HYPOTHESIS TESTING IN THE PRESENCE OF MULTIPLE SAMPLES UNDER DENSITY RATIO MODELS

Song Cai[1], Jiahua Chen[2] and James V. Zidek[2]

[1]Carleton University and [2]University of British Columbia

Supplementary Material

This supplementary material presents the detailed proofs of the theorems in the paper "Hypothesis testing in the presence of multiple samples under density ratio models".

S1 Some notations and facts

We first introduce more notations applicable to $k = 0, \ldots, m$. Recall that $\varphi_k(\boldsymbol{\theta}, x) = \exp\{\alpha_k + \boldsymbol{\beta}_k^{\intercal}\boldsymbol{q}(x)\}$. We write

$$\mathcal{L}_{n,k}(\boldsymbol{\theta}, x) = -\log\Big\{\sum_{r=0}^{m}\hat{\lambda}_r\varphi_r(\boldsymbol{\theta}, x)\Big\} + \Big\{\alpha_k + \boldsymbol{\beta}_k^{\intercal}\boldsymbol{q}(x)\Big\}$$

with $\hat{\lambda}_r = n_r/n$ being the sample proportion. Hence, the DEL $l_n(\boldsymbol{\theta}) = \sum_{k,j}\mathcal{L}_{n,k}(\boldsymbol{\theta}, x_{kj})$ where the summation is over all possible (k, j). Let $\mathcal{L}_k(\boldsymbol{\theta}, x)$ be the "population" version of $\mathcal{L}_{n,k}(\boldsymbol{\theta}, x)$ by replacing $\hat{\lambda}_r$ with its limit ρ_r in the above definition. Let $\boldsymbol{e}_k$ be a vector of length m with the k^{th} entry being 1 and the others being 0s, and let $\delta_{ij} = 1$ when $i = j$, and 0 otherwise.

Recall the definitions (3.6) of $\boldsymbol{h}(\boldsymbol{\theta},\, x)$, $s(\boldsymbol{\theta},\, x)$ and $H(\boldsymbol{\theta},\, x)$. The first order derivatives of $\mathcal{L}_k(\boldsymbol{\theta},\, x)$ can be written as

$$\begin{aligned}\partial \mathcal{L}_k(\boldsymbol{\theta},\, x)/\partial \boldsymbol{\alpha} &= (1-\delta_{k0})\boldsymbol{e}_k - \boldsymbol{h}(\boldsymbol{\theta},\, x)/s(\boldsymbol{\theta},\, x),\\ \partial \mathcal{L}_k(\boldsymbol{\theta},\, x)/\partial \boldsymbol{\beta} &= \{\partial \mathcal{L}_k(\boldsymbol{\theta},\, x)/\partial \boldsymbol{\alpha}\} \otimes \boldsymbol{q}(x).\end{aligned} \tag{S1.1}$$

Similarly, we have

$$\begin{aligned}\partial^2 \mathcal{L}_k(\boldsymbol{\theta},\, x)/\partial \boldsymbol{\alpha}\partial \boldsymbol{\alpha}^{\intercal} &= -H(\boldsymbol{\theta},\, x)/s(\boldsymbol{\theta},\, x),\\ \partial^2 \mathcal{L}_k(\boldsymbol{\theta},\, x)/\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^{\intercal} &= -\Big\{H(\boldsymbol{\theta},\, x)/s(\boldsymbol{\theta},\, x)\Big\} \otimes \Big\{\boldsymbol{q}(x)\boldsymbol{q}^{\intercal}(x)\Big\},\\ \partial^2 \mathcal{L}_k(\boldsymbol{\theta},\, x)/\partial \boldsymbol{\alpha}\partial \boldsymbol{\beta}^{\intercal} &= -\Big\{H(\boldsymbol{\theta},\, x)/s(\boldsymbol{\theta},\, x)\Big\} \otimes \boldsymbol{q}^{\intercal}(x).\end{aligned} \tag{S1.2}$$

