Supplemental material: Multiple testing of conditional independence hypotheses using information-theoretic approach

Małgorzata Łazęcka[1,2[0000−0003−0975−4274]] and Jan Mielniczuk[1,2[0000−0003−2621−2303]]

[1] Institute of Computer Science
Polish Academy of Sciences
Warsaw, Jana Kazimierza 5, 01-248, Poland
{malgorzata.lazecka,jan.mielniczuk}@ipipan.waw.pl
[2] Faculty of Mathematics and Information Science
Warsaw University of Technology
Warsaw, Koszykowa 75, 00-662 Poland

1 Proof of Lemma 1

Lemma 1. *We have (i)*

$$JMI = 0 \iff H_0 = \cap_i H_{0i} \text{ holds.}$$

(ii) The following representation is valid

$$JMI = I(X,Y) + \frac{1}{p}\sum_{i=1}^{p}(I(X,Z_i|Y) - I(X,Z_i))$$

(iii)

$$\frac{1}{2}\sum_{i=1}^{p}\Big(\sum_{x,y,z_i}|p(x,y,z_i) - p(x|z_i)p(y|z_i)p(z_i)|\Big)^2 \leq p \times JMI \leq \sum_{i=1}^{p}\log(\chi_i^2+1)$$

and both inequalities are tight when H_0 holds.

Proof. The proof of (i) is almost immediate in the view of our short discussion of properties of CMI. Namely, as the summands in the definition of JMI are non-negative, JMI equals 0 only when all $I(X,Y|Z_i)$ are 0 and this is equivalent to $X \perp Y|Z_i$ for $i = 1,\ldots,p$ in the view of information inequality (see Cover and Thomas (2006)). The proof of (ii) follows by applying chain rule (4) twice.

In order to prove (iii) observe that concavity of logarithmic function and Jensen's inequality imply for any $i = 1,\ldots,p$

$$\begin{aligned} I(X,Y|Z_i) &= \sum_{x,y,z_i} p(x,y,z_i)\log\Big(\frac{p(x,y|z_i)}{p(x|z_i)p(y|z_i)}\Big) \leq \log\Big(\sum_{x,y,z_i}\frac{p^2(x,y,z_i)}{p(x|z_i)p(y|z_i)p(z_i)}\Big) \\ &= \log\Big(\sum_{x,y,z_i}\frac{(p(x,y,z_i) - p(x|z_i)p(y|z_i)p(z_i))^2}{p(x|z_i)p(y|z_i)p(z_i)} + 1\Big) = \log(\chi_i^2+1). \end{aligned}$$

Summing over $i = 1,\ldots,p$ yields the RHS inequality. LHS is a direct consequence of Pinsker inequality (cf. e.g. Tsybakov (2009)).

2 Supplementary figures

Figures 1 and 2 show the behaviour of the true asymptotic distribution of $\widehat{JMI}$ and its estimate.The left column depicts boxplots of sorted eigenvalues $\lambda_i(\hat{M})$ and compares them with $\lambda_i(M)$. The middle column compares averaged CDFs corresponding to $\lambda_i(\widehat{M})$ and 90% confidence bands for the true CDF based on them with the true asymptotic CDF and the empirical CDF based on $\widehat{JMI}$ values. The right column shows actual type I errors versus the assumed level α based on $N = 5000$ repetitions of the experiment. In Figure 1, we show again the results from the main text for Model B and estimated type I errors and we added remaining results for Model A and C. In Figure 2 we show the results for independent variables X, Y and $Z_1, Z_2, \ldots, Z_p$ sampled from Bernoulli distribution with probability of success 0.5 for $p = 3, 5, 7$. The results are based on $N = 5000$ experiments.

Figure 3 shows actual type I errors versus the assumed level α based on $N = 5000$ repetitions of the experiment.

In Figure 4 the behaviour of the power of the considered procedures is compared against one of the model's varying parameters when the remaining ones are held fixed and the significance level is set to $\alpha = 0.05$. The plots for the models A and C are also presented in the main text. The parameters are chosen in such a way that the strength of the dependence between the variables allow to easily compare the power of the testing procedures. Parameters p and s control number of variables and number of significant variables respectively (in Models A and B $p = s$). Through parameters α_x, α_y (Model A, A'), α_z (Model B, B') and q (Model C(q)) we control the strength of dependence. The parameter q_{XY} (Model C(q, q_{XY})) is not easily interpretable, but due to its presence the H_0 is violated. In the supplement we show results for Model B that were moved from the main text (Figure 4) and additional plots (Figures 5-7) with different values of the parameters from those considered in Figure 4 (Figure 3 in the main text). Figures 5-7 compare the performance of the procedures for smaller (left column) and larger (right column) parameter values than in Figure 4 (middle column).

