MODEL FREE MULTIVARIATE REDUCED-RANK REGRESSION WITH CATEGORICAL PREDICTORS

C. Messan Setodji and Lexin Li

RAND Corporation and North Carolina State University

Supplementary Material

This note contains proofs for Corollary 1 and Propositions 3 and 4.

S1. Proof of Proposition 3

With i.i.d. observations $\{(\boldsymbol{Y}_{i_w}, \boldsymbol{X}_{i_w}) : i = 1, \ldots, n_w\}$ and $n = \sum_{w=1}^{c} n_w$, the sample estimates of the quantities in $\boldsymbol{\Sigma}_\bullet^{1/2}\boldsymbol{\beta}^*\boldsymbol{\Omega}^{-1/2}$ can be expressed as follows.

$$
\begin{aligned}
\hat{\boldsymbol{\Sigma}}_w &= n_w^{-1}\sum_{i_w=1}^{n_w}(\boldsymbol{X}_{i_w} - \bar{\boldsymbol{X}}_w)(\boldsymbol{X}_{i_w} - \bar{\boldsymbol{X}}_w)^\intercal,\\
\hat{\boldsymbol{\sigma}}_{k_w} &= n_w^{-1}\sum_{i_w=1}^{n_w}(\boldsymbol{X}_{i_w} - \bar{\boldsymbol{X}}_w)(Y_{ik_w} - \bar{Y}_{\bullet k_w}),\\
\hat{\varepsilon}_{ik_w} &= (Y_{ik_w} - \bar{Y}_{\bullet k_w}) - \hat{\boldsymbol{\beta}}_{ik_w}^\intercal(\boldsymbol{X}_{i_w} - \bar{\boldsymbol{X}}_w),\\
\hat{\boldsymbol{\varepsilon}}_{iw} &= (\hat{\varepsilon}_{i1_w}, \hat{\varepsilon}_{i2_w}, \cdots, \hat{\varepsilon}_{ir_w})^\intercal,\\
\hat{\boldsymbol{\Sigma}}_\bullet &= \sum_{w=1}^{c}\hat{a}_w^2\hat{\boldsymbol{\Sigma}}_w = \frac{1}{n}\sum_{w=1}^{c}n_w\hat{\boldsymbol{\Sigma}}_w,\\
\hat{\boldsymbol{\Omega}}_w &= n_w^{-1}\sum_{i_w=1}^{n_w}\hat{\boldsymbol{\varepsilon}}_{iw}\hat{\boldsymbol{\varepsilon}}_{iw}^\intercal,
\end{aligned}
$$

where $\bar{\boldsymbol{X}}_w$ and $\bar{Y}_{\bullet k_w}$ are the sample average of $\boldsymbol{X}_{i_w}$ and Y_{ik_w}, $i = 1, 2, \cdots, n_w$, and $\hat{\boldsymbol{\beta}}_{k_w} = \hat{\boldsymbol{\Sigma}}_w^{-1}\hat{\boldsymbol{\sigma}}_{k_w}$, $\hat{\boldsymbol{\beta}}_w = (\hat{\boldsymbol{\beta}}_{1_w}, \ldots, \hat{\boldsymbol{\beta}}_{r_w})$, $\hat{a}_w = (n_w/n)^{1/2}$, and $\hat{\boldsymbol{\beta}}^* = (\hat{a}_1\hat{\boldsymbol{\beta}}_1, \ldots, \hat{a}_c\hat{\boldsymbol{\beta}}_c)$.

It follows from Eaton and Tyler (1994) that the asymptotic distribution of $\hat{\Lambda}_d$ is the same as that of $\Lambda_d = \mathrm{trace}(\boldsymbol{U}\boldsymbol{U}^\intercal) = \mathrm{vec}(\boldsymbol{U})^\intercal\mathrm{vec}(\boldsymbol{U})$, where

$$
\boldsymbol{U} = \sqrt{n}\boldsymbol{\Gamma}^\intercal(\hat{\boldsymbol{\Sigma}}_\bullet^{1/2}\hat{\boldsymbol{\beta}}^*\hat{\boldsymbol{\Omega}}^{-1/2} - \boldsymbol{\Sigma}_\bullet^{1/2}\boldsymbol{\beta}^*\boldsymbol{\Omega}^{-1/2})\boldsymbol{\Psi}.
$$

