Biometrika (2016), **xx**, x, *pp*. 1–9

Printed in Great Britain

Supplementary Material for Indirect multivariate response linear regression

BY AARON J. MOLSTAD AND ADAM J. ROTHMAN

School of Statistics, University of Minnesota, Minneapolis, Minnesota 55455, U.S.A.

molst029@umn.edu arothman@umn.edu

1. PROOF OF PROPOSITION 2

In our proof of Proposition 2, we use the matrix inequality

$$\|A^{(1)}A^{(2)}A^{(3)} - B^{(1)}B^{(2)}B^{(3)}\| \leq \sum_{j=1}^{3} \|A^{(j)} - B^{(j)}\| \prod_{k\neq j} \|B^{(k)}\| + \sum_{j=1}^{3} \|B^{(j)}\| \prod_{k\neq j} \|A^{(k)} - B^{(k)}\| + \prod_{j=1}^{3} \|A^{(j)} - B^{(j)}\|. \quad (1)$$

Bickel & Levina (2008) used (1) to prove their Theorem 3.

Proof of Proposition 2. From (A2) in the proof of Proposition 1, $\Sigma_{*E}^{-1} = \Sigma_{*YY}^{-1} + \eta_* \Delta_*^{-1} \eta_*^{\mathrm{T}}$. Define $\hat{\Sigma}_E^{-1} = \hat{\Sigma}_{YY}^{-1} + \hat{\eta}\hat{\Delta}^{-1}\hat{\eta}^{\mathrm{T}}$. Applying (1),

$$\begin{aligned}
\|\hat{\beta} - \beta_*\| =& \|\hat{\Delta}^{-1}\hat{\eta}^{\mathrm{T}}\hat{\Sigma}_E - \Delta_*^{-1}\eta_*^{\mathrm{T}}\Sigma_{*E}\| \\
\leq& \|\hat{\Delta}^{-1} - \Delta_*^{-1}\|\|\eta_*\|\|\Sigma_{*E}\| + \|\hat{\eta} - \eta_*\|\|\Delta_*^{-1}\|\|\Sigma_{*E}\| + \|\hat{\Sigma}_E - \Sigma_{*E}\|\|\Delta_*^{-1}\|\|\eta_*\| \\
&+ \|\Delta_*^{-1}\|\|\hat{\eta} - \eta_*\|\|\hat{\Sigma}_E - \Sigma_{*E}\| + \|\eta_*\|\|\hat{\Delta}^{-1} - \Delta_*^{-1}\|\|\hat{\Sigma}_E - \Sigma_{*E}\| \\
&+ \|\Sigma_{*E}\|\|\hat{\Delta}^{-1} - \Delta_*^{-1}\|\|\hat{\eta} - \eta_*\| + \|\hat{\eta} - \eta_*\|\|\hat{\Delta}^{-1} - \Delta_*^{-1}\|\|\hat{\Sigma}_E - \Sigma_{*E}\|. \quad (2)
\end{aligned}$$

We will show that the third term in (2) dominates the others. We continue by deriving its bound. Employing a matrix identity used by Cai et al. (2010), we write $\hat{\Sigma}_E - \Sigma_{*E} = \Sigma_{*E}(\Sigma_{*E}^{-1} -$

$\hat{\Sigma}_E^{-1})\hat{\Sigma}_E$, so

$$\|\hat{\Sigma}_E - \Sigma_{*E}\| \leq \|\hat{\Sigma}_E\|\|\Sigma_{*E}\|\|\hat{\Sigma}_E^{-1} - \Sigma_{*E}^{-1}\|. \tag{3}$$

