

---

# Robust Statistical Comparison of Random Variables with Locally Varying Scale of Measurement (Supplementary Material)

---

Christoph Jansen<sup>1</sup>   Georg Schollmeyer<sup>1</sup>   Hannah Blocher<sup>1</sup>   Julian Rodemann<sup>1</sup>   Thomas Augustin<sup>1</sup>

<sup>1</sup>Department of Statistics, Ludwig-Maximilians-Universität, Munich, Bavaria, Germany

In the following, we give supplementary information and material to the main paper. This includes all mathematical proofs of the propositions and corollaries established in the main paper (Part A), further details on the implementation and reproducibility (Part B), further calculations for the robustified test statistics (Part C), and further analyses of the applications in the main paper (Part D). If not explicitly stated otherwise, from now on, all references to equations, propositions, etc. refer to the main part of the paper.

## A PROOFS OF THE RESULTS IN THE MAIN PAPER

### A.1 PROOFS FOR PROPOSITION 1 AND 2: BOUNDED PREFERENCE SYSTEMS

We start by proving Propositions 1 resp. 2 from Sections 2 resp. 4 that state that checking consistency resp. GSD simplifies if the underlying preference system is bounded.

**Proposition 1** *Let  $\mathcal{A} = [A, R_1, R_2]$  be a bounded preference system. Then  $\mathcal{A}$  is consistent iff it is 0-consistent.*

**Proof.** If  $\mathcal{A}$  is 0-consistent, then it is obviously also consistent, since every normalized representation is in particular a representation. For the other direction, assume  $\mathcal{A}$  to be consistent. Choose  $u \in \mathcal{U}_{\mathcal{A}}$  arbitrarily and denote by  $a_*$ ,  $a^*$  the  $R_1$ -minimal resp.  $R_1$ -maximal elements satisfying  $(a^*, a_*) \in P_{R_1}$ . From the latter we know that  $u(a^*) > u(a_*)$ . Thus, the function

$$\tilde{u} : A \rightarrow [0, 1], \quad a \mapsto \frac{u(a) - u(a_*)}{u(a^*) - u(a_*)}$$

is well-defined. Moreover, one easily verifies that  $\tilde{u} \in \mathcal{U}_{\mathcal{A}}$ , and  $u(a_*) = 0$ , and  $u(a^*) = 1$ . Thus, we can conclude that  $\tilde{u} \in \mathcal{N}_{\mathcal{A}}$ , which – by definition – implies 0-consistency.  $\square$

**Proposition 2** *If  $\mathcal{A}$  is consistent and bounded with  $a_*$ ,  $a^*$  as before, then  $(X, Y) \in R_{(\mathcal{A}, \pi)}$  iff*

$$\forall u \in \mathcal{N}_{\mathcal{A}} : \mathbb{E}_{\pi}(u \circ X) \geq \mathbb{E}_{\pi}(u \circ Y).$$

**Proof.** The direction  $\Rightarrow$  follows trivially by observing  $\mathcal{N}_{\mathcal{A}} \subseteq \mathcal{U}_{\mathcal{A}}$ . For the direction  $\Leftarrow$ , assume that it holds  $\forall u \in \mathcal{N}_{\mathcal{A}} : \mathbb{E}_{\pi}(u \circ X) \geq \mathbb{E}_{\pi}(u \circ Y)$ . Choose  $u \in \mathcal{U}_{\mathcal{A}}$  arbitrarily. With the same argument as given in the proof of Proposition 1, we know that then  $\tilde{u} \in \mathcal{N}_{\mathcal{A}}$ , where  $\tilde{u}$  is defined as in the proof of Proposition 1. Since  $\tilde{u}$  is a positive (affine) linear transformation of  $u$ , we know that  $\mathbb{E}_{\pi}(u \circ X) \geq \mathbb{E}_{\pi}(u \circ Y)$  if and only if  $\mathbb{E}_{\pi}(\tilde{u} \circ X) \geq \mathbb{E}_{\pi}(\tilde{u} \circ Y)$ . Since the latter is true by assumption (utilizing  $\tilde{u} \in \mathcal{N}_{\mathcal{A}}$ ), the first also is true. As  $u$  was chosen arbitrarily, this completes the proof.  $\square$

### A.2 PROOFS OF PROPOSITIONS 3 AND 4: COMPUTATIONS FOR THE PERMUTATION TEST

We now give proofs for Propositions 3 resp. 4 from Section 5.3 that concern the computation of the maximum regularization strength resp. the computation of the (regularized) test statistic for the permutation-test.

**Proposition 3** For samples  $\mathbf{x}$  and  $\mathbf{y}$  of the form (6) and (7) and  $\varepsilon \in [0, 1]$ , we consider the linear program

$$\xi \longrightarrow \max_{(v_1, \dots, v_s, \xi)}$$

with constraints  $(v_1, \dots, v_s, \xi) \in C(\mathbf{x}, \mathbf{y})$ . Denote by  $\xi^*$  its optimal value. It then holds  $\delta_\varepsilon(\omega_0) = \varepsilon \cdot \xi^*$ .

**Proof.** The Proposition follows from standard results on linear optimization and the fact that  $C(\mathbf{x}, \mathbf{y})$  is compact. Set  $I := \{\ell : (z_\ell, a^*) \in I_{R_1}\}$  and define the vector  $\underline{v} := (0, 1, v_3, \dots, v_s, 0) \in [0, 1]^{s+1}$  by  $v_\ell = 1$  if  $\ell \in I$  and  $v_\ell = 0$  otherwise. One then easily verifies that  $\underline{v}$  is an admissible solution to the above linear program. Since  $C(\mathbf{x}, \mathbf{y})$  is compact, this implies the existence of an optimal solution. Denote thus by  $\underline{v}^* := (0, 1, v_3^*, \dots, v_s^*, \xi^*)$  an arbitrary optimal solution. We have to show that

$$\xi^* = \sup\{\xi : \mathcal{N}_{\mathcal{A}_{\omega_0}}^\xi \neq \emptyset\} =: c.$$

Assume, for contradiction, the above equality does not hold. We distinguish two cases:

*Case 1:*  $\xi^* < c$ . Then, one easily verifies that for any function  $u \in \mathcal{N}_{\mathcal{A}_{\omega_0}}^c$  the vector  $(u(z_1), \dots, u(z_s), c)$  defines an admissible solution to the above linear program with an objective value of  $c$ . This contradicts the optimality of  $\underline{v}^*$ .

*Case 2:*  $\xi^* > c$ . Then, setting  $u : (\mathbf{X}\mathbf{Y})_{\omega_0} \rightarrow [0, 1]$  with  $u(z_\ell) := v_\ell^*$  defines an element of  $\mathcal{N}_{\mathcal{A}_{\omega_0}}^{\xi^*}$ , contradicting that  $c$  is the largest number for which  $\mathcal{A}_{\omega_0}$  is  $c$ -consistent.

Thus, we have that  $c = \xi^*$ , implying  $\delta_\varepsilon(\omega_0) = \varepsilon \cdot \xi^*$ .  $\square$

**Proposition 4** For samples  $\mathbf{x}$  and  $\mathbf{y}$  of the form (6) and (7) and  $\varepsilon \in [0, 1]$ , we consider the following linear program

$$\sum_{\ell=1}^s v_\ell \cdot \left( \frac{|\{i: x_i = z_\ell\}|}{n} - \frac{|\{i: y_i = z_\ell\}|}{m} \right) \longrightarrow \min_{(v_1, \dots, v_s)}$$

with constraints  $(v_1, \dots, v_s) \in C_{\varepsilon\xi^*}(\mathbf{x}, \mathbf{y})$ , where  $\xi^*$  denotes the optimal value of (8). Denote by  $opt_\varepsilon(\mathbf{x}, \mathbf{y})$  its optimal value. It then holds:

- i)  $opt_\varepsilon(\mathbf{x}, \mathbf{y}) = d_{\mathbf{X}, \mathbf{Y}}^\varepsilon(\omega_0)$ .
- ii) There is in-sample GSD of  $X$  over  $Y$  if and only if  $opt_0(\mathbf{x}, \mathbf{y}) \geq 0$ .