(Here $\boldsymbol{U}$ and $\boldsymbol{\Sigma}_{\bullet}^{1/2}\boldsymbol{\beta}^{*}\boldsymbol{\Omega}^{-1/2}$ correspond to Eaton and Tyler's $\boldsymbol{Z}_n$ and $\boldsymbol{B}$, respectively, in their equations (4.4) and (4.1).) Consequently, it is sufficient to prove that vec($\boldsymbol{U}$) is asymptotically normally distributed, with mean 0 and covariance matrix $\boldsymbol{\Delta}$. Note that

$$\boldsymbol{\Gamma}^{\mathsf{T}}\boldsymbol{\Sigma}_{\bullet}^{1/2}\boldsymbol{\beta}^{*}=0 \qquad \text{and} \qquad \boldsymbol{\beta}^{*}\boldsymbol{\Omega}^{-1/2}\boldsymbol{\Psi}=0. \tag{S1.1}$$

The matrix $\boldsymbol{U}$ can be expanded as

$$\begin{aligned}\boldsymbol{U}=\sqrt{n}\boldsymbol{\Gamma}^{\mathsf{T}}\{&\boldsymbol{\Sigma}_{\bullet}^{1/2}\boldsymbol{\beta}^{*}(\hat{\boldsymbol{\Omega}}^{-1/2}-\boldsymbol{\Omega}^{-1/2})+(\hat{\boldsymbol{\Sigma}}_{\bullet}^{1/2}-\boldsymbol{\Sigma}_{\bullet}^{1/2})\boldsymbol{\beta}^{*}\boldsymbol{\Omega}^{-1/2}\\&+\boldsymbol{\Sigma}_{\bullet}^{1/2}(\hat{\boldsymbol{\beta}}^{*}-\boldsymbol{\beta}^{*})\boldsymbol{\Omega}^{-1/2}\}\boldsymbol{\Psi}+O_p(n^{-1/2})\end{aligned}$$

By (S1.1) the first and second terms are 0, so we have

$$\boldsymbol{U}=\sqrt{n}\boldsymbol{\Gamma}^{\mathsf{T}}\boldsymbol{\Sigma}_{\bullet}^{1/2}(\hat{\boldsymbol{\beta}}^{*}-\boldsymbol{\beta}^{*})\boldsymbol{\Omega}^{-1/2}\boldsymbol{\Psi}+O_p(n^{-1/2}) \tag{S1.2}$$

and the limiting distribution of $\boldsymbol{U}$ is the same as that of $\boldsymbol{U}_0=\sqrt{n}\boldsymbol{\Gamma}^{\mathsf{T}}\boldsymbol{\Sigma}_{\bullet}^{1/2}(\hat{\boldsymbol{\beta}}^{*}-\boldsymbol{\beta}^{*})\boldsymbol{\Omega}^{-1/2}\boldsymbol{\Psi}$. Li, Cook and Chiaromonte(2003) and Cook and Setodji(2003) proved that

$$\hat{a}_w\sqrt{n}(\hat{\boldsymbol{\beta}}_{j_w}-\boldsymbol{\beta}_{j_w})=\sqrt{n_w}(\hat{\boldsymbol{\beta}}_{j_w}-\boldsymbol{\beta}_{j_w})=n_w^{-1/2}\boldsymbol{\Sigma}_w^{-1/2}\sum_{i=1}^{n_w}\boldsymbol{Z}_{i_w}\varepsilon_{ij_w}+O_p(n_w^{-1/2}).$$

Consequently,

$$\begin{aligned}\sqrt{n_w}(\hat{\boldsymbol{\beta}}_{j_w}-\boldsymbol{\beta}_{j_w}) &= n_w^{-1/2}\boldsymbol{\Sigma}_w^{-1/2}\sum_{i=1}^{n_w}\boldsymbol{Z}_{i_w}\varepsilon_{ij_w}+O_p(n_w^{-1/2}),\\ \sqrt{n_w}(\hat{\boldsymbol{\beta}}_{w}-\boldsymbol{\beta}_{w}) &= n_w^{-1/2}\boldsymbol{\Sigma}_w^{-1/2}\sum_{i=1}^{n_w}\left(\boldsymbol{Z}_{i_w}\varepsilon_{i1_w},\boldsymbol{Z}_{i_w}\varepsilon_{2j_w},\ldots,\boldsymbol{Z}_{i_w}\varepsilon_{ir_w}\right)+O_p(n_w^{-1/2}),\\ \sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) &= \Big(\sqrt{n_1}(\hat{\boldsymbol{\beta}}_1-\boldsymbol{\beta}_1),\sqrt{n_2}(\hat{\boldsymbol{\beta}}_2-\boldsymbol{\beta}_2),\ldots,\sqrt{n_c}(\hat{\boldsymbol{\beta}}_c-\boldsymbol{\beta}_c)\Big).\end{aligned}$$

