IS THE DISTRIBUTION OF RESOLVABLE UNCERTAINTY TYPE I EXTREME VALUE? A TEST FOR RANDOM COEFFICIENT MODELS USING CHOICE PROBABILITIES

ROMUALD MÉANGO[1]

[1] *University of Oxford and CESifo*

Abstract. Probabilistic stated choices allow respondents to express uncertainty about their intended decisions. They are typically paired with random-coefficient models (RCM) that impose Type I extreme value (EV1) resolvable uncertainty. This paper proposes a test of these assumptions based on the restrictions imposed by the EV1 distribution on the population distribution of interquantile ranges of ex ante returns. Our test is rejected in four studies published in leading economics journals. In these applications, EV1-based estimators often perform reasonably well for median willingness-to-pay (WTP), but are less reliable for distributional objects such as the full WTP distribution.

Keywords: random utility model; choice probability; stated preferences; resolvable uncertainty; willingness-to-pay.
JEL codes: C12, C21, D84.

Correspondence address: Romuald Méango, Department of Economics, University of Oxford, Manor Road Building, Manor Road, Oxford, OX1 3UQ, United Kingdom; email: romuald.meango@economics.ox.ac.uk. I acknowledge the support from the Higher Studies Fund - Travel Fund for the Economics of Developing Countries of the University of Oxford. I thank Esther Mirjam Girsberger, Marc Henry, Ismaël Mourifié, Martin Weidner, Basit Zafar, and participants to various seminars for useful discussions and comments at different stages of this project. All remaining errors are mine.

Date: This version: June 18, 2026.

1. Introduction

A recent development in stated preference research is the use of probabilistic stated choices, which allow respondents to express uncertainty about intended decisions (Juster, 1966; Manski, 1999). Blass et al. (2010) propose a random-coefficient model (RCM) to analyse choice probabilities for hypothetical scenarios. Following the logic of mixed logit models (McFadden and Train, 2000), they assume Type I extreme value (EV1) distributed resolvable uncertainty, which provides a convenient estimating equation. This assumption, now standard in the literature (e.g. Arcidiacono et al., 2020; Aucejo et al., 2023; Boneva et al., 2022; Wiswall and Zafar, 2018, 2021), is rarely tested despite being central to identification and estimation.

The main contribution of this paper is to provide and implement a test for the EV1 assumption in RCMs using choice probabilities. The test makes use of a definition of ex ante returns, say S, the minimum pecuniary compensation needed by the agent to choose a given option. Under the RCM framework and the assumption of an EV1 distribution of the resolvable uncertainty, the conditional distribution of S is a logistic distribution, and its interquantile range (IQR) is invariant up to a transformation.

Section 2 derives the invariance result and introduces a corollary to Theorem 1 in Méango and Girsberger (2026)(MG2026 hereafter) that identifies the distribution of IQR in the population nonparametrically. Section 3 exploits the invariance to derive a test of the validity of the EV1 assumption and a weaker symmetry restriction. Section 4 applies the method to four existing studies published in leading economic journals.

2. RCM of probabilistic choice and identification of the IQR

Consider a binary choice set $\{0, 1\}$.[1] A RCM describes the utility of agent i for option j as

$$U_{ij} = \frac{1}{\sigma_i} \left(\gamma_i Y_{ij} + X_{ij}\beta_i + \varepsilon_{ij} \right), \quad j = 0, 1, \tag{1}$$

with $\gamma_i, \sigma_i > 0$; Y_{ij} is the *numeraire* (e.g., income), which helps to translate preference parameters into willingness-to-pay; X_{ij} are the choice characteristics, and β_i is a vector that describes individual-specific preferences. In standard models, the taste shock ϵ_{ij} is known to the agent but not to the analyst; in incomplete stated scenarios, the agent learns ϵ_{ij} only at decision time. This *resolvable uncertainty* is precisely what probabilistic answers can capture.

In the RCM framework, it is easy to show that the ex ante returns of option 1 over option 0 are

$$S(w, \eta_i, \nu_i) = (y_1 - y_0) + (x_1 - x_0)\frac{\beta_i}{\gamma_i} + \nu_i \tag{2}$$

where $w = (y_1, y_0, x_1, x_0)$, $\eta_i = (\beta_i, \gamma_i, \sigma_i)$, and $\nu_i = (\varepsilon_{i1} - \varepsilon_{i0})/\gamma_i$. Under the EV1 assumption ($\varepsilon_{ij} \sim EV1$), $S(w, \eta_i, \nu_i)$ conditional on η_i is logistic with a standard deviation

[1] The framework applies to polychotomous choices as well, with a minor change of notation (MG2026).

$\sigma_{S,i} = \sigma_i/\gamma_i$. This implies the interquantile range

$$IQR_{S,i}(\tau, 0.5) = \sigma_{S,i}\ell(\tau), \qquad \text{with } \ell(\tau) = \left|\log\left(\frac{\tau}{1-\tau}\right)\right|. \tag{3}$$

Hence, the population distribution of $IQR_{S,i}(\tau, 0.5)/\ell(\tau)$ is invariant in τ. This invariance is the *key restriction* tested in this paper. We also test the weaker assumption that ν_i is symmetrically distributed; that is, $IQR_{S,i}(\tau, 0.5) = IQR_{S,i}(1-\tau, 0.5)$ for all $\tau \in (0, 1)$.

