Supplement to "Rationalizing rational expectations: Characterizations and tests"

Xavier D'Haultfoeuille
CREST-ENSAE

Christophe Gaillac
CREST-ENSAE and TSE

Arnaud Maurel
Department of Economics, Duke University, NBER, and IZA

In this Online Appendix, we first establish additional results with aggregate shocks. Then we discuss the robustness of linear regressions for testing rational expectations (RE) when expectations and realizations of the variable of interest are jointly observed but measured with errors. Third, we consider tests when only rounded expectations are observed. Fourth, we develop tests when the two samples are not representative of the same population. Fifth, we present additional simulations, with covariates. Sixth, we display additional material on the application. The last section gathers all remaining proofs.

S1. Additional results with aggregate shocks

S1.1 *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 $\widetilde{Y}_c = Dq(\widetilde{Y}, c) + (1-D)\psi$. 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 $\mathbb{E}[q(Y, c_0(X)) \mid X] = \mathbb{E}[\psi \mid X]$. When $q(y, c) = y - c$ or $q(y, c) = y/c$, we get respectively $c_0(X) = \mathbb{E}(Y \mid X) - \mathbb{E}(\psi \mid X)$ and $c_0(X) = \mathbb{E}(Y \mid X)/\mathbb{E}(\psi \mid X)$, and $\widehat{c}$ is easy to compute using nonparametric estimators of $\mathbb{E}(Y \mid X)$ and $\mathbb{E}(\psi \mid X)$.

Because in Proposition 3(ii) we do not test for a moment equality anymore, $m(D_i, \widetilde{Y}_i, X_i, g, y)$ reduces to $m_1(D_i, \widetilde{Y}_{c,i}, X_i, g, y)$. We let hereafter $\overline{m}_n(g, y) = \sum_{i=1}^n m_1(D_i, \widetilde{Y}_{c,i}, X_i, g, y)/n$. In the test statistic T, we replace, for $(y, g) \in \mathcal{Y} \times \bigcup_{r \geq 1} \mathcal{G}_r$, $\overline{\Sigma}_n(g, y)$ by $\overline{\Sigma}_n(g, y) = \widehat{\Sigma}_n(g, y) + \epsilon \text{Diag}(\widehat{\mathbb{V}}(\widetilde{Y}_{\widehat{c}}), \widehat{\mathbb{V}}(\widetilde{Y}_{\widehat{c}}))$, where $\widehat{\Sigma}_n(g, y)$ and $\widehat{\mathbb{V}}(\widetilde{Y}_{\widehat{c}})$ are respectively the

Xavier D'Haultfoeuille: xavier.dhaultfoeuille@ensae.fr
Christophe Gaillac: christophe.gaillac@tse-fr.eu
Arnaud Maurel: arnaud.maurel@duke.edu

sample covariance matrix of $\sqrt{n}\overline{m}_n(g,y)$ and the empirical variance of $\widetilde{Y}_{\widehat{c}}$. The last difference with the test considered in Section 3 is that when using the bootstrap to compute the critical value, we also have to reestimate c_0 in the bootstrap sample.

We obtain in this context a result similar to Theorem 2 above, under the regularity conditions stated in Assumption S1. We let hereafter $C_s([0,1]^{d_X})$ denote the space of continuously differentiable functions of order s on $[0,1]^{d_X}$ that have a finite norm $\|c\|_{s,\infty} = \max_{|\mathbf{k}|\leq s} \sup_{x\in[0,1]^{d_X}} |c^{(\mathbf{k})}(x)|$. We also let, for any function f on a set $\mathcal{G}$, $\|f\|_{\mathcal{G}} = \sup_{x\in\mathcal{G}} |f(x)|$. Finally, when the distribution of $(D, \widetilde{Y}, X)$ is F, K_F denotes the asymptotic covariance kernel of $n^{-1/2}\mathrm{Diag}(\mathbb{V}(\widetilde{Y}_{c_0}))^{-1/2}\overline{m}$.

ASSUMPTION S1.   (i) $\widehat{c}$ and c_0 belong to $C_s([0,1]^{d_X})$, with $s \geq d_X$. Moreover, $\|\widehat{c} - c_0\|_{[0,1]^{d_X}} = o_P(1)$.

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

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

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

  (v) *$\inf_{F\in\mathcal{F}} \mathbb{V}_F[\widetilde{Y}_c^2] > 0$ 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, $\widehat{\mathbb{V}}_F[\widetilde{Y}_{\widehat{c}}^2]$ is a consistent estimator of $\mathbb{V}_F[\widetilde{Y}_c^2]$.*

Part (i) imposes some regularity conditions on c_0 and its nonparametric estimator $\widehat{c}$. It is possible to check such regularity conditions on $\widehat{c}$ with kernel or series estimators of $\mathbb{E}(Y \mid X)$ and $\mathbb{E}(\psi \mid X)$. 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 S1 shows that under these conditions, the test has asymptotically correct size.

PROPOSITION S1.   *Suppose that $r_n \to \infty$ and that Assumptions 3 and S1 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 S1, modifying directly the proof of Proposition 2.

S1.2   *Impossibility results with more flexible effects of aggregate shocks*

We show here that restrictions in the way aggregate shocks affect the outcome are needed to be able to reject RE with F_Y and F_ψ. We consider for that purpose the following model:

$$Y = \sum_{k=0}^{K} C_k V^k + \varepsilon, \tag{S1}$$

where V is $\mathcal{I}$-measurable and the individual shock ε satisfies $E[\varepsilon \mid \mathcal{I}] = 0$. The vector $C := (C_0, \ldots, C_K)'$ represents aggregate shocks, which is assumed to be independent of $\mathcal{I}$, with support $\mathbb{R}^{K+1}$. We also assume that $E(C) = (0, 1, 0, \ldots, 0)'$, so that $V = \mathbb{E}[Y \mid \mathcal{I}]$ and under RE, $\psi = V$. Let $Q_c(y) = \sum_{k=0}^{K} c_k y^k$. Then $\mathbb{E}(Y \mid C = c, \mathcal{I}) = Q_c(V)$ and under RE, we have

$$\mathbb{E}(Y \mid C = c, \mathcal{I}) = Q_c(\psi).$$

Hence, as in Section 2.2.3, we consider the following hypothesis:

$\text{H}_{0SK}:$ there exist random variables (Y', ψ'), a sigma-algebra $\mathcal{I}'$ and $c \in \mathbb{R}^{K+1}$

such that $\sigma(\psi') \subset \mathcal{I}'$, $Y' \sim Y$, $\psi' \sim \psi$ and $\mathbb{E}[Y' \mid \mathcal{I}'] = Q_c(\psi')$.