Using the triangle inequality and (1),

$$\begin{aligned}
\|\hat{\Sigma}_E^{-1} - \Sigma_{*E}^{-1}\| &\leq \|\hat{\Sigma}_{YY}^{-1} - \Sigma_{*YY}^{-1}\| + \|\hat{\eta}\hat{\Delta}^{-1}\hat{\eta}^{\mathrm{T}} - \eta_*\Delta_*^{-1}\eta_*^{\mathrm{T}}\| \\
&\leq \|\hat{\Sigma}_{YY}^{-1} - \Sigma_{*YY}^{-1}\| + 2\|\hat{\eta} - \eta_*\|\|\Delta_*^{-1}\|\|\eta_*\| + \|\hat{\Delta}^{-1} - \Delta_*^{-1}\|\|\eta_*\|^2 \\
&\quad + 2\|\eta_*\|\|\hat{\Delta}^{-1} - \Delta_*^{-1}\|\|\hat{\eta} - \eta_*\| + \|\Delta_*^{-1}\|\|\hat{\eta} - \eta_*\|^2 + \|\hat{\eta} - \eta_*\|^2\|\hat{\Delta}^{-1} - \Delta_*^{-1}\| \\
&= O_P\left(c_n + a_n\|\eta_*\|\|\Delta_*^{-1}\| + b_n\|\eta_*\|^2\right). \qquad (4)
\end{aligned}$$

Since $\varphi_{\min}(\Sigma_{*YY}^{-1}) \geq K$ and Δ_*^{-1} is positive definite, Weyl's eigenvalue inequality implies that $\varphi_{\min}(\Sigma_{*E}^{-1}) \geq K$ so

$$\|\Sigma_{*E}\| = \varphi_{\min}^{-1}(\Sigma_{*E}^{-1}) \leq 1/K. \tag{5}$$

Also,

$$\|\hat{\Sigma}_E\| = \varphi_{\min}^{-1}(\hat{\Sigma}_E^{-1}) = O_P(1) \tag{6}$$

because $\varphi_{\min}(\Sigma_{*E}^{-1}) \geq K$, $\hat{\Sigma}_E$ is positive definite, and $a_n\|\eta_*\|\|\Delta_*^{-1}\| + b_n\|\eta_*\|^2 + c_n = o(1)$ in (4). Using (4), (5), and (6), in (3),

$$\|\hat{\Sigma}_E - \Sigma_{*E}\| = O_P\left(a_n\|\eta_*\|\|\Delta_*^{-1}\| + b_n\|\eta_*\|^2 + c_n\right).$$

We then see that the third term in (2) dominates and

$$\begin{aligned}
\|\hat{\beta} - \beta_*\| &= O_P\left\{\left(a_n\|\eta_*\|\|\Delta_*^{-1}\| + b_n\|\eta_*\|^2 + c_n\right)\|\eta_*\|\|\Delta_*^{-1}\|\right\} \\
&= O_P\left(a_n\|\eta_*\|^2\|\Delta_*^{-1}\|^2 + b_n\|\eta_*\|^3\|\Delta_*^{-1}\| + c_n\|\eta_*\|\|\Delta_*^{-1}\|\right).
\end{aligned}$$

2. ADDITIONAL SIMULATION STUDIES

2·1. *Sparse inverse regression elliptical t-distribution simulations*

For 200 independent replications, we generated n independent copies of the random vector $(X^{\mathrm{T}}, Y^{\mathrm{T}})^{\mathrm{T}}$ which has the $p+q$-variate elliptical t-distribution with ν degrees of freedom and parameters $\mu_* \in \mathbb{R}^{p+q}$ and $\Sigma_* \in \mathbb{S}_+^{p+q}$. This parameterization was used by Muirhead (2005). We set $\mu_* = 0$ and we picked the entries in Σ_* by specifying Σ_{*XX}, Σ_{*XY}, and Σ_{*YY} defined through the partition of Σ_* used in Section 2.1. The (i,j)th element of Σ_{*YY} was $\rho_Y^{|i-j|}$ and the (i,j)th element of Δ_* was $0{\cdot}7^{|i-j|}$. We set $\Sigma_{*XY}^{\mathrm{T}} = \Sigma_{*YY}\eta_*$, and $\Sigma_{*XX} = \Delta_* + \eta_*\Sigma_{*XY}^{\mathrm{T}}$, where η_* was generated as it was in Section 5.1 with entry-wise nonzero probability s_*.

Results when $\nu = 3$ and $\nu = 10$ are displayed in Fig. 1 and Fig. 2, respectively. When $n = 100$, $p = 60$, and $q = 60$, the relative performance of I_1 was similar to the normal data generating model studied in Section 5.1 for both values of ν. When $n = 50$, $p = 200$, $q = 200$, I_1 performed worse than both ridge regression estimators except when $\nu = 10$ and $\rho_Y = 0{\cdot}9$.