Defining $\boldsymbol{R}_w=n_w^{-1/2}\boldsymbol{\Sigma}_w^{-1/2}\sum_{i=1}^{n_w}\left(\boldsymbol{Z}_{i_w}\varepsilon_{i1_w},\boldsymbol{Z}_{i_w}\varepsilon_{2j_w},\ldots,\boldsymbol{Z}_{i_w}\varepsilon_{ir_w}\right)$ and $\boldsymbol{R}=(\boldsymbol{R}_1,\boldsymbol{R}_2,\ldots,\boldsymbol{R}_c)$, so that $\boldsymbol{U}_0=\boldsymbol{\Gamma}^{\mathsf{T}}\boldsymbol{\Sigma}_{\bullet}^{1/2}\boldsymbol{R}\boldsymbol{\Omega}^{-1/2}\boldsymbol{\Psi}$, and

$$\text{vec}(\boldsymbol{U}_0)=[(\boldsymbol{\Psi}^{\mathsf{T}}\boldsymbol{\Omega}^{-1/2})\otimes(\boldsymbol{\Gamma}^{\mathsf{T}}\boldsymbol{\Sigma}_{\bullet}^{1/2})]\text{vec}(\boldsymbol{R}),$$

By the central limit theorem, we then have

$$\text{vec}(\boldsymbol{U}_0)\xrightarrow{\mathcal{L}}\mathcal{N}(0,\boldsymbol{\Delta}),$$

where

$$\mathbf{\Delta} = \sum_{w=1}^{c} [(\Psi_w^\intercal \Omega_w^{-1/2}) \otimes (\mathbf{\Gamma}^\intercal \mathbf{\Sigma}_\bullet^{1/2} \mathbf{\Sigma}_w^{-1/2})](\mathrm{Var}(\boldsymbol{T}_w))[(\Omega_w^{-1/2}\Psi_w) \otimes (\mathbf{\Sigma}_w^{-1/2}\mathbf{\Sigma}_\bullet^{1/2}\mathbf{\Gamma})].$$

Note that $\mathrm{Var}(\boldsymbol{T}_w) = \mathrm{E}(\boldsymbol{T}_w\boldsymbol{T}_w^\intercal) - \mathrm{E}(\boldsymbol{T}_w)\mathrm{E}(\boldsymbol{T}_w)^\intercal$. Because $\boldsymbol{\varepsilon}_w$ is an OLS residual of the regression within the subgroup w, we have $\mathrm{Cov}(\boldsymbol{\varepsilon}_w, \boldsymbol{Z}_w) = 0$ and so $\mathrm{E}(\boldsymbol{T}_w) = 0$. From this we can write $\mathrm{Var}(\boldsymbol{T}_w) = \mathrm{E}(\boldsymbol{T}_w\boldsymbol{T}_w^\intercal)$ and

$$\mathbf{\Delta} = \sum_{w=1}^{c} [(\Psi_w^\intercal \Omega_w^{-1/2}) \otimes (\mathbf{\Gamma}^\intercal \mathbf{\Sigma}_\bullet^{1/2} \mathbf{\Sigma}_w^{-1/2})](\mathrm{E}(\boldsymbol{T}_w\boldsymbol{T}_w^\intercal))[(\Omega_w^{-1/2}\Psi_w) \otimes (\mathbf{\Sigma}_w^{-1/2}\mathbf{\Sigma}_\bullet^{1/2}\mathbf{\Gamma})].$$