The following proposition is a negative result on the possibility to test for H_{0SK}.

PROPOSITION S2. *Suppose that F_Y and F_ψ are continuous with supports that are bounded intervals. For any $\eta > 0$, there exists $K > 0$ and F, with $\sup_{u \in \mathbb{R}} |F(u) - F_\psi(u)| < \eta$, such that H_{0SK} holds with Y and $\widetilde{\psi} \sim F$ (instead of ψ).*

Proposition S2 states that as K grows large, the set of cdfs F_Y and F_ψ satisfying H_{0SK} (and thus RE in Model (S1)) becomes arbitrarily close, for the Kolmogorov–Smirnov metric, to the set of of cdfs F_Y and F_ψ that do not satisfy H_{0SK}. In other words, $\bigcup_{K \in \mathbb{N}} \text{H}_{0SK}$ is dense in the set of all continuous cdfs having bounded interval as supports. When combined with Theorem 2 in Bertanha and Moreira (2020), this implies that there does not exist any almost-surely continuous test of $\bigcup_{K \in \mathbb{N}} \text{H}_{0SK}$ that has nontrivial power.

A similar, negative result holds if aggregate shocks are allowed to vary with respect to unobserved, individual-specific variables. For instance, shocks may be sector-specific, but sectors may be unobserved in the data. To show such an impossibility result, consider the following model:

$$Y = q(C, U) + V + \varepsilon,$$

where both U and V are $\mathcal{I}$-measurable, C is an aggregate shock independent of $\mathcal{I}$ and the individual shock ε satisfies $E[\varepsilon \mid \mathcal{I}] = 0$. Thus, aggregate shocks affect the outcome in an additive way, but heterogeneously across individuals, depending on their U, which is assumed to be unobserved by the econometrician and can thus depend on V in a flexible way. We assume without loss of generality that $E[q(C, U) \mid \mathcal{I}] = 0$, so that $\psi = V$ under RE. Let us also assume that $q(u, c) = \sum_{k=0}^{K} c_k u^k$ and $U = \xi V$, with $\xi > 0$, $\xi \perp\!\!\!\perp V$ and $\mathbb{E}[\xi^k] < \infty$ for all $k \leq K$. Let $C'_k = E[\xi^k] C_k$ if $k \neq 1$, $C'_1 = E[\xi] C_1 - 1$ and $C' = (C'_0, \ldots, C'_K)'$. Then, under RE,

$$\mathbb{E}[Y \mid C' = c', \mathcal{I}] = \sum_{k=0}^{K} c'_k \psi^k.$$

Moreover, if $\text{Supp}(C) = \mathbb{R}^{K+1}$, we also have $\text{Supp}(C') = \mathbb{R}^{K+1}$, and no constraint is imposed on c'.[1] As a result, we are led again to test H_{0SK}, and the same negative result as above holds.

S2. Tests based on linear regressions with measurement errors

We suppose here to observe both $(\widehat{Y}, \widehat{\psi})$ satisfying (1). In this framework, we study the restrictions that RE entail on the coefficient β of the (theoretical) linear regression of $\widehat{Y}$ on $\widehat{\psi}$.

PROPOSITION S3. 1. *For any values of* $(\mathbb{V}(\widehat{Y}), \mathbb{V}(\widehat{\psi}), \text{Cov}(\widehat{Y}, \widehat{\psi}))$ *such that* $\mathbb{V}(\widehat{Y}) > \mathbb{V}(\widehat{\psi})$, *there exists a DGP compatible with this triple, satisfying* (1), *for which RE hold and such that* $\varepsilon + \xi_Y \perp\!\!\!\perp \psi$ *and* F_{ξ_ψ} *dominates at the second-order* $F_{\xi_Y + \varepsilon}$.

2. *If* $\beta < 1 - 1/(1 + \underline{\lambda})$ *for some* $\underline{\lambda} \geq 0$, *there exists no DGP compatible with this value of* β, *satisfying* (1), *for which RE hold and such that* $\text{corr}(\xi_\psi, \xi_Y + \varepsilon) \geq 0$ *and* $\mathbb{V}(\psi)/\mathbb{V}(\xi_\psi) \geq \underline{\lambda}$.

The first result is a negative one. It implies that without further restrictions than those already imposed in Proposition 4, the regression of $\widehat{Y}$ on $\widehat{\psi}$ does not bring any additional restriction related to RE. The second result, on the other hand, shows that if one assumes a positive correlation between ξ_ψ and $\xi_Y + \varepsilon$ and a lower bound on the signal-to-noise ratio $\mathbb{V}(\psi)/\mathbb{V}(\xi_\psi)$, then β is bounded from below under RE. The restriction $\text{corr}(\xi_\psi, \xi_Y + \varepsilon) \geq 0$ seems reasonable. First, given that the shocks ε cannot be anticipated, it is natural to assume that $\text{corr}(\xi_\psi, \varepsilon) = 0$. It then follows that the assumption $\text{corr}(\xi_\psi, \xi_Y + \varepsilon) \geq 0$ holds if the measurement errors on Y and ψ are positively correlated. This would typically happen, for instance, if individuals report their expectations and realized earnings omitting in both cases some components of their earnings, or if they instead overstate their realized earnings, and their expectations accordingly.