2·2. *Additional sparse inverse regression simulations*

Figure 3 displays additional side-by-side boxplots of the observed model errors for the simulation study described in Section 5·1. We see that I_1 generally outperforms the competitors. However, when $n = 50$, $p = 200$, $q = 200$, and the responses were marginally uncorrelated, there was a small number of replications in which both ridge regression estimators performed better than I_1 did.

2·3. *Additional Non-normal forward regression simulations*

In Fig. 4, we display additional side-by-side boxplots of the observed model errors for the simulation study described in Section 5·2. When $n = 100$, $p = 60$, and $q = 60$, I_1 outperformed all non-oracle competitors. When $n = 50$, $p = 200$, $q = 200$, and the responses were marginally

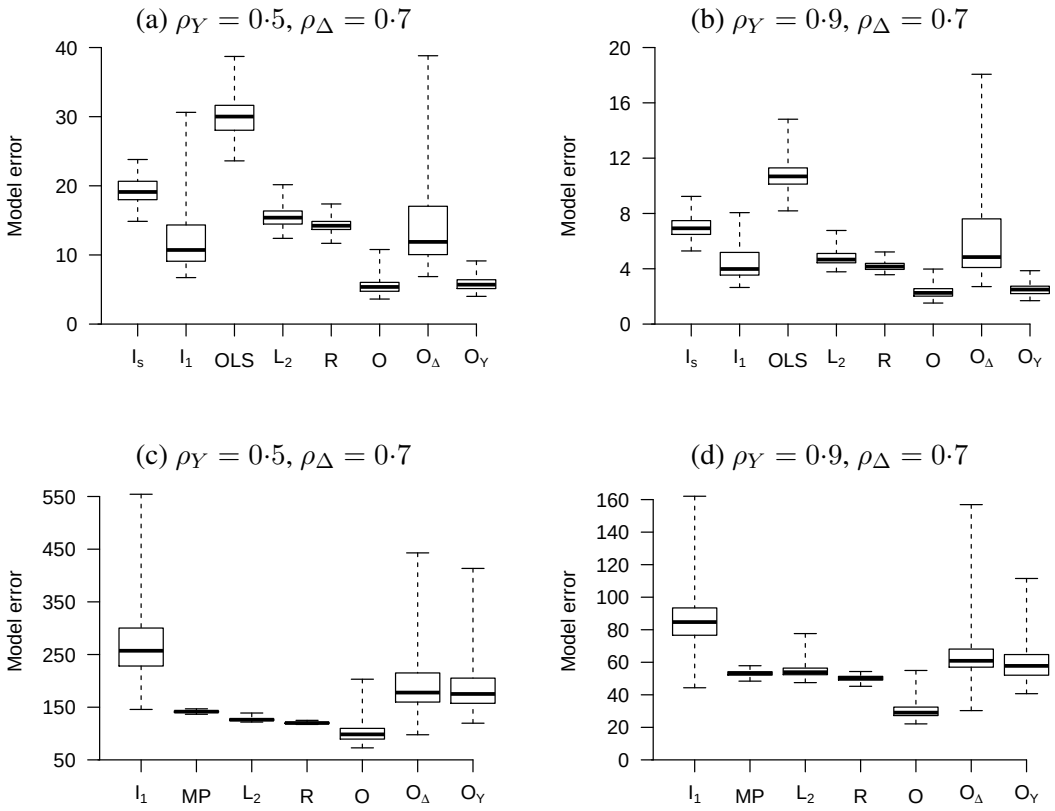

Fig. 1. Boxplots of the observed model errors from 200 replications when the data generating model from Section 2·1 is used. In (a) and (b), $n = 100, p = 60, q = 60$, $s_* = 0{\cdot}1$, and $\nu = 3$. In (c) and (d), $n = 50, p = 200$, $q = 200$, $s_* = 0{\cdot}03$, and $\nu = 3$.

uncorrelated, I_1 outperformed the non-oracle competitors except for a small number of replications.

2·4. *Additional reduced-rank inverse regression simulation*

In Fig. 5 (a)-(c), we show additional side-by-side boxplots of observed model errors for the simulation study described in Section 5·3. When responses are marginally uncorrelated, I_{RR} outperforms the direct likelihood-based reduced-rank regression estimator, both of which performed better than the part oracle estimators $O_{R\Delta}$ and O_{RY}.