Figure 8 shows ROC-type curves for all three procedures considered in the article. ROC-type curves are based on two models: the one for which H_0 holds and the second for which H_1 is true, and the report *the actual* type I error and the power approximated by means of simulations for varying α. In this way y values of three ROC curves for the fixed x value correspond to the power for *the same* actual type I error.

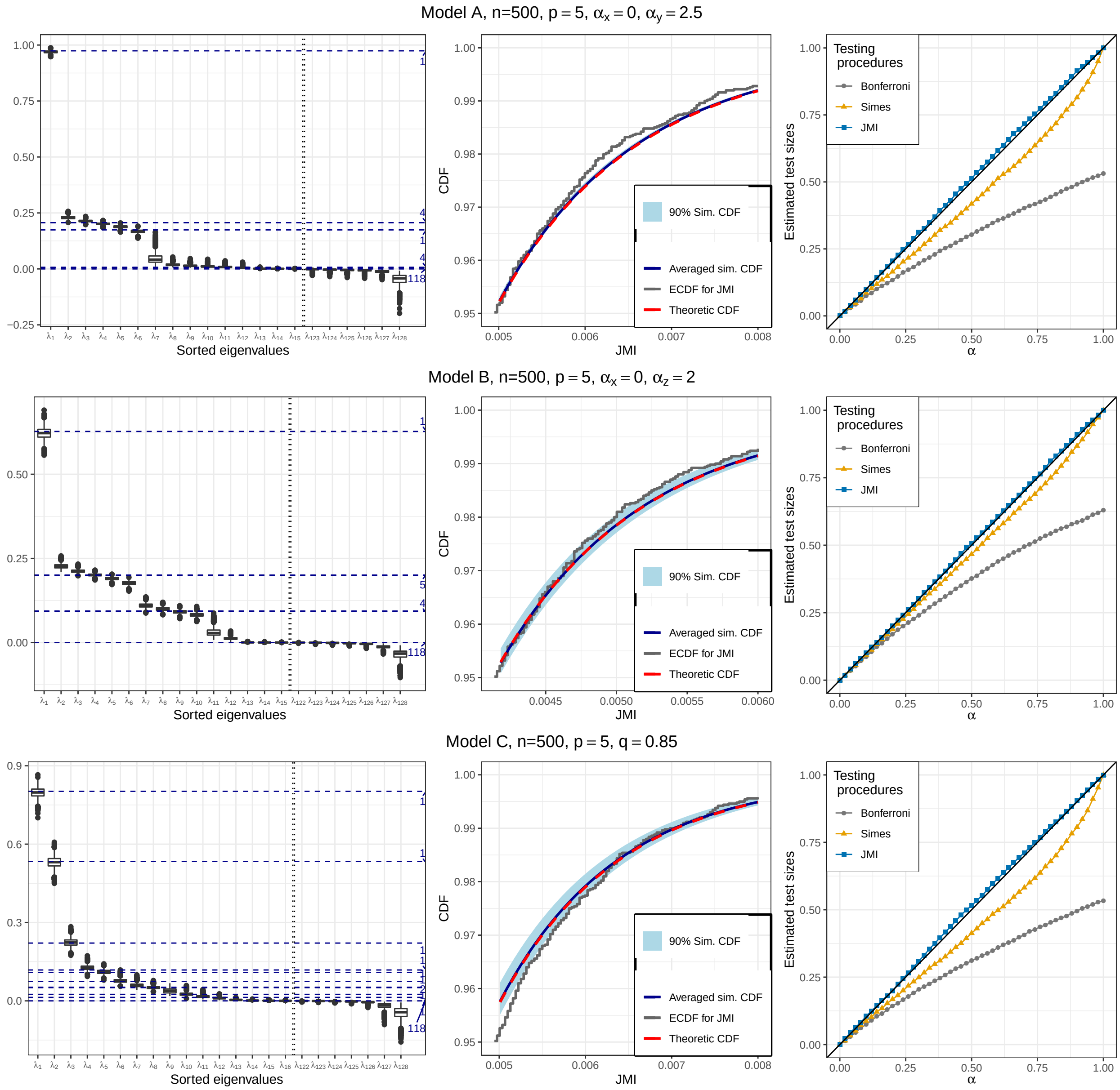

Fig. 1: Left panel: Box-plots of the empirical values $\lambda_i(\widehat{M}), i = 1, \ldots, 128$ for Models A, B and C based on $N = 5000$ repetitions. True eigenvalues are marked by the horizontal lines and are equal to

- 0.974 (multiplicity 1), 0.206 (multiplicity 4), 0.174 (multiplicity 1), 0.006 (multiplicity 4), 0 (multiplicity 118) for Model A,
- 0.627 (multiplicity 1), 0.2 (multiplicity 5), 0.093 (multiplicity 4), 0 (multiplicity 118) for Model B,
- 0.013, 0.025, 0.051, 0.052, 0.074, 0.109, 0.118, 0.221, 0.534, 0.802, (multiplicity 1), 0 (multiplicity 118) for Model C.