This proposition just focuses on the linear regression of $\widehat{Y}$ on $\widehat{\psi}$, since this regression has been very often used to test for RE. This means, however, that there may in principle be additional restrictions on the joint distribution of $(\widehat{Y}, \widehat{\psi})$ implied by RE.

S3. 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

---

[1] $E[q(C, U) \mid \mathcal{I}] = 0$ implies that $E[C_k] = 0$ for $k = 0, \dots, K$, but it does not restrict the set of possible c'_k.

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.[2] In such cases, we only observe (ψ_L, ψ_U), with $\psi_L \leq \psi \leq \psi_U$. For a thorough discussion of this issue, and especially of how to infer rounding practices, see Manski and Molinari (2010).

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

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

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

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

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

ASSUMPTION S2. $\mathbb{E}(|Y|) < \infty$, $\mathbb{E}(|\psi_L|) < \infty$ and $\mathbb{E}(|\psi_U|) < \infty$.

PROPOSITION S4. *Suppose that Assumption S2 holds. First, if $\mathbb{E}[\psi_L] \leq \mathbb{E}[Y] \leq \mathbb{E}[\psi_U]$, there exists a unique $F^* \in \mathcal{F}_B$ such that $\delta_{F^*} = 0$. Second, the following statements are equivalent:*

(i) H_{0B} *holds.*

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

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

---

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

S4. Tests with sample selection in the datasets

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

$$D \perp\!\!\!\perp (Y, \psi) \mid X. \tag{S2}$$

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

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

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

PROPOSITION S5. *Suppose that* (S2) *and Assumption* 1 *hold. Then* H_{0X} *is equivalent to*

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

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

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 a series logit estimator, for instance. Validity of such a test would follow using very similar arguments as for the test with aggregate shocks considered above.

S5. Simulations with covariates

We consider here simulations including covariates. The DGP is similar to that considered in Section 4. Specifically, we assume that $Y = \rho\psi + \sqrt{X}\varepsilon$, with $\rho \in [0, 1]$, $\psi \sim \mathcal{N}(0, 1)$, $X \sim \text{Beta}(0.1, 10)$ and

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

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, $\mathbb{E}[Y \mid X] = \mathbb{E}[\psi \mid X]$, 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 $n_\psi = n_Y = n \in \{400; 800; 1,200; 1,600; 3,200\}$, use 500 bootstrap simulations to compute the critical value, and rely on 800 Monte Carlo replications for each value of ρ and n. We use the same parameters $p = 0.05$ and $b_0 = 0.3$ as above.

Figure S1 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.
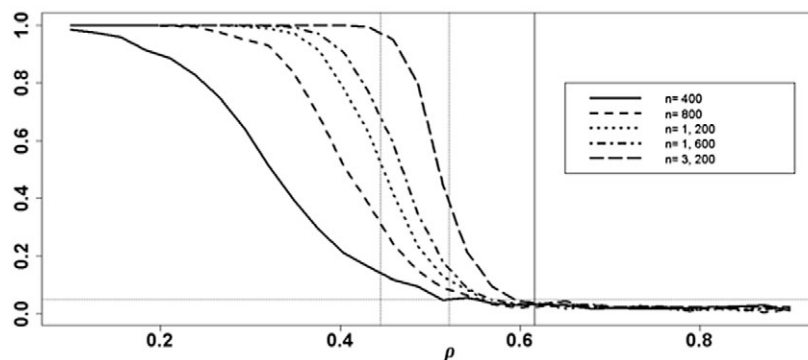
FIGURE S1. Power curves for the test with covariates. *Notes:* the dotted vertical lines correspond to the theoretical limit for the rejection of the null hypothesis for test based on variance ($\rho \simeq 0.445$), our test without covariates ($\rho \simeq 0.521$) and our tests with covariates ($\rho = 0.616$). The dotted horizontal line corresponds to the 5% level.

S6. ADDITIONAL MATERIAL ON THE APPLICATION

S6.1 *Effect of the winsorization on the RE test*

TABLE SI. Full test of RE with different levels of winsorization.

	Winsorization Level		
	p-Value		
	0.95	0.97	0.99
All	<0.001	<0.001	0.002
Women	<0.001	<0.001	<0.001
Men	0.210	0.238	0.326
White	0.015	0.023	0.037
Minorities	0.005	0.005	0.020
College degree	0.129	0.143	0.194
No college degree	0.015	0.014	0.012
High numeracy	0.013	0.017	0.034
Low numeracy	0.022	0.027	0.029
Tenure ≤6 months	0.001	0.003	0.009
Tenure >6 months	0.074	0.094	0.122

Note: We test H_{0S} with $q(y, c) = y/c$, using 5000 bootstrap simulations to compute the critical values. Distributions of realized earnings (Y) and earnings beliefs (ψ) are both Winsorized at either the 0.95, 0.97, or 0.99 quantile.

TABLE SII. Logit model of attrition.

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

Note: 1565 observations. Standard errors in parentheses.

S6.2 *Possibly endogenous attrition in the survey*

In addition to measurement errors, another potential issue when using the linked data (Y, ψ) is that attrition may be related to Y itself. This would create a sample selection issue that would invalidate the direct test, even absent any measurement errors. To explore this possibility, Table SII below reports the estimation results from a logit model of attrition on earnings beliefs, gender, race/ethnicity, college degree attainment, numeracy test score, tenure, and a (linear) time trend. The main takeaway from this table is that earnings beliefs ψ are significantly associated with attrition, even after controlling for this extensive set of characteristics. This result suggests that individuals for whom we observe both earnings expectations and realizations are likely to earn more than those who are not followed across the two waves. Along the same lines, a Kolmogorov–Smirnov test rejects at the 1% level the equality of the distributions of realized earnings between the whole sample and the subsample that would be used for the direct test. Similarly, we reject the equality of the distributions of expected earnings between these two samples. These results indicate that, in this context, the direct RE test is likely to be misleading. Conversely, attrition is unlikely to be an issue with our test, since we use in each wave the observations of all respondents.[3]

