heterogeneous beliefs and school choice mechanisms. Working paper, Princeton University, Industrial Relations Section.

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**, 473–522.

Kuchler, T. and Zafar, B. (2017), 'Personal experiences and expectations about aggregate outcomes'. Working paper.

Landier, A., Ma, Y. and Thesmar, D. (2017), New experimental evidence on expectations formation. Working paper.

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

MaCurdy, T. (1982), 'The use of time series processes to model the error structure of earnings in a longitudinal data analysis', *Journal of Econometrics* **18**(1), 83–114.

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**, 219–231.

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

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

Navarro, S. and Zhou, J. (2017), 'Identifying agent's information sets: an application to a lifecycle model of schooling, consumption and labor supply', *Review of Economic Dynamics* **25**, 58–92.

Pacini, D. (2017), 'Two-sample least squares projection', *Econometrics Review* pp. 1–29.

Pass, B. (2013), 'Optimal transportation with infinitely many marginals', *Journal of Functional Analysis* **264**, 947–963.

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

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

Pollard, D. (1990), Empirical processes: theory and applications, *in* 'NSF-CBMS regional conference series in probability and statistics', pp. i–86.

Rachev, S. T. and Rüschendorf, L. (1998), *Mass Transportation Problems: Volume I: Theory*, Vol. 1, 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**, 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. (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.

Suehiro, D., Hatano, K., Kijima, S., Takimoto, E. and Nagano, K. (2012), Online prediction under submodular constraints, *in* 'International Conference on Algorithmic Learning Theory', pp. 260–274.

Tversky, A. and Kahneman, D. (1992), 'Advances in prospect theory: Cumulative representation of uncertainty', *Journal of Risk and Uncertainty* **5**, 297–323.

Van der Klaauw, W. (2012), 'On the use of expectations data in estimating structural dynamic choice models', *Journal of Labor Economics* **30**, 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**, 21–42.

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

Van der Vaart, A. and Wellner, J. (1996), *Weak Convergence*, Springer.

Villani, C. (2008), *Optimal transport: old and new*, Vol. 338, Springer Science & Business Media.

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

Wolpin, K. I. (2013), *The limits of inference without theory*, MIT Press.

Zafar, B. (2011), '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 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 $E(q(Y, c_0(X))|X) = E(\psi|X)$. When $q(y,c) = y - c$ or $q(y,c) = y/c$, we get respectively $c_0(X) = E(Y|X) - E(\psi|X)$ and $c_0(X) = E(Y|X)/E(\psi|X)$, and $\widehat{c}$ is easy to compute using nonparametric estimators of $E(Y|X)$ and $E(\psi|X)$.

We let hereafter $\overline{m}_n(h,y) = \sum_{i=1}^n m\left(D_i, \widetilde{Y}_{c,i}, X_i, h, y\right)/n$. In the test statistic T, we replace, for $(y,h) \in \mathcal{Y} \times \cup_{r\geq 1}\mathcal{H}_r$, $\overline{\Sigma}_n(h,y)$ by $\overline{\Sigma}_n(h,y) = \widehat{\Sigma}_n(h,y) + \epsilon \text{Diag}\left(\widehat{\mathbb{V}}\left(\widetilde{Y}_{\hat{c}}\right), \widehat{\mathbb{V}}\left(\widetilde{Y}_{\hat{c}}\right)\right)$, where $\widehat{\Sigma}_n(h,y)$ and $\widehat{\mathbb{V}}\left(\widetilde{Y}_{\hat{c}}\right)$ are respectively the sample covariance matrix of $\sqrt{n}\overline{m}_n(h,y)$ and the empirical variance of $\widetilde{Y}_{\hat{c}}$.

We obtain in this context a result similar to Theorem 2 above, under the regularity conditions stated in Assumption 5. 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 $\|c\|_{s,\infty} := \max_{|\boldsymbol{k}|\leq s} \sup_{x\in[0,1]^{d_X}} \left|c^{(\boldsymbol{k})}(x)\right|$. Also, 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 5 *(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 $\mathcal{Y} \times [-C, C]$ and $\sup_{(y,c)\in\mathcal{Y}\times[-C,C]} |q(y,c)| \leq M_0$;

(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 $\sup_{F\in\mathcal{F}_0} Q_{F,1} \leq \overline{Q}_1 < +\infty$. Alson $F_{q(\psi,c(X))}$, $F_{q(Y,c(X))}$ 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, $\widehat{\mathbb{V}}_F\left[\widetilde{Y}_{\hat{c}}^2\right]$ is a consistent estimator of $\mathbb{V}_F\left[\widetilde{Y}_c^2\right]$.

Part (i) imposes some regularity condition 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 $E(Y|X)$ and $E(\psi|X)$. 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 5 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.[23] In such case,s 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).

Testable implications of rational expectations take different forms in this case. In this paper we propose a test of the null hypothesis that the objective conditional expectation of Y concurs with at least one distribution of ψ that is compatible with the observed bounds. Formally, we consider the following null hypothesis:

$$\begin{aligned}\mathrm{H}_{0B} : \exists (Y', \psi', \mathcal{I}') :\ & \sigma(\psi') \subset \mathcal{I},\ Y' = Y \text{ when } D = 1,\ \psi' \in [\psi_L, \psi_U], \\ & D \perp\!\!\!\perp (Y', \psi') \text{ and } \mathbb{E}(Y'|\mathcal{I}') = \psi'.\end{aligned}$$

A naive extension of the previous results would suggest that one should test whether the condition $E[Y'|\psi'] = \psi'$ holds, for an infinite number of possible distributions of beliefs $F_{\psi'}$ such that $F_{\psi_U} \leq F_{\psi'} \leq F_{\psi_L}$. Importantly, we show in the following proposition that it is in fact sufficient to check that this condition holds for a particular distribution $F_{\psi'}$ satisfying the constraints $F_{\psi_U} \leq F_{\psi'} \leq F_{\psi_L}$. We define hereafter the random variable $\psi^b = \psi_U \mathbb{1}\{\psi_U < b\} + (b \vee \psi_L)\mathbb{1}\{\psi_U \geq b\}$, which is distributed according to $F_{\psi^b}(t) = F_{\psi_U}(t)\mathbb{1}\{t < b\} + F_{\psi_L}(t)\mathbb{1}\{t \geq b\}$, and let $\Delta^b(y) = \int_{-\infty}^{y} F_Y(t) - F_{\psi^b}(t)dt$, and $\delta^b = E(Y) - E(\psi^b)$.

Proposition 6 *Suppose that $\mathbb{E}(|Y|) < +\infty$, $\mathbb{E}(|\psi_L|) < +\infty$ and $E(|\psi_U|) < +\infty$. The following statements are equivalent:*

(i) H_{0B} holds.

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

(ii) $\mathbb{E}[\psi_L] \leq \mathbb{E}[Y] \leq \mathbb{E}[\psi_U]$ *and* $\Delta^{b_0}(y) \geq 0$ *for all* $y \in \mathbb{R}$*, for the unique* b_0 *satisfying* $\delta^{b_0} = 0$*.*

Note that b_0 exists if and only if $\mathbb{E}[\psi_L] \leq \mathbb{E}[Y] \leq \mathbb{E}[\psi_U]$.