The central panel: values of theoretical CDF, the empirical CDF of $\widehat{JMI}$ and the average of CDFs corresponding to $\lambda_i(\hat{M})$ for the values of JMI grater than 0.95 quantile of $\widehat{JMI}$. Right panel: Actual type I errors against assumed level α.

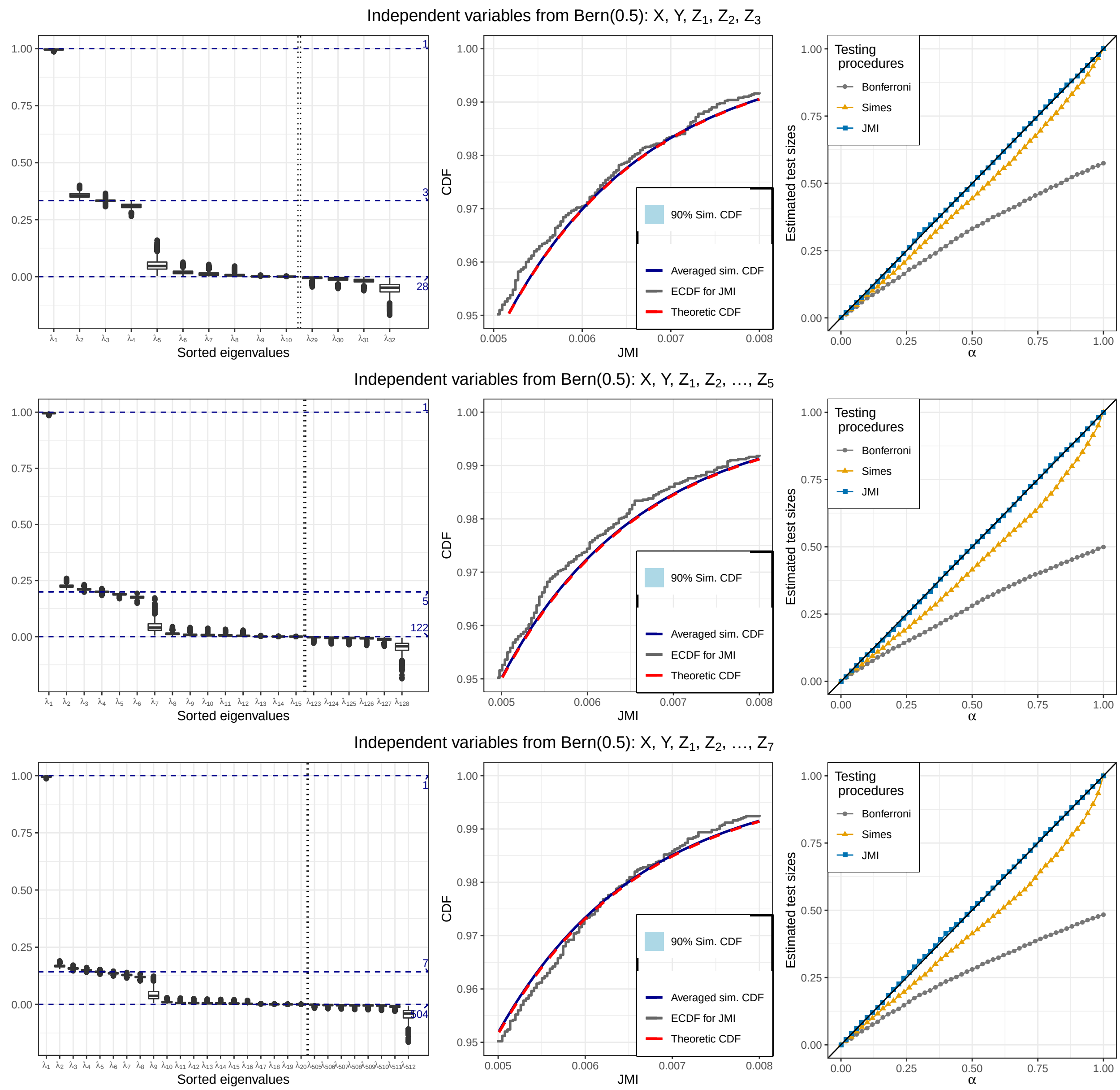

Fig. 2: Left panel: Box-plots of the empirical values $\lambda_i(\hat{M})$, $i = 1, \ldots, 2^{p+2}$ for a model, in which X, Y, $Z_1, Z_2, \ldots, Z_p$ for $p = 3, 5, 7$ are iid Bernoulli variables with probability of success 0.5 based on $N = 5000$ repetitions. True eigenvalues are marked by the horizontal lines and are equal to