S7. Proofs

S7.1 *Notation and preliminaries*

For any set $\mathcal{G}$, let us denote by $l^\infty(\mathcal{G})$ the collection of all uniformly bounded real functions on $\mathcal{G}$ equipped with the supremum norm $\|f\|_{\mathcal{G}} = \sup_{x \in \mathcal{G}} |f(x)|$. Denote by $L^2(F)$ the square integrable space with respect to the measure associated with F, and let $\| \cdot \|_{F,2}$ be the corresponding norm. We let $N(\epsilon, \mathcal{T}, L_2(F))$ denote the minimal number of ϵ-balls with respect to $\| \cdot \|_{F,2}$ needed to cover $\mathcal{T}$. An ϵ-bracket (with respect to F) is a pair of real functions (l, u) such that $l \leq u$ and $\|u - l\|_{F,2} \leq \epsilon$. Then, for any set of real functions $\mathcal{M}$, we let $N_{[]}(\epsilon, \mathcal{M}, L_2(F))$ denote the minimum number of ϵ-brackets needed to cover $\mathcal{M}$. We denote by $\mathcal{G} = (\bigcup_{r \geq 1} \mathcal{G}_r)$. 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 $\sup_{F \in \mathcal{F}_0} \mathbb{P}_F(|U_n| >$

---

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

$M) < \epsilon$ 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 $(\widetilde{Y}_i, D_i, X_i)_{i=1,\dots,n}$. Convergence in distribution conditional on $(\widetilde{Y}_i, D_i, X_i)_{i=1,\dots,n}$ is denoted by $\to_{d^*}$.

S7.2 *Proof of Theorem 2*

(i) This is a particular case of Proposition S1 below, with $q(Y, c_0) = Y$. The proof is therefore omitted.

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

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

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

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

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

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

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

We have, by definition of T,

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

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

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

Remark that $\mathcal{M}_0$ is a particular case of classes $\mathcal{M}$ defined in (S6) below. Then, by the proof of Proposition S1 below, Assumptions PS1 and PS2 in AS are satisfied. Thus, the

assumptions of Lemma D.2 in AS hold as well. This entails that Assumptions PS4 and PS5 in AS hold. Namely, there exists a Gaussian process ν_{F_0} such that

- $\nu_{n,F_0} \to_d \nu_{F_0}$ and $\nu^*_{n,F_0} \to_{d^*} \nu_{F_0}$;

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

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

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

$$T \to_d \sup_{y \in \mathcal{Y}} \sum_{(a,r) \in A_r \times \mathbb{N}} \mathcal{S}\big(\nu_{F_0}(y, g_{a,r}) + k_{F_0}(y, g_{a,r}), K_{F_0}(y, g_{a,r}) + \epsilon I_2\big)$$

$$= S := \sup_{y \in \mathcal{Y}} \sum_{(a,r):(y,g_{a,r}) \in \mathcal{L}_{F_0}} q(r)\mathcal{S}\big(\nu_{F_0}(y, g_{a,r}), K_{F_0}(y, g_{a,r}) + \epsilon I_2\big),$$

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

2. The cdf H of S is continuous and strictly increasing in the neighborhood of any of its quantile of order $1 - \alpha > 1/2$ First, the cdf H of S is a convex functional of the Gaussian process ν_{F_0}. Then, as in the proof of Lemma B3 in Andrews and Shi (2013), we can use Theorem 11.1 of Davydov, Lifshits, and Smorodina (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 $r > 0$,

$$H(r) \geq \mathbb{P}\bigg(\sup_{y \in \mathcal{Y}} \sum_{(a,r):(y,g_{a,r}) \in \mathcal{L}_{F_0}} q(r)\mathcal{S}\big(\nu_{F_0}(y, g_{a,r}), K_{F_0}(y, g_{a,r}) + \epsilon I_2\big) < r\bigg)$$

$$\geq \mathbb{P}\bigg(\sup_{j \in \{1,2\},(y,a,r):(y,g_{a,r}) \in \mathcal{L}_{F_0}} \big|(K_{2,F_0,j,j}(y, g_{a,r}) + \epsilon)^{-1/2} \nu_{F_0,j}(y, g_{a,r})\big| < \frac{\sqrt{r/2}}{Q}\bigg)$$

$$> 0,$$

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

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

$$\mathbb{P}(S > 0) = 1 - \mathbb{P}\bigg(\sup_{y \in \mathcal{Y}} \sum_{(a,r):(y,g_{a,r}) \in \mathcal{L}_{F_0}} q(r)\mathcal{S}\big(\nu_{F_0}(y, g_{a,r}), K_{F_0}(y, g_{a,r}) + \epsilon I_2\big) = 0\bigg)$$

$$\geq 1 - \mathbb{P}\big(\nu_{F_0,1}(y_0, g_0) \leq 0, \nu_{F_0,2}(y, g_0) = 0\big)$$

$$\geq 1 - \min\big\{\mathbb{P}\big(\nu_{F_0,1}(y_0, g_0) \leq 0\big), \mathbb{P}\big(\nu_{F_0,2}(y_0, g_0) = 0\big)\big\}$$

$$\geq 1/2. \tag{S3}$$

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

Finally, using that H is strictly increasing on $(0, \infty)$, (S3) 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 $T^* \rightarrow_{d^*} S$ in distribution, Step 2 and Lemma 21.2 in Van der Vaart (2000), we have that for $\eta > 0$, $c^*_{n,\alpha} \rightarrow_{d^*} c(1 - \alpha + \eta) + \eta$, where $c(1 - \alpha + \eta)$ is the $(1 - \alpha + \eta)$-th quantile of the distribution of S. Because $T \rightarrow_d S$ and H is continuous at $c(1 - \alpha + \eta) + \eta > 0$, we obtain that

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

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

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

S7.3 *Proof of Proposition* S1

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

$$\nu_{n,F}(y, g) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathrm{Diag}\big(\widehat{\mathbb{V}}_F(\widetilde{Y}_{\widehat{c}})\big)^{-1/2}\big(m(D_i, \widetilde{Y}_{\widehat{c},i}, X_i, g, y) - E_{F,\widehat{c}}\big),$$