The conclusion then follows. □

S2. Proof of Corollary 1

With either assumptions (a) or (b), we have $\mathrm{Var}(\boldsymbol{T}_w) = \Omega_w \otimes \mathbf{I}_p$, and thus,

$$\begin{aligned}
\mathbf{\Delta} &= \sum_{w=1}^{c} [(\Psi_w^\intercal \Omega_w^{-1/2}) \otimes (\mathbf{\Gamma}^\intercal \mathbf{\Sigma}_\bullet^{1/2} \mathbf{\Sigma}_w^{-1/2})](\mathrm{Var}(\boldsymbol{T}_w))[(\Omega_w^{-1/2}\Psi_w) \otimes (\mathbf{\Sigma}_w^{-1/2}\mathbf{\Sigma}_\bullet^{1/2}\mathbf{\Gamma})] \\
&= \sum_{w=1}^{c} [(\Psi_w^\intercal \Omega_w^{-1/2}) \otimes (\mathbf{\Gamma}^\intercal \mathbf{\Sigma}_\bullet^{1/2} \mathbf{\Sigma}_w^{-1/2})](\Omega_w \otimes \mathbf{I}_p)[(\Omega_w^{-1/2}\Psi_w) \otimes (\mathbf{\Sigma}_w^{-1/2}\mathbf{\Sigma}_\bullet^{1/2}\mathbf{\Gamma})] \\
&= \sum_{w=1}^{c} [(\Psi_w^\intercal \Omega_w^{-1/2}\Omega_w) \otimes (\mathbf{\Gamma}^\intercal \mathbf{\Sigma}_\bullet^{1/2} \mathbf{\Sigma}_w^{-1/2}\mathbf{I}_p)][(\Omega_w^{-1/2}\Psi_w) \otimes (\mathbf{\Sigma}_w^{-1/2}\mathbf{\Sigma}_\bullet^{1/2}\mathbf{\Gamma})] \\
&= \sum_{w=1}^{c} [(\Psi_w^\intercal \Omega_w^{1/2}) \otimes (\mathbf{\Gamma}^\intercal \mathbf{\Sigma}_\bullet^{1/2} \mathbf{\Sigma}_w^{-1/2})][(\Omega_w^{-1/2}\Psi_w) \otimes (\mathbf{\Sigma}_w^{-1/2}\mathbf{\Sigma}_\bullet^{1/2}\mathbf{\Gamma})] \\
&= \sum_{w=1}^{c} [(\Psi_w^\intercal \Omega_w^{1/2})(\Omega_w^{-1/2}\Psi_w)] \otimes [(\mathbf{\Gamma}^\intercal \mathbf{\Sigma}_\bullet^{1/2} \mathbf{\Sigma}_w^{-1/2})(\mathbf{\Sigma}_w^{-1/2}\mathbf{\Sigma}_\bullet^{1/2}\mathbf{\Gamma})] \\
&= \sum_{w=1}^{c} [\Psi_w^\intercal \Psi_w] \otimes [\mathbf{\Gamma}^\intercal \mathbf{\Sigma}_\bullet^{1/2} \mathbf{\Sigma}_w^{-1}\mathbf{\Sigma}_\bullet^{1/2}\mathbf{\Gamma}].
\end{aligned}$$

If in addition $\mathbf{\Sigma}_\bullet = \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \ldots = \mathbf{\Sigma}_c$, then we have $\mathbf{\Delta} = \mathbf{I}_{(p-d)(rc-d)}$, and consequently $\hat{\Lambda}_d$ converges to a chi-squared distribution with $(p-d)(rc-d)$ degrees of freedom. □

S3. Proof of Propostion 4

The joint asymptotic distribution of the $p-d$ smallest singular values of $\sqrt{n}\boldsymbol{M}$ is the same as the distribution of the singular values of the matrix $\boldsymbol{V} = \sqrt{n}\tilde{\boldsymbol{\Gamma}}^{\top}(\hat{\boldsymbol{M}} - \boldsymbol{M})\tilde{\boldsymbol{\Psi}}$ (Eaton and Tyler, 1994). This implies that the asymptotic distribution of $\tilde{\Lambda}_d$ is the same as that of $\mathrm{vec}(\boldsymbol{V})^{\top}\mathrm{vec}(\boldsymbol{V})$, i.e., the sum of the squared elements of $\boldsymbol{V}$. Consequently, it is sufficient to show that $\mathrm{vec}(\boldsymbol{V})$ is asymptotically normally distributed with mean 0 and covariance matrix $\tilde{\boldsymbol{\Delta}}$.