- 1 (multiplicity 1), 0.333 (multiplicity 3), 0 (multiplicity 28) for $p = 3$,
- 1 (multiplicity 1), 0.2 (multiplicity 5), 0 (multiplicity 122) for $p = 5$,
- 1 (multiplicity 1), 0.143 (multiplicity 5), 0 (multiplicity 504) for $p = 7$.

The central panel: values of theoretical CDF, the empirical CDF of $\widehat{JMI}$ and the average of CDFs corresponding to $\lambda_i(\hat{M})$ for the values of JMI greater than 0.95 quantile of $\widehat{JMI}$. Right panel: Actual type I errors against assumed level α. The setting corresponds to Model A with parameters: $\alpha_x = 0$, $\alpha_y = 0$ and Model B with parameters $\alpha_z = 0$ and any α_x for $p = 3, 5, 7$.

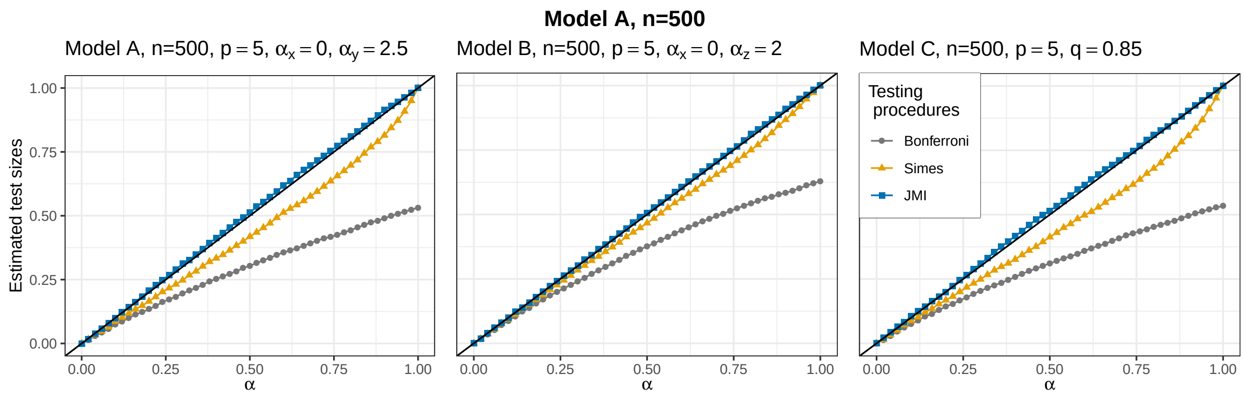

Fig. 3: Actual type I errors against assumed level α in models A, B and C based on $N = 5000$ simulations. Parameters of the models indicated in the header.