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

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

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

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

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

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

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

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

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

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

1. We first show that

$$\sup_{F \in \mathcal{F}_0} \sup_{g \in \bigcup_{r \ge 1} \mathcal{G}_r, y \in \mathcal{Y}} \left\| \nu_{n,F}(y, g) - \overline{\nu}_{n,F}(y, g) \right\|_\infty = o_P(1), \tag{S4}$$

$$\sup_{F \in \mathcal{F}_0} \sup_{g \in \bigcup_{r \ge 1} \mathcal{G}_r, y \in \mathcal{Y}} \left\| \nu_{n,F}^*(y, g) - \overline{\nu}_{n,F}^*(y, g) \right\|_\infty = o_{P^*}(1). \tag{S5}$$

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

1. Proof of (S4)–(S5)   We apply the uniform version over $F \in \mathcal{F}_0$ of Theorem 3 in Chen, Linton, and Van Keilegom (2003) to a general class of functions to which pertain the moment condition m (see (2), with $\widetilde{Y}$ replaced here by $\widetilde{Y}_c = Dq(\widetilde{Y}, c) + (1 - D)\psi$ and without the moment equality m_2). Hence, it suffices to verify that Assumptions (3.2) and (3.3) of Theorem 3 in Chen, Linton, and Van Keilegom (2003) are satisfied. Let us introduce, for any $0 < N_1 < M_1$, the classes of functions

$$\mathcal{M}_1 = \left\{ f_{c,y,\phi,g}(\widetilde{y}, x) = \phi(y - q(\widetilde{y}, c(x)))^+ g(x), \right.$$
$$\left. (c, y, \phi, g) \in \mathcal{C}_s([0,1]^{d_X}) \times \mathcal{Y} \times [N_1, M_1] \times \mathcal{G} \right\},$$
$$\mathcal{M}_2 = \left\{ f_{c,y,\phi,g}(\widetilde{y}, x) = \phi(y - \widetilde{y})^+ g(x), \right.$$
$$\left. (c, y, \phi, g) \in \mathcal{C}_s([0,1]^{d_X}) \times \mathcal{Y} \times [N_1, M_1] \times \mathcal{G} \right\}, \tag{S6}$$
$$\mathcal{M} = \left\{ f_{c,y,\phi_1,\phi_2,g}(\widetilde{y}, x, d) = (dg_{c,y,\phi_1,g} - (1-d)q_{c,y,\phi_2,g})(\widetilde{y}, x), g \in \mathcal{M}_1, q \in \mathcal{M}_2, \right.$$
$$\left. (c, y, \phi_1, \phi_2, g) \in \mathcal{C}_s([0,1]^{d_X}) \times \mathcal{Y} \times [N_1, M_1]^2 \times \mathcal{G} \right\}.$$

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

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

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

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

tions S1(i) and S1(v),

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

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

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

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

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

Next, by Assumption S1(iv), we obtain

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

Finally, we have

$$\mathbb{E}\left[\left|\mathbb{1}\{Y \le q^I(y', c(X))\} - \mathbb{1}\{\widetilde{y} \le q^I(y', c'(X))\}\right|\right]$$
$$\le \mathbb{E}\left[\mathbb{1}\{Y \le q^I(y', c(X)) - Q_{F,2}\delta\} - \mathbb{1}\{\widetilde{y} \le q^I(y', c(X)) + Q_{F,2}\delta\}\right]$$
$$\le \mathbb{E}\left[F_{Y|X}\left(q^I(y', c(X)) - Q_{q^I}\delta \mid X\right) - F_{Y|X}\left(q^I(y', c(X)) + Q_{q^I}\delta \mid X\right)\right]$$
$$\le 2Q_{F,1}Q_{q^I}\delta,$$

where Q_{q^I} is the Lipschitz constant of q^I. Thus, by Assumption S1, there exists $Q > 0$ such that

$$\sup_{F \in \mathcal{F}_0} \mathbb{E}\left[\sup_{\|v-v'\| \leq \delta} \left| f_v(\widetilde{Y}, X, D) - f_{v'}(\widetilde{Y}, X, D) \right|^2 \right] \leq Q\delta. \tag{S7}$$

Therefore, the class $\mathcal{M}$ satisfies Condition (3.2) of Theorem 3 in Chen, Linton, and Van Keilegom (2003) uniformly in $F \in \mathcal{F}_0$. Moreover, the class $\mathcal{G}$ is manageable and thus Donsker (see Lemma 3 in Andrews and Shi (2013)). Finally, by Remark 3(ii) in Chen, Linton, and Van Keilegom (2003), $\mathcal{C}_s([0,1]^{d_X})$ is also Donsker. Then, $\mathcal{C}_s([0,1]^{d_X})$, $\mathcal{Y}$, $[N_1, M_1]$, and $\mathcal{G}$ satisfy Condition (3.3) of Theorem 3 in Chen, Linton, and Van Keilegom (2003). The result follows by Theorem 3 in Chen, Linton, and Van Keilegom (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 S1(iii) and the proof of Lemma 7.2(a) in AS, PS1 is satisfied replacing B by $\max(M, M_0)$ in the proof of Lemma 7.2(a) in AS.