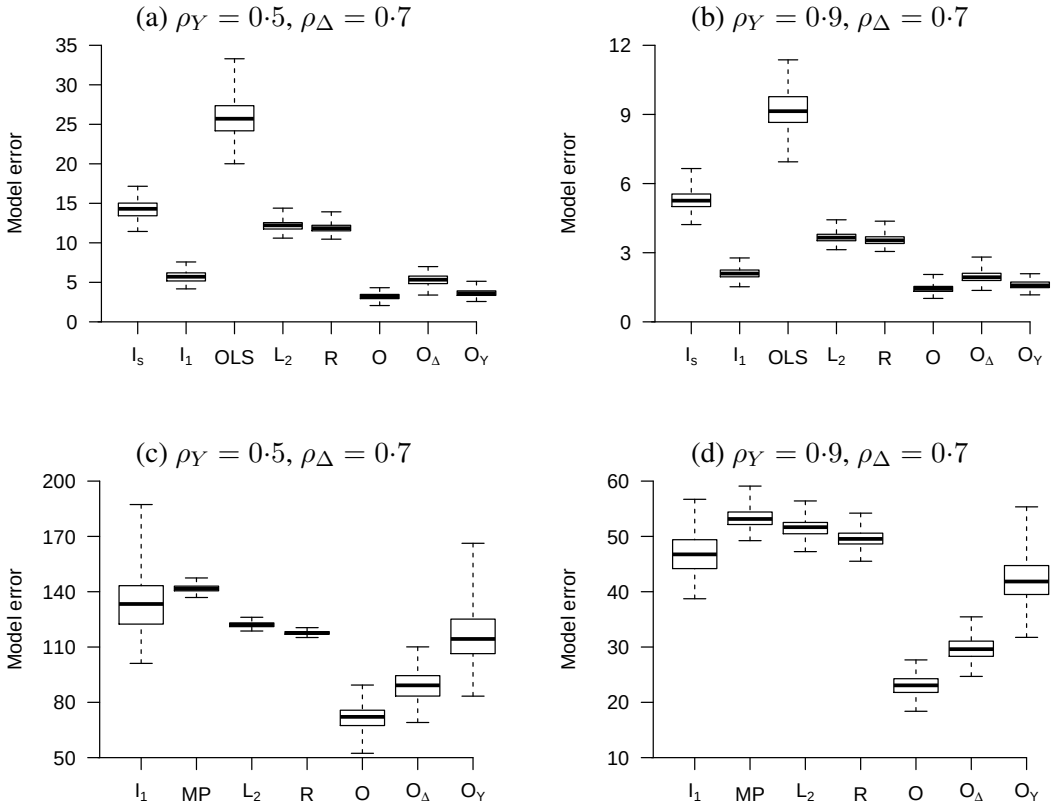

Fig. 2. Boxplots of the observed model errors from 200 replications where (a), (b) $n = 100, p = 60, q = 60, s_* = 0{\cdot}1$, $\nu = 10$; (c), (d) $n = 50, p = 200$, $q = 200$, $s_* = 0{\cdot}03$, $\nu = 10$; and the data generating model from Section 2·1 is used.

2·5. *Additional reduced-rank forward regression simulation*

In Fig. 5 (d)-(f), we display additional side-by-side boxplots of the observed model errors for the simulation study described in Section 5·4. In each setting, I_{RR} and I_{ML} performed similarly to RR. Both I_{RR} and I_{ML} outperformed the part oracle estimators as well. Results displayed in Fig. 5 are consistent with those from Section 5·4. This suggests that the indirect estimators I_{ML} and I_{RR} are competitive with the direct likelihood-based reduced-rank regression estimator even when the inverse regression error precision matrix is not sparse.