The algebraic expressions of the derivatives of $\mathcal{L}_{n,k}(\boldsymbol{\theta},\, x)$ are similar to those of $\mathcal{L}_k(\boldsymbol{\theta},\, x)$, only with ρ_r replaced by the sample proportion $\hat{\lambda}_r$. Note that all entries of $\boldsymbol{h}(\boldsymbol{\theta},\, x)$ are non-negative, and $s(\boldsymbol{\theta},\, x)$ exceeds the sum of all entries of $\boldsymbol{h}(\boldsymbol{\theta},\, x)$. Thus, $\|\boldsymbol{h}(\boldsymbol{\theta},\, x)/s(\boldsymbol{\theta},\, x)\| \le 1$ in terms of Euclidean norm, and the absolute value of each entry of $H(\boldsymbol{\theta},\, x)/s(\boldsymbol{\theta},\, x)$ is bounded by 1. By examining the algebraic expressions closely, this result implies

$$\begin{aligned}\left|\frac{\partial^2 \mathcal{L}_{n,k}(\boldsymbol{\theta},\, x)}{\partial \theta_i \partial \theta_j}\right| &\le 1 + \boldsymbol{q}^{\intercal}(x)\boldsymbol{q}(x)\\ \left|\frac{\partial^3 \mathcal{L}_{n,k}(\boldsymbol{\theta},\, x)}{\partial \theta_i \partial \theta_j \partial \theta_k}\right| &\le \{1 + \boldsymbol{q}^{\intercal}(x)\boldsymbol{q}(x)\}^{3/2},\end{aligned} \tag{S1.3}$$

where θ_i denotes the i^{th} entry of $\boldsymbol{\theta}$.

We also observed the following important relationships between the first and second order derivatives of $\mathcal{L}_k(\boldsymbol{\theta}, x)$:

$$\mathrm{E}_0\left\{\frac{\partial \mathcal{L}_0(\boldsymbol{\theta}^*, x)}{\partial \boldsymbol{\alpha}}\right\} = -\rho_0^{-1} U_{\boldsymbol{\alpha}\boldsymbol{\alpha}} \mathbf{1}_m, \quad \mathrm{E}_0\left\{\frac{\partial \mathcal{L}_0(\boldsymbol{\theta}^*, x)}{\partial \boldsymbol{\beta}}\right\} = -\rho_0^{-1} U_{\boldsymbol{\beta}\boldsymbol{\alpha}} \mathbf{1}_m, \tag{S1.4}$$

and, for $k = 1, \ldots, m$,

$$\begin{aligned}
\mathrm{E}_k\left\{\frac{\partial \mathcal{L}_k(\boldsymbol{\theta}^*, x)}{\partial \boldsymbol{\alpha}}\right\} &= \rho_k^{-1} U_{\boldsymbol{\alpha}\boldsymbol{\alpha}} \boldsymbol{e}_k, \quad E_k\left\{\frac{\partial \mathcal{L}_k(\boldsymbol{\theta}^*, x)}{\partial \boldsymbol{\alpha}} \boldsymbol{q}^\intercal(x)\right\} = \rho_k^{-1} U_{\boldsymbol{\alpha}\boldsymbol{\beta}}(\boldsymbol{e}_k \otimes I_d), \\
\mathrm{E}_k\left\{\frac{\partial \mathcal{L}_k(\boldsymbol{\theta}^*, x)}{\partial \boldsymbol{\beta}}\right\} &= \rho_k^{-1} U_{\boldsymbol{\beta}\boldsymbol{\alpha}} \boldsymbol{e}_k, \quad E_k\left\{\frac{\partial \mathcal{L}_k(\boldsymbol{\theta}^*, x)}{\partial \boldsymbol{\beta}} \boldsymbol{q}^\intercal(x)\right\} = \rho_k^{-1} U_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{e}_k \otimes I_d).
\end{aligned} \tag{S1.5}$$

The assumption that $\int \exp\{\boldsymbol{\beta}_k^\intercal \boldsymbol{q}(x)\} dF_0 < \infty$ for $\boldsymbol{\theta}$ in a neighbourhood of $\boldsymbol{\theta}^*$ implies that the moment generating function of $\boldsymbol{q}(x)$ with respect to each F_k, exists in a neighbourhood of $\mathbf{0}$. Hence, all finite order moments of $\boldsymbol{q}(x)$ with respect to each F_k are finite. This fact and inequalities (S1.3) reveal that the second and third order derivatives of $l_n(\boldsymbol{\theta})$ are bounded by an integrable function.

Under the assumption of Theorem 1 that $\int \boldsymbol{Q}(x) \boldsymbol{Q}^\intercal(x) dF_0$ is positive definite, the information matrix U given by (3.7) is positive definite. As a reminder, $\boldsymbol{Q}(x) = (1, \boldsymbol{q}^\intercal(x))^\intercal$.

S2 Proof of Theorem 1

Under the null hypothesis (3.4), we show that the DELR statistic is approximated by a quadratic form that has a chi-square limiting distribution.