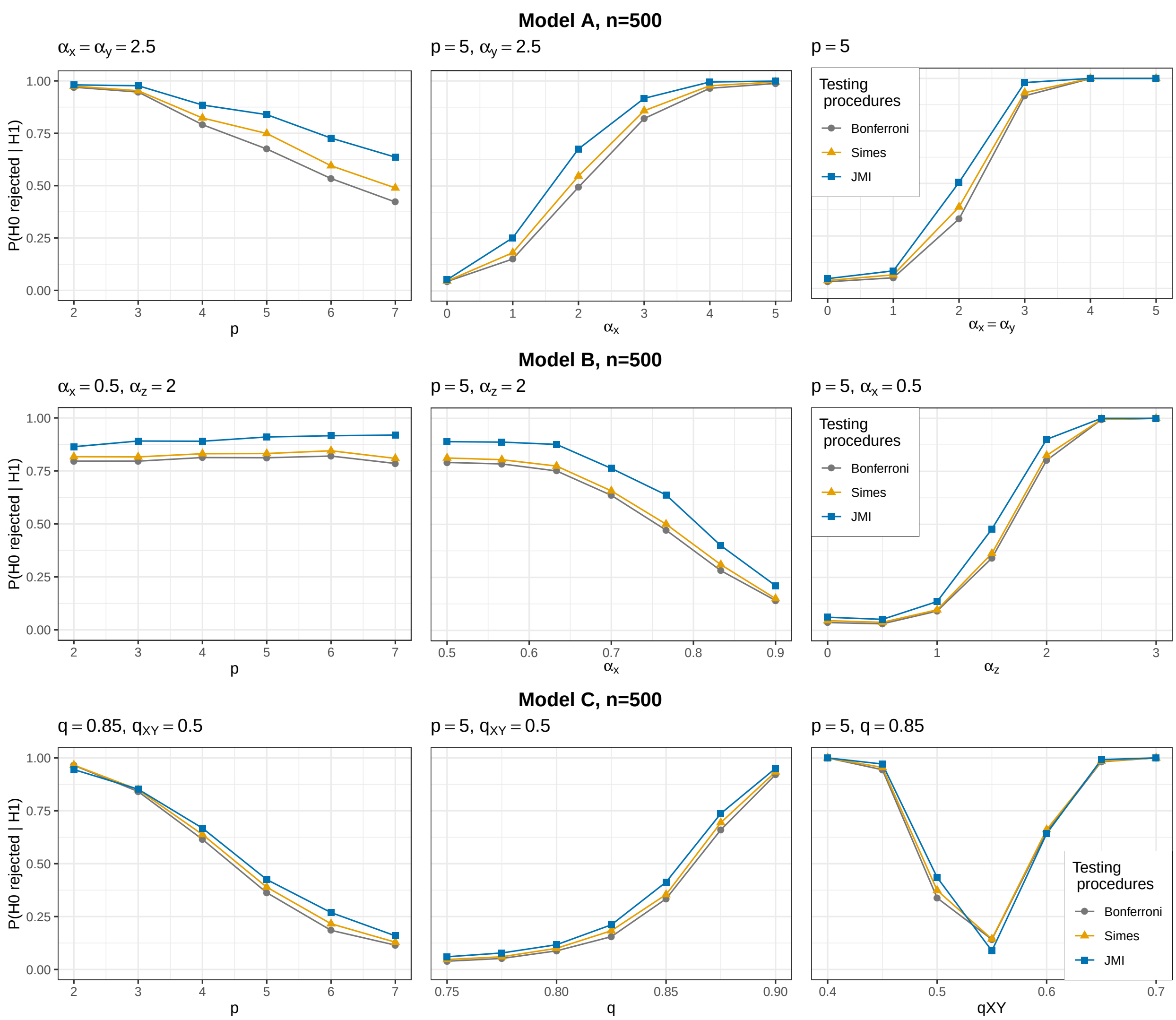

Fig. 4: Power against the changing number of variables and the parameter values for models A, B and C based on $N = 1000$ simulations.