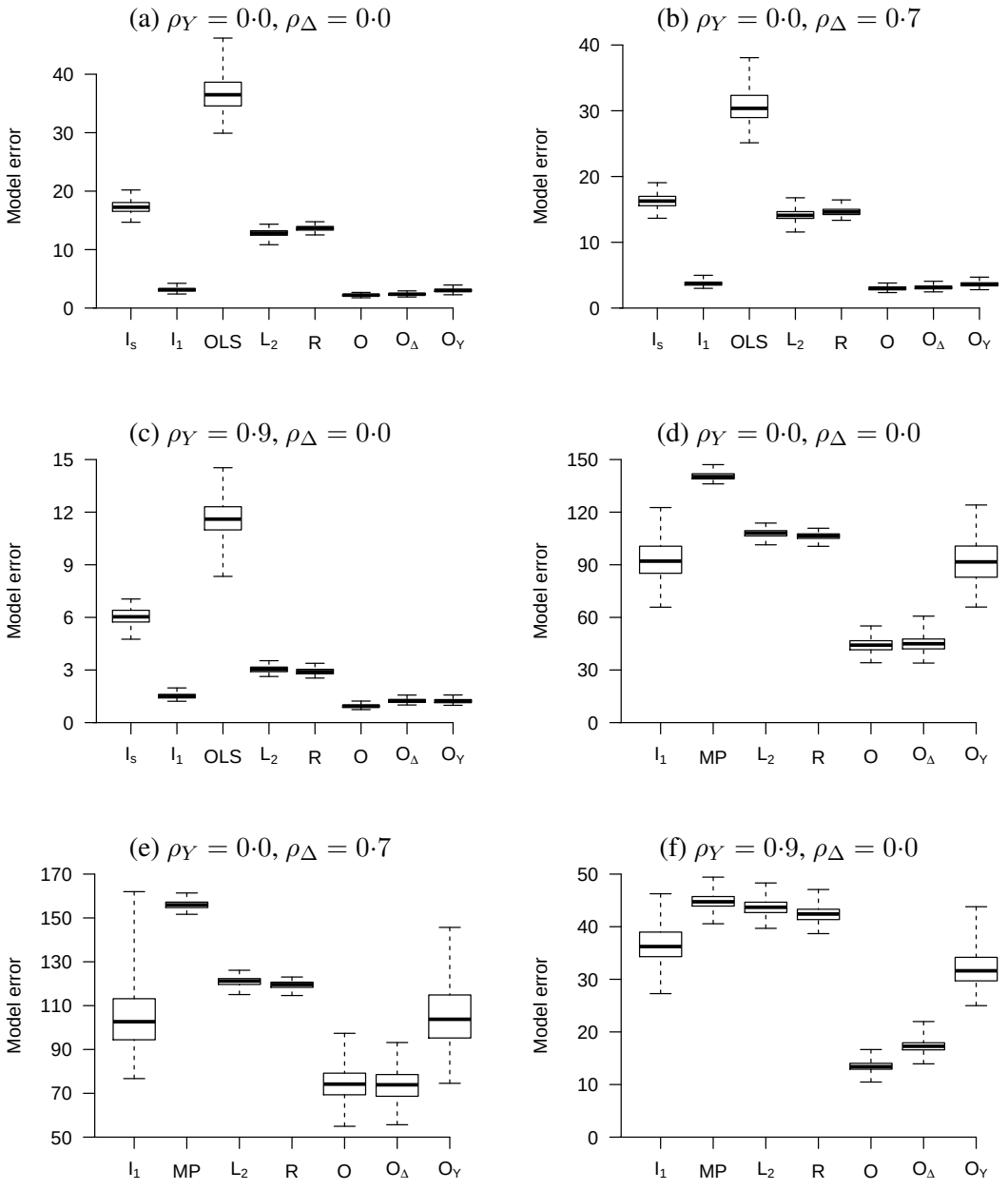

Fig. 3. Boxplots of the observed model errors from 200 replications where the data generating model from Section 5·1 was used. In (a)-(c), $n = 100, p = 60,\ q = 60$, and $s_* = 0{\cdot}1$. In (d)-(f), $n = 50, p = 200,\ q = 200$, and $s_* = 0{\cdot}03$.