We first give two key lemmas.

Let $T = \rho_0^{-1}\mathbf{1}_m\mathbf{1}_m^\intercal + \mathrm{diag}\{\rho_1^{-1}, \ldots, \rho_m^{-1}\}$ and $W = \mathrm{diag}\{T,\, 0_{md\times md}\}$. Put $\boldsymbol{v} = n^{-1/2}\partial l_n(\boldsymbol{\theta}^*)/\partial\boldsymbol{\theta}$. Let $\mathrm{E}(\cdot)$ be the usual expectation operator and $\mathrm{E}_k(\cdot)$ be the expectation operator respect F_k.

Lemma 1 (Asymptotic properties of the score function)**.** *Under the conditions of Theorem 1, $E\boldsymbol{v} = \mathbf{0}$ and $\boldsymbol{v}$ is asymptotically multivariate normal with mean $\mathbf{0}$ and covariance matrix $V = U - UWU$.*

Proof of Lemma 1. Denote $\boldsymbol{\mu}_k = \mathrm{E}_k\{\partial\mathcal{L}_{n,k}(\boldsymbol{\theta}^*,\, x)/\partial\boldsymbol{\theta}\}$. We can verify that

$$\mathrm{E}\boldsymbol{v} = n^{1/2}\sum_{k=0}^{m}\hat{\lambda}_k\boldsymbol{\mu}_k = \mathbf{0}.$$

Hence, we have

$$\boldsymbol{v} = \sum_{k=0}^{m}\hat{\lambda}_k^{1/2}\Big\{n_k^{-1/2}\sum_{j=1}^{n_k}\Big(\partial\mathcal{L}_{n,k}(\boldsymbol{\theta}^*,\, x_{kj})/\partial\boldsymbol{\theta} - \boldsymbol{\mu}_k\Big)\Big\}.$$

Clearly, each term in curly brackets is a centered sum of iid random variables with finite covariance matrices. Thus, they are all asymptotically normal with appropriate covariance matrices. In addition, these terms are independent of each other, $\hat{\lambda}_k = n_k/n$ are non-random with a limit ρ_k. Therefore, the linear combination is also asymptotically normal.

What left is to verify the form of the asymptotic covariance matrix. The asymptotic covariance matrix of each term in curly brackets is given by

$$V_k = \mathrm{E}_k \left\{ (\partial \mathcal{L}_k(\boldsymbol{\theta}^*, x)/\partial \boldsymbol{\theta})(\partial \mathcal{L}_k(\boldsymbol{\theta}^*, x)/\partial \boldsymbol{\theta}^\intercal) \right\} - \boldsymbol{\mu}_k \boldsymbol{\mu}_k^\intercal,$$

and hence the overall asymptotic variance matrix is $V = \sum_{k=0}^{m} \rho_k V_k$. In addition, it is easy to verify that

$$\sum_{k=0}^{m} \rho_k \mathrm{E}_k \left\{ (\partial \mathcal{L}_k(\boldsymbol{\theta}^*, x)/\partial \boldsymbol{\theta})(\partial \mathcal{L}_k(\boldsymbol{\theta}^*, x)/\partial \boldsymbol{\theta}^\intercal) \right\} = U$$

and we also find $\sum_{k=0}^{m} \rho_k \boldsymbol{\mu}_k \boldsymbol{\mu}_k^\intercal = UWU$ by (S1.4) and (S1.5). Thus, $V = U - UWU$ and this completes the proof. □

Lemma 2 (Quadratic form decomposition formula)**.** *Let* $\boldsymbol{z}^\intercal = (\boldsymbol{z}_1^\intercal, \boldsymbol{z}_2^\intercal)$ *be a vector of length* $m+n$*, partitioned in agreement with* m *and* n*, and* Σ *be a* $(m+n) \times (m+n)$ *a nonsingular matrix with partition*

$$\Sigma = \begin{pmatrix} \underset{m\times m}{A} & \underset{m\times n}{B} \\ \underset{n\times m}{B^\intercal} & \underset{n\times n}{C} \end{pmatrix}.$$

When A *is nonsingular, so is* $C - B^\intercal A^{-1} B$ *and*

$$\boldsymbol{z}^\intercal \Sigma^{-1} \boldsymbol{z} = \left(\boldsymbol{z}_2 - B^\intercal A^{-1} \boldsymbol{z}_1 \right)^\intercal \left(C - B^\intercal A^{-1} B \right)^{-1} \left(\boldsymbol{z}_2 - B^\intercal A^{-1} \boldsymbol{z}_1 \right) + \boldsymbol{z}_1^\intercal A^{-1} \boldsymbol{z}_1.$$

One can verify the above conclusion directly or refer to Theorem 8.5.11 of Harville 2008.