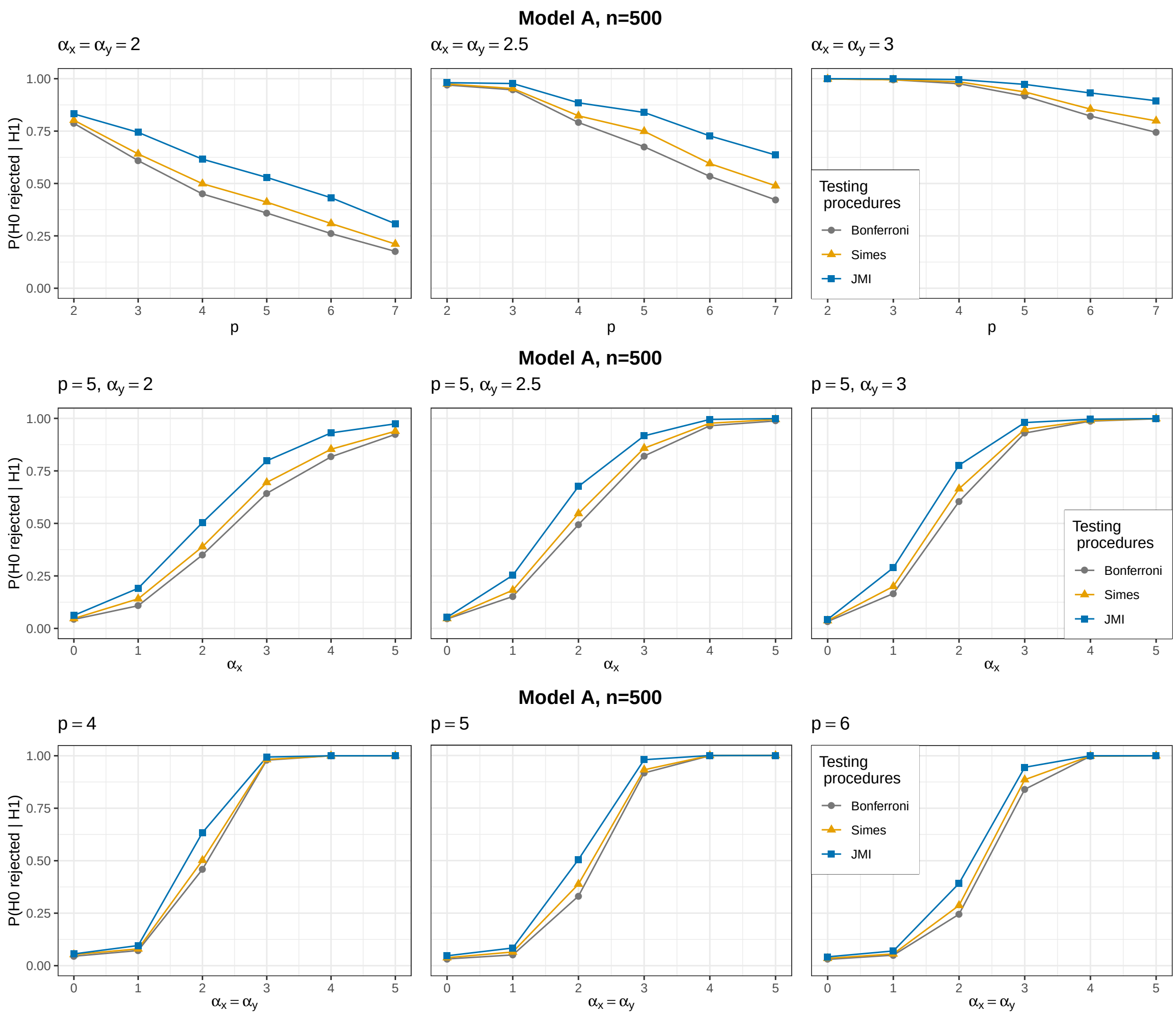

Fig. 5: Power against the changing number of variables and the parameter values for model A based on $N = 1000$ simulations.