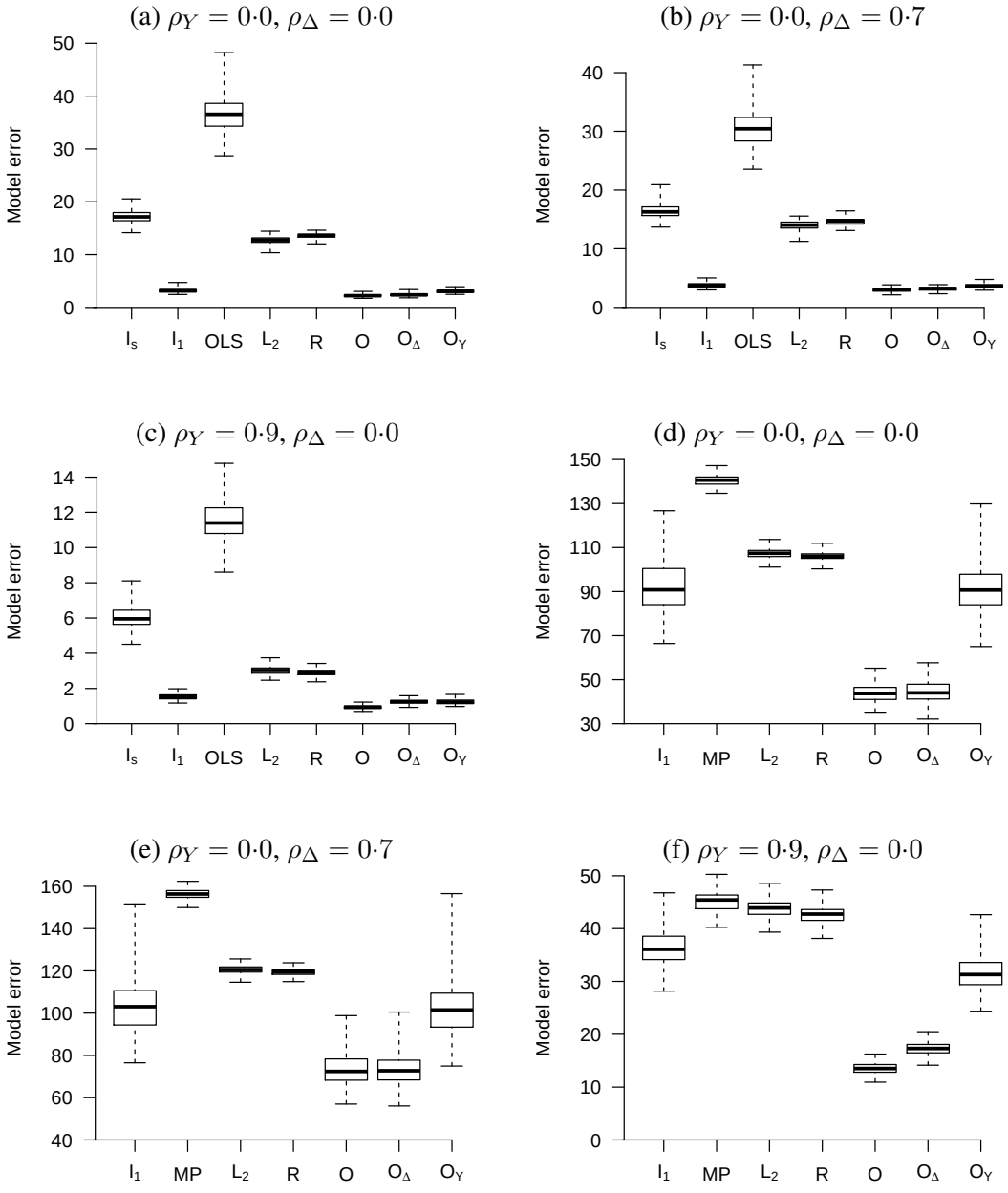

Fig. 4. Boxplots of the observed model errors from 200 replications when the data generating model from Section 5·2 is used. In (a)-(c), $n = 100, p = 60$, $q = 60$, and $s_* = 0{\cdot}1$. In (d)-(f), $n = 50, p = 200$, $q = 200$, and $s_* = 0{\cdot}03$.