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

$$\widetilde{Y}_{n,c(X_{n,i})} = D_{n,i} q(Y_{n,i}, c(X_{n,i})) + (1 - D_{n,i})\psi_{n,i},$$

$$W_{n,i} = \frac{D_{n,i}}{\mathbb{E}_{F_n}[D_{n,i}]} - \frac{1 - D_{n,i}}{\mathbb{E}_{F_n}[1 - D_{n,i}]},$$

$$\sigma_{F_n}^2 = \mathbb{V}_{F_n}(\widetilde{Y}_{n,c(X_{n,i})}).$$

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

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

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

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

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

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

Reasoning as for the class $\mathcal{M}$ defined in (S6), and using the last equation of the proof of Theorem 3 in Chen, Linton, and Van Keilegom (2003, p. 1607), we have that for $\epsilon > 0$,

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

with $\epsilon' = (\epsilon/(2Q))^2$ and Q defined in (S7). Using Theorem 2.7.1, p. 155 in Van der Vaart and Wellner (1996), there exists a constant Q_2 depending only on s, d_X, and $[0,1]^{d_X}$ such that

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

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

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

This yields

$$\ln\big(N_{[\cdot]}\big(\epsilon, \mathcal{M}', \|\cdot\|_2\big)\big) \leq (6 + Q_2) \max\big(-\ln(\epsilon'), \epsilon'^{-d_X/s}\big) + \ln(Q_3 Q_4). \tag{S8}$$

Let $\odot$ denote element-by-element product and $\mathcal{D}(\epsilon | \alpha \odot U_1'(\omega)|, \alpha \odot \mathcal{T}_{1,n,\omega}')$ denote random packing numbers. By (A.1) in Andrews (1994, p. 2284), we have

$$\sup_{\omega \in \Omega, n \geq 1, \alpha \in \mathbb{R}_+^n} \mathcal{D}\big(\epsilon \big| \alpha \odot U_1'(\omega)\big|, \alpha \odot \mathcal{T}_{1,n,\omega}'\big) \leq \sup_{F \in \mathcal{F}_0} N\left(\frac{\epsilon}{2}, \mathcal{M}', \|\cdot\|_2\right)$$

$$\leq \sup_{F \in \mathcal{F}_0} N_{[\cdot]}\big(\epsilon, \mathcal{M}', \|\cdot\|_2\big),$$

where the second inequality follows as in, for example, Van der Vaart and Wellner (1996, p. 84). Then (S8) ensures (see Definition 7.9 in Pollard (1990, p. 38)) that

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

where $\lambda(\epsilon) = \exp((6 + Q_2) \max(-2\ln(\epsilon/(2Q)), (\epsilon/(2Q))^{-2d_X/s}) + \ln(Q_3 Q_4))$. Moreover, by using $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ for all $a, b \geq 0$,

$$\int_0^1 \sqrt{\ln\big(\lambda(\epsilon)\big)} \, d\epsilon \leq \sqrt{6 + Q_2} \int_0^1 \big[\max\big(-2\ln\big(\epsilon/(2Q)\big), \big(\epsilon/(2Q)\big)^{-2d_X/s}\big)\big]^{1/2} \, d\epsilon + \sqrt{\ln(Q_3 Q_4)}$$
$$< \infty.$$

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 S1(iii) and follow the proof of Lemma 7.2(b) in AS where we replace Y by $q(Y, c(X))$ and B by $\max(M, M_0)$. The result follows.

S7.4 *Proof of Proposition* S2

Hereafter, we let $[\underline{\psi}, \overline{\psi}]$ (resp., $[\underline{y}, \overline{y}]$) denote the support of ψ (resp., of Y). As in Lemma 1, H_{0SK} holds if and only if there exists a pair of random variables (Y', ψ') and c such that $Y' \sim Y$, $\psi' \sim \psi$ and $\mathbb{E}[Y' \mid \psi'] = Q_c(\psi')$. Now, if Q_c is strictly increasing on $[\underline{\psi}, \overline{\psi}]$, we have $\mathbb{E}[Y' \mid \psi'] = Q_c(\psi')$ if and only if $\mathbb{E}[Y' \mid Q_c(\psi')] = Q_c(\psi')$. In view of Theorem 1, the latter is equivalent to F_Y being a mean-preserving spread of $F_{Q_c(\psi')}$. Therefore, the proposition holds if for any $\eta > 0$, there exists K, $c \in \mathbb{R}^{K+1}$ and F such that (i) Q_c is strictly increasing on $[\underline{\psi}, \overline{\psi}]$; (ii) $\sup_{y \in \mathbb{R}} |F_\psi(y) - F(y)| < \eta$; (iii) F_Y is mean-preserving spread of $F_{Q_c(\widetilde{\psi})}$, with $\widetilde{\psi} \sim F$.

Fix $\eta > 0$. Since F_Y is continuous on $[\underline{y}, \overline{y}]$, it is uniformly continuous on this set. Hence, there exists η' such that

$$|y - y'| < \eta' \quad \Rightarrow \quad |F_Y(y) - F_Y(y')| < \eta. \tag{S9}$$

By assumption, $F_Y^{-1} \circ F_\psi$ is increasing and continuous. Then, by Theorem 9 in Mulansky and Neamtu (1998), there exists a sequence $(P_n)_{n \in \mathbb{N}}$ of increasing polynomials on $[\underline{\psi}, \overline{\psi}]$ satisfying $P_n(\underline{\psi}) = \underline{y}$ and $P_n(\overline{\psi}) = \overline{y}$ and converging uniformly to $F_Y^{-1} \circ F_\psi$. Hence, there exists P_{n_0} such that

$$\sup_{y \in [\underline{\psi}, \overline{\psi}]} | P_{n_0}(y) - F_Y^{-1} \circ F_\psi(y) | < \eta'. \tag{S10}$$

Let K be the degree of P_{n_0} and $c \in \mathbb{R}^K$ denote the vector of coefficients of P_{n_0}, so that $Q_c = P_{n_0}$. Q_c is a nonconstant polynomial, which is increasing on $[\underline{\psi}, \overline{\psi}]$. Hence, its derivative vanishes a finite number of times and Q_c is actually strictly increasing. Hence, Condition (i) above holds. Moreover, combining (S10) with (S9), we obtain

$$\sup_{y \in [\underline{\psi}, \overline{\psi}]} |F_Y \circ Q_c(y) - F_\psi(y)| < \eta.$$

Now, let $F := F_Y \circ Q_c$ on $[\underline{\psi}, \overline{\psi}]$, $F(y) := 0$ for all $y < \underline{\psi}$ and $F(y) := 1$ for all $y > \overline{\psi}$. Then F is continuous and increasing, with limit 0 and 1, respectively, at $-\infty$ and ∞. Thus, it is a cdf and Condition (ii) above holds. Finally, let $\widetilde{\psi} \sim F$. We have, for any $y \in [\underline{y}, \overline{y}]$,

$$P\big(Q_c(\widetilde{\psi}) \leq y\big) = F \circ Q_c^{-1}(y) = F_Y(y).$$

This implies that $F_{Q_c(\widetilde{\psi})}$ is a mean-preserving spread of F_Y. The result follows.