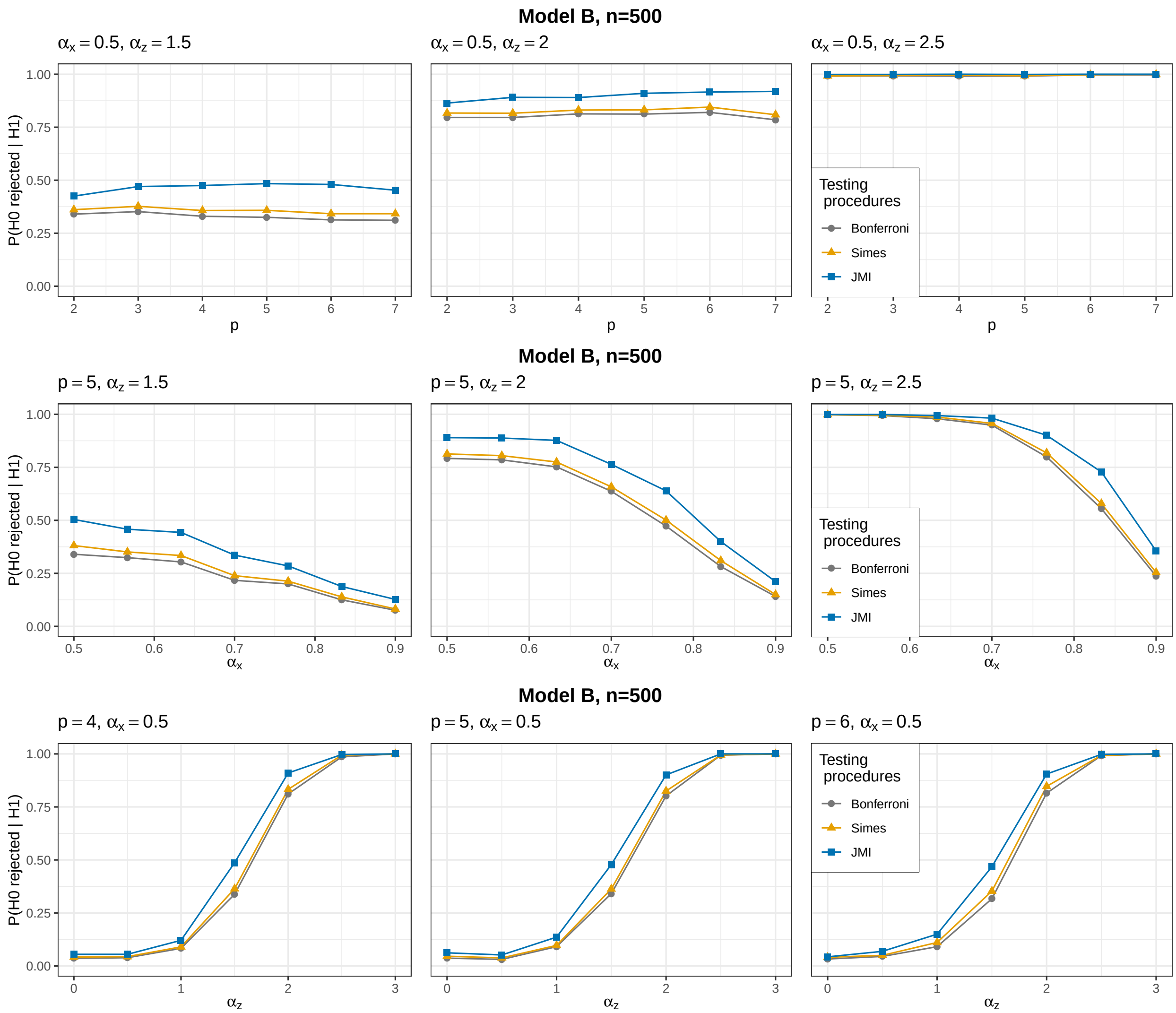

Fig. 6: Power against the changing number of variables and the parameter values for model B based on $N = 1000$ simulations.

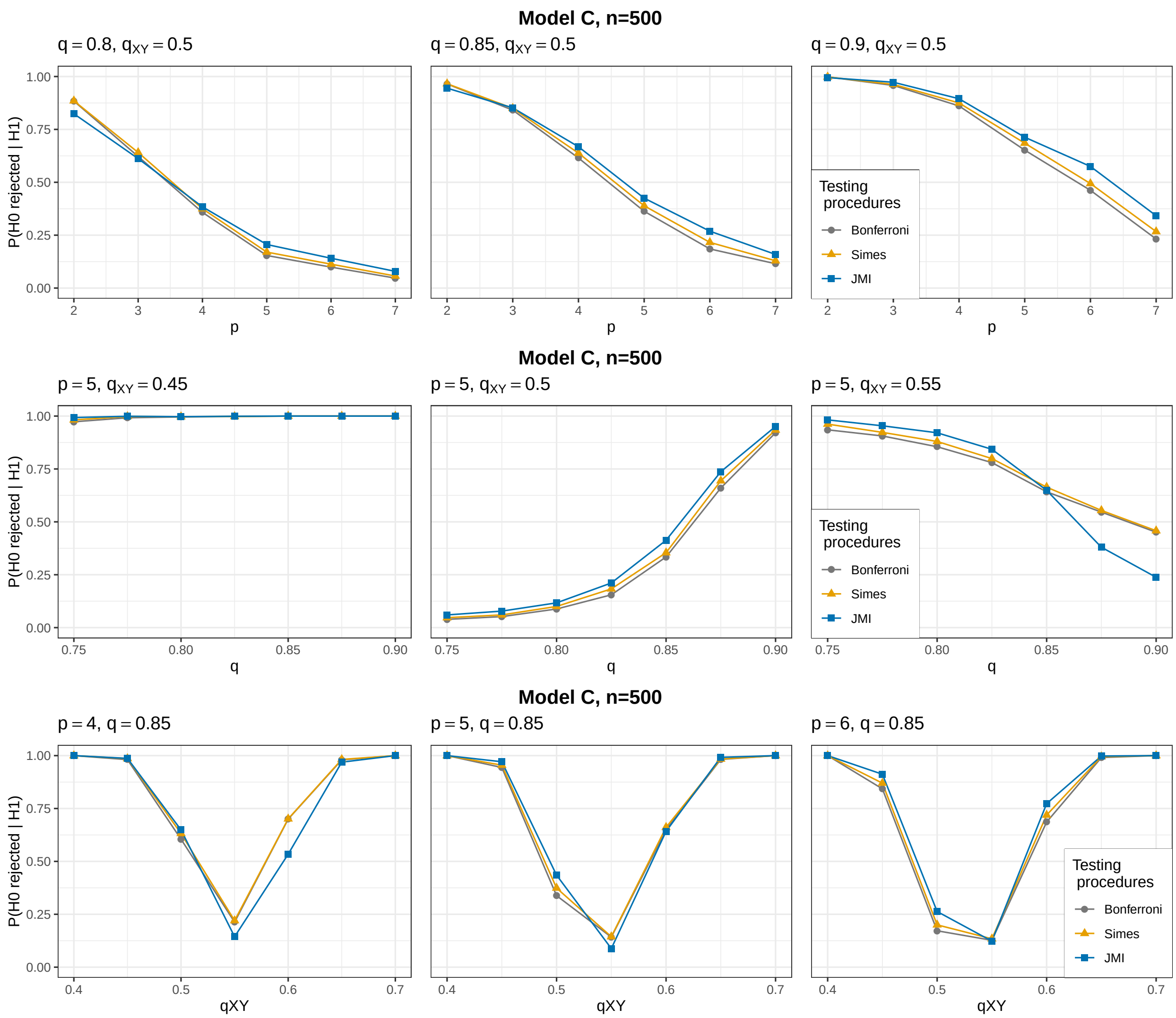

Fig. 7: Power against the changing number of variables and the parameter values for model C based on $N = 1000$ simulations.