Let $\boldsymbol{N}_w = (\boldsymbol{M}_{1_w}, \ldots, \boldsymbol{M}_{r_w})$. Since $\tilde{\boldsymbol{\Gamma}}^{\top}\boldsymbol{M}_{k_w} = 0$ for any k and w,

$$\boldsymbol{V} = \sqrt{n}\tilde{\boldsymbol{\Gamma}}^{\top}(\hat{\boldsymbol{M}} - \boldsymbol{M})\tilde{\boldsymbol{\Psi}} = \sum_{w=1}^{c}\sqrt{n_w}\tilde{\boldsymbol{\Gamma}}^{\top}(\hat{\boldsymbol{N}}_w - \boldsymbol{N}_w)\tilde{\boldsymbol{\Psi}}_w \equiv \sum_{w=1}^{c}\sqrt{n_w}\boldsymbol{V}_w$$

where $\hat{\boldsymbol{N}}_w$ is the sample estimate of $\boldsymbol{N}_w$. This implies $\mathrm{vec}(\boldsymbol{V}) = \sum_{w=1}^{c}\mathrm{vec}(\boldsymbol{V}_w)$. Since the $\boldsymbol{V}_w$'s are mutually independent, we can determine the limiting distribution of just one $\mathrm{vec}(\boldsymbol{V}_w)$, and then add them to obtain our desired result.

Note that $\mathrm{vec}(\boldsymbol{V}_w)$ can be rewritten as $\mathrm{vec}(\boldsymbol{V}_w) = (\tilde{\boldsymbol{\Psi}}_w^{\top} \otimes \tilde{\boldsymbol{\Gamma}}^{\top})\mathrm{vec}(\hat{\boldsymbol{N}} - \boldsymbol{N})$. Cook and Li (2004) showed that, with $\xi_{k_w}^{(i)}$ and $\xi_w^{(i)}$ denoting the ith observation of the random variable $\xi_{k_w} \in \mathbb{R}^{p^2}$ and $\xi_w \in \mathbb{R}^{rp^2}$ respectively,

$$\mathrm{vec}(\hat{\boldsymbol{M}}_{k_w} - \boldsymbol{M}_{k_w}) = \boldsymbol{G}_{k_w} \times \left(\frac{1}{n_w}\sum_{i=1}^{n_w}\xi_{k_w}^{(i)}\right) + O_p(n_w^{-1}).$$

In the multivariate setting, this leads to

$$\mathrm{vec}(\hat{\boldsymbol{N}}_w - \boldsymbol{N}_w) = \boldsymbol{G}_w \times \left(\frac{1}{n_w}\sum_{i=1}^{n_w}\xi_w^{(i)}\right) + O_p(n_w^{-1}).$$

Since each $\boldsymbol{G}_w$ is a $rp^2 \times rp^2$ constant matrix, $\sqrt{n}\mathrm{vec}(\hat{\boldsymbol{N}}_w - \boldsymbol{N}_w)$ converges in distribution to a rp^2-dimensional multivariate normal with mean 0 and covariance matrix $\boldsymbol{G}_w\mathrm{E}(\xi_w\xi_w^{\top})\boldsymbol{G}_w^{\top}$. Consequently, $\sqrt{n}\tilde{\boldsymbol{\Gamma}}^{\top}(\hat{\boldsymbol{M}} - \boldsymbol{M})\tilde{\boldsymbol{\Psi}}$ converges to a multivariate normal distribution with mean 0 and covariance matrix $\tilde{\boldsymbol{\Delta}} = \sum_{w=1}^{c}(\tilde{\boldsymbol{\Psi}}_w \otimes \tilde{\boldsymbol{\Gamma}})^{\top}\boldsymbol{G}_w\mathrm{E}(\xi_w\xi_w^{\top})\boldsymbol{G}_w^{\top}(\tilde{\boldsymbol{\Psi}}_w \otimes \tilde{\boldsymbol{\Gamma}})$. This completes the proof. □

Additional References

Eaton, M.L. and Tyler, D. (1994). The asymptotic distribution of singular values with application to canonical correlations and correspondence analysis., *Journal of Multivariate Analysis*, **50**, 238–264.