Proof of Theorem 1. We first work on quadratic expansions of $l_n(\hat{\boldsymbol{\theta}})$ and $l_n(\tilde{\boldsymbol{\theta}})$ under the null model. The difference of the two quadratic forms is then shown to have a chi-square limiting distribution.

Recall $\boldsymbol{v} = n^{-1/2}\partial l_n(\boldsymbol{\theta}^*)/\partial\boldsymbol{\theta}$. By expanding $l_n(\boldsymbol{\theta})$ at $\boldsymbol{\theta}^*$, we get

$$l_n(\boldsymbol{\theta}) = l_n(\boldsymbol{\theta}^*) + \sqrt{n}\boldsymbol{v}^\intercal(\boldsymbol{\theta} - \boldsymbol{\theta}^*) - (1/2)n(\boldsymbol{\theta} - \boldsymbol{\theta}^*)^\intercal U_n(\boldsymbol{\theta} - \boldsymbol{\theta}^*) + \epsilon_n$$

where $\epsilon_n = O_p(n^{-1/2})$ when $\boldsymbol{\theta} - \boldsymbol{\theta}^* = O_p(n^{-1/2})$ because the third derivative is bounded by an integrable function shown in (S1.3). Ignoring ϵ_n, the leading term in this expansion is maximized when

$$\boldsymbol{\theta} - \boldsymbol{\theta}^* = n^{-1/2}U_n^{-1}\boldsymbol{v} + o_p(n^{-1/2}).$$

At the same time, the DEL $l_n(\boldsymbol{\theta})$ is by definition maximized at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, and $\hat{\boldsymbol{\theta}}$ is known to be root-n consistent (Chen and Liu 2013 and Zhang 2002), hence

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* = n^{-1/2}U_n^{-1}\boldsymbol{v} + o_p(n^{-1/2}) = n^{-1/2}U^{-1}\boldsymbol{v} + o_p(n^{-1/2}),$$

which leads to

$$l_n(\hat{\boldsymbol{\theta}}) = l_n(\boldsymbol{\theta}^*) + (1/2)\boldsymbol{v}^\intercal U^{-1}\boldsymbol{v} + o_p(1). \tag{S2.1}$$

Next, we work on an expansion for $l_n(\tilde{\boldsymbol{\theta}})$ under the null model. Recall that $\boldsymbol{\beta}$ is part of $\boldsymbol{\theta}$. We express the null hypothesis $\boldsymbol{g}(\boldsymbol{\beta}) = \mathbf{0}$ in another equivalent form. Let $\boldsymbol{\beta}^*$ represent a null model. Recall that $\boldsymbol{g} : \mathbb{R}^{md} \to \mathbb{R}^q$ is thrice differentiable in a neighbourhood of $\boldsymbol{\beta}^*$ with full rank Jacobian matrix $\bigtriangledown = \partial\boldsymbol{g}(\boldsymbol{\beta}^*)/\partial\boldsymbol{\beta}$. When $q < md$, by the implicit function theorem (Zorich, 2004, 8.5.4, Theorem 1), there exists a unique function $\mathcal{G}$: $\mathbb{R}^{md-q} \to$

$\mathbb{R}^{md}$, such that $\boldsymbol{g}(\boldsymbol{\beta}) = 0$ if and only if $\boldsymbol{\beta} = \mathcal{G}(\boldsymbol{\gamma})$ for some $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ in a corresponding neighbourhoods of $\boldsymbol{\beta}^*$ and $\boldsymbol{\gamma}^*$ respectively. In addition, $\mathcal{G}$ is also thrice differentiable in a neighbourhood of $\boldsymbol{\gamma}^*$, and its Jacobian is

$$J = \partial \mathcal{G}(\boldsymbol{\gamma}^*)/\partial \boldsymbol{\gamma} = (-(\bigtriangledown_1^{-1} \bigtriangledown_2)^\intercal,\, I_{md-q})^\intercal.$$

This Jacobian is the same as the matrix J in Theorem 2. When $q = md$, by the inverse function theorem (Zorich, 2004, 8.6.1, Theorem 1), $\boldsymbol{g}$ is invertible at $\boldsymbol{\beta}^*$, i.e. $\boldsymbol{\beta}^* = \boldsymbol{g}^{-1}(\mathbf{0})$. Hence, in this case, $\boldsymbol{g}$ defines a simple hypothesis testing problem with $\boldsymbol{\beta}$ being fully specified to be $\boldsymbol{g}^{-1}(\mathbf{0})$ in the null.