S7.5 *Proof of Proposition* S3

1. We consider for that purpose $(\psi^*, \xi_\psi^*, \xi_Y^*, \varepsilon^*) \sim \mathcal{N}(m, \Sigma)$, potentially different from the true $(\psi, \xi_\psi, \xi_Y, \varepsilon)$, and let

$$\widehat{\psi}^* = \psi^* + \xi_\psi^*,$$
$$\widehat{Y}^* = \psi^* + \varepsilon^* + \xi_Y^*.$$

We then fix (m, Σ) so that the DGP satisfies all the restrictions specified in the propositions, and in particular, $(\mathbb{V}(\widehat{Y}^*), \mathbb{V}(\widehat{\psi}^*), \mathrm{Cov}(\widehat{Y}^*, \widehat{\psi}^*)) = (\mathbb{V}(\widehat{Y}), \mathbb{V}(\widehat{\psi}), \mathrm{Cov}(\widehat{Y}, \widehat{\psi}))$. First, letting $m = (m_1, m_2, m_3, m_4)'$, we impose $m_2 = m_3 = m_4 = 0$, and set all the nondiagonal terms of Σ, except $\Sigma_{23} = \mathrm{Cov}(\xi_\psi^*, \xi_Y^*)$, equal to zero. Then $(\widehat{Y}^*, \widehat{\psi}^*, \psi^*)$ satisfy (1) and RE hold (considering $\mathcal{I} = \sigma(\psi^*)$ and $Y^* = \psi^* + \varepsilon^*$). We fix below $\Sigma_{22} \in [0, \mathbb{V}(\widehat{\psi})]$. Then let $\Sigma_{11} = \mathbb{V}(\widehat{\psi}) - \Sigma_{22}$ and $\Sigma_{33} = \mathbb{V}(\widehat{Y}) - \mathbb{V}(\widehat{\psi}) + \Sigma_{22}$ and $\Sigma_{44} = 0$, so that $(\mathbb{V}(\widehat{Y}^*), \mathbb{V}(\widehat{\psi}^*)) = (\mathbb{V}(\widehat{Y}), \mathbb{V}(\widehat{\psi}))$. Also, because $\mathbb{V}(\widehat{Y}) > \mathbb{V}(\widehat{\psi})$, $\mathbb{V}(\xi_\psi^*) < \mathbb{V}(\xi_Y^* + \varepsilon^*)$ and $F_{\xi_\psi^*}$ dominates at the second-order $F_{\xi_Y^* + \varepsilon^*}$.

Now, we fix Σ_{22}. Let $a = \mathbb{V}(\widehat{Y}) - \mathbb{V}(\widehat{\psi})$ and $c = \mathrm{Cov}(\widehat{Y} - \widehat{\psi}, \widehat{\psi})$. Then, by Cauchy–Schwarz inequality,

$$c^2 \le \mathbb{V}(\widehat{\psi})\mathbb{V}(\widehat{Y} - \widehat{\psi}) = \mathbb{V}(\widehat{\psi})(a - 2c).$$

This means that there exists $\sigma^2 \in [0, \mathbb{V}(\widehat{\psi})]$ such that

$$c^2 \le \sigma^2(a - 2c). \tag{S11}$$

Let $\Sigma_{22} = \sigma^2$ and $\Sigma_{23} = c + \Sigma_{22}$. Then, by construction,

$$\begin{aligned}
\mathrm{Cov}(\widehat{Y}^*, \widehat{\psi}^*) &= \Sigma_{11} + \Sigma_{23} \\
&= \mathbb{V}(\widehat{\psi}) - \Sigma_{22} + \Sigma_{22} + c \\
&= \mathrm{Cov}(\widehat{Y}, \widehat{\psi}).
\end{aligned}$$

Moreover, in view of (S11) and by definition of Σ_{22} and Σ_{33},

$$\begin{aligned}
\Sigma_{23}^2 &= c^2 + 2c\Sigma_{22} + \Sigma_{22}^2 \\
&\le (a - 2c)\Sigma_{22} + 2c\Sigma_{22} + \Sigma_{22}^2 \\
&= \Sigma_{33}\Sigma_{22}.
\end{aligned}$$

In other words, Σ is a proper covariance matrix.

2. Let $\lambda = \mathbb{V}(\psi)/\sigma_{\xi_\psi}^2$. If (1) and RE hold, $\mathrm{Cov}(\xi_\psi, \varepsilon + \xi_Y) \ge 0$ and $\lambda \ge \underline{\lambda}$, we obtain

$$\begin{aligned}
\beta - 1 &= \frac{\mathrm{Cov}(\widehat{Y} - \widehat{\psi}, \widehat{\psi})}{\mathbb{V}(\widehat{\psi})} \\
&= \frac{\mathrm{Cov}(\varepsilon + \xi_Y - \xi_\psi, \xi_\psi)}{\sigma_{\xi_\psi}^2(1 + \lambda)} \\
&\ge -\frac{1}{1 + \underline{\lambda}}.
\end{aligned}$$

The result follows.