**Proof.** i) By definition and Proposition 2, we know that  $\mathcal{N}_{\mathcal{A}_{\omega_0}}^{\varepsilon\xi^*} \neq \emptyset$ . As these sets are nested with decreasing  $\xi$ -value and we have  $\varepsilon\xi^* \leq \xi^*$ , this implies that also  $\mathcal{N}_{\mathcal{A}_{\omega_0}}^{\varepsilon\xi^*} \neq \emptyset$ . Hence, we can choose  $u \in \mathcal{N}_{\mathcal{A}_{\omega_0}}^{\varepsilon\xi^*}$ . One then easily verifies that the vector  $(u(z_1), \dots, u(z_s))$  defines an admissible solution to the above linear program. Since  $C_{\varepsilon\xi^*}(\mathbf{x}, \mathbf{y})$  is compact, this implies the existence of an optimal solution. Thus, denote by  $\underline{v}^* := (v_1^*, \dots, v_s^*)$  an arbitrary such optimal solution. If we then define  $u : (\mathbf{X}\mathbf{Y})_{\omega_0} \rightarrow [0, 1]$  with  $u(z_\ell) := v_\ell^*$ , then one easily verifies that  $u \in \mathcal{N}_{\mathcal{A}_{\omega_0}}^{\varepsilon\xi^*}$  and that

$$opt_\varepsilon(\mathbf{x}, \mathbf{y}) = \sum_{z \in (\mathbf{X}\mathbf{Y})_{\omega_0}} u(z) \cdot (\hat{\pi}_X^{\omega_0}(\{z\}) - \hat{\pi}_Y^{\omega_0}(\{z\})) \quad (1)$$

(to see this, note that the right side of the equation is a simple reformulation of the objective function with  $\underline{v}^*$  plugged-in).

We have to show that

$$opt_\varepsilon(\mathbf{x}, \mathbf{y}) = d_{\mathbf{X}, \mathbf{Y}}^\varepsilon(\omega_0).$$

Assume, for contradiction, the above equality does not hold. We distinguish two cases:

*Case 1:*  $opt_\varepsilon(\mathbf{x}, \mathbf{y}) > d_{\mathbf{X}, \mathbf{Y}}^\varepsilon(\omega_0)$ . This would imply that there exists an  $u' \in \mathcal{N}_{\mathcal{A}_{\omega_0}}^{\varepsilon\xi^*}$  that – if it was set in the right-hand side of the above Equation (1) (in the supplementary material) instead of  $u$  – would produce a value strictly smaller than  $opt_\varepsilon(\mathbf{x}, \mathbf{y})$ . This contradicts the optimality of  $\underline{v}^*$ , since every  $u' \in \mathcal{N}_{\mathcal{A}_{\omega_0}}^{\varepsilon\xi^*}$  produces an admissible solution to the linear program with objective value given by the right-hand side of the above Equation (1).

*Case 2:*  $opt_\varepsilon(\mathbf{x}, \mathbf{y}) < d_{\mathbf{X}, \mathbf{Y}}^\varepsilon(\omega_0)$ . This would be an immediate contradiction to the above Equation (1) (in the supplementary material), since  $d_{\mathbf{X}, \mathbf{Y}}^\varepsilon(\omega_0)$  is by definition the infimum over all the expressions on the equation's right-hand side.

This completes the proof of i). To see ii), note that i) implies  $opt_0(\mathbf{x}, \mathbf{y}) = d_{\mathbf{X}, \mathbf{Y}}^0(\omega_0)$ . Thus, we have  $opt_0(\mathbf{x}, \mathbf{y}) \geq 0$  if and only if  $d_{\mathbf{X}, \mathbf{Y}}^0(\omega_0) \geq 0$ , which – by definition – is true if and only if there is in-sample GSD of  $X$  over  $Y$ .  $\square$

### A.3 PROOFS OF PROPOSITION 5 AND 6: COMPUTATIONS FOR ROBUSTIFIED TESTING

We now give proofs of Proposition 5 resp. 6 from Section 6 concerning the computation of the robustified test statistic resp. its simplification under the special case of a  $\gamma$ -contamination model (with  $\gamma \in [0, 1]$ ).

**Proposition 5** For samples  $\mathbf{x}$  and  $\mathbf{y}$  of the form (6) and (7),  $\varepsilon \in [0, 1]$ , and  $(\pi_1, \pi_2) \in \mathcal{E}(\mathcal{M}_X^{\omega_0}) \times \mathcal{E}(\mathcal{M}_Y^{\omega_0})$ , we consider the following linear program:

$$\sum_{\ell=1}^s v_\ell \cdot (\pi_1(\{z\}) - \pi_2(\{z\})) \longrightarrow \min_{(v_1, \dots, v_s)}$$

with constraints  $(v_1, \dots, v_s) \in C_{\varepsilon \xi^*}(\mathbf{x}, \mathbf{y})$ , where  $\xi^*$  denotes the optimal value of (8). Denote by  $opt_\varepsilon(\mathbf{x}, \mathbf{y}, \pi_1, \pi_2)$  its optimal value and by  $\underline{opt}_\varepsilon(\mathbf{x}, \mathbf{y})$  the minimal optimum over all combinations of  $(\pi_1, \pi_2) \in \mathcal{E}(\mathcal{M}_X^{\omega_0}) \times \mathcal{E}(\mathcal{M}_Y^{\omega_0})$ . It then holds:

i)  $\underline{opt}_\varepsilon(\mathbf{x}, \mathbf{y}) = \underline{d}_{\mathbf{X}, \mathbf{Y}}^\varepsilon(\omega_0)$ .

ii) There is in-sample GSD of  $X$  over  $Y$  for any  $\pi$  with  $\hat{\pi}_X^{\omega_0} \in \mathcal{M}_X^{\omega_0}$  and  $\hat{\pi}_Y^{\omega_0} \in \mathcal{M}_Y^{\omega_0}$  if  $\underline{opt}_\varepsilon(\mathbf{x}, \mathbf{y}) \geq 0$ .

**Proof.** i) Since nothing in the proof of Proposition 4 hinges on the concrete structure of the involved empirical image measures, Proposition 4 is still valid if we replace  $\hat{\pi}_X^{\omega_0}$  and  $\hat{\pi}_Y^{\omega_0}$  by arbitrary  $\pi_1 \in \mathcal{M}_X^{\omega_0}$  and  $\pi_2 \in \mathcal{M}_Y^{\omega_0}$ , respectively. This specifically implies

$$opt_\varepsilon(\mathbf{x}, \mathbf{y}, \pi_1, \pi_2) = \inf_{u \in \mathcal{N}_{\mathcal{A}\omega_0}^{\delta_\varepsilon(\omega_0)}} \sum_{z \in (\mathbf{X}\mathbf{Y})_{\omega_0}} u(z) \cdot (\pi_1(\{z\}) - \pi_2(\{z\})). \quad (2)$$

In order to show i), we now need to verify that

$$\inf_{(\pi_1, \pi_2) \in \mathcal{E}(\mathcal{M}_X^{\omega_0}) \times \mathcal{E}(\mathcal{M}_Y^{\omega_0})} opt_\varepsilon(\mathbf{x}, \mathbf{y}, \pi_1, \pi_2) = \underline{d}_{\mathbf{X}, \mathbf{Y}}^\varepsilon(\omega_0).$$

Due to the above Equation (2) (in the supplementary material) and the fact that iterated infima can be equivalently replaced by one global infimum, we know that

$$\inf_{(\pi_1, \pi_2) \in \mathcal{M}_X^{\omega_0} \times \mathcal{M}_Y^{\omega_0}} opt_\varepsilon(\mathbf{x}, \mathbf{y}, \pi_1, \pi_2) = \underline{d}_{\mathbf{X}, \mathbf{Y}}^\varepsilon(\omega_0). \quad (3)$$

We then can compute:

$$\begin{aligned} \underline{d}_{\mathbf{X}, \mathbf{Y}}^\varepsilon(\omega_0) &\stackrel{(3)}{=} \inf_{(\pi_1, \pi_2) \in \mathcal{M}_X^{\omega_0} \times \mathcal{M}_Y^{\omega_0}} opt_\varepsilon(\mathbf{x}, \mathbf{y}, \pi_1, \pi_2) \\ &= \inf_{(\pi_1, \pi_2) \in \mathcal{M}_X^{\omega_0} \times \mathcal{M}_Y^{\omega_0}} \inf_{u \in \mathcal{N}_{\mathcal{A}\omega_0}^{\delta_\varepsilon(\omega_0)}} \sum_{z \in (\mathbf{X}\mathbf{Y})_{\omega_0}} u(z) \cdot (\pi_1(\{z\}) - \pi_2(\{z\})) \\ &= \inf_{u \in \mathcal{N}_{\mathcal{A}\omega_0}^{\delta_\varepsilon(\omega_0)}} \inf_{(\pi_1, \pi_2) \in \mathcal{M}_X^{\omega_0} \times \mathcal{M}_Y^{\omega_0}} \sum_{z \in (\mathbf{X}\mathbf{Y})_{\omega_0}} u(z) \cdot (\pi_1(\{z\}) - \pi_2(\{z\})) \\ &\stackrel{(*)}{=} \inf_{u \in \mathcal{N}_{\mathcal{A}\omega_0}^{\delta_\varepsilon(\omega_0)}} \left( \inf_{\pi_1 \in \mathcal{M}_X^{\omega_0}} \sum_{z \in (\mathbf{X}\mathbf{Y})_{\omega_0}} u(z) \cdot \pi_1(\{z\}) - \sup_{\pi_2 \in \mathcal{M}_Y^{\omega_0}} \sum_{z \in (\mathbf{X}\mathbf{Y})_{\omega_0}} u(z) \cdot \pi_2(\{z\}) \right) \\ &\stackrel{(**)}{=} \inf_{u \in \mathcal{N}_{\mathcal{A}\omega_0}^{\delta_\varepsilon(\omega_0)}} \left( \inf_{\pi_1 \in \mathcal{E}(\mathcal{M}_X^{\omega_0})} \sum_{z \in (\mathbf{X}\mathbf{Y})_{\omega_0}} u(z) \cdot \pi_1(\{z\}) - \sup_{\pi_2 \in \mathcal{E}(\mathcal{M}_Y^{\omega_0})} \sum_{z \in (\mathbf{X}\mathbf{Y})_{\omega_0}} u(z) \cdot \pi_2(\{z\}) \right) \\ &= \inf_{(\pi_1, \pi_2) \in \mathcal{E}(\mathcal{M}_X^{\omega_0}) \times \mathcal{E}(\mathcal{M}_Y^{\omega_0})} \inf_{u \in \mathcal{N}_{\mathcal{A}\omega_0}^{\delta_\varepsilon(\omega_0)}} \sum_{z \in (\mathbf{X}\mathbf{Y})_{\omega_0}} u(z) \cdot (\pi_1(\{z\}) - \pi_2(\{z\})) \\ &= \inf_{(\pi_1, \pi_2) \in \mathcal{E}(\mathcal{M}_X^{\omega_0}) \times \mathcal{E}(\mathcal{M}_Y^{\omega_0})} opt_\varepsilon(\mathbf{x}, \mathbf{y}, \pi_1, \pi_2) \end{aligned}$$

Here,  $(\star)$  follows since – for  $u$  fixed – the infimum of the differences of the two sums is attained if the first sum is smallest possible and the second sum is largest possible (note that all sums involved are finite). Further,  $(\star\star)$  follows since – again for  $u$  fixed – the sums are linear functions on the compact sets  $\mathcal{M}_X^{\omega_0}$  resp.  $\mathcal{M}_Y^{\omega_0}$  and, therefore, attain their optima on  $\mathcal{E}(\mathcal{M}_X^{\omega_0})$  resp.  $\mathcal{E}(\mathcal{M}_Y^{\omega_0})$ . The fifth and sixth equalities are just reversing the computation done in the first three equalities.

To see ii), note that i) implies  $\underline{opt}_0(\mathbf{x}, \mathbf{y}) = \underline{d}_{\mathbf{X}, \mathbf{Y}}^0(\omega_0)$ . Thus,  $\underline{opt}_0(\mathbf{x}, \mathbf{y}) \geq 0$  if and only if  $\underline{d}_{\mathbf{X}, \mathbf{Y}}^0(\omega_0) \geq 0$ . But – by definition – the latter is true if and only if

$$\inf_{u \in \mathcal{N}_{\mathcal{A}\omega_0}^0} \sum_{z \in (\mathbf{X}\mathbf{Y})_{\omega_0}} u(z) \cdot (\pi_1(\{z\}) - \pi_2(\{z\})) \geq 0$$

for all  $(\pi_1, \pi_2) \in \mathcal{M}_X^{\omega_0} \times \mathcal{M}_Y^{\omega_0}$ . This obviously implies in-sample GSD of  $X$  over  $Y$  for any  $\pi$  with  $\hat{\pi}_X^{\omega_0} \in \mathcal{M}_X^{\omega_0}$  and  $\hat{\pi}_Y^{\omega_0} \in \mathcal{M}_Y^{\omega_0}$ , since  $\mathcal{N}_{\mathcal{A}\omega_0}^0 = \mathcal{N}_{\mathcal{A}\omega_0}$ .  $\square$

**Proposition 6** Consider again the situation of Proposition 5 with the additional assumption that  $\mathcal{M}_X^{\omega_0}$  and  $\mathcal{M}_Y^{\omega_0}$  are of the form (11) with extreme points as in (12). It then holds:

$$\underline{opt}_{\varepsilon}(\mathbf{x}, \mathbf{y}) = \underline{opt}_{\varepsilon}(\mathbf{x}, \mathbf{y}, \pi_*, \pi^*)$$

where

$$\pi_* = \gamma \delta_{a_*} + (1 - \gamma) \hat{\pi}_X^{\omega_0}$$

and

$$\pi^* = \gamma \delta_{a_*} + (1 - \gamma) \hat{\pi}_Y^{\omega_0}.$$

**Proof.** By again utilizing Equation (2) (of the supplementary material), the claim modifies to showing that

$$\underline{opt}_{\varepsilon}(\mathbf{x}, \mathbf{y}) = \inf_{u \in \mathcal{N}_{\mathcal{A}\omega_0}^{\delta_{\varepsilon}(\omega_0)}} \sum_{z \in (\mathbf{X}\mathbf{Y})_{\omega_0}} u(z) \cdot (\pi_*(\{z\}) - \pi^*(\{z\})).$$

Since, by Proposition 2, we know that  $\underline{d}_{\mathbf{X}, \mathbf{Y}}^{\varepsilon}(\omega_0) = \underline{opt}_{\varepsilon}(\mathbf{x}, \mathbf{y})$  and  $\underline{d}_{\mathbf{X}, \mathbf{Y}}^{\varepsilon}(\omega_0)$  is by definition the infimum over all the expressions on the right-hand side, the direction  $\leq$  is immediate. So, it remains to show the direction  $\geq$ . To do so, choose  $(\pi_1, \pi_2) \in \mathcal{M}_X^{\omega_0} \times \mathcal{M}_Y^{\omega_0}$  arbitrarily. Since both  $\mathcal{M}_X^{\omega_0}$  and  $\mathcal{M}_Y^{\omega_0}$  are of the form (11), we then know that there exist probability measures  $\nu_1$  and  $\nu_2$  such that

$$\pi_1 = \gamma \cdot \nu_1 + (1 - \gamma) \cdot \hat{\pi}_X^{\omega_0}$$

and

$$\pi_2 = \gamma \cdot \nu_2 + (1 - \gamma) \cdot \hat{\pi}_Y^{\omega_0}.$$

Here, we utilized the fact that credal sets of the form (11) can be equivalently characterized as

$$\mathcal{M}_Z^{\omega} = \left\{ \pi : \pi \geq (1 - \gamma) \cdot \hat{\pi}_Z^{\omega} \right\} = \left\{ \gamma \cdot \nu + (1 - \gamma) \cdot \hat{\pi}_Z^{\omega} : \nu \text{ probability measure} \right\}.$$

For  $u \in \mathcal{N}_{\mathcal{A}\omega_0}^{\delta_{\varepsilon}(\omega_0)}$  fixed (but arbitrary), we then can compute:

$$\begin{aligned} \sum_{z \in (\mathbf{X}\mathbf{Y})_{\omega_0}} u(z) \cdot \pi_1(\{z\}) &= \gamma \cdot \sum_{z \in (\mathbf{X}\mathbf{Y})_{\omega_0}} u(z) \cdot \nu_1(\{z\}) + (1 - \gamma) \sum_{z \in (\mathbf{X}\mathbf{Y})_{\omega_0}} u(z) \cdot \hat{\pi}_X^{\omega_0}(\{z\}) \\ &\geq \gamma \cdot u(a_*) + (1 - \gamma) \sum_{z \in (\mathbf{X}\mathbf{Y})_{\omega_0}} u(z) \cdot \hat{\pi}_X^{\omega_0}(\{z\}) \\ &= \gamma \cdot \sum_{z \in (\mathbf{X}\mathbf{Y})_{\omega_0}} u(z) \cdot \delta_{a_*}(\{z\}) + (1 - \gamma) \sum_{z \in (\mathbf{X}\mathbf{Y})_{\omega_0}} u(z) \cdot \hat{\pi}_X^{\omega_0}(\{z\}) \\ &= \sum_{z \in (\mathbf{X}\mathbf{Y})_{\omega_0}} u(z) \cdot \pi_*(\{z\}) \end{aligned}$$

Analogous reasoning yields:

$$\sum_{z \in (\mathbf{XY})_{\omega_0}} u(z) \cdot \pi_2(\{z\}) \leq \sum_{z \in (\mathbf{XY})_{\omega_0}} u(z) \cdot \pi^*(\{z\})$$

Putting the two together, we arrive at:

$$\sum_{z \in (\mathbf{XY})_{\omega_0}} u(z) \cdot (\pi_1(\{z\}) - \pi_2(\{z\})) \geq \sum_{z \in (\mathbf{XY})_{\omega_0}} u(z) \cdot (\pi_*(\{z\}) - \pi^*(\{z\}))$$

As  $\pi_1, \pi_2$ , and  $u$  were chosen arbitrarily, the inequality remains valid for the infimum, i.e.

$$\inf_{(\pi_1, \pi_2, u) \in \mathcal{M}_X^{\omega_0} \times \mathcal{M}_Y^{\omega_0} \times \mathcal{N}_{\mathcal{A}\omega_0}^{\delta_\varepsilon(\omega_0)}} \sum_{z \in (\mathbf{XY})_{\omega_0}} u(z) \cdot (\pi_1(\{z\}) - \pi_2(\{z\})) \geq \inf_{u \in \mathcal{N}_{\mathcal{A}\omega_0}^{\delta_\varepsilon(\omega_0)}} \sum_{z \in (\mathbf{XY})_{\omega_0}} u(z) \cdot (\pi_*(\{z\}) - \pi^*(\{z\}))$$

Observing that the left side of this inequality by definition equals  $\underline{d}_{\mathbf{X}, \mathbf{Y}}^\varepsilon(\omega_0)$  and, therefore, by Proposition 4, also  $\underline{opt}_\varepsilon(\mathbf{x}, \mathbf{y})$  completes the direction  $\geq$  and thus the proof.  $\square$

#### A.4 PROOFS OF PROPOSITIONS 7 AND 8: MULTI-DIMENSIONAL SPACES

Finally, we give proofs of Propositions 7 and 8 from Section 7 concerning several different characterizing properties of the GSD-order for the special case of preferences systems arising from multi-dimensional spaces with differently scaled dimensions. For this, recall that in Section 4 for a preference system  $\mathcal{A}$  and a probability measure  $\pi$  we defined

$$\mathcal{F}_{(\mathcal{A}, \pi)} := \left\{ X \in A^\Omega : u \circ X \in \mathcal{L}^1(\Omega, \mathcal{S}_1, \pi) \forall u \in \mathcal{U}_{\mathcal{A}} \right\}.$$

This definition is needed for stating the next proposition.

**Proposition 7** *Let  $\pi$  be a probability measure on  $(\Omega, \mathcal{S}_1)$ , and  $X = (\Delta_1, \dots, \Delta_r), Y = (\Lambda_1, \dots, \Lambda_r) \in \mathcal{F}_{(\text{pref}(\mathbb{R}^r), \pi)}$ , where the first  $0 \leq z \leq r$  dimensions of  $\text{pref}(\mathbb{R}^r)$  are of cardinal scale. Then, the following holds:*

- i)  $\text{pref}(\mathbb{R}^r)$  is consistent.
- ii) If  $z = 0$ , then  $R_{(\text{pref}(\mathbb{R}^r), \pi)}$  coincides with (first-order) stochastic dominance w.r.t.  $\pi$  and  $R_1^*$  (short:  $FSD(R_1^*, \pi)$ ).
- iii) If  $(X, Y) \in R_{(\text{pref}(\mathbb{R}^r), \pi)}$  and  $\Delta_j, \Lambda_j \in \mathcal{L}^1(\Omega, \mathcal{S}_1, \pi)$  for all  $j = 1, \dots, r$ , then
  - I.  $\mathbb{E}_\pi(\Delta_j) \geq \mathbb{E}_\pi(\Lambda_j)$  for all  $j = 1, \dots, r$ , and
  - II.  $(\Delta_j, \Lambda_j) \in FSD(\geq, \pi)$  for all  $j = z + 1, \dots, r$ .

*Additionally, in the special case where all components of  $X$  are jointly independent and all components of  $Y$  are jointly independent, properties I. and II. imply  $(X, Y) \in R_{(\text{pref}(\mathbb{R}^r), \pi)}$  (i.e. also the converse implication holds).*

**Proof.** i) Let  $\alpha_1, \dots, \alpha_r \in \mathbb{R}^+$  and  $\phi_{z+1}, \dots, \phi_r : \mathbb{R} \rightarrow \mathbb{R}$  strictly isotone functions. Define  $u : \mathbb{R}^r \rightarrow \mathbb{R}$  by setting

$$u(x) := \sum_{s=1}^z \alpha_s \cdot x_s + \sum_{s=z+1}^r \alpha_s \cdot \phi_s(x_s).$$

Then one easily verifies that  $u$  defines a representation of  $\text{pref}(\mathbb{R}^r)$ , proving its consistency.

ii) Assume  $z = 0$ , i.e. all considered dimensions are purely ordinal. We claim that for  $\mathcal{A}_0 := [\mathbb{R}^r, R_1^*, \emptyset]$  it holds  $\mathcal{U}_{\text{pref}(\mathbb{R}^r)} = \mathcal{U}_{\mathcal{A}_0}$ . The direction  $\subseteq$  is trivial, so assume  $u \in \mathcal{U}_{\mathcal{A}_0}$  arbitrary. It suffices to show that  $u$  represents arbitrary pairs of pairs in  $R_2^*$ . As  $R_2^*$  is antisymmetric for  $z = 0$ , this reduces to show that  $u$  strictly represents arbitrary pairs of pairs in  $P_{R_2^*}$ . So, let  $((v, w), (x, y)) \in P_{R_2^*}$ . This means that for all  $j \in \{1, \dots, r\}$  we have  $v_j \geq x_j \geq y_j \geq w_j$  and that there is  $j_0 \in \{1, \dots, r\}$  such that either  $v_{j_0} > x_{j_0}$  or  $y_{j_0} > w_{j_0}$ . Together, this implies  $u(v) > u(x) \geq u(y) \geq u(w)$  or  $u(v) \geq u(x) \geq u(y) > u(w)$ , either way implying  $u(v) - u(w) > u(x) - u(y)$ . Thus  $u \in \mathcal{U}_{\text{pref}(\mathbb{R}^r)}$ . As  $R_{(\mathcal{A}_0, \pi)}$  coincides with (first-order) stochastic dominance by definition and we have  $\mathcal{U}_{\text{pref}(\mathbb{R}^r)} = \mathcal{U}_{\mathcal{A}_0}$  also  $R_{(\text{pref}(\mathbb{R}^r), \pi)}$  coincides with (first-order) stochastic dominance.

iii) Let  $(X, Y) \in R_{(\text{pref}(\mathbb{R}^r), \pi)}$ . We start by showing I, so choose  $j \in \{1, \dots, r\}$  arbitrary. By part i) of the proof, for every  $n \in \mathbb{N}$ , the function  $u_n : \mathbb{R}^r \rightarrow \mathbb{R}$  defined by

$$u_n(x) := x_j + \frac{1}{n} \cdot \sum_{s \neq j} x_s$$

is a representation of  $\text{pref}(\mathbb{R}^r)$ , that is  $u_n \in \mathcal{U}_{\text{pref}(\mathbb{R}^r)}$ . Thus, by our assumption  $(X, Y) \in R_{(\text{pref}(\mathbb{R}^r), \pi)}$ , we know that we have  $\mathbb{E}_\pi(u_n \circ X) \geq \mathbb{E}_\pi(u_n \circ Y)$ . This implies (by the linearity of the expectation operator)

$$\mathbb{E}_\pi(\Delta_j) + \frac{1}{n} \cdot \sum_{s \neq j} \mathbb{E}_\pi(\Delta_s) \geq \mathbb{E}_\pi(\Lambda_j) + \frac{1}{n} \cdot \sum_{s \neq j} \mathbb{E}_\pi(\Lambda_s).$$

Letting  $n \rightarrow \infty$  on both sides gives  $\mathbb{E}_\pi(\Delta_j) \geq \mathbb{E}_\pi(\Lambda_j)$ .

We use a very similar argument to see II: Choose  $j \in \{z+1, \dots, r\}$  arbitrarily and let  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  be strictly isotone. By part i) of the proof, for every  $n \in \mathbb{N}$ , the function  $u'_n : \mathbb{R}^r \rightarrow \mathbb{R}$  defined by

$$u'_n(x) := \phi(x_j) + \frac{1}{n} \cdot \sum_{s \neq j} x_s$$

is a representation of  $\text{pref}(\mathbb{R}^r)$ , that is  $u_n \in \mathcal{U}_{\text{pref}(\mathbb{R}^r)}$ . Thus, by our assumption  $(X, Y) \in R_{(\text{pref}(\mathbb{R}^r), \pi)}$ , we know that we have  $\mathbb{E}_\pi(u_n \circ X) \geq \mathbb{E}_\pi(u_n \circ Y)$ . This implies (by the linearity of the expectation operator)

$$\mathbb{E}_\pi(\phi \circ \Delta_j) + \frac{1}{n} \cdot \sum_{s \neq j} \mathbb{E}_\pi(\Delta_s) \geq \mathbb{E}_\pi(\phi \circ \Lambda_j) + \frac{1}{n} \cdot \sum_{s \neq j} \mathbb{E}_\pi(\Lambda_s).$$

Letting  $n \rightarrow \infty$  gives  $\mathbb{E}_\pi(\phi \circ \Delta_j) \geq \mathbb{E}_\pi(\phi \circ \Lambda_j)$ . As  $\phi$  was chosen arbitrarily, this implies  $(\Delta_j, \Lambda_j) \in \text{FSD}(\geq, \pi)$ .

To see the addition to part iii), let  $X = (\Delta_1, \dots, \Delta_r)$  and  $Y = (\Lambda_1, \dots, \Lambda_r)$  have both jointly independent components, respectively, and let I. and II. of iii) be true. Let furthermore  $u \in \mathcal{U}_{\text{pref}(\mathbb{R}^r)}$  be an arbitrary utility function that represents the preference system  $\text{pref}(\mathbb{R}^r)$ . We now show that  $\mathbb{E}_\pi(u \circ X) \geq \mathbb{E}_\pi(u \circ Y)$  holds: Because of independence we can compute the expectations of  $u \circ X$  and  $u \circ Y$  by using Fubini's theorem. To prove the inequality, we first integrate over the ordinal part and use isotonicity of  $u$  in every integration. Then we integrate over the cardinal parts and iteratively use the fact that the corresponding functions are representing the corresponding cardinal subsystem built by the components we did not integrate over before. Formally, we arrive at:

$$\begin{aligned} \mathbb{E}_\pi(u \circ X) &= \int_{\Omega} u \circ X d\pi \\ &\stackrel{(\text{ind.})}{=} \int_{\Delta_1(\Omega)} \cdots \int_{\Delta_r(\Omega)} u(\delta_1, \dots, \delta_z, \delta_{z+1}, \dots, \delta_r) d\pi_{\Delta_r} \dots d\pi_{\Delta_{z+1}} d\pi_{\Delta_z} \dots d\pi_{\Delta_1} \\ &\stackrel{(\star)}{\geq} \int_{\Delta_1(\Omega)} \cdots \int_{\Lambda_r(\Omega)} u(\delta_1, \dots, \delta_z, \lambda_{z+1}, \dots, \lambda_r) d\pi_{\Lambda_r} \dots d\pi_{\Lambda_{z+1}} d\pi_{\Delta_z} \dots d\pi_{\Delta_1} \\ &\stackrel{(\star\star)}{\geq} \int_{\Lambda_1(\Omega)} \cdots \int_{\Lambda_r(\Omega)} u(\lambda_1, \dots, \lambda_z, \lambda_{z+1}, \dots, \lambda_r) d\pi_{\Lambda_r} \dots d\pi_{\Lambda_{z+1}} d\pi_{\Lambda_z} \dots d\pi_{\Lambda_1} \\ &\stackrel{(\text{ind.})}{=} \mathbb{E}_\pi(u \circ Y) \end{aligned}$$

Here,  $(\star)$  is valid because, for fixed cardinal components,  $u$  is isotone in every ordinal component and we have first order stochastic dominance, which means that the iterated integrals gets smaller if one switches from  $\pi_{\Delta_k}$  to  $\pi_{\Lambda_k}$ .

Similarly,  $(\star\star)$  is valid because e.g., for the mapping

$$\psi : \mathbb{R}^{z-1} \rightarrow \mathbb{R} \quad , \quad (\delta_1, \dots, \delta_{z-1}) \mapsto \int_{\Delta_z(\Omega)} u(\delta_1, \dots, \delta_r) d\pi_{\Delta_z}$$

is a positive (affine) linear transformation w.r.t. the corresponding subsystem. □

**Corollary 1** If  $\mathcal{C} = [C, R_1^c, R_2^c]$  is a bounded subsystem of  $\text{pref}(\mathbb{R}^r)$  and  $X, Y \in \mathcal{F}_{(\mathcal{C}, \pi)}$ , then  $\mathcal{C}$  is 0-consistent and ii) and iii) from Prop. 7 hold, if we replace  $R_{(\text{pref}(\mathbb{R}^r), \pi)}$  by  $R_{(\mathcal{C}, \pi)}$ ,  $FSD(R_1^*, \pi)$  by  $FSD(R_1^c, \pi)$ , and  $(X, Y) \in R_{(\text{pref}(\mathbb{R}^r), \pi)}$  by  $\forall u \in \mathcal{N}_{\mathcal{C}} : \mathbb{E}_{\pi}(u \circ X) \geq \mathbb{E}_{\pi}(u \circ Y)$ .

**Proof.** As, according to Proposition 7 i), we know that  $\text{pref}(\mathbb{R}^r)$  is consistent, the same holds true for all of its subsystems. Hence,  $\mathcal{C}$  is consistent. Since  $\mathcal{C}$  is assumed to be bounded, it then is 0-consistent by Proposition 1. The rest of the Corollary follows, since – by Proposition 2 – for bounded preference systems it suffices to check for dominance only over all normalized representations.  $\square$

**Proposition 8** Let  $z = 1$  and denote by  $\mathcal{U}_{\text{sep}}$  the set of all  $u : \mathbb{R}^r \rightarrow \mathbb{R}$  such that, for  $(x_2, \dots, x_r) \in \mathbb{R}^{r-1}$  fixed, the function  $u(\cdot, x_2, \dots, x_r)$  is strictly increasing and (affine) linear and such that, for  $x_1 \in \mathbb{R}$  fixed, the function  $u(x_1, \cdot, \dots, \cdot)$  is strictly isotone w.r.t. the the componentwise partial order on  $\mathbb{R}^{r-1}$ . Then  $\mathcal{U}_{\text{sep}} = \mathcal{U}_{\text{pref}(\mathbb{R}^r)}$ .

**Proof.** First, let  $u \in \mathcal{U}_{\text{pref}(\mathbb{R}^r)}$ . One easily verifies that, for  $x_- := (x_2, \dots, x_r) \in \mathbb{R}^{r-1}$  fixed, the preference system  $Z := [\mathbb{R}, R_1^{x_-}, R_2^{x_-}]$ , where  $R_1^{x_-} := \geq$  and  $R_2^{x_-}$  is defined by

$$\left\{ \left( (t, u), (v, w) \right) : \left( \left( \begin{pmatrix} t \\ x_- \end{pmatrix}, \begin{pmatrix} u \\ x_- \end{pmatrix} \right), \left( \begin{pmatrix} v \\ x_- \end{pmatrix}, \begin{pmatrix} w \\ x_- \end{pmatrix} \right) \right) \in R_2^* \right\}$$

is a complete positive-difference structure in the sense of Krantz et al. [1971, Definition 1, p. 147]. According to Krantz et al. [1971, Theorem 1, p. 147] this implies that any two representations of  $Z$  are positive (affine) linear transformations of each other. But it is immediate that both  $u(\cdot, x_2, \dots, x_r)$  and  $\text{id}_{\mathbb{R}}(\cdot)$  are representations of  $Z$ . Thus,  $u(\cdot, x_2, \dots, x_r) = \alpha \cdot \text{id}_{\mathbb{R}}(\cdot) + \beta$  for some  $\alpha \in \mathbb{R}^+$  and  $\beta \in \mathbb{R}$ , proving the first claim of this direction. The second claim – i.e., the strict isotony of the function  $u(x_1, \cdot, \dots, \cdot)$  w.r.t. the the componentwise partial order on  $\mathbb{R}^{r-1}$  for fixed  $x_1 \in \mathbb{R}$  – is also immediate. Thus,  $u \in \mathcal{U}_{\text{sep}}$ .

For the other direction, assume that  $u \in \mathcal{U}_{\text{sep}}$ . It follows directly from the assumptions that  $u$  is strictly isotone w.r.t.  $R_1^*$ . To see that  $u$  also strictly represents  $R_2^*$ , choose  $((x, y), (x', y')) \in R_2^*$  arbitrary. We have two cases:

*Case 1:*  $((x, y), (x', y')) \in I_{R_2^*}$ . This implies that  $x_1 - y_1 = x'_1 - y'_1$  and therefore also  $x_1 - x'_1 = y_1 - y'_1$ . Moreover, one easily verifies that the restriction of  $R_2^*$  to the ordinal dimensions is antisymmetric. Since we have that  $x_-$  componentwise dominates  $x'_-$  and vice versa and that  $y_-$  componentwise dominates  $y'_-$  and vice versa, this antisymmetry then implies that  $x_- = x'_-$  and  $y_- = y'_-$ . Therefore, there are common  $\alpha_1, \alpha_2 \in \mathbb{R}^+$  and  $\beta_1, \beta_2 \in \mathbb{R}$  such that

$$\begin{aligned} u(x) &= \alpha_1 \cdot x_1 + \beta_1 & , & & u(x') &= \alpha_1 \cdot x'_1 + \beta_1 \\ u(y) &= \alpha_2 \cdot y_1 + \beta_2 & , & & u(y') &= \alpha_2 \cdot y'_1 + \beta_2 \end{aligned}$$

Moreover, observe that  $\alpha_1 = \alpha_2$ , since otherwise there would be  $x^* \in \mathbb{R}$  with  $u(x^*, x_-) < u(x^*, y_-)$ , which is not possible, since  $u$  is strictly isotone w.r.t.  $R_1^*$ . Define

$$D := (u(x) - u(y)) - (u(x') - u(y')).$$

Simple computations then yield

$$D = \alpha_1 \cdot (x_1 - x'_1) - \alpha_2 \cdot (y_1 - y'_1) = (x_1 - x'_1) \cdot (\alpha_1 - \alpha_2)$$

which, as  $\alpha_1 = \alpha_2$ , implies  $D = 0$ .

*Case 2:*  $((x, y), (x', y')) \in P_{R_2^*}$ . This implies  $x_- \geq x'_- \geq y'_- \geq y_-$ , where  $\geq$  is to be understood componentwise. Using the same argument as seen before, this implies that there exists a  $\alpha \in \mathbb{R}^+$  and  $\beta_1, \beta_2, \beta_3, \beta_4 \in \mathbb{R}$  such that

$$\begin{aligned} u(x) &= \alpha \cdot x_1 + \beta_1 & , & & u(x') &= \alpha \cdot x'_1 + \beta_3 \\ u(y) &= \alpha \cdot y_1 + \beta_2 & , & & u(y') &= \alpha \cdot y'_1 + \beta_4 \end{aligned}$$

Thus, computing  $D$  defined as above yields:

$$D = \alpha \cdot ((x_1 - y_1) - (x'_1 - y'_1)) + \beta_1 - \beta_2 - \beta_3 + \beta_4$$

*Sub-Case 2.1:*  $x_1 - y_1 > x'_1 - y'_1$ . Observe that, as  $u$  is isotone w.r.t.  $R_1^*$ , we have that  $u(y'_1, y'_-) \geq u(y'_1, y_-)$ . However, this implies  $\beta_4 \geq \beta_2$ . Analogous reasoning yields  $\beta_1 \geq \beta_3$ . Using the assumptions of the sub-case, this implies  $D > 0$ .

*Sub-Case 2.2:*  $x_1 - y_1 = x'_1 - y'_1$ . Using the case assumption, this implies that either  $x_- > x'_-$  or  $y'_- > y_-$ , where the  $>$  is to be understood as the strict part of the componentwise  $\geq$ . As  $u$  is strictly isotone w.r.t.  $R_1^*$ , this implies that either  $u(y'_1, y'_-) > u(y'_1, y_-)$  or  $u(x'_1, x_-) > u(x'_1, x'_-)$ , which itself implies either  $\beta_4 > \beta_2$  or  $\beta_1 > \beta_3$ . As we know  $\beta_4 \geq \beta_2$  and  $\beta_1 \geq \beta_3$ , this, together with the sub-case assumption, implies  $D > 0$ .  $\square$

## B DETAILS ON IMPLEMENTATION AND REPRODUCIBILITY

In Section 8.1 we stated that the implementation of the constraint matrix has worst-case complexity  $\mathcal{O}(s^4)$ . This worst case occurs when everything in  $R_1^*$  and  $R_2^*$  is comparable and then

$$s \cdot (s - 1) + (s \cdot (s - 1)) \cdot ((s \cdot (s - 1)) - 1) = s^4 - 2s^3 + s^2$$

many pairwise comparisons have to be considered. Note that we omit the reflexive part of the pre-orders  $R_1^*$  and  $R_2^*$ .

We are interested in the non-regularized test statistic as well as the regularized test statistic with  $\varepsilon \in \{0.25, 0.5, 0.75, 1\}$ , see Section 8. For all these cases, we compute the test statistics based on the sample, as well as 1000 times on a permuted version of that sample. Note that the linear programs for computing the test statistics based on the permuted data are identical to that for the non-permuted data except for the objective function, see Section 5.2. In Section C (in the supplementary material), we prove that the robustified test statistics are a shift of the non-robustified test statistic. Thus, the robustified test statistics are immediately given.

The simulation is based on a random sample of the data set. Two of the data sets and the corresponding R-code can be found here:

[https://github.com/hannahblo/Robust\\_GSD\\_Tests](https://github.com/hannahblo/Robust_GSD_Tests)

The data set used for the poverty analysis (ALLBUS) is freely accessible, but registration in the corresponding online portal is needed.<sup>1</sup>

For the computation of the linear programs, we used the R interface of Gurobi optimizer, which is documented in Gurobi Optimization, LLC [2020]. This is a commercial solver that offers free academic licenses<sup>2</sup>. In particular, the computation of linear programs is faster than using the free and open source solvers known to us, see Meindl and Templ [2012]. We also used the R-packages *purrr*, *dplyr*, *slam*, *readr*, *tidyr*, *forecast*, *ggplot2*, *reshape2*, *tidyverse*, *ggridges*, *latex2exp*, *RColorBrewer*, *rcartocolor* and *foreign* for our implementation, see Mailund [2022], Yarberrry and Yarberrry [2021], Wickham et al. [2022], Hornik et al. [2022], Wickham et al. [2023], Hyndman et al. [2023], Wickham and Chang [2014], Wickham [2022], Wickham and RStudio [2022], Wilke [2022], Meschiari [2022], Neuwirth [2022], Nowosad [2022], R Core Team et al. [2022].

The computation was done for

- ALLBUS data set, see GESIS [2018], on a commodity desktop laptop with a 8-core Intel(R) Core(TM) i7-8665U CPU @ 1.90GHz processor and 16 GB RAM in R version 4.2.2.
- dermatology data set, see Demiroz et al. [1998] accessed via Dua and Graff [2017], on a commodity desktop computer with a 32-core Intel(R) Xeon(R) CPU E5-2650 v2 @ 2.60GHz processor and 64 GB RAM in R version 4.2.1
- German credit data set, see Dua and Graff [2017], on a commodity desktop laptop with a 8-core Intel(R) Core(TM) i7-10510U CPU @ 1.80GHz processor and 16 GB RAM in R version 4.2.2.

## C CALCULATIONS FOR ROBUSTIFIED TEST STATISTICS

In Section 8 we show a graph visualizing the fraction of resamples in favor of **non**-rejection of  $H_0$  (i.e., the p-values) as a function of the size of the contamination  $\gamma$  of the underlying linear-vacuous model (see Figure 3). We will briefly show here how the exact function is calculated. For general (polyhedral) credal sets, a resample  $I$  is in favor of rejection of  $H_0$  under the robustified resampling scheme, if  $\underline{d}_{\mathbf{X}, \mathbf{Y}}^\varepsilon(\omega_0) > \bar{d}_I^\varepsilon$ . Hence, the fraction of resamples in favor of rejection of  $H_0$  is given by

$$\frac{1}{N} \cdot \sum_{I \in \mathcal{I}_N} \mathbb{1}_{\{\underline{d}_{\mathbf{X}, \mathbf{Y}}^\varepsilon(\omega_0) > \bar{d}_I^\varepsilon\}}$$

where  $N$  denotes the number of resamples and  $\mathcal{I}_N$  is the corresponding set of resamples. In the special case that the credal sets involved are  $\gamma$ -contamination models, we can use Proposition 6 (and a slight variation of it with  $\pi_*$  and  $\pi^*$  in reversed

<sup>1</sup>Further information on the survey and the data set itself can be found here: [https://search.gesis.org/research\\_data/ZA5240](https://search.gesis.org/research_data/ZA5240) (accessed: Febr 16, 2023)

<sup>2</sup>Further details can be found here: <https://www.gurobi.com/academia/academic-program-and-licenses/> (accessed: Febr 16, 2023)



roles) to obtain

$$\underline{d}_{\mathbf{X}, \mathbf{Y}}^\varepsilon(\omega_0) = (1 - \gamma) \cdot d_{\mathbf{X}, \mathbf{Y}}^\varepsilon(\omega_0) - \gamma$$

and

$$\bar{d}_I^\varepsilon = (1 - \gamma) \cdot d_I^\varepsilon + \gamma$$

and, therefore, the condition in the indicator above is satisfied if and only if

$$d_{\mathbf{X}, \mathbf{Y}}^\varepsilon(\omega_0) - d_I^\varepsilon > \frac{2\gamma}{(1 - \gamma)}.$$

Finally, if we interpret  $\varepsilon$  as a function parameter, then we can write the fraction of resamples in favor of **non**-rejection of  $H_0$  (i.e., the observed p-values) as a function of the size  $\gamma$  of the contamination of the underlying linear-vacuous model:

$$f_\varepsilon(\gamma) := 1 - \frac{1}{N} \cdot \sum_{I \in \mathcal{I}_N} \mathbb{1}_{\{d_{\mathbf{X}, \mathbf{Y}}^\varepsilon(\omega_0) - d_I^\varepsilon > \frac{2\gamma}{(1 - \gamma)}\}}.$$

## D FURTHER DETAILS ON THE APPLICATIONS

### D.1 DATA SETS

We applied our analysis to three different data sets:

- For the poverty analysis, see Section 8, we used the ALLBUS data set. The data set is described by GESIS [2018] and Breyer and Danner [2015]. As mentioned already in the previous section, the data set is freely accessible, but only after registration in the corresponding online portal: [https://search.gesis.org/research\\_data/ZA5240](https://search.gesis.org/research_data/ZA5240) (accessed: 08.02.2023). Please download the file ZA5240\_v2-2-0.sav (5.31MB) there.

The analysis was done on a sample consisting of 100 female and 100 male observations.

- We analyzed the dermatology data set, see Demiroz et al. [1998] accessed via Dua and Graff [2017].  
The analysis was performed on a sample of 46 individuals with family history of eryhemato-squamous disease and 100 individuals without.
- We analyzed the German credit data set, see Dua and Graff [2017].  
The analysis was performed on a sample of 100 credit risks classified as good and 100 credit risks classified as poor individuals.

### D.2 APPLICATION ON CREDIT DATA

We focus on three variables (features) in the German credit data set Dua and Graff [2017]: credit amount (numeric), credit history (ordinal, 5 levels ranging from “delay in paying off in the past” to “all credits paid back duly”) and employment status (ordinal, 5 levels ranging from “unemployed” to “present employment longer than 7 years”). We use a subsample with  $n = m = 100$  high-risk applicants and low-risk applicants each. We are interested in the hypothesis that high-risk applicants are dominated by low-risk applicants w.r.t. GSD. The test results (see Figures 1 and 2 in the supplementary material) can be interpreted analogously to Section 8: For  $\varepsilon \in \{0.75, 1\}$  we reject for the common significance level of  $\alpha \approx 0.05$ . This time, we do not reject in case of  $\varepsilon = 0.5$ .

Similar to the example of poverty analysis in Section 8, rejecting  $H_0$  does not necessarily mean that high-risk applicants are dominated by low-risk applicants. They could also be incomparable, see also Section 5. However, our tests with reversed variables give no evidence of incomparability: The observed p-values for all these reversed tests are all 1.

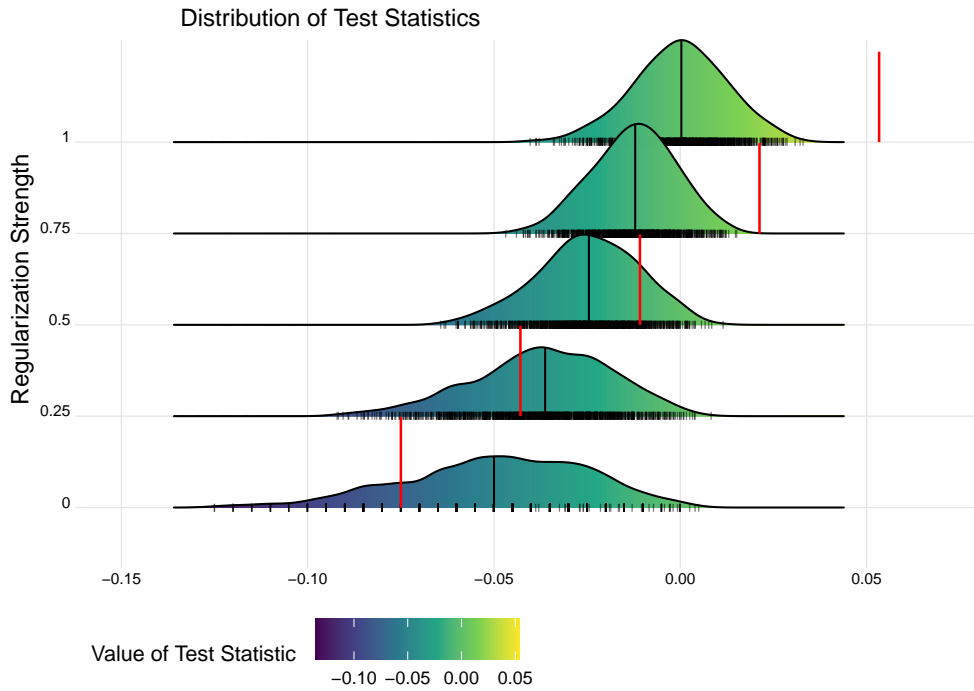


Figure 1: Distributions of  $d_I^\varepsilon$  with  $\varepsilon \in \{0, 0.25, 0.5, 0.75, 1\}$  obtained from  $N = 1000$  resamples of Credit data. Black stripes show exact positions of  $d_I^\varepsilon$  values. Vertical black line marks median. Red line shows value of the respective observed test statistics  $d_{\mathbf{X}, \mathbf{Y}}^\varepsilon(\omega)$ .

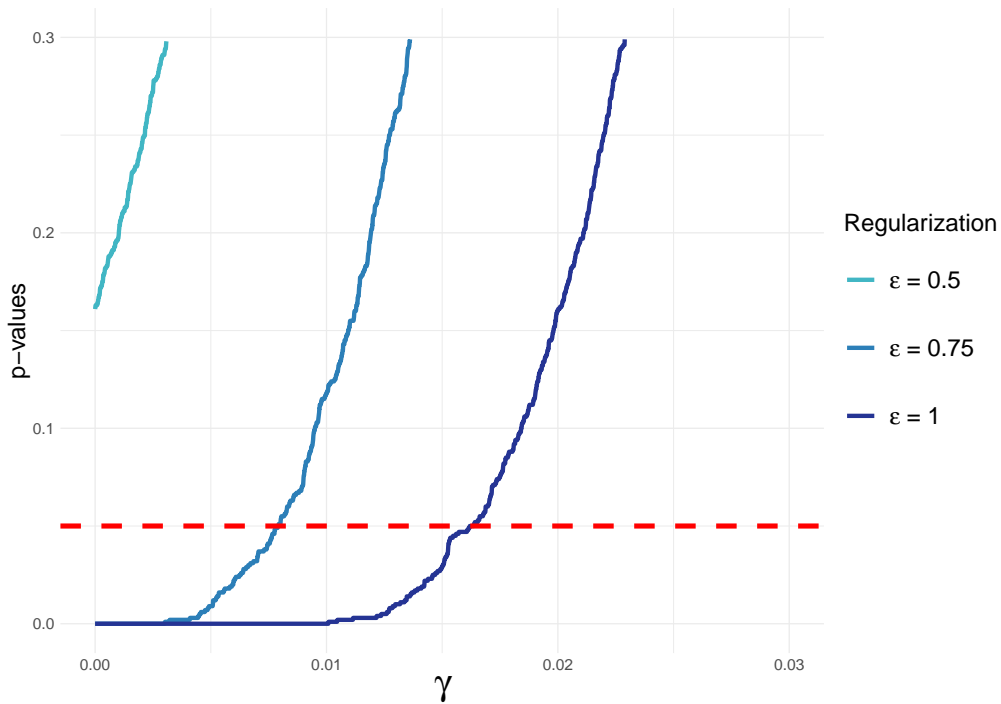


Figure 2: P-values as function of the contamination  $\gamma$  (see Supp. C) for tests with different regularization strength  $\varepsilon$  performed nd on credit data set. Dotted red line marks significance level  $\alpha = 0.05$ .

### D.3 APPLICATION ON DERMATOLOGICAL DATA

We focus on three variables (features) in the dermatology data set Demiroz et al. [1998], Dua and Graff [2017]: age of skin (numeric), the intensity of itching (ordinal, 4 levels ranging from “no itching” to “strong itching”) and erythema (redness of skin) (ordinal, 4 levels again ranging from no to highest intensity). We use a subsample with  $n = 46$  patients with a family history of eryhemato-squamous disease and  $m = 100$  without. We are interested in the hypothesis that patients without a family history of the disease are dominated by patients without a family history with respect to GSD. The test results (see Figures 3 and 4 in the supplementary material) can be interpreted analogously to Section 8: For  $\varepsilon \in \{0.75, 1\}$  we again reject for the common significance level of  $\alpha \approx 0.05$ . However, the p-values are much higher than in the other two applications, see also Figure 4 (in the supplementary material).

Similar to the example of poverty analysis in Section 8, rejecting  $H_0$  does not necessarily mean that patients with a family history of eryhemato-squamous disease are dominated by patients without. They could also be incomparable; see also Section 5. However, our tests with reversed variables give no evidence of incomparability: The observed p-values for all these reversed tests are all 1.

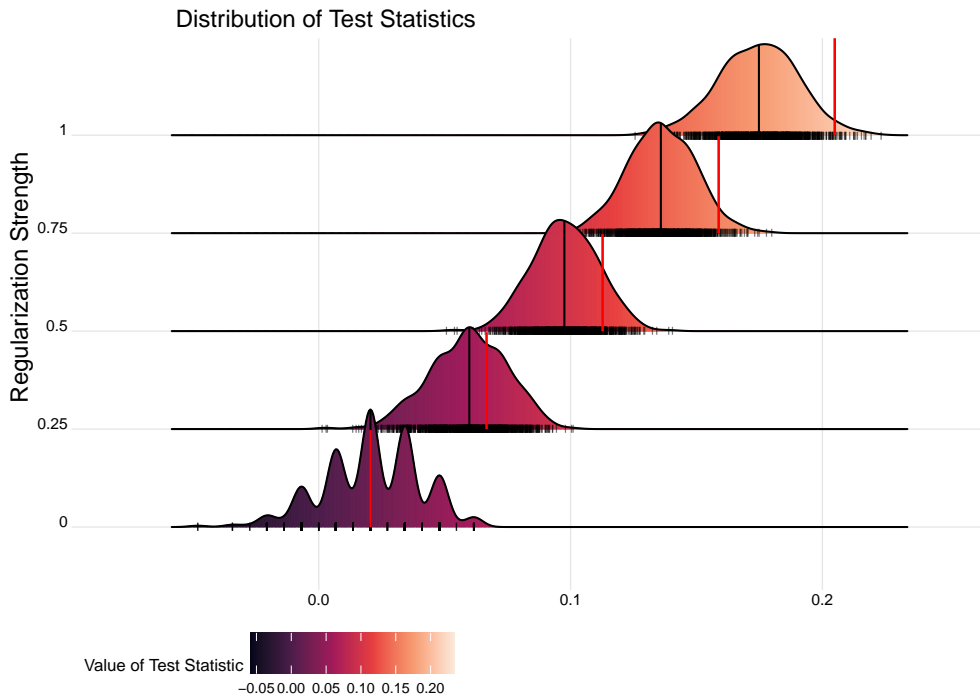


Figure 3: Distributions of  $d_I^\varepsilon$  with  $\varepsilon \in \{0, 0.25, 0.5, 0.75, 1\}$  obtained from  $N = 1000$  resamples of dermatology data. Black stripes show exact positions of  $d_I^\varepsilon$  values. Vertical black line marks median. Red line shows value of the respective observed test statistics  $d_{\mathbf{X}, \mathbf{Y}}^\varepsilon(\omega)$ .

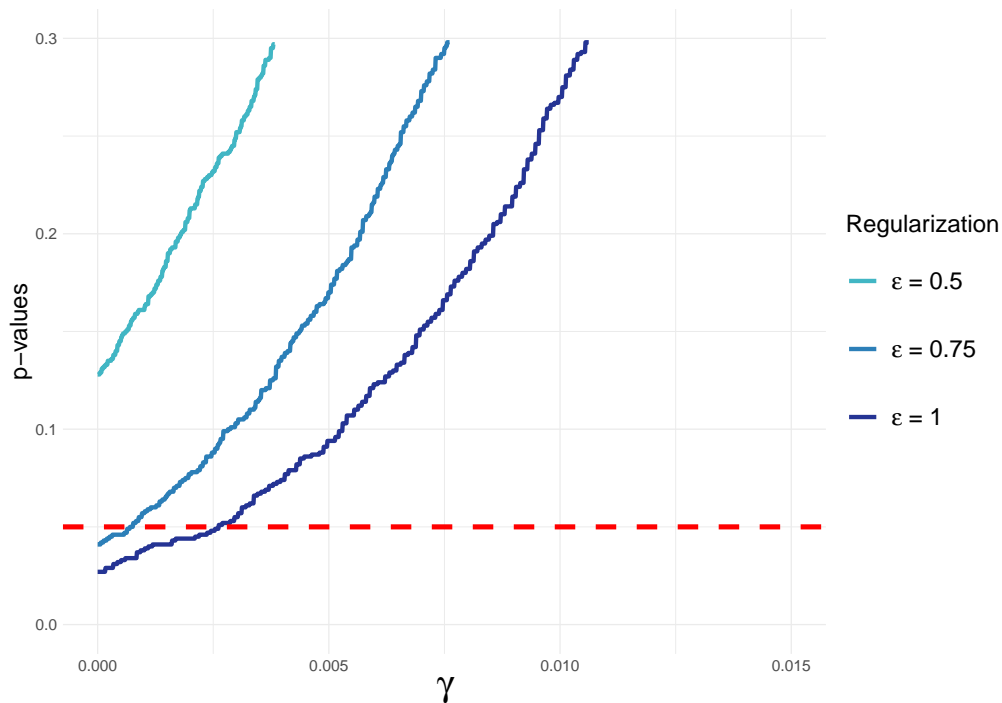


Figure 4: P-values as function of the contamination  $\gamma$  (see Supp. C) for tests with different regularization strength  $\varepsilon$  performed on Dermatology data set. The dotted red line marks significance level  $\alpha = 0.05$ .

## References

- Gurobi Optimization, LLC. Gurobi optimizer reference manual, 2020. URL [https://www.gurobi.com/wp-content/plugins/hd\\_documentations/documentation/9.0/refman.pdf](https://www.gurobi.com/wp-content/plugins/hd_documentations/documentation/9.0/refman.pdf). [Online; Accessed 08.02.2023].
- B. Breyer and D. Danner. Skala zur Erfassung des Lebenssinns (ALLBUS). In *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS) (GESIS – Leibniz-Institut für Sozialwissenschaften)*, volume 10, 2015.
- G. Demiroz, H. Govenir, and N. Ilter. Learning differential diagnosis of Erythematous-Squamous diseases using voting feature intervals. *Artif Intell Med*, 13:147–165, 1998.
- D. Dua and C. Graff. UCI machine learning repository, 2017. <http://archive.ics.uci.edu/ml>.
- GESIS. Allgemeine Bevölkerungsumfrage der Sozialwissenschaften ALLBUS 2014. GESIS Datenarchiv, Köln. ZA5240 Datenfile Version 2.2.0, <https://doi.org/10.4232/1.13141>, 2018.
- K. Hornik, D. Meyer, and C. Buchta. Package ‘slam’, October 2022. URL <https://cran.r-project.org/web/packages/slam/slam.pdf>. [Online; Accessed 08.02.2023].
- R. Hyndman, G. Athanasopoulos, C. Bergmeir, G. Caceres, L. Chhay, K. Kuroptev, M. O’Hara-Wild, F. Petropoulos S. Razbash, E. Wang, F. Yasmeeen, F. Garza, D. Girolimetto, R. Ihaka, R Core Team, D. Reid, D. Shaub, Y. Tang, X. Wang, and Z. Zhou. Package ‘forecast’, January 2023. URL <https://cran.r-project.org/web/packages/forecast/forecast.pdf>. [Online; Accessed 09.02.2023].
- D. Krantz, R. Luce, P. Suppes, and A. Tversky. *Foundations of Measurement. Volume I: Additive and Polynomial Representations*. Academic Press, 1971.
- T. Mailund. Functional programming: purrr. In *R 4 Data Science Quick Reference: A Pocket Guide to APIs, Libraries, and Packages*, pages 89–110. Springer, 2022.

- B. Meindl and M. Templ. Analysis of commercial and free and open source solvers for linear optimization problems. *Eurostat and Statistics Netherlands within the project ESSnet on common tools and harmonised methodology for SDC in the ESS*, 20, 2012.
- S. Meschiari. Package ‘latex2exp’, November 2022. URL <https://cran.r-project.org/web/packages/latex2exp/latex2exp.pdf>. [Online: Accessed 09.02.2023].
- E. Neuwirth. Package ‘rcolorbrewer’, October 2022. URL <https://cran.r-project.org/web/packages/RColorBrewer/RColorBrewer.pdf>. [Online: Accessed 09.02.2023].
- J. Nowosad. Package ‘rcartocolor’, October 2022. URL <https://cran.r-project.org/web/packages/rcartocolor/rcartocolor.pdf>. [Online: Accessed 09.02.2023].
- R Core Team, R. Bivand, V. Carey, S. DebRoy, S. Eglen, R. Guha, S. Herbrandt, N. Lewin-Koh, M. Myatt, M. Nelson, B. Pfaff, B. Quistorff, F. Warmerdam, S. Weigand, and Inc. Free Software Foundation. Package ‘foreign’, December 2022. URL <https://cran.r-project.org/web/packages/foreign/foreign.pdf>. [Online: Accessed 09.02.2023].
- H. Wickham. Package ‘reshape’, October 2022. URL <https://cran.r-project.org/web/packages/reshape/reshape.pdf>. [Online: Accessed 09.02.2023].
- H. Wickham and W. Chang. Package ‘ggplot2’, December 2014. URL <https://cran.microsoft.com/snapshot/2015-01-06/web/packages/ggplot2/ggplot2.pdf>. [Online: Accessed 09.02.2023].
- H. Wickham and RStudio. Package ‘tidyverse’, October 2022. URL <https://cran.r-project.org/web/packages/tidyverse/tidyverse.pdf>. [Online: Accessed 09.02.2023].
- H. Wickham, J. Hester, R. Francois, J. Bryan, and S. Bearrows. Package ‘readr’, October 2022. URL <https://cran.r-project.org/web/packages/readr/readr.pdf>. [Online; Accessed 08.02.2023].
- H. Wickham, D. Vaughan, M. Girlich, K. Ushey, and PBC Posit. Package ‘tidyr’, January 2023. URL <https://www.vps.fmvz.usp.br/CRAN/web/packages/tidyr/tidyr.pdf>. [Online: Accessed 09.02.2023].
- C. O. Wilke. Package ‘ggridges’, October 2022. URL <https://cran.r-project.org/web/packages/ggridges/ggridges.pdf>. [Online: Accessed 09.02.2023].
- W. Yarberry and W. Yarberry. Dplyr. *CRAN Recipes: DPLYR, Stringr, Lubridate, and RegEx in R*, pages 1–58, 2021.