We first look at the case of $q < md$. With the above representaion of the null model, the DRM parameter under the null hypothesis is $\boldsymbol{\theta} = (\boldsymbol{\alpha},\, \mathcal{G}(\boldsymbol{\gamma}))$. Hence, we may write the likelihood function under null model as

$$\ell_n(\boldsymbol{\alpha},\, \boldsymbol{\gamma}) = l_n(\boldsymbol{\alpha},\, \mathcal{G}(\boldsymbol{\gamma})).$$

Let $(\tilde{\boldsymbol{\alpha}},\, \tilde{\boldsymbol{\gamma}})$ be the maximum point of $\ell_n(\boldsymbol{\alpha}, \boldsymbol{\gamma})$. Clearly, $\ell_n(\boldsymbol{\alpha},\, \boldsymbol{\gamma})$ has the same properties as $l_n(\boldsymbol{\theta})$ and $\ell_n(\tilde{\boldsymbol{\alpha}},\, \tilde{\boldsymbol{\gamma}})$ has a similar expansion as (S2.1). Partition $\boldsymbol{v}$ into $\boldsymbol{v}_1 = n^{-1/2}\partial l_n(\boldsymbol{\theta}^*)/\partial \boldsymbol{\alpha}$ and $\boldsymbol{v}_2 = n^{-1/2}\partial l_n(\boldsymbol{\theta}^*)/\partial \boldsymbol{\beta}$. Note that

$$n^{-1/2}\partial \ell_n(\boldsymbol{\alpha}^*, \boldsymbol{\gamma}^*)/\partial \boldsymbol{\alpha} = n^{-1/2}\partial l_n(\boldsymbol{\theta}^*)/\partial \boldsymbol{\alpha} = \boldsymbol{v}_1.$$

By the chain rule,

$$n^{-1/2}\partial \ell_n(\boldsymbol{\alpha}^*, \boldsymbol{\gamma}^*)/\partial \boldsymbol{\gamma} = n^{-1/2} J^\intercal \{\partial l_n(\boldsymbol{\theta}^*)/\partial \boldsymbol{\beta})\} = J^\intercal \boldsymbol{v}_2. \tag{S2.2}$$

Similarly, the new information matrix is found to be

$$\tilde{U} = \begin{pmatrix} I_m & 0 \\ 0 & J^\intercal \end{pmatrix} \begin{pmatrix} U_{\boldsymbol{\alpha\alpha}} & U_{\boldsymbol{\alpha\beta}} \\ U_{\boldsymbol{\beta\alpha}} & U_{\boldsymbol{\beta\beta}} \end{pmatrix} \begin{pmatrix} I_m & 0 \\ 0 & J \end{pmatrix} = \begin{pmatrix} U_{\boldsymbol{\alpha\alpha}} & U_{\boldsymbol{\alpha\beta}}J \\ J^\intercal U_{\boldsymbol{\beta\alpha}} & J^\intercal U_{\boldsymbol{\beta\beta}}J \end{pmatrix}.$$

Consequently, we have

$$l_n(\tilde{\boldsymbol{\theta}}) = \ell_n(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\gamma}}) = \ell_n(\boldsymbol{\alpha}^*, \boldsymbol{\gamma}^*) + (1/2)(\boldsymbol{v}_1^\intercal, \boldsymbol{v}_2^\intercal J)\tilde{U}^{-1}(\boldsymbol{v}_1^\intercal, \boldsymbol{v}_2^\intercal J)^\intercal + o_p(1).$$

Combining (S2.1) and the above expansion, and noticing that $\ell_n(\boldsymbol{\alpha}^*, \boldsymbol{\gamma}^*) = l_n(\boldsymbol{\theta}^*)$, we have

$$R_n = 2\{l_n(\hat{\boldsymbol{\theta}}) - l_n(\tilde{\boldsymbol{\theta}})\} = \boldsymbol{v}^\intercal U^{-1}\boldsymbol{v} - (\boldsymbol{v}_1^\intercal, \boldsymbol{v}_2^\intercal J)\tilde{U}^{-1}(\boldsymbol{v}_1^\intercal, \boldsymbol{v}_2^\intercal J)^\intercal + o_p(1).$$