To exploit this restriction, we characterise the distribution of $S(w, \eta_i, \nu_i)$ even without the EV1 assumption. We use the framework of MG2026.[2]

Given a hypothetical scenario $W_i = (Y_{i1}, Y_{i0}, X_{i1}, X_{i0})$, respondent i states

$$P_i = m(W_i, \eta_i) = \Pr\left(S(W_i, \eta_i, \nu_i) \geq 0 \mid \eta_i\right). \tag{4}$$

The mapping $w \mapsto m(w, \eta)$ defines the *stated demand function*. The second equality comes from reinterpreting the elicited choice probability: the probability of choosing option 1 is the probability that the return S is positive. In line with the literature, the stated choice experiment is construed as a *ceteris paribus* experiment (e.g. Dominitz and Manski, 1996) where the respondents report their choice as if W_i were exogenous.

Theorem 1 in MG2026 showed that under the random assignment of the scenarios, we can recover the population distribution of quantiles of $S(\cdot)$ from the joint distribution of (P, W). Corollary 1 in the Supplemental Appendix shows that, using the same logic, one can also recover the population distribution of IQRs. The precise statement and proof of this result are available in the Supplemental Appendix.

3. Test Procedure

The restrictions imposed by the EV1 and symmetry assumptions in RCMs can be expressed as unconditional moment equalities. Denote by $IQR_S(\tau, 0.5; W, \eta)$ the interquantile range of ex ante returns given a hypothetical scenario W and unobserved heterogeneity η. Let:

$$G_\tau(y) = \Pr(IQR_S(\tau, 0.5; W, \eta) \leq y\ell(\tau)).$$

For a finite collection $T = \{\tau_1, \ldots, \tau_K\}$:

$$H_0^{EV1}: \quad G_\tau(\cdot) = G_{\tau'}(\cdot), \quad \forall \tau, \tau' \in T, \tag{5}$$

$$H_0^{SYM}: \quad G_\tau(\cdot) = G_{1-\tau}(\cdot), \quad \forall \tau \in T. \tag{6}$$

The hypotheses imply a set of unconditional moment equalities. Following Andrews and Soares (2010), we construct a test statistic based on differences in estimated $G_\tau(y)$ across quantiles. The logic of the test procedure is the same for both tests, so we discuss the test for the first hypothesis only.

[2] MG2026 develops a general identification strategy for quantiles, mean returns, willingness-to-pay, and related policy objects using probabilistic stated choice experiments under nonparametric assumptions; the present manuscript adds a diagnostic test of EV1 and symmetry using interquantile ranges.

Define for a finite collection $\mathcal{Y}$ in $\mathbb{R}^+$ the long vector M obtained by collecting all differences $G_\tau(y) - G'_\tau(y)$ for $\tau \neq \tau', \tau, \tau' \in \mathcal{T}$ and $y \in \mathcal{Y}$. Let $\dim(M)$ denote its dimension. Equation (12) implies that:

$$S(M, \Omega) := (M'\Omega M)^{1/2} = 0 \tag{7}$$

for any conformable, positive, semi-definite matrix Ω.

The proposed test statistic estimates $Q_{P|W}$ by linear quantile regression and uses it to recover empirical counterparts of $G_\tau(y)$. With the empirical version of $G_\tau(\cdot)$, $\widehat{G}_{\tau,n}(y)$, it constructs $\widehat{M}_n$, the empirical counterpart of M. For Ω, it uses the inverse of the variance-covariance matrix of the random vector M, a matrix of size $\dim(M) \times \dim(M)$. The empirical counterpart $\widehat{\Omega}_n$ is constructed using a bootstrap sample. Finally, the critical value is obtained using the plug-in method in Section 7 of Andrews and Soares (2010).

Before turning to the empirics, it is important to emphasise that although the main restriction is derived from the EV1 assumption, a rejection of the test should be regarded as the rejection of at least one of the assumptions of the framework: EV1, RCM representation, and/or ceteris paribus experiment. We provide a step-by-step explanation of the implementation in a Supplemental Appendix. An example Matlab routine is available online.

4. Four empirical applications

This section applies the test to four stated choice experiments. The first, Wiswall and Zafar (2018) (WZ2018), studies preferences for job attributes among 247 NYU students who choose among three unlabelled jobs in eight scenarios. The second, MG2026, surveys 587 Ivorian students choosing between public- and private-sector jobs across five scenarios. The third, Koşar et al. (2022)(KRV2022), uses data from 1,861 US residents (SCE panel) who report migration probabilities across up to 22 counterfactual locations. The final, Aucejo et al. (2023)(AFZ2023), elicits valuations of in-person instruction and social activities from 1,150 ASU students in 42 scenarios, with tuition cost as the numeraire.