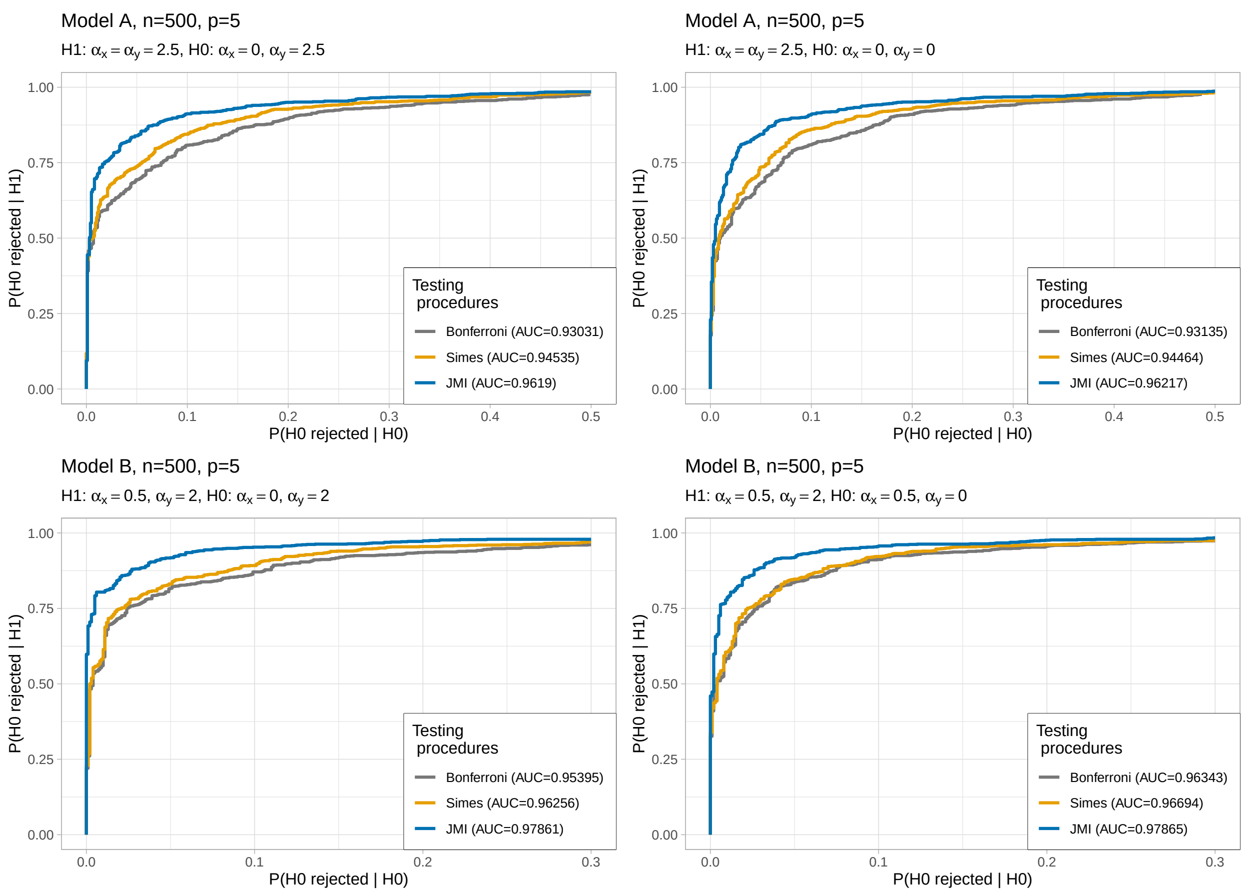

Fig. 8: ROC-type curves and the corresponding values of AUC for models A, B and C for H_0 and H_1 indicated in headers. In the first column null hypotheses indicates that X is independent of Y and all Z_i, in the second all variables are pairwise independent. The results are based on $N = 1000$ tests for the data generated from a model, in which H0 holds and $N = 1000$ tests for the data sampled from a model, in which H0 is violated. In both cases sample sizes equal $n = 500$. Plots in the first column are also shown in the main text.

References

1. Cover, T.M., Thomas, J.A.: Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing). Wiley-Interscience (2006)
2. Tsybakov, S.: Introduction to Nonparametric Estimation. Springer, New York (2009)