Applying Lemma 2 to the two quadratic forms on the right hand side (RHS) of the above expansion, we get

$$\begin{aligned} &\boldsymbol{v}^\intercal U^{-1}\boldsymbol{v} = \boldsymbol{\xi}^\intercal \Lambda^{-1}\boldsymbol{\xi} + \boldsymbol{v}_1^\intercal U_{\boldsymbol{\alpha\alpha}}^{-1}\boldsymbol{v}_1, \\ &(\boldsymbol{v}_1^\intercal, \boldsymbol{v}_2^\intercal J)\tilde{U}^{-1}(\boldsymbol{v}_1^\intercal, \boldsymbol{v}_2^\intercal J)^\intercal = \boldsymbol{\xi}^\intercal J(J^\intercal \Lambda J)^{-1} J^\intercal \boldsymbol{\xi} + \boldsymbol{v}_1^\intercal U_{\boldsymbol{\alpha\alpha}}^{-1}\boldsymbol{v}_1, \end{aligned} \tag{S2.3}$$

where $\boldsymbol{\xi} = (-U_{\boldsymbol{\beta\alpha}}U_{\boldsymbol{\alpha\alpha}}^{-1},\ I_{md})\boldsymbol{v}$ and $\Lambda = U_{\boldsymbol{\beta\beta}} - U_{\boldsymbol{\beta\alpha}}U_{\boldsymbol{\alpha\alpha}}^{-1}U_{\boldsymbol{\alpha\beta}}$ is defined in Theorem 2. We then obtain the following expansion

$$R_n = 2\{l_n(\hat{\boldsymbol{\theta}}) - l_n(\tilde{\boldsymbol{\theta}})\} = \boldsymbol{\xi}^\intercal\{\Lambda^{-1} - J(J^\intercal \Lambda J)^{-1} J^\intercal\}\boldsymbol{\xi} + o_p(1). \tag{S2.4}$$

Recall that, by Lemma 1, $\boldsymbol{v}$ is asymptotically $N(\mathbf{0}, U - UWU)$, so $\boldsymbol{\xi}$ is asymptotic normal with mean $\mathbf{0}$ and covariance matrix $(-U_{\boldsymbol{\beta\alpha}}U_{\boldsymbol{\alpha\alpha}}^{-1},\ I_{md})(U - UWU)(-U_{\boldsymbol{\beta\alpha}}U_{\boldsymbol{\alpha\alpha}}^{-1},\ I_{md})^\intercal = \Lambda$, where the last equality is obtained using the expression of W given in Lemma 1.

The last step is to verify the quadratic form in the above expansion of R_n has the claimed limiting distribution. We can easily check that

$$\Lambda^{1/2}\{\Lambda^{-1} - J(J^{\intercal}\Lambda J)^{-1}J^{\intercal}\}\Lambda^{1/2}$$

is idempotent. Moreover, the trace of the above idempotent matrix is found to be q. Therefore, by Theorem 5.1.1 of Mathai (1992), the quadratic form in expansion (S2.4), and hence also R_n, has a χ^2_q limiting distribution.

The above proof is applicable to $q < md$. When $q = md$, the value of $\boldsymbol{\beta}$ is fully specified. Hence, the maximization under null is solely with respect to $\boldsymbol{\alpha}$ and we easily find

$$l_n(\tilde{\boldsymbol{\theta}}) = l_n(\boldsymbol{\theta}^*) + (1/2)\boldsymbol{v}_1^{\intercal}U_{\boldsymbol{\alpha}\boldsymbol{\alpha}}^{-1}\boldsymbol{v}_1 + o_p(1).$$

This, along with the expansion (S2.1) of $l_n(\hat{\boldsymbol{\theta}})$ and expression (S2.3), implies that $R_n = \boldsymbol{\xi}^{\intercal}\Lambda^{-1}\boldsymbol{\xi} + o_p(1)$. Just as the proof for the case of $q < md$, the limiting distribution of the above R_n is seen to be χ^2_{md}. □

S3 Proof of Theorem 2

We first sketch out the proof of Theorem 2. Let $\boldsymbol{\beta}^*$ be a specific parameter value under the null hypothesis and $\{F_k\}$ be the corresponding distribution functions. Let $\{G_k\}$ be the set of distribution functions satisfying the DRM with parameter given by $\boldsymbol{\beta}_k = \boldsymbol{\beta}_k^* + n_k^{-1/2}\boldsymbol{c}_k$, $k = 1, \ldots, m$, and $G_0 =$