These four applications are broadly aligned with the proposed framework, though their suitability varies. Table C.1 in the Online Appendix discusses how closely each application matches the assumptions. In all cases, the rejection of the test is formally a rejection of at least one component of the maintained framework. The rejection is most directly informative about EV1 when the ceteris paribus condition is credible, separability restrictions are modest, and conditional quantiles can be estimated reliably.

WZ2018 and KRV2022 provide explicit within-scenario ceteris paribus instructions: alternatives differ only in the listed attributes and are otherwise identical. Thus, the validity of the ceteris paribus relies on the separability of respondents' utility between listed and non-listed attributes, as in the RCM. By contrast, MG2026 provide an explicit ceteris paribus instruction across scenarios. The RCM is not required, but there remains a concern about whether respondents internalise the instruction. AFZ2023 do not instruct respondents to hold all unspecified factors fixed. Instead, identification relies on within-student variation across COVID-related states and tuition levels, plus the assumption that outside options are constant across scenarios conditional on COVID status.

Both in WZ2018 and AFZ2023, the scenario support is limited, and the estimation of conditional quantiles relies strongly on the RCM framework. Similarly, since different attributes are presented in different blocks, KRV2022 also relies on the RCM. In MG2026, job attributes are drawn randomly on a pre-specified support, generating richer joint variation.

Extreme stated probabilities are important in KRV2022 and AFZ2023. These mass points at 0 or 100 can make the estimated conditional quantile functions less stable, especially at low or high quantiles.

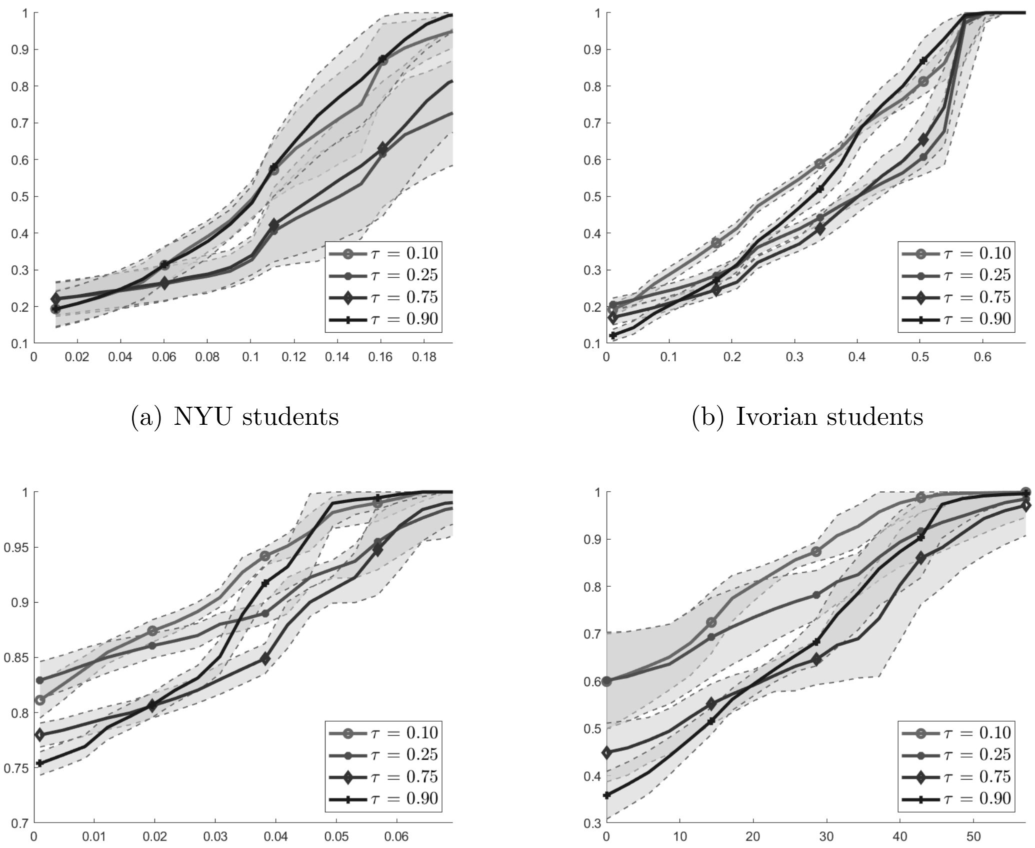
Overall, the data from MG2026 appear to provide the clearest case for interpreting rejection as evidence against the EV1 component of the maintained RCM framework. In the other applications, the test remains informative, but rejection should be interpreted more cautiously as evidence against the relevant jointly maintained framework.

Figure 4.1 displays the estimated cumulative distribution functions $\widehat{G}_\tau(y)$ and the pointwise confidence region following the procedure of MG2026. We can make three remarks: First, we refute the assumption 'that respondents make the same assumption subjectively' (Blass et al., 2010, p.424), that is, $\sigma_i = \sigma$, for all i. Second, the invariance property seems to be violated as the distributions appear to differ significantly. Third, symmetry seems to hold in the WZ2018 sample and not in the other samples. In the former, the elements of each pair $(G_{0.10}, G_{0.90})$ and $(G_{0.25}, G_{0.75})$ are very close. This should be expected because the hypothetical jobs in the WZ2018 sample are not labelled and are a priori symmetric.