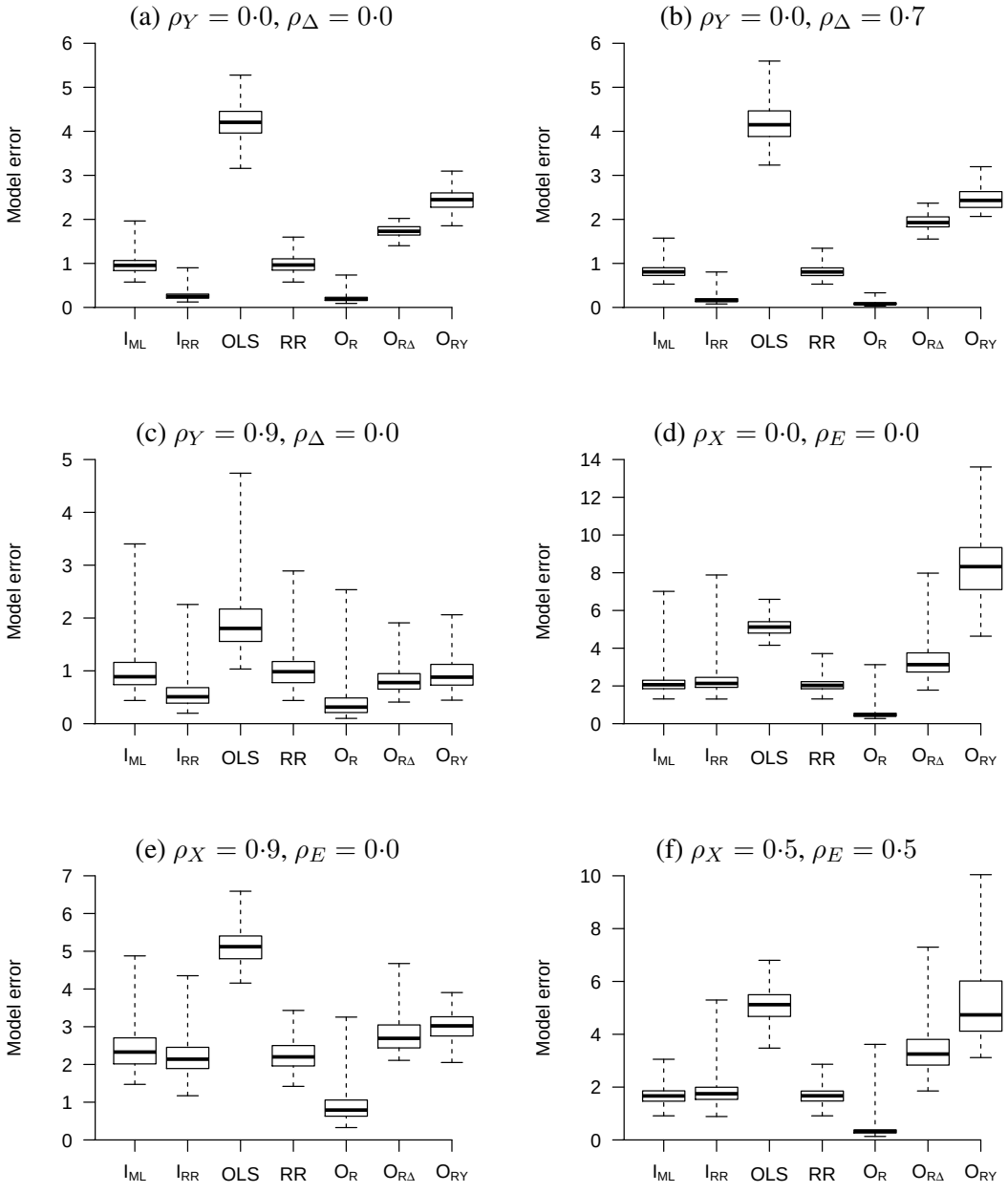

Fig. 5. Boxplots of the observed model errors from 200 replications when $n = 100, p = 20,\ q = 20$. In (a) – (c), the data generating model from Section 5·3 was used. In (d) – (f), the data generating model from Section 5·4 was used.

REFERENCES

BICKEL, P. J. & LEVINA, E. (2008). Regularized estimation of large covariance matrices. *Annals of Statistics* **36**, 199–227.

CAI, T. T., ZHANG, C.-H. & ZHOU, H. H. (2010). Optimal rates of convergence for covariance matrix estimation. *Annals of Statistics* **38**, 2118–2144.

MUIRHEAD, R. J. (2005). *Aspects of multivariate statistical theory*. Hoboken, New Jersey: John Wiley & Sons, Inc.