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 independant 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:

$$D \perp\!\!\!\perp (Y, \psi)|X. \tag{15}$$

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 yield the following proposition:

Proposition 7 *Suppose that* (15) *and Assumption 1 hold. Then* H_{0X} *is equivalent to*

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

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

This proposition shows that under sample selection, we can build a statistical test of H_{0X} akin to that developed in Section 3, by merely estimating nonparametrically $p(X)$. We could consider for that purpose, as in e.g., Hirano et al. (2003), a series logit estimator. 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

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

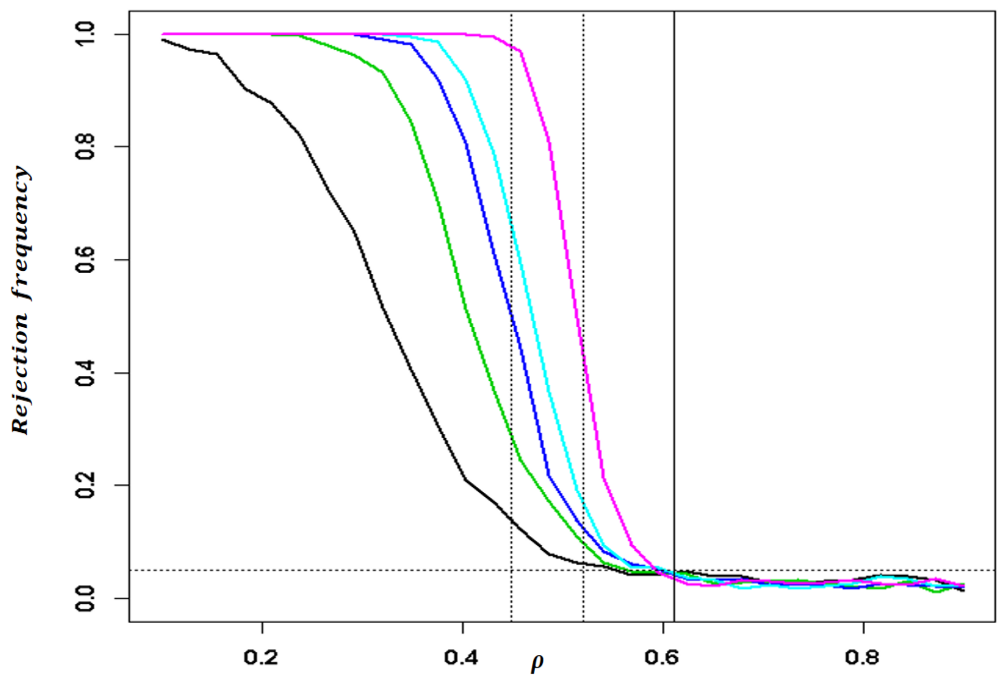
with $\rho \in [0, 1]$, $\psi \sim \mathcal{N}(0, 1)$, $X \sim \mathrm{Beta}(0.1, 10)$ and

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

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 $n_\psi = n_Y = n \in \{400, 800, 1200, 1600, 3200\}$, 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 7 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.



Note: The curves from right to left correspond to $n = 400, 800, 1200, 1600$ and 3200. The dotted vertical lines correspond to the theoretical limit for the rejection of the test based on variance ($\rho \simeq 0.445$) and our test without covariates ($\rho \simeq 0.521$). The plain vertical line at $\rho = 0.616$ corresponds to the same limit for our test with covariates.

Figure 7: Power curves for the test with covariates.

E Additional material on the application

E.1 Details on the life-cycle consumption model

We briefly illustrate how we can compute the optimal consumption path with quadratic preferences and the individual expectations of future income. For simplicity of exposition, we assume here that there is no retirement. Let us denote by

$$v_t(A_{i,t+1}, e^{z_{i,t}}) = \max_{C_{i,t}} u(C_{i,t}) + \mathcal{E}\left[\beta\xi_{t+1} v_{t+1}(A_{i,t+2}, e^{z_{i,t+1}})|\mathcal{I}_t\right],$$

the recursive form of the problem with the two state variables $(A_{i,t+1}, e^{z_{i,t}})$. Denoting by C^* the consumption reference point, at $t = T$ we get, using that $v_T(A_{i,T+1}, e^{z_{i,T}}) = u(C_{i,T})$, where $A_{i,T+1} = 0$ (all assets are consumed in the last period),

$$\begin{aligned} v(A_{i,T}, e^{z_{i,T-1}}) &= \max_{C_{i,T-1}} u(C_{i,T-1}) + \mathcal{E}\left[\beta\xi_T u\left((1+r)A_{i,T} + \widetilde{Y}_{i,T}\right)\Big|\mathcal{I}_{i,T-1}\right] \\ &= \frac{1}{2}\left(\max_{C_{i,T-1}} (C^* - C_{i,T-1})^2 + \beta\xi_T \mathcal{E}\left[\left(C^* - (1+r)A_{i,T} - \widetilde{Y}_{i,T}\right)^2\Big|\mathcal{I}_{i,T-1}\right]\right) \end{aligned} \tag{16}$$

and the first-order condition yields, using (13) and $A_{i,T+1} = 0$,

$$(C^* - C_{i,T-1}) = \beta\xi_T\left(C^* - (1+r)A_{i,T} - \mathcal{E}\left[\widetilde{Y}_{i,T}\Big|\mathcal{I}_{i,T-1}\right]\right). \tag{17}$$

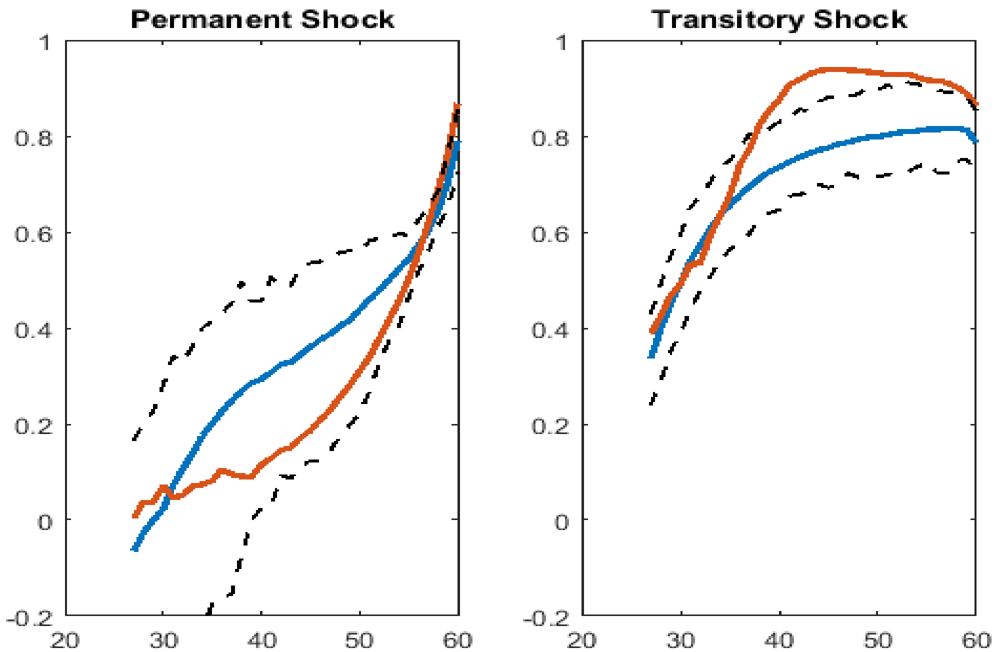
As long as $\mathcal{E}\left[\widetilde{Y}_{i,T}\Big|\mathcal{I}_{i,T-1}\right]$ is known, this allows us to compute the optimal consumption rule at date $T-1$, $C_{i,T-1}(A_{i,T}, e^{z_{i,T-1}})$, as a function of the state variables. Then, by induction, and using the following equations, for all $0 \leq t \leq T$,

$$\begin{aligned} \partial_1 v(A_{i,t+1}, e^{z_{i,t}}) &= (1+r)\beta\xi_{t+1}\mathcal{E}\left[u'\left(C_{i,t+1}\left(A_{i,t+2}, e^{z_{i,t+1}}\right)\right)\Big|\mathcal{I}_t\right] \\ u'(C_{i,t}) &= \partial_1 v(A_{i,t+1}, e^{z_{i,t}}) \end{aligned}$$