The test is conducted for a coarse (10 points) and fine (20 points) grid of the support of quantiles, T, a coarse (25 points) and fine (50 points) grid of the support of *numéraire*, and using two different variance-covariance matrices (one diagonal and one full) proposed by (Andrews and Soares, 2010, see matrices S_1 and S_2, p.126-127). The results are very consistent. Table 4.1 presents the test statistic and the associated critical values at conventional levels for the finer grids and the full variance-covariance matrix. The test unequivocally rejects the RCM + EV1 framework in all four samples at conventional levels. As expected, the symmetry assumption is not rejected for the NYU sample but is rejected in the remaining three samples.

5. Concluding remarks

A rejection of the test does not necessarily imply that EV1-based methods are substantively misleading for every target parameter in every application. Table D.1 in the supplemental material computes the median willingness-to-pay (WTP) in the population for at least two attributes in each of the empirical applications using five estimation



(a) NYU students

(b) Ivorian students

(c) US residents (SCE panel)

(d) ASU Students

FIGURE 4.1. Estimates of the population distributions $G_\tau(y)$, for $\tau = 0.1, 0.25, 0.75, 0.90$, with 90% confidence intervals

Note: The figures display the estimated cumulative distribution functions $\widehat{G}_\tau(\cdot)$ and the pointwise confidence region following the procedure of MG2026. The solid lines represent the estimated distributions, and the shaded regions represent the confidence regions.

procedures: the LAD estimation of Blass et al. (2010), the individual-specific LAD estimator of WZ2018, the individual-specific quasi-maximum likelihood estimator of AFZ2023, the quantile regression estimator of MG2026, and the maximum score (MS) estimator proposed by Blass et al. (2010) and implemented as in Florios and Skouras (2008). The first three procedures rely on the EV1 assumption.

We test equality with the MG2026 estimator since it relies on the weakest set of assumptions. The estimators based on EV1 perform relatively well in general for estimating the median WTP (4 rejections at 10% out of 18 possible), except in the case of KRV2022 (5 rejections out of 6 possible).

The performance is worse when the goal is to estimate the full distribution of WTP. Figure 5.1 illustrates this. For the NYU sample, it compares women's WTP distributions for allowing part-time work at the same wage. Estimates based on AFZ2023, which rely on the EV1 assumption, are contrasted with those from MG2026. Visually, the distributions differ, with parametric estimates being more dispersed than the semiparametric ones. A

Sample	H_0	Test Statistic	Critical values 10%	5%	1%	Decision
NYU	EV1	51.09	35.83	36.18	36.88	Reject at 1%
	Symmetry	13.64	16.49	16.81	17.37	Do not reject
Ivoirian	EV1	223.19	37.64	38.01	38.68	Reject at 1%
	Symmetry	20.36	18.11	18.44	19.06	Reject at 1%
SCE	EV1	55.91	35.86	36.19	36.87	Reject at 1%
	Symmetry	21.83	16.80	17.12	17.72	Reject at 1%
ASU	EV1	68.24	37.29	37.61	38.33	Reject at 1%
	Symmetry	18.72	17.23	17.54	18.14	Reject at 1%

Notes: The Table presents the test statistic, the critical values at conventional levels, and the test decision. The simulated critical values use $L = 10,000$ draws for the normal distribution. The test uses 500 replications to compute the standard error of $\widehat{\Sigma}_n$.

TABLE 4.1. Test results

formal test rejects the null hypothesis of equality at conventional levels for the WTP for part-time work, the other WTP in Table D.1, and when using the WZ2018 procedure.

These findings suggest that researchers should exercise caution and, depending on their context and target parameter, rely on identification strategies that remain valid under weaker assumptions. Available options include, for discrete stated choices, the special regressor approach of Lewbel et al. (2011), and for probabilistic stated choices, the quantile regression approach of Méango and Girsberger (2026).

REFERENCES

Andrews, D. W. and Soares, G. (2010). Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection. *Econometrica*, 78(1):119–157.

Arcidiacono, P., Hotz, V. J., Maurel, A., and Romano, T. (2020). Ex Ante Returns and Occupational Choice. *Journal of Political Economy*, 128(12):4475–4522.

Aucejo, E. M., French, J., and Zafar, B. (2023). Estimating Students' Valuation for College Experiences. *Journal of Public Economics*, 224:104926.

Blass, A. A., Lach, S., and Manski, C. F. (2010). Using Elicited Choice Probabilities to Estimate Random Utility Models: Preferences for Electricity Reliability. *International Economic Review*, 51(2):421–440.

Boneva, T., Golin, M., and Rauh, C. (2022). Can Perceived Returns Explain Enrollment Gaps in Postgraduate Education? *Labour Economics*, 77:101998. European Association of Labour Economists, World Conference EALE/SOLE/AASLE, Berlin, Germany, 25 – 27 June 2020.