F_0. When the samples are generated from the $\{G_k\}$, we still have that the DELR statistic is approximated by the quadratic form on the RHS of (S2.4). The limiting distribution of R_n is therefore determined by that of $\boldsymbol{v} = n^{-1/2}\partial l_n(\boldsymbol{\theta}^*)/\partial\boldsymbol{\theta}$. According to Le Cam's third lemma (van der Vaart 2000, 6.7), $\boldsymbol{v}$ has a specific limiting distribution under the $\{G_k\}$ if $\boldsymbol{v}$ and $\sum_{k,j}\log\{dG_k(x_{kj})/dF_k(x_{kj})\}$, under the $\{F_k\}$, are jointly normal with a particular mean and variance structure. The core of the proof then is to establish that structure.

For each $k = 0, \ldots, m$, let $\mathrm{Var}_k(\cdot)$ and $\mathrm{Cov}_k(\cdot, \cdot)$ be the variance and covariance operators with respect to F_k, respectively.

Lemma 3. *Under the conditions of Theorem 1 and the distribution functions $\{G_k\}$, $\boldsymbol{v}$ is asymptotically normal with mean*

$$\boldsymbol{\tau} = \sum_{k=1}^{m}\sqrt{\rho_k}\,Cov_k\{\partial\mathcal{L}_k(\boldsymbol{\theta}^*, x)/\partial\boldsymbol{\theta}, \boldsymbol{q}^{\intercal}(x)\}\boldsymbol{c}_k$$

and covariance matrix $V = U - UWU$ as given in Lemma 1.

Proof of Lemma 3. We first expand $\boldsymbol{w}_k = \sum_{j=1}^{n_k}\log\{\mathrm{d}G_k(x_{kj})/\mathrm{d}F_k(x_{kj})\}$. Notice that

$$\begin{aligned} dG_k(x)/dF_k(x) &= \exp\{\alpha_k + \boldsymbol{\beta}_k\boldsymbol{q}(x)\}/\exp\{\alpha_k^* + \boldsymbol{\beta}_k^*\boldsymbol{q}(x)\} \\ &= \exp\{\alpha_k - \alpha_k^* + n_k^{-1/2}\boldsymbol{c}_k\boldsymbol{q}(x)\}. \end{aligned}$$

Because α_k and α_k^* are normalization constants, we have

$$\exp\{\alpha_k^* - \alpha_k\} = \int \exp\{\alpha_k^* + (\boldsymbol{\beta}_k^{*\intercal} + n_k^{-1/2}\boldsymbol{c}_k^\intercal)\boldsymbol{q}(x)\} dF_0(x).$$

Ignoring terms of order $n^{-3/2}$ and higher, it leads to

$$\begin{aligned}\exp\{\alpha_k^* - \alpha_k\} &= \int \exp\{n_k^{-1/2}\boldsymbol{c}_k^\intercal\boldsymbol{q}(x)\} \exp\left\{\boldsymbol{\alpha}_k^* + \boldsymbol{\beta}_k^{*\intercal}\boldsymbol{q}(x)\right\} dF_0(x)\\ &\approx \int \left\{1 + n_k^{-1/2}\boldsymbol{c}_k^\intercal\boldsymbol{q}(x) + (2n_k)^{-1}(\boldsymbol{c}_k^\intercal\boldsymbol{q}(x))^2\right\} dF_k(x).\end{aligned}$$

Denote $\boldsymbol{\nu}_k = \mathrm{E}_k\boldsymbol{q}(x)$. Then, it is further simplified to

$$\exp\{\alpha_k^* - \alpha_k\} \approx 1 + n_k^{-1/2}\boldsymbol{c}_k^\intercal\boldsymbol{\nu}_k + (2n_k)^{-1}\boldsymbol{c}_k^\intercal\mathrm{E}_k(\boldsymbol{q}^2(x))\boldsymbol{c}_k.$$

Hence, ignoring a $O(n^{-3/2})$ term, we have

$$\log\{dG_k(x)/dF_k(x)\} \approx n_k^{-1/2}\boldsymbol{c}_k\boldsymbol{q}(x) - \log\{1 + n_k^{-1/2}\boldsymbol{c}_k^\intercal\boldsymbol{\nu}_k + (2n_k)^{-1}\boldsymbol{c}_k^\intercal\mathrm{E}_k(\boldsymbol{q}^2(x))\boldsymbol{c}_k\}.$$