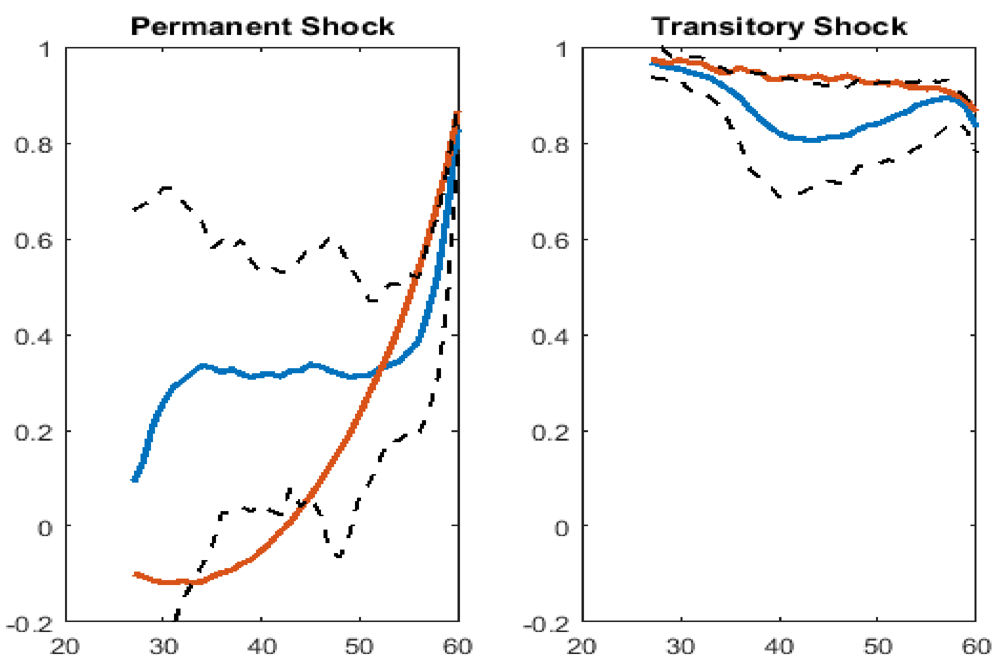
we can compute the consumption rule $C_{i,t}(A_{i,t+1}, e^{z_{i,t}})$ at date t.

E.2 Additional results on the life-cycle consumption model

Figure 8: Age profiles of insurance coefficients.



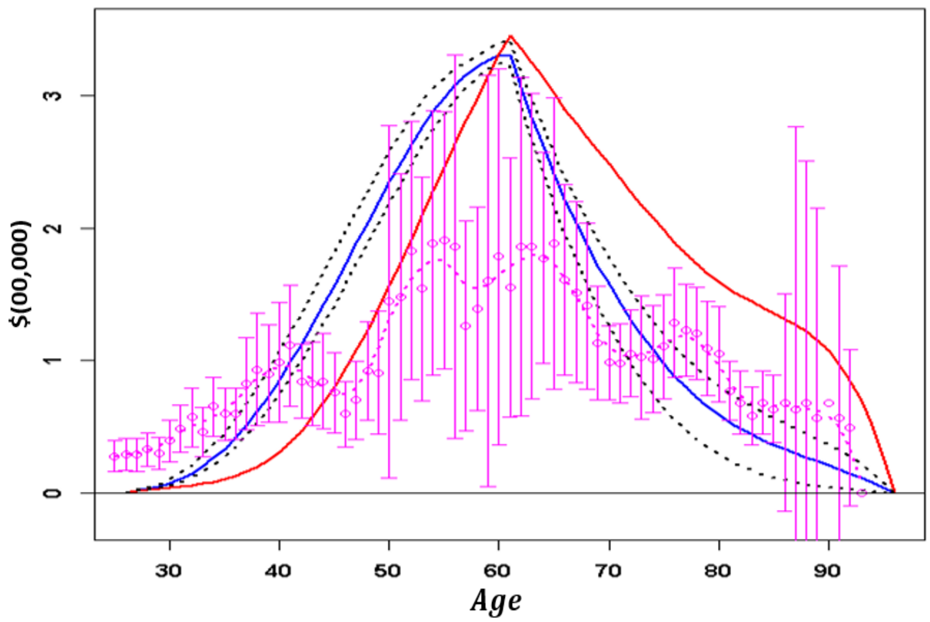
(a) With borrowing constraints



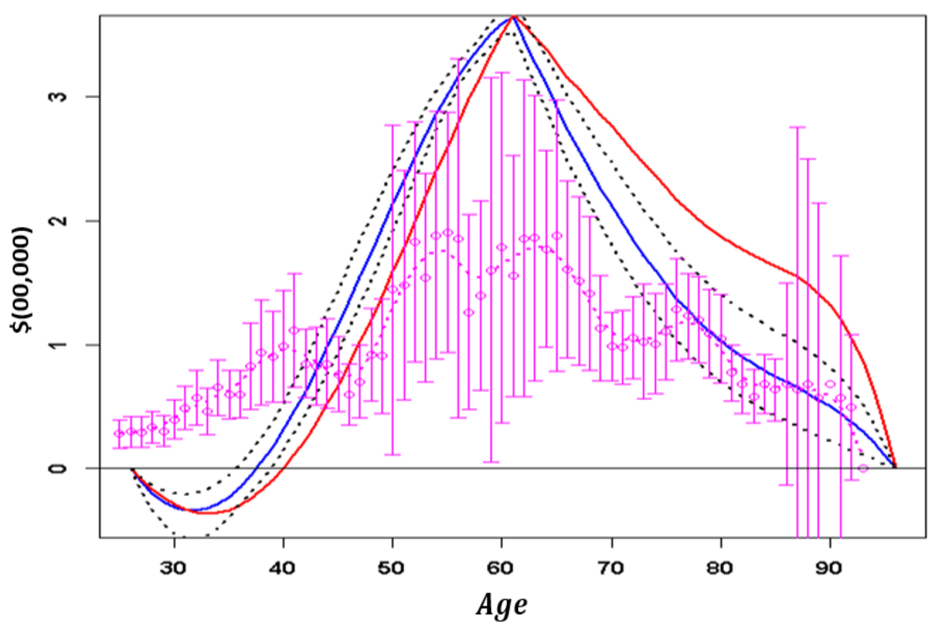
(b) Without borrowing constraints

Notes: the curves in red (resp. in blue) correspond to insurance coefficients under RE (resp. minimal deviations from RE). The dotted black curves are the 0.025 and 0.975 quantiles of the blue line, taking into account the randomness of $\widehat{g}^*$. They are obtained using 200 bootstrap samples.

Figure 9: Average lifetime net worth profiles.