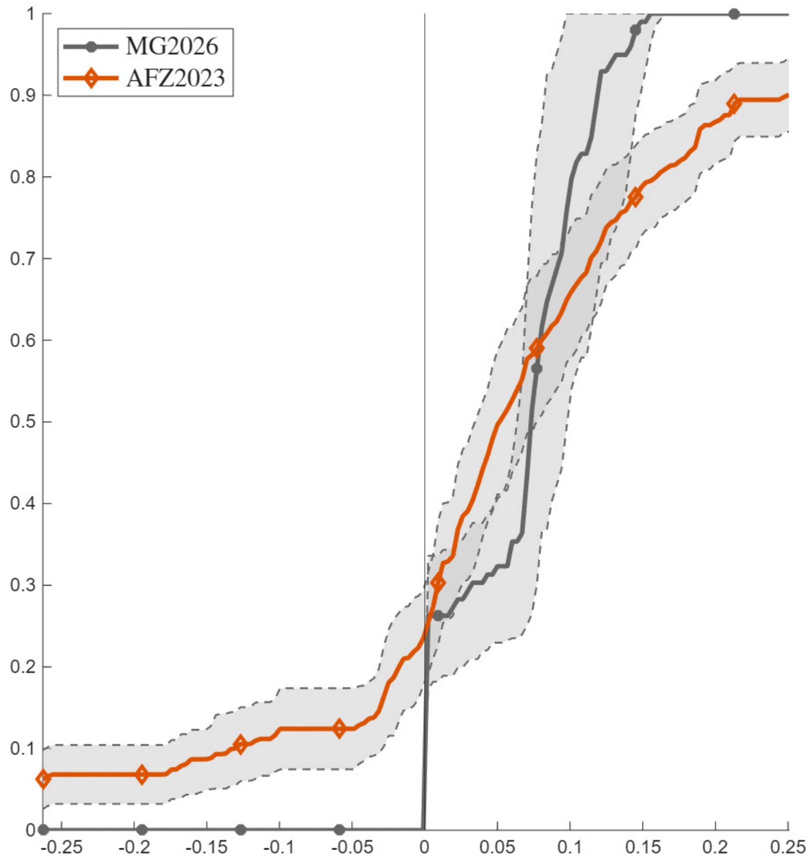


FIGURE 5.1. Comparing estimates of women's WTP distributions for adding the option of working part-time.

Notes: The estimated distribution WZ2018 uses the methodology of Aucejo et al. (2023). The estimated distribution MG2026 uses Theorem 2 for $qWTP$ and $\tau = 0.5$. The shaded regions represent the 90% confidence regions.

Chernozhukov, V., Fernández-Val, I., Newey, W., Stouli, S., and Vella, F. (2020). Semiparametric Estimation of Structural Functions in Nonseparable Triangular Models. *Quantitative Economics*, 11(2):503–533.

Dominitz, J. and Manski, C. F. (1996). Eliciting Student Expectations of the Return to Schooling. *Journal of Human Resources*, 31(1):1–26.

Florios, K. and Skouras, S. (2008). Exact computation of max weighted score estimators. *Journal of Econometrics*, 146(1):86–91.

Juster, F. T. (1966). Consumer Buying Intentions and Purchase Probability: an Experiment in Survey Design. *Journal of the American Statistical Association*, 61(315):658–696.

Karr, A. F. (1993). *Probability*. Springer New York, New York, NY.

Koşar, G., Ransom, T., and Van der Klaauw, W. (2022). Understanding Migration Aversion Using Elicited Counterfactual Choice Probabilities. *Journal of Econometrics*, 231(1):123–147.

Lewbel, A., McFadden, D., and Linton, O. (2011). Estimating Features of a Distribution from Binomial Data. *Journal of Econometrics*, 162(2):170–188.

Manski, C. F. (1999). Analysis of Choice Expectations in Incomplete Scenarios. *Journal of Risk and Uncertainty*, 19:49–66.

McFadden, D. and Train, K. (2000). Mixed MNL Models for Discrete Response. *Journal of applied Econometrics*, 15(5):447–470.

Méango, R. and Girsberger, E. M. (2026). Just Ask Them Twice: Choice Probabilities and Identification of Ex Ante Returns and Willingness-to-pay. *The Economic Journal*, page ueag018.

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.

Wiswall, M. and Zafar, B. (2021). Human Capital Investments and Expectations about Career and Family. *Journal of Political Economy*, 129(5):1361–1424.

Online Appendix

Appendix A. Statement and Proof of Corollary 1

Definition 1. *Define by* (P, W) *the random vector of stated choice and scenario attributes observed by the analyst. We denote with* $Q_{P|W}$ *the conditional (on* W*) quantile of the variable* P*.*

We assume that $F_{\nu|\eta}$ *is a continuous distribution and define:* $F_{S,i}(s; W) := \Pr(S(W, \eta_i, \nu_i) \leq s|\eta_i)$ *and* $Q_{S,i}(\tau; W) := \inf\{s : F_{S,i}(s; W) \geq \tau\}$*, respectively the individual-specific distribution of returns and its quantile.*

Furthermore, for any $0 < \tau_1 < \tau_2 < 1$*, the IQR for an individual with scenario* W *and unobserved characteristic* η *is defined by:*

$$IQR_S(\tau_1, \tau_2; W, \eta) = Q_S(\tau_2; W, \eta) - Q_S(\tau_1; W, \eta) \tag{8}$$

Finally, denote $t(s, w) = (y_1 - s, y_0, x_1, x_0)$*.*

Corollary 1. *If* $W \perp\!\!\!\perp \eta$*, the following holds:*

For any real value s *such that* $t(s, w) \in \mathcal{W}$ *and* $\tau_1, \tau_2 \in [0, 1]$*, let:*

$$A^{\tau_2-\tau_1}(w, a) = \int_{\mathcal{S}} 1\left\{1 - \tau_2 \leq Q_{P|W}(a|t(s, w)) \leq 1 - \tau_1\right\} ds \tag{9}$$

The population distribution of IQR is identified by:

$$\Pr\left(IQR_S(\tau_1, \tau_2; W, \eta) \leq y\right) = \int_{\mathcal{W}} \int_0^1 1\left\{A^{\tau_2-\tau_1}(w, a) \leq y\right\} da\, dF_W(w). \tag{10}$$

The following provides a proof of Corollary 1: MG2026 establishes that $\Pr(S(w, \eta, \nu) \leq s|\eta) = 1 - m(t(s, W), \eta)$. $\Pr(S(w, \eta, \nu) \leq s|\eta)$ represents a cumulative distribution function, and one can derive the conditional quantiles, say $Q_S(\tau; w, \eta)$, as in Chernozhukov et al. (2020) (cf. also Karr, 1993, pp. 113-114). Let $\mathcal{S}$ denote the support of S:

$$Q_S(\tau; w, \eta) \;=\; \int_{\mathcal{S}} \{[1 - m(t(s, w), \eta)] \leq \tau\} - 1\{s \leq 0\} ds \tag{11}$$

and, by equation (8), the IQR is:

$$IQR_S(\tau_1, \tau_2; w, \eta) = \int_{\mathcal{S}} 1\left\{1 - \tau_2 \leq m(t(s, w), \eta) \leq 1 - \tau_1\right\} ds$$

which is a strictly increasing functional of m. The stated demand function $m(w, \eta)$ is not directly identified because η is not identified. However, its quantile treatment response (QTR) function $Q_{m(w,\eta)}(a)$ corresponds to $Q_{P|W}(a|w)$, which is observed. In fact, by independence between W and η, for any $a \in (0, 1)$ and any $w \in \mathcal{W}$, $Q_{m(w,\eta)}(a) = Q_{m(W,\eta)|W}(a|w) = Q_{P|W}(a|w)$. Thus, $A^{\tau_2-\tau_1}(w, a)$ that replaces $Q_{P|W}(a|t(s, w))$ for $m(t(s, w), \eta)$ represents the QTR for the $IQR_S(w, \eta; \tau_1, \tau_2)$. Therefore, the population distribution of IQR can be rewritten as in Equation (10). This completes the proof.

Appendix B. Details of the implementation of the test procedure

Let $G_\tau(y) = \Pr(IQR_S(\tau, 0.5; W, \eta) \leq y\ell(\tau))$. Consider a collection of $\mathcal{T} = \{\tau_1, \ldots, \tau_K\}$.[3] The null hypothesis for the EV1 assumption is:

$$H_0 : G_\tau(\cdot) = G_{\tau'}(\cdot), \text{ for any } \tau, \tau' \in \mathcal{T} \tag{12}$$

The null hypothesis for the symmetry assumption is:

$$H_0 : G_\tau(\cdot) = G_{1-\tau}(\cdot), \text{ for any } \tau \in \mathcal{T}. \tag{13}$$

The logic of the test procedure is the same for both tests, so we discuss the test for the first hypothesis only. Define for a finite collection $\mathcal{Y}$ in $\mathbb{R}^+$ the long vector M obtained by collecting all differences $G_\tau(y) - G'_\tau(y)$ for $\tau \neq \tau', \tau, \tau' \in \mathcal{T}$ and $y \in \mathcal{Y}$. Let $\dim(M)$ denote its dimension. Equation (12) implies that:

$$S(M, \Omega) := (M'\Omega M)^{1/2} = 0 \tag{14}$$

for any conformable, positive, semi-definite matrix Ω.

Let $\{P_i, W_i\}_{i=1,\ldots,N}$ be the i.i.d. sample observed by the analyst. Denote by $\{P_i^b, W_i^b\}_{i=1,\ldots,N}$ a bootstrap sample obtained by resampling. The test procedure is as follows:

Step 1. Estimate the quantile treatment response of P given W, $Q_{P|W}(.|.)$ in the original data $\widehat{Q}_{P|W}(.|.)$ and in the bootstrap sample $\widehat{Q}^b_{P|W}(.|.)$. In the empirical applications, $\widehat{Q}_{\ell(P)|W}(a|w) = r(w) * \hat{\beta}_a$ is estimated by performing several quantile regressions, where $r(x)$ is the vector of differences in choice attributes.

Step 2. For each $\tau \in \mathcal{T}, \tau > 0.5, y \in \mathcal{Y}$, estimate the empirical counterpart of the population distribution of IQR defined by Equation (10).

$$\widehat{G}_{\tau,n}(y) = \frac{1}{N} \sum_i \sum_{k=1}^{K_a} \delta_a 1 \left\{ \widehat{A}_n^{\tau-0.5}(W_i, a_k) \leq y\ell(\tau) \right\} \tag{15}$$

$$\widehat{A}_n^{\tau-0.5}(w, a) = \sum_s \delta_s 1 \left\{ 1 - \tau \leq \widehat{Q}_{P|W}(a|t(s, w)) \leq 0.5 \right\} \tag{16}$$

The first sum is on a finite grid of the unit interval with K_a points $\{a_1, \ldots, a_{K_a}\}$ and a step width of δ_a. The sum over s in the expression of $\widehat{A}^{\tau-0.5}(w, a)$ is on a fine grid of $[0, K_S \times \delta_s]$ where K_S is a large integer and a step width of $\delta_s > 0$. There are similar expressions for $\tau > 0.5$.

Step 3. Construct the matrix $\widehat{M}_n$, the matrix collecting all differences $\widehat{G}_{\tau,n}(y) - \widehat{G}_{\tau',n}(y)$ for $\tau \neq \tau', \tau, \tau' \in \mathcal{T}$ and $y \in \mathcal{Y}$ for the original and the bootstrap sample. Stacking the bootstrap sample of $\widehat{M}_n^b$ in a matrix $\widehat{\boldsymbol{M}}_n$ of dimension $B \times \dim(M)$, obtain the empirical estimate $\widehat{\Sigma}_n = \widehat{\text{cov}}\left(\widehat{\boldsymbol{M}}_n\right)$. Finally, estimate $S(\widehat{M}_n, \widehat{\Sigma}_n^{-1})$ of dimension $\dim(M) \times \dim(M)$.

[3] In practice, because answers to choice probabilities questions tend to be rounded to multiples of 5 or 10, we do not need to consider a continuum of τ.

Step 4. Simulate L random variables $Z_l \sim N(0, \widehat{\Sigma}_n)$ and compute $S(Z_l, \widehat{\Sigma}_n^{-1})$. Compute the critical value $c\left(\widehat{\Sigma}_n^{-1}, 1-\alpha\right)$ as the $(1-\alpha)$-quantile of the simulated collections $\left\{S\left(Z_l, \widehat{\Sigma}_n^{-1}\right)\right\}_{l=1}^{L}$.

Step 5. Reject H_0 if $S\left(\widehat{M}_n, \widehat{\Sigma}_n^{-1}\right) > c\left(\widehat{\Sigma}_n^{-1}, 1-\alpha\right)$.

The results of Andrews and Soares (2010) imply that the proposed test procedure is valid and has good power properties under usual regularity assumptions.

Appendix C. Comparison of the appropriateness of the empirical applications

Table C.1 summarises the four empirical applications and how closely each matches the assumptions behind the proposed test. In all cases, the rejection of the test is formally a rejection of at least one component of the maintained framework: the RCM representation, the EV1 restriction on resolvable uncertainty, and the ceteris paribus interpretation. The rejection is most directly informative about EV1 when the ceteris paribus condition is credible, separability restrictions are modest, and conditional quantiles can be estimated reliably.

Application	Scenario design	Ceteris paribus	Variation/Reliance on RCM	Interpretation of rejection
WZ2018	Two blocks of eight hypothetical job-choice scenarios with three unlabelled jobs. Job attributes are experimentally varied in the constructed scenarios, but all respondents receive the same scenarios in the same order.	Explicit within-scenario ceteris paribus instruction: jobs differ only in listed attributes and are otherwise identical. Validity of the ceteris paribus framework relies on separability of the utility between listed and non-listed attributes.	Eight point of support for each block implying moderate joint variation. Responses are rounded to multiples of 5 or 10, but there is little excessive mass at 0, 50, or 100. Extreme probabilities concerns appear relatively limited.	Informative about the RCM+EV1 framework under a credible within-scenario ceteris paribus design, while formally still a rejection of the joint framework.
MG2026	Five public/private-sector job-offer scenarios. Each scenario presents one public-sector and one private-sector offer. The first scenario equates the main non-sector attributes; in the remaining scenarios, job attributes are randomly drawn from a pre-specified support.	Explicit ceteris paribus instruction across scenarios: only the indicated characteristics vary between scenarios, while other job features remain fixed. A concern is whether respondents fully internalize this instruction or infer omitted attributes from previous scenarios.	The design provides joint variations in the wage and non-wage job attributes (more than 300 point on the support), thus, no reliance on RCM. Rounding at the nearest 10 percent but no conspicuous use of extreme values or 0.5.	This is the cleanest application for interpreting rejection as evidence against the EV1 component of the maintained RCM framework. Formally, however, rejection remains a rejection of the joint RCM + EV1 + ceteris paribus framework.
KRV2022	Stated residential-location choices including the current-location option. Scenarios are organised in blocks; respondents are randomly assigned to four of six blocks, with 15 possible combinations of blocks, normally yielding 16 scenarios per respondent	Explicit within-scenario ceteris paribus instruction: locations differ only in listed attributes and are otherwise identical. Validity of the ceteris paribus framework relies on separability between listed and non-listed attributes.	Because attributes are not the same in blocks, the joint estimation relies on separability of the utility (RCM). Extreme responses are important and (bootstrap) quantile regression sometimes fails at low/high quantiles.	Rejection should be read as evidence against the joint RCM + EV1 + ceteris paribus/separability framework. The large sample size somewhat compensates for the sensitivity of the quantile estimation at low and high quantiles, yielding informative critical values.
AFZ2023	Six Fall-2020 states of the world crossed with seven tuition levels, yielding 42 observations per student. Scenarios vary COVID status, instruction mode, campus social life, and vaccine availability; the design is not a full factorial randomization.	Does not instruct respondents to hold all unspecified factors fixed. Identification relies on within-student scenario and cost variation, plus the assumption that the outside option is constant across scenarios conditional on COVID status.	Scenario support is structured and incomplete. There are many extreme responses: many students are certain they would return under some scenarios, and 58% assign zero probability to at least one scenario; (bootstrap) quantile regression sometimes fails at low quantiles.	Rejection should be interpreted especially cautiously as evidence against the joint RCM + EV1 + scenario/outside-option framework as applied to the full AFZ2023 design.

Notes: The table summarises how closely each application matches the assumptions behind the proposed test.

TABLE C.1. Assessment of the identifying assumptions across empirical applications

Appendix D. Comparison of Estimators

Estimators	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
	0.068	-0.013	0.462	-0.011	4.922	5.000	-0.013	-0.170
	(0.021)	(0.002)	(0.267)	(0.003)	(2.501)	(3.002)	(0.001)	(0.008)
BLM2010	[0.777]	[0.849]	[0.741]	[0.947]	[**0.080**]	[0.104]	[**0.000**]	[**0.000**]
	0.052	-0.051	0.450	-0.005	0.503	0.539	-0.012	-0.091
	(0.010)	(0.034)	(0.268)	(0.003)	(0.053)	(0.056)	(0.001)	(0.004)
WZ2018	[**0.050**]	[0.260]	[0.940]	[0.963]	[0.699]	[0.348]	[**0.000**]	[**0.000**]
	0.761	-1.101	0.024	-0.011	0.227	0.455	-0.002	-0.076
	(0.285)	(1.015)	(0.019)	(0.005)	(0.026)	(0.045)	(0.003)	(0.003)
AFZ2023	[**0.014**]	[0.284]	[**0.097**]	[0.950]	[0.486]	[0.519]	[0.464]	[**0.000**]
	0.073	-0.013	0.424	-0.007	0.846	0.284	-0.004	-0.039
	(0.012)	(0.001)	(0.242)	(0.055)	(0.887)	(0.262)	(0.000)	(0.001)
MG2026	–	–	–	–	–	–	–	–
	0.055	-0.010	-0.246	0.003	–	–	–	–
	–	–	–	–	–	–	–	–
MS	–	–	–	–	–	–	–	–

Notes: The table computes the median willingness-to-pay (WTP) in the population for two attributes in each of the empirical applications using five estimation procedures: the LAD estimation of Blass et al. (2010)(BLM2010), the individual-specific LAD estimator of Wiswall and Zafar (2018)(WZ2018), the individual-specific quasi-maximum likelihood estimator of Aucejo et al. (2023)(AFZ2023), the quantile regression estimator of (Méango and Girsberger, 2026)(MG2026), and the maximum score (MS) estimator proposed by Blass et al. (2010) and implemented as in Florios and Skouras (2008). Each cell reports the point estimate of the Median WTP for one unit increase in the attribute, standard error in parentheses, and p-value in square brackets for the test of equality between the point estimate of the respective method and the point estimate from MG2026. Columns (1)–(8) correspond to: (1) WZ2018 – Part-time work; (2) WZ2018 – One additional hour; (3) MG2026 – Job security; (4) MG2026 – One additional hour; (5) AFZ2023 – In person teaching; (6) AFZ2023 – Social interaction; (7) KRV2022 – more crime; (8) KRV2022 – More distance. Where applicable, standard errors are obtained by bootstrap with 200 replications. Due to the computational requirements, we subsample at most 2,000 observations for the MS estimator. We are unable to compute standard error for the MS estimator (see Blass et al., 2010). Unfortunately, the MS estimator does not converge when applied to KRV2022 or AFZ2023. This is probably the result of a very skewed distribution of answers.

Table D.1. Comparison of WTP estimates between different methods