S7.6 *Proof of Proposition* S4

We first prove that if $\mathbb{E}[\psi_L] \le \mathbb{E}[Y] \le \mathbb{E}[\psi_U]$, there exists a unique $F^* \in \mathcal{F}_B$ such that $\delta_{F^*} = 0$. First, suppose that $F^b \ne F^{b'}$ and, without loss of generality, $b > b'$. Then $\psi^b \le \psi^{b'}$,

implying that $F^b(y) \leq F^{b'}(y)$ for all y. Moreover, the inequality is strict for at least one y. As a result, $\mathbb{E}(\psi^b) > \mathbb{E}(\psi^{b'})$. In other words, there is at most one $F^* \in \mathcal{F}_B$ such that $\delta_{F^*} = 0$. If $\mathbb{E}[\psi_L] = \mathbb{E}[Y]$ or $\mathbb{E}[\psi_U] = \mathbb{E}[Y]$, such a solution also exists by taking $b = -\infty$ and $b = \infty$, respectively. Now, suppose that $\mathbb{E}[\psi_L] < \mathbb{E}[Y] < \mathbb{E}[\psi_U]$. For all $\infty > b > b' > -\infty$,

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

As a result, $|\psi^b - \psi^{b'}| \leq |b - b'|$. This implies that $\widetilde{\delta} : b \mapsto \mathbb{E}[\psi^b]$ is continuous. Moreover, $\lim_{b \to -\infty} \widetilde{\delta}(b) = \mathbb{E}[\psi_L] < \mathbb{E}(Y)$ and $\lim_{b \to \infty} \widetilde{\delta}(b) = \mathbb{E}[\psi_U] > \mathbb{E}(Y)$. By the intermediate value theorem, there exists b^* such that $\widetilde{\delta}(b^*) = \mathbb{E}(Y)$. Hence, there exists $F^* \in \mathcal{F}_B$ such that $\delta_{F^*} = 0$. The first part of Proposition S4 follows.

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

Now, let us prove that (i) implies (ii). Let us denote by $\mathcal{D}$ the set of all the cdfs for ψ such that H_{0B} holds. By Theorem 1, these are cdfs F satisfying $F_{\psi_U} \leq F \leq F_{\psi_L}$, $\delta_F = 0$ and dominating at the second-order F_Y. We show below that all $F \in \mathcal{D}$ are dominated at the second order by F^*. Then, because $F_{\psi_U} \leq F^* \leq F_{\psi_L}$ and $\int y \, dF^*(y) = \int y \, dF_Y(y)$, $\mathcal{D}$ is not empty only if F^* dominates at the second-order F_Y. The result then follows by Theorem 1.

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

$$F^* = \arg\min_{F_\psi \in \mathcal{D}} \int_{-\infty}^t F_\psi(y) \, dy. \tag{S12}$$

First, if $F^* = F^{-\infty}$, we have for all $F \neq F^*$, $F(y) \leq F_{\psi_L}(y) = F^*(y)$ for all y, with strict inequality for some y. Then $\delta_F > \delta_{F^*} = 0$ and $\mathcal{D} = \{F^*\}$, implying that (S12) holds. Similarly, (S12) holds if $F^* = F^\infty$.

Suppose now that $F^* = F^{b_0}$ for some $b_0 \in \mathbb{R}$. Because $F_{\psi_U}(y) \leq F_\psi(y)$ for all $y < b_0$ and all $F_\psi \in \mathcal{D}$, (S12) holds for all $t < b_0$. We now prove that (S12) holds also for $t \geq b_0$. First, suppose that $t \geq \max(b_0, 0)$. For all $F_\psi \in \mathcal{D}$, $\int y \, dF_Y(y) = \int y \, dF_\psi(y) \, dy$. As a result, by Fubini's theorem,

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

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

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

and thus (S12) holds for $t \geq \max(b_0, 0)$. Now, if $b_0 < 0$ and $t \in (b_0, 0)$, we have

$$-\left(\int_{-\infty}^{t} F^*(y)\,dy + \int_{t}^{0} F^*(y)\,dy\right) + \int_{0}^{\infty}\left(1 - F^*(y)\right)dy$$

$$= -\left(\int_{-\infty}^{t} F_\psi(y)\,dy + \int_{t}^{0} F_\psi(y)\,dy\right) + \int_{0}^{\infty}\left(1 - F_\psi(y)\right)dy.$$

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

$$-\int_{t}^{0} F^*(y)\,dy + \int_{0}^{\infty}\left(1 - F^*(y)\right)dy \leq -\int_{t}^{0} F_\psi(y)\,dy + \int_{0}^{\infty}\left(1 - F_\psi(y)\right)dy.$$

Therefore, the result also follows in this case.

References

Andrews, D. (1994), "Empirical process methods in econometrics." *Handbook of Econometrics*, 4, 2247–2294. [15]

Andrews, D. and X. Shi (2013), "Inference based on conditional moment inequalities." *Econometrica*, 81 (2), 609–666. [10, 14]

Bertanha, M. and M. J. Moreira (2020), "Impossible inference in econometrics: Theory and applications." *Journal of Econometrics*, 218, 247–270. [3]

Chen, X., O. Linton, and I. Van Keilegom (2003), "Estimation of semiparametric models when the criterion function is not smooth." *Econometrica*, 71 (5), 1591–1608. [12, 14, 15]

Davydov, Y. A., M. A. Lifshits, and N. V. Smorodina (1998), *Local Properties of Distributions of Stochastic Functionals*. American Mathematical Society. [10]

Manski, C. (2004), "Measuring expectations." *Econometrica*, 72 (5), 1329–1376. [4]

Manski, C. and F. Molinari (2010), "Rounding probabilistic expectations in surveys." *Journal of Business & Economic Statistics*, 28 (2), 219–231. [5]

Mulansky, B. and M. Neamtu (1998), "Interpolation and approximation from convex sets." *Journal of Approximation Theory*, 92 (1), 82–100. [16]

Pollard, D. (1990), *Empirical Processes: Theory and Applications*. NSF-CBMS Regional Conference Series in Probability and Statistics, i–86, Institute of Mathematical Statistics and the American Statistical Association. [12, 15]

Van der Vaart, A. (2000), *Asymptotic Statistics*. Cambridge University Press. [11]

Van der Vaart, A. and J. Wellner (1996), "Weak convergence." In *Weak Convergence and Empirical Processes*, 16–28, Springer. [15]