(a) With borrowing constraints



(b) Without borrowing constraints

Notes: the curves in red (resp. in blue) correspond to net worth under RE (resp. with deviation from RE). The purple points correspond to the lifetime net worth for single individuals in the Survey of Consumer Finance of 1992. The dotted black curves are the 0.025 and 0.975 quantiles of the blue line, taking into account the randomness of $\widehat{g}^M$. They are obtained using 200 bootstrap samples.

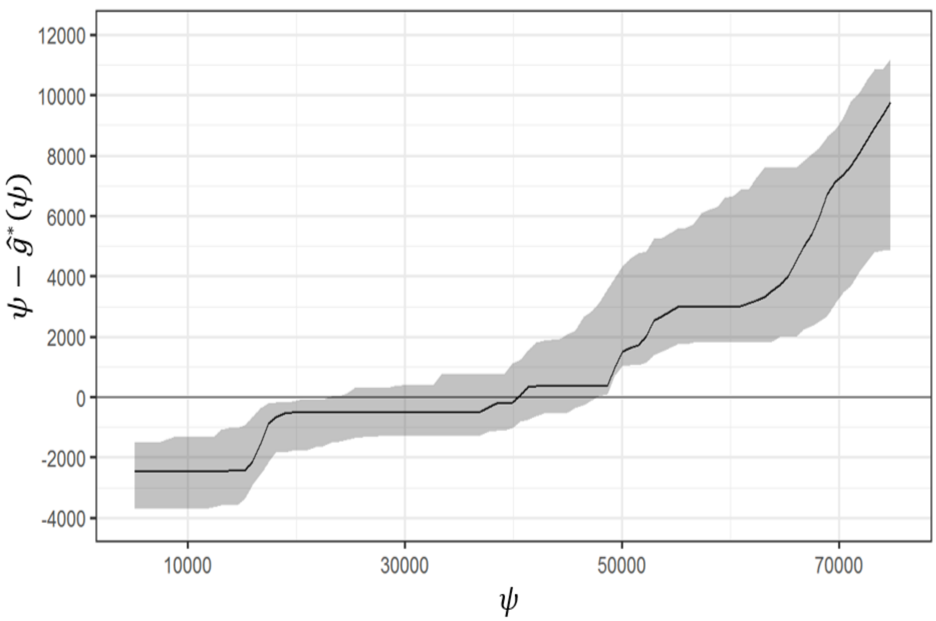


Figure 10: Average estimated function $\psi - \widehat{g}^{*}(\psi)$ (plain black) for annual earnings for those who do not have a College degree. The shaded grey area corresponds to the 95% bootstrap pointwise confidence interval. All results are displayed in 2015 US dollars.

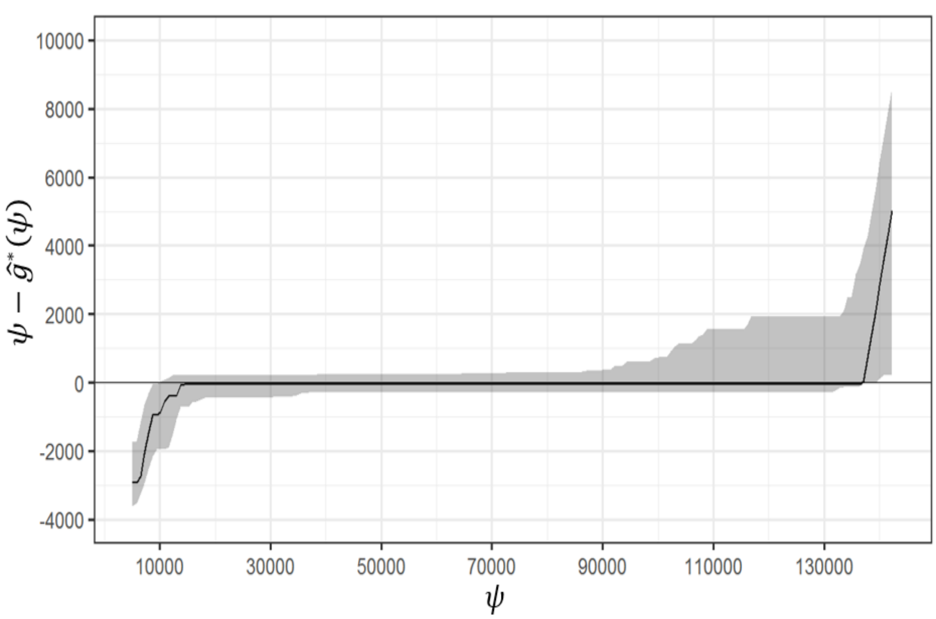


Figure 11: Average estimated function $\psi - \widehat{g}^{*}(\psi)$ (plain black) for annual earnings for those who have a College degree. The shaded grey area corresponds to the 95% bootstrap pointwise confidence interval. All results are displayed in 2015 US dollars.

F Proofs

F.1 Notation and preliminaries

For any set $\mathcal{H}$, let us denote by $l^\infty(\mathcal{H})$ the collection of all uniformly bounded real functions on $\mathcal{H}$ equipped with the supremum norm $\|f\|_{\mathcal{H}} = \sup_{x\in\mathcal{H}}|f(x)|$. 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 $\mathcal{H} = (\cup_{r\geq 1}\mathcal{H}_r)$. For $x \in \mathbb{R}^d$, $d > 1$, we denote by $\|x\|_\infty = \max_{j=1,\dots,d}|x|$.

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 exists 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.

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,

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

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

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

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 $E(Y'|\psi', X) = \psi'$. 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 $E(Y'|\mathcal{I}') = \psi'$. 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 $F_{\xi_Y + \varepsilon}$,

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

As a result, for all y,

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

Moreover, $\mathbb{E}\left(\widehat{Y}\right) = \mathbb{E}\left(\widehat{\psi}\right)$. By Theorem 1 again, $\widehat{Y}$ 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^*.

First, 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

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

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

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

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

With this notation, we have, by definition of T,

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

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

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

Remark first that this class is a particular case of classes $\mathcal{M}$ defined in (27) 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 $(y,h) \in \mathcal{Y} \times \mathcal{H}_r$, $\overline{K}_{n,F_0}(y,h) \to_P K_{F_0}(y,h) + \epsilon I_2$ and $K^*_{n,F_0}(y,h) \to_{P^*} K_{F_0}(y,h) + \epsilon I_2$, where I_2 is the 2×2 identity matrix.

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

Thus, using Equation (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,

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

where the equality follows by definition of $\mathcal{S}$ and $k_{F_0}(y,h)$. Similarly, using Assumption PS5 and (D.11) in AS, replacing T by T^* and quantities $\nu_{n,F_0}(y,h_{a,r})$ and $K_{n,F_0}(y,h_{a,r})$ 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 $r > 0$,

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

where $Q = \sum_{(a,r):(y,h_{a,r}) \in \mathcal{L}_{F_0}} q(r) < \infty$ and we use Problem 11.3 of Davydov et al. (1998) p.79 for the last inequality. Thus, $r > \underline{r}$ 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 $(y_0, h_0) \in \mathcal{L}_{F_0}$ such that either either $K_{F_0,11}(y_0,h_0) > 0$ or $K_{F_0,2}(y_0,h_0) > 0$. Then

$$\begin{aligned} \mathbb{P}(S > 0) &= 1 - \mathbb{P}\left(\sup_{y \in \mathcal{Y}} \sum_{(a,r):(y,h_{a,r}) \in \mathcal{L}_{F_0}} q(r) \mathcal{S}\left(\nu_{F_0}(y, h_{a,r}), K_{F_0}(y, h_{a,r}) + \epsilon I_2\right) = 0\right) \\ &\geq 1 - \mathbb{P}\left(\nu_{F_0,1}(y_0,h_0) \leq 0,\ \nu_{F_0,2}(y,h_0) = 0\right) \\ &\geq 1 - \min\left\{\mathbb{P}\left(\nu_{F_0,1}(y_0,h_0) \leq 0\right), \mathbb{P}\left(\nu_{F_0,2}(y_0,h_0) = 0\right)\right\} \\ &\geq 1/2. \end{aligned} \tag{18}$$

The first inequality holds by definition of the supremum and because $\mathcal{S}$ is nonnegative. To obtain the last inequality, note that either $\nu_{F_0,1}(y_0,h_0)$ is non-degenerate, in which case the first probability is 1/2 (since $\nu_{F_0,1}(y_0,h_0)$ is normal with zero mean), or $\nu_{F_0,2}(y_0,h_0)$ is non-degenerate, in which case the second probability is 0.