Write $\boldsymbol{\sigma}_k = \mathrm{Var}_k(\boldsymbol{q}(x))$. Expanding the logarithmic term on the RHS, we get

$$\begin{aligned}&\log\{1 + n_k^{-1/2}\boldsymbol{c}_k^\intercal\boldsymbol{\nu}_k + (2n_k)^{-1}\boldsymbol{c}_k^\intercal\mathrm{E}_k(\boldsymbol{q}^2(x))\boldsymbol{c}_k\}\\ =&n_k^{-1/2}\boldsymbol{c}_k^\intercal\boldsymbol{\nu}_k + (2n_k)^{-1}\boldsymbol{c}_k^\intercal\mathrm{E}_k(\boldsymbol{q}^2(x))\boldsymbol{c}_k - n_k\boldsymbol{c}_k^\intercal\{\boldsymbol{\nu}_k\boldsymbol{\nu}_k^\intercal\}\boldsymbol{c}_k + O(n^{-3/2})\\ =&n_k^{-1/2}\boldsymbol{c}_k^\intercal\boldsymbol{\nu}_k + (2n_k)^{-1}\boldsymbol{c}_k^\intercal\boldsymbol{\sigma}_k\boldsymbol{c}_k + O(n^{-3/2}).\end{aligned}$$

Therefore

$$\log\{dG_k(x)/dF_k(x)\} = n_k^{-1/2}\boldsymbol{c}_k^\intercal\{\boldsymbol{q}(x) - \boldsymbol{\nu}_k\} - (2n_k)^{-1}\boldsymbol{c}_k^\intercal\boldsymbol{\sigma}_k\boldsymbol{c}_k + O(n^{-3/2}).$$

Summing over j, we get, for each k,

$$\begin{aligned}\boldsymbol{w}_k &= \sum_{j=1}^{n_k} \log\{\mathrm{d}G_k(x_{kj})/\mathrm{d}F_k(x_{kj})\} \\ &= n_k^{-1/2}\boldsymbol{c}_k^\intercal \sum_{j=1}^{n_k}\{\boldsymbol{q}(x_{kj}) - \boldsymbol{\nu}_k\} - (1/2)\boldsymbol{c}_k^\intercal \boldsymbol{\sigma}_k \boldsymbol{c}_k + O(n^{-1/2}).\end{aligned}$$

When $k = 0$, we have $\boldsymbol{c}_0 = 0$.

Recall that $l_n(\boldsymbol{\theta}^*) = \sum_{k,j} \mathcal{L}_{n,k}(\boldsymbol{\theta}^*, x_{kj})$ and $\hat{\lambda}_k = n_k/n$ whose limit is ρ_k, we have

$$\begin{pmatrix} \boldsymbol{v} \\ \sum_k \boldsymbol{w}_k \end{pmatrix} \approx \sum_{k=0}^{m} \frac{1}{\sqrt{n_k}} \sum_{j=1}^{n_k} \begin{pmatrix} \sqrt{\rho_k}\,\{\partial \mathcal{L}_{n,k}(\boldsymbol{\theta}^*, x_{kj})/\partial\boldsymbol{\theta} - \boldsymbol{\mu}_k\} \\ \boldsymbol{c}_k^\intercal\{\boldsymbol{q}(x_{kj}) - \boldsymbol{\nu}_k\} \end{pmatrix} - \sum_{k=0}^{m} \begin{pmatrix} \boldsymbol{0} \\ \frac{1}{2}\boldsymbol{c}_k^\intercal \boldsymbol{\sigma}_k \boldsymbol{c}_k \end{pmatrix},$$

which is seen to be jointly asymptotically normal under the null distributions $\{F_k\}$. The corresponding mean vector and variance matrix are given by

$$\left(\boldsymbol{0}^\intercal,\ -\frac{1}{2}\sum_k \boldsymbol{c}_k^\intercal \boldsymbol{\sigma}_k \boldsymbol{c}_k\right)^\intercal \text{ and } \begin{pmatrix} V & \boldsymbol{\tau} \\ \boldsymbol{\tau}^\intercal & \sum_k \boldsymbol{c}_k^\intercal \boldsymbol{\sigma}_k \boldsymbol{c}_k \end{pmatrix},$$

where $\boldsymbol{\tau}$ is the one given in the Lemma. Because the second entry of the mean vector equals negative half of the lower-right entry of the covariance matrