

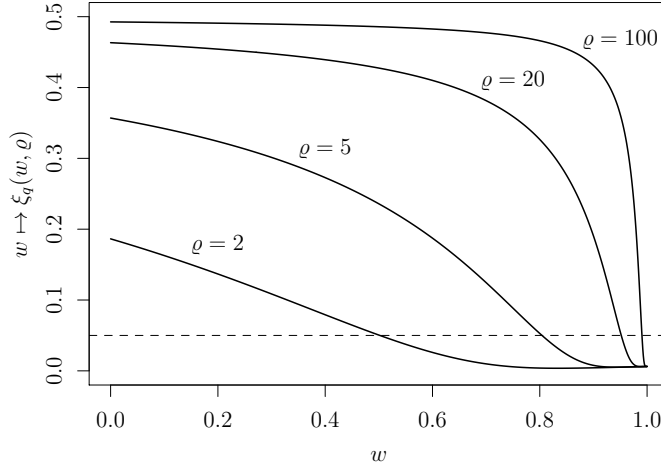


FIGURE 1. Solid lines show the size bound $\xi_q(w, \rho)$ at $q = 20$ control observations as a function of the weight w for different values of the maximal heterogeneity ρ . The dashed line equals .05.

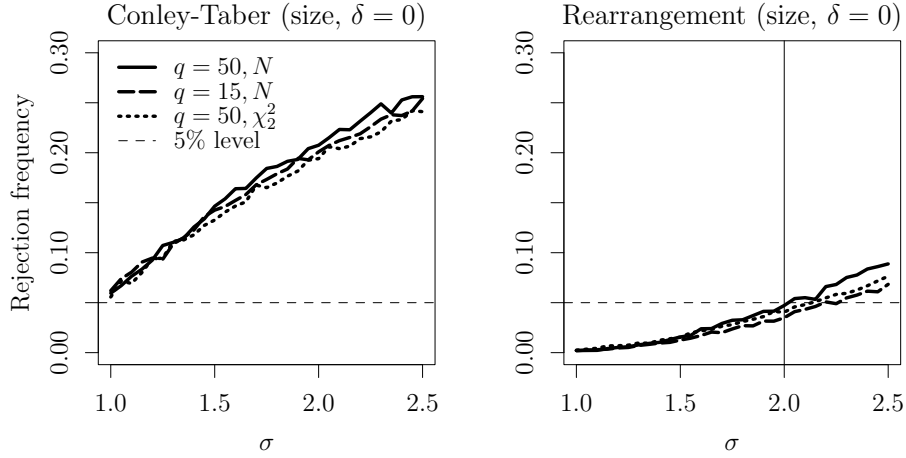


FIGURE 2. Rejection frequencies of a true null as a function of the heterogeneity σ for the Conley-Taber test (left) and the rearrangement test (right) with (i) $q = 50$ control clusters and normal errors (solid lines), (ii) $q = 15$ and normal errors (long-dashed), and (iii) $q = 50$ and chi-squared errors (dotted). The short-dashed line equals .05. The rearrangement test uses $\rho = 2$ (vertical line).

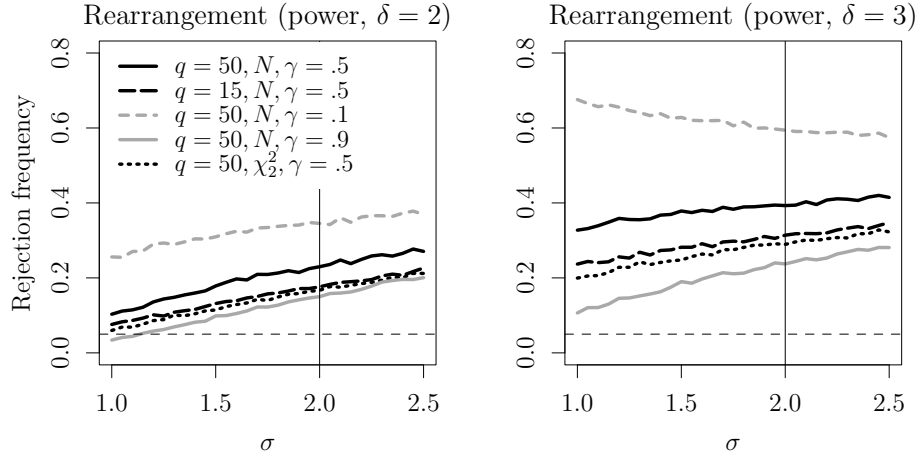


FIGURE 3. Rejection frequencies of the rearrangement test ($\varrho = 2$) under the alternative as a function of the heterogeneity σ at $\delta = 2$ (left) and $\delta = 3$ (right) with (i) and (ii) as in Figure 2, (iii) is (i) with weak time dependence $\gamma = .1$ (short-dashed grey), (iv) is (i) with strong time dependence $\gamma = .9$ (solid grey) (v) is (i) with chi-squared errors (dotted). The short-dashed line equals .05.

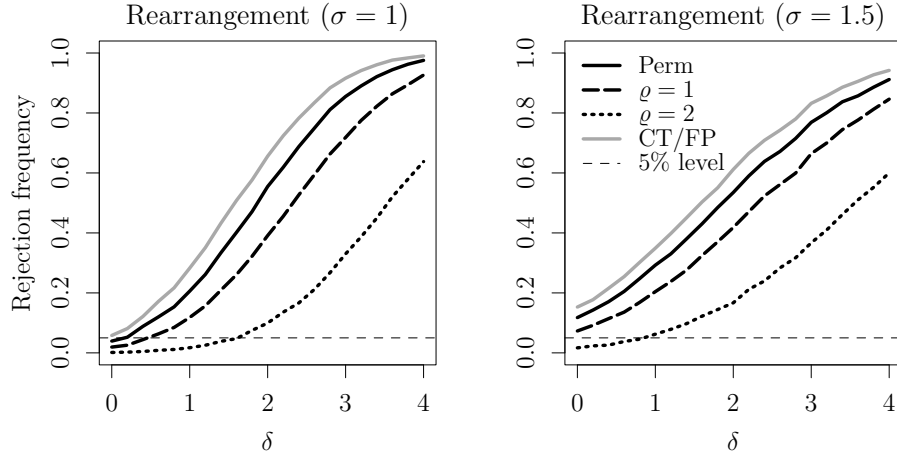


FIGURE 4. Rejection frequencies of (i) the permutation test (solid lines), the rearrangement test with (ii) $\varrho = 1$ (dashed), and (iii) $\varrho = 2$ (dotted) as a function of the treatment effect δ at $\sigma = 1$ (left) and $\sigma = 1.5$ (right). Grey lines are the Conley-Taber/Ferman-Pinto test with correctly (left) and incorrectly (right) specified heterogeneity. The null hypothesis is true at $\delta = 0$.