Finally, using that H is strictly increasing on $(0,\infty)$, (18) 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 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, we can apply Theorem E.1 in AS and the result follows. □

F.8 Proof of Theorem 3.

For any positive convex function ρ, we let

$$W_\rho(F,G)=\inf_{F_{U,V},U\sim F,V\sim G}\mathbb{E}\left[\rho\left(|U-V|\right)\right].$$

We also define

$$\mathcal{G}=\left\{G \text{ cdf}: \int_{-\infty}^{y}G(t)dt\le\int_{-\infty}^{y}F_Y(t)dt\ \forall y\in\mathbb{R},\ \int ydG(y)=\int ydF_Y(y)\right\}.$$

The proof is divided in three steps. First, we prove that the initial infimum is equal to $\inf_{G\in\mathcal{G}}W_\rho(F_\psi,G)$. Second, we prove that there is a unique G^* that reaches this infimum for all convex function $\rho:\mathbb{R}^+\to\mathbb{R}^+$ such that $\rho(0)=0$. Third, we prove that there is a unique function g^* such that (5) holds, and that this function is increasing.

1. $\inf_{(Y',\psi',\psi'')\in\boldsymbol{\Psi}}\mathbb{E}[\rho(|\psi'-\psi''|)]=\inf_{G\in\mathcal{G}}W_\rho(F_\psi,G)$**.**

First, by definition of W_ρ and because for all $(Y',\psi',\psi'')\in\boldsymbol{\Psi}$ $F_{\psi'}=F_\psi$, we have

$$\inf_{(Y',\psi',\psi'')\in\boldsymbol{\Psi}}\mathbb{E}[\rho(|\psi'-\psi''|)]=\inf_{G:\exists(Y',\psi',\psi'')\in\boldsymbol{\Psi}:\,F_{\psi''}=G}W_\rho(F_\psi,G).$$

Thus, it remains to prove that

$$\mathcal{G}'\equiv\left\{G:\exists(Y',\psi',\psi'')\in\boldsymbol{\Psi}:\ F_{\psi''}=G\right\}=\mathcal{G}. \tag{19}$$

First, let $G \in \mathcal{G}'$. Let $(Y', \psi', \psi'') \in \mathbf{\Psi}$ be such that $F_{\psi''} = G$. By definifion of Ψ, we have $\mathbb{E}(Y'|\psi'') = \psi''$ and $F_{Y'} = F_Y$. Therefore, by implication (i)⇒ (ii) of Theorem 1 applied to Y' and ψ'', $G = F_{\psi''} \in \mathcal{G}$. Hence, $\mathcal{G}' \subset \mathcal{G}$. Conversely, let $G \in \mathcal{G}$. Then, by implication (ii)⇒ (i) of Theorem 1, there exists (Y', ψ'') such that $Y' \sim Y$, $F_{\psi''} = G$ and $\mathbb{E}(Y'|\psi'') = \psi''$. Define $\psi' = \psi$. Then $(Y', \psi', \psi'') \in \Psi$ and $G \in \mathcal{G}'$. Equation (19) follows.

2. There exists a unique G^* such that for all ρ, $W_\rho(F_\psi, G^*) = \inf_{G\in\mathcal{G}} W_\rho(F_\psi, G)$.

Because F_ψ has no atom, the distribution of $H^{-1} \circ F_\psi(\psi)$ is H, for any cdf H. Hence, the set $\left\{F_{g(\psi)}, g \text{ measurable}\right\}$ is actually the set of all cdf's. Then, by Proposition 3.1 and Remark 3.2 in Gozlan et al. (2018), we have, for any convex function $\rho : \mathbb{R}^+ \to \mathbb{R}^+$ such that $\rho(0) = 0$,

$$\inf_{G\in\mathcal{G}} W_\rho(F_\psi, G) = \inf_{F_{Y'},\psi' :\ F_{Y'}=F_Y, F_{\psi'}=F_\psi} \mathbb{E}\left[\rho\left(\left|\psi' - \mathbb{E}\left[Y'|\psi'\right]\right|\right)\right]. \tag{20}$$

Third, by Theorem 1.4 in Gozlan et al. (2018), there exists a distribution G^* such that

1. For all f convex, $\int f dG^* \leq \int f dF_Y$;
2. for any convex function $\rho : \mathbb{R}^+ \to \mathbb{R}^+$ such that $\rho(0) = 0$,

$$\inf_{F_{Y'},\psi' :\ F_{Y'}=F_Y, F_{\psi'}=F_\psi} \mathbb{E}\left[\rho\left(\left|\psi' - \mathbb{E}\left[Y'|\psi'\right]\right|\right)\right] = W_\rho(F_\psi, G^*). \tag{21}$$

By, e.g., Proposition 2.3 in Gozlan et al. (2018), Point (1) is equivalent to G^* satisfying (iii) in Theorem 1. Therefore, in view of Theorem 1, we have $G^* \in \mathcal{G}$. Combining (20) and (21), we obtain:

$$G^* \in \arg\min_{G\in\mathcal{G}} W_\rho(F_\psi, G).$$

Now, $\mathcal{G}$ is convex. Moreover, by Lemma 3.2.1 of Pass (2013) and because F_ψ has no atom, $G \mapsto W_\rho(F_\psi, G)$ is strictly convex for $\rho(x) = x^2$. Therefore, G^* is the unique minimizer of $G \mapsto W_\rho(F_\psi, G)$ for this ρ. It is therefore the unique $G \in \mathcal{G}$ minimizing $W_\rho(F_\psi, G)$ for all convex function $\rho : \mathbb{R}^+ \mapsto \mathbb{R}^+$ such that $\rho(0) = 0$.

3. There exists a unique g^* such that $\mathbb{E}[\rho(|\psi - g^*(\psi)|)] = \inf_{(Y',\psi',\psi'')\in\mathbf{\Psi}} \mathbb{E}[\rho(|\psi' - \psi''|)]$ and g^* is increasing.

Let $g^* = G^{*-1} \circ F_\psi$. g^* is increasing. We now prove that it satisfies the equality above. First, by construction, $F_{g^*(\psi)} = G^*$. Moreover, by e.g., Theorem 5.26 of Villani (2008), g^* is the unique function satisfying

$$\mathbb{E}\left[\rho(|\psi - g^*(\psi)|)\right] = W_\rho(F_\psi, G^*). \tag{22}$$

This equation, together with the first and second steps, imply that

$$\mathbb{E}\left[\rho(|\psi - g^*(\psi)|)\right] = \inf_{(Y',\psi',\psi'')\in\mathbf{\Psi}} \mathbb{E}[\rho(|\psi' - \psi''|)]. \tag{23}$$

Now, consider $g \neq g^*$ such that $F_{g(\psi)} = G^*$. By unicity of g^* satisfying (22), we have $\mathbb{E}[\rho(|\psi - g(\psi)|)] > W_\rho(F_\psi, G^*)$. Finally, if $g \neq g^*$ is such that $F_{g(\psi)} = G \neq G^*$ for some $G \in \mathcal{G}$, we have, taking $\rho(x) = x^2$,

$$\mathbb{E}[\rho(|\psi - g(\psi)|)] \geq W_\rho(F_\psi, G) > W_\rho(F_\psi, G^*).$$

Therefore, g^* is the unique function satisfying (23) for all convex function $\rho : \mathbb{R}^+ \mapsto \mathbb{R}^+$ such that $\rho(0) = 0$.

F.9 Proof of Theorem 4.

First, in Step 1 of the proof of their Theorem 1.4, Gozlan et al. (2018) show that $\widehat{G}_L^*$, defined as the empirical distribution of $\left(\widetilde{\psi}_1, ..., \widetilde{\psi}_L\right)$ satisfies

$$\widehat{G}_L^* = \arg\min_{G \in \mathcal{G}} W_2(\widehat{F}_\psi, G),$$

where $\widehat{F}_\psi$ denotes the empirical cdf of ψ and for any $q \geq 1$, $W_p(F, G) = W_{\rho_q}(F, G)^{1/q}$ with $\rho_q(x) = |x|^q$. Given the definition of g^*, we also have $\widehat{g}^* = \widehat{G}_L^{*-1} \circ \widehat{F}_\psi$. Moreover, $\widehat{F}_\psi(x)$ converges almost surely to $F_\psi(x)$.

Let us focus hereafter on the event of probability one for which $\widehat{F}_\psi$ and $\widehat{F}_Y$ converges to F_ψ and F_Y, respectively, for the W_2 distance. On this event, consider any subsequence of $\left(\widehat{G}_L^*\right)_{L \in \mathbb{N}}$. Following Step 2 of the proof of Theorem 1.4 in Gozlan et al. (2018), but replacing $|x|$ by x^2 and using the fact that $\mathbb{E}(\psi^2) < +\infty$ and $\mathbb{E}(Y^2) < +\infty$, there exists a further subsequence converging for the W_2 distance. Moreover, the corresponding limit $\widetilde{G}$ satisfies, for all convex function $\rho : \mathbb{R}^+ \to \mathbb{R}^+$ such that $\rho(0) = 0$,

$$W_\rho\left(F_\psi, \widetilde{G}\right) = \inf_{F_{Y', \psi'}:\ F_{Y'} = F_Y, F_{\psi'} = F_\psi} \mathbb{E}\left[\rho\left(\left|\psi' - \mathbb{E}\left[Y'|\psi'\right]\right|\right)\right].$$

Hence, by the proof of Theorem 3, $W_\rho\left(F_\psi, \widetilde{G}\right) = W_\rho(F_\psi, G^*)$. Because

$$G^* = \arg\min_{G \in \mathcal{G}} W_2(F_\psi, G),$$

we have $\widetilde{G} = G^*$. Hence, any subsequence of $\left(\widehat{G}_L^*\right)_{L \in \mathbb{N}}$ admits a converging further subsequence converging to G^*. This implies that almost surely, $\left(\widehat{G}_L^*\right)_{L \in \mathbb{N}}$ converges to G^* for the W_2 distance. Because convergence for the W_2 distance implies weak convergence, $\left(\widehat{G}_L^*\right)_{L \in \mathbb{N}}$ converges weakly to G^*, almost surely. But then, by Lemma 21.2 in Van der Vaart (2000), $\left(\widehat{G}_L^{-1*}(x)\right)_{L \in \mathbb{N}}$ converges almost surely to $G^{*-1}(x)$, for all x that is a continuity point of G^{*-1}.

Finally, let us prove the almost sure convergence of $\widehat{g}^*(t)$ to $g^*(t)$ for all t that is a continuity point of g^* and such that $F_\psi(t) \in (0, 1)$. Fix $\varepsilon > 0$ and let us prove that for all L large enough, $|\widehat{g}^*(t) - g^*(t)| < \varepsilon$ with probability one. Because $F_\psi(t)$ is a continuity point of G^{*-1}, there exists $\delta > 0$ such that for all u satisfying $|u - F_\psi(t)| < \delta$, $|G^{*-1}(u) - G^{*-1}(F_\psi(t))| < \varepsilon/2$. It is easy to see that the set of points of discontinuity of G^{*-1} is at most countable. Thus, there exists $\eta \in (0, \delta)$ such that $F_\psi(t) + \eta$ and $F_\psi(t) - \eta$ are continuity points of G^{*-1}. Moreover, with probability one and for all L large enough, $|\widehat{F}_\psi(t) - F_\psi(t)| \leq \eta$. Then, for all L large enough and with probability one,

$$\widehat{G}_L^{*-1} \circ \widehat{F}_\psi(t) \leq \widehat{G}_L^{*-1} \circ [F_\psi(t) + \eta].$$

Because $F_\psi(t) + \eta$ is a continuity points of G^{*-1}, we have by what precedes that for all L large enough and with probability one,

$$\begin{aligned}\widehat{G}_L^{*-1} \circ \widehat{F}_\psi(t) &\leq \widehat{G}_L^{*-1} \circ [F_\psi(t) + \eta] \\ &\leq G^{*-1} \circ [F_\psi(t) + \eta] + \varepsilon/2 \\ &\leq G^{*-1} \circ F_\psi(t) + \varepsilon.\end{aligned}$$

Similarly, for all L large enough and with probability one, $\widehat{G}_L^{*-1} \circ \widehat{F}_\psi(t) \geq G^{*-1} \circ F_\psi(t) - \varepsilon$. The result follows by definition of $\widehat{g}^*(t)$.

F.10 Proof of Theorem 5.

Note first that because $F_{\mathbb{E}[Y|\mathcal{I}]}$ is continuous, $F_{\mathbb{E}[Y|\mathcal{I}]}(\mathbb{E}[Y|\mathcal{I}])$ is uniformly distributed (see,e.g. Van der Vaart, 2000, p.305). In turn, this implies that the cdf of $h^M(\mathbb{E}[Y|\mathcal{I}])$ is F_ψ. Hence, $(h^M(\mathbb{E}[Y|\mathcal{I}]), \mathbb{E}[Y|\mathcal{I}]) \in \Psi^M$. If for all (ψ', ψ''), $E[\rho(|\psi' - \psi''|)] = +\infty$, Equality (9) holds. If not, let $(\psi', \psi'') \in \Psi^M$ be such that $E[\rho(|\psi' - \psi''|)] < +\infty$. 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),

$$\begin{aligned}E[\rho(|\psi' - \psi''|)] &\geq \int_0^1 \rho\left(\left|F_\psi^{-1}(u) - F_{\mathbb{E}(Y|\mathcal{I}^M)}^{-1}(u)\right|\right) du. \\ &= \int \rho\left(\left|F_\psi^{-1} \circ F_{\mathbb{E}[Y|\mathcal{I}]}(v) - F_{\mathbb{E}[Y|\mathcal{I}]}^{-1} \circ F_{\mathbb{E}[Y|\mathcal{I}]}(v)\right|\right) dF_{\mathbb{E}[Y|\mathcal{I}]}(v) \\ &= \mathbb{E}\left[\rho\left(\left|h^M(\mathbb{E}[Y|\mathcal{I}]) - F_{\mathbb{E}[Y|\mathcal{I}]}^{-1} \circ F_{\mathbb{E}[Y|\mathcal{I}]}(\mathbb{E}[Y|\mathcal{I}])\right|\right)\right].\end{aligned} \tag{24}$$

Finally, note that $F_{\mathbb{E}[Y|\mathcal{I}]}^{-1} \circ F_{\mathbb{E}[Y|\mathcal{I}]}(v) < v$ 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, $F_{\mathbb{E}[Y|\mathcal{I}]}^{-1} \circ F_{\mathbb{E}[Y|\mathcal{I}]}(\mathbb{E}[Y|\mathcal{I}]) = \mathbb{E}[Y|\mathcal{I}]$ almost surely. Combined with Equation (24), this implies (9).

Now, let us suppose that ρ is strictly convex and let $(\psi', \mathbb{E}[Y|\mathcal{I}]) \in \mathbf{\Psi}^M$ satisfy (9). 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 $(\psi', \mathbb{E}[Y|\mathcal{I}])$ is equal to that of $(h^M(\mathbb{E}[Y|\mathcal{I}]), \mathbb{E}[Y|\mathcal{I}])$. Hence, conditional on $\mathbb{E}[Y|\mathcal{I}]$, ψ' is degenerate and equal to $h^M(\mathbb{E}[Y|\mathcal{I}])$. The result follows.

F.11 Proof of Theorem 6.

The functional $F \mapsto F^{-1}(t)$ is continuous with respect to the supremum norm at every F such that $F^{-1}(t)$ is continuous. Therefore, by the Glivenko-Cantelli and continuous mapping theorems, $\widehat{F}_{\psi}^{-1} \circ F_{\mathbb{E}[Y|\mathcal{I}]}(t)$ converges almost surely to $h^M(t)$ for all t such that $F_{\mathbb{E}[Y|\mathcal{I}]}(t)$ is a continuity point of F_{ψ}^{-1}. The second result follows from, e.g. Corollary 21.5 in Van der Vaart (2000). Finally, the third follows from the functional delta method for bootstrap (see, e.g., Van der Vaart, 2000, example 23.11).

F.12 Proof of Proposition 5

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

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

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 $(d, y', x, h, y, c, j) \in \{0,1\} \times \mathcal{Y} \times [0,1]^{d_X} \times \mathcal{H}_r \times \mathcal{Y} \times \mathcal{C}_s\left([0,1]^{d_X}\right) \times \{1,2\}$

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

  where $\delta > 0$ and for some functions M_j;

- **PS2**: for all sequence $F_n \in \mathcal{F}$ and $j \in \{1,2\}$, the i.i.d triangular array of processes

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

  are manageable with respect to some envelopes functions U_1 and U_2 (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

$$\sup_{F \in \mathcal{F}_0} \sup_{h \in \cup_{r \geq 1}\mathcal{H}_r, y \in \mathcal{Y}} \left\|\nu_{n,F}(y,h) - \overline{\nu}_{n,F}(y,h)\right\|_\infty = o_P(1), \tag{25}$$
$$\sup_{F \in \mathcal{F}_0} \sup_{h \in \cup_{r \geq 1}\mathcal{H}_r, y \in \mathcal{Y}} \left\|\nu^*_{n,F}(y,h) - \overline{\nu}^*_{n,F}(y,h)\right\|_\infty = o_{P^*}(1). \tag{26}$$

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 (25)-(26).

We prove these results for each coordinates of $\nu_{n,F}(y,h)$ and $\nu_{n,F}^*(y,h)$ separately. For that purpose, 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 conditions m (see (2), with $\widetilde{Y}$ replaced here by $\widetilde{Y}_c = Dq\left(\widetilde{Y}, c\right) + (1-D)\psi$). As the proof for the second coordinate is similar and much simpler than for the first one, we give a detailed proof only for the latter. 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

$$\mathcal{M}_1 = \left\{f_{c,y,\phi,h}\left(\widetilde{y},x\right) = \phi\left(y - q\left(\widetilde{y},c(x)\right)\right)^+ h(x), (c,y,\phi,h) \in \mathcal{C}_s\left([0,1]^{d_X}\right) \times \mathcal{Y} \times [N_1, M_1] \times \mathcal{H}\right\}, \tag{27}$$

$$\mathcal{M}_2 = \left\{f_{c,y,\phi,h}\left(\widetilde{y},x\right) = \phi\left(y - \widetilde{y}\right)^+ h(x),\ (c,y,\phi,h) \in \mathcal{C}_s\left([0,1]^{d_X}\right) \times \mathcal{Y} \times [N_1, M_1] \times \mathcal{H}\right\},$$

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

Note that ϕ_1, ϕ_2, and c in the class $\mathcal{M}$ denote components of m_1 that are estimated.

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

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

For $v := (c,y,\phi_1,\phi_2,h), v' := (c',y',\phi_1',\phi_2',h') \in \mathcal{C}_s\left([0,1]^{d_X}\right) \times \mathcal{Y} \times [N_1, M_1]^2 \times \mathcal{H}$ and $(\widetilde{y},x,d) \in \mathcal{Y} \times [0,1]^{d_X} \times \{0,1\}$, we have, by the triangular inequality and Assumptions 5-(i) and 5-(v),

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

Since $q(\widetilde{y},.)$ is Lipschitz and by convexity of $x \mapsto x^2$, we obtain

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

for some constant $K_q > 0$. Fix $\delta > 0$. If $\|v - v'\| \leq \delta$, then

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

Next, by Assumption 5-(iv),

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

Finally,

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

where Q_{q^I} is the Lipschitz constant of q^I. Thus, by Assumption 5, there exists $Q > 0$ such that

$$
\sup_{F \in \mathcal{F}_0} \mathbb{E}\left[\sup_{\|v - v'\| \leq \delta}\left| f_v\left(\widetilde{Y}, X, D\right) - f_{v'}\left(\widetilde{Y}, X, D\right)\right|^2\right] \leq Q\delta. \tag{28}
$$

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 $\mathcal{H}$ 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 $\mathcal{H}$ 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 5 (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 $\mathcal{F}$, we have to consider sequences F_n in $\mathcal{F}$ that are cdfs of $(D_{n,i}, Y_{n,i}, X_{n,i})_{i=1\dots n}$. We also define

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

Note that by Assumption 3-(iii), $\sigma_{F_n}^2 \geq \overline{\sigma} > 0$ 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 $0 < N_1 < M_1 := 1/\inf_F \sigma_F^2$, the i.i.d triangular array of processes

$$\mathcal{T}_{1,n,\omega} = \Big\{W_{n,i}\phi\left(y - \widetilde{Y}_{n,c(X_{n,i})}\right)^+ h(X_{n,i}),\ (c,y,\phi,h) \in \mathcal{C}_s\left([0,1]^{d_X}\right) \times \mathcal{Y} \times [N_1, M_1] \times \mathcal{H},$$
$$i \leq n, n \geq 1\Big\}$$
$$\mathcal{T}_{2,n,\omega} = \left\{W_{n,i}\phi\widetilde{Y}_{n,c(X_{n,i})}h(X_{n,i}),\ (c,y,\phi,h) \in \mathcal{C}_s\left([0,1]^{d_X}\right) \times \mathcal{Y} \times [N_1, M_1] \times \mathcal{H},\ i \leq n, n \geq 1\right\},$$

are manageable with respect to some envelopes functions U_1 and U_2. Lemma 3 in Andrews and Shi (2013) shows that the processes $\{h(X_{n,i}),\ h \in \mathcal{H},\ i \leq n, n \geq 1\}$ are manageable with respect to the constant function 1. Then, using Lemma D.5 in AS, it remains to show that

$$\mathcal{T}'_{1,n,\omega} = \left\{W_{n,i}\phi\left(y - \widetilde{Y}_{n,c(X_{n,i})}\right)^+,\ (c,y,\phi) \in \mathcal{C}_s\left([0,1]^{d_X}\right) \times \mathcal{Y} \times [N_1, M_1], i \leq n, n \geq 1\right\}$$
$$\mathcal{T}'_{2,n,\omega} = \left\{W_{n,i}\phi\widetilde{Y}_{n,c(X_{n,i})},\ (c,y,\phi) \in \mathcal{C}_s\left([0,1]^{d_X}\right) \times \mathcal{Y} \times [N_1, M_1], i \leq n, n \geq 1\right\},$$

are manageable with respect to some envelopes. For such envelopes, we can consider respectively $U'_1(\omega) = (M_0 + M)/(\overline{\sigma}\epsilon_0)$ and $U'_2(\omega) = M_0/(\overline{\sigma}\epsilon_0)$. We only prove manageability of $\mathcal{T}'_{1,n,\omega}$, as the reasoning is similar for $\mathcal{T}'_{2,n,\omega}$. Let us define

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

Reasoning as for the class $\mathcal{M}$ defined in (27), 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 is defined in (28). 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

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

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

$$N\left(\epsilon', [N_1, M_1]^2, |\cdot|\right) \leq Q_3\epsilon'^{-4} \text{ and } N\left(\epsilon', \mathcal{Y}, |\cdot|\right) \leq Q_4\epsilon'^{-2}. \tag{29}$$

This yields that,

$$\log\left(N_{[\cdot]}\left(\epsilon, \mathcal{M}', \|\cdot\|_2\right)\right) \leq (6 + Q_2)\max\left(-\log(\epsilon'), \epsilon'^{-d_X/s}\right) + \log(Q_3 Q_4). \tag{30}$$

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

$$\sup_{\omega\in\Omega, n\geq 1,\ \alpha\in\mathbb{R}^n_+} \mathcal{D}\left(\epsilon\left|\alpha \odot U'_1(\omega)\right|, \alpha \odot \mathcal{T}'_{1,n,\omega}\right) \leq \sup_{F\in\mathcal{F}_0} N\left(\frac{\epsilon}{2}, \mathcal{M}', \|\cdot\|_2\right),$$
$$\leq \sup_{F\in\mathcal{F}_0} N_{[\cdot]}\left(\epsilon, \mathcal{M}', \|\cdot\|_2\right), \tag{31}$$

where the second inequality follows as in e.g., Van der Vaart and Wellner (1996, p.84). Then, (30) 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 $\lambda(\epsilon) = \exp\left((6+Q_2)\max\left(-2\log\left(\epsilon/(2Q)\right), (\epsilon/(2Q))^{-2d_X/s}\right) + \log(Q_3 Q_4)\right)$ and using that $\sqrt{a+b}\leq\sqrt{a}+\sqrt{b}$ for $a,b\geq 0$,

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

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 5 (iii) and follow the proof of Lemma 7.2 (b) in AS where we replace Y by $q(Y, c(X))$ and B by $\max(M, M_0)$, which yields the result.

F.13 Proof of Proposition 6

(ii) $\Rightarrow$ (i). By construction and Theorem 1, Y and ψ^{b_0} satisfy H_0. Moreover, $\psi^{b_0}\in[\psi_L,\psi_U]$. Therefore, H_{0B} holds as well.

(i) $\Rightarrow$ (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 $F_{\psi_U}\leq F\leq F_{\psi_L}$, $\int ydF(y) = \int ydF_Y(y)$ and dominating at the second order F_Y. We show below that all $F\in\mathcal{D}$ are dominated at the second order by $F_{\psi^{c_0}}$. Then, because $F_{\psi_U}\leq F_{\psi^{c_0}}\leq F_{\psi_L}$ and $\int ydF_{\psi^{c_0}}(y) = \int ydF_Y(y)$, $\mathcal{D}$ is not empty only if $F_{\psi^{c_0}}$ dominates at the second order F_Y. The result then follows by Theorem 1.

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

$$F_{\psi^{c_0}} = \text{argmin}_{F_\psi\in\mathcal{D}}\int_{-\infty}^t F_\psi(y)dy. \tag{32}$$

Because $F_{\psi_U}(y)\leq F_\psi(y)$ for all $y<c_0$ and all $F_\psi\in\mathcal{D}$, we have, for all $t<c_0$,

$$\int_{-\infty}^t F_{\psi^{c_0}}(y)dy \leq \int_{-\infty}^t F_\psi(y)dy.$$

We now prove that (32) holds also for $t\geq c_0$.
First suppose that $t\geq c_0\vee 0$. For all $F_\psi\in\mathcal{D}$, $\int ydF_Y(y) = \int ydF_\psi(y)dy$. As a result, by Fubini's theorem,

$$\begin{aligned}
&-\int_{-\infty}^0 F_{\psi^{c_0}}(y)dy + \int_0^t \left(1-F_{\psi^{c_0}}(y)\right)dy + \int_t^\infty \left(1-F_{\psi^{c_0}}(y)\right)dy \\
&= -\int_{-\infty}^0 F_\psi(y)dy + \int_0^t \left(1-F_\psi(y)\right)dy + \int_t^\infty \left(1-F_\psi(y)\right)dy.
\end{aligned}$$

Because $F_\psi \le F_{\psi_L} = F_{\psi^{c_0}}$ on $[c_0, +\infty]$, this implies that

$$-\int_{-\infty}^{0} F_{\psi^{c_0}}(y)dy + \int_0^t \left(1 - F_{\psi^{c_0}}(y)\right) dy \ge -\int_{-\infty}^{0} F_\psi(y)dy + \int_0^t \left(1 - F_\psi(y)\right) dy$$

and thus (32) holds for $t \ge c_0 \vee 0$. Now, if $c_0 < 0$ and $t \in (c_0, 0)$, we have

$$\begin{aligned} & -\left(\int_{-\infty}^{t} F_{\psi^{c_0}}(y)dy + \int_t^0 F_{\psi^{c_0}}(y)dy\right) + \int_0^\infty \left(1 - F_{\psi^{c_0}}(y)\right) dy \\ &= -\left(\int_{-\infty}^{t} F_\psi(y)dy + \int_t^0 F_\psi(y)dy\right) + \int_0^\infty \left(1 - F_\psi(y)\right) dy. \end{aligned}$$

Using agin $F_\psi \le F_{\psi_L} = F_{\psi^{c_0}}$ on $[t, +\infty)$ yields

$$-\int_t^0 F_{\psi^{c_0}}(y)dy + \int_0^\infty \left(1 - F_{\psi^{c_0}}(y)\right) dy \le -\int_t^0 F_\psi(y)dy + \int_0^\infty \left(1 - F_\psi(y)\right) dy.$$

Therefore, the result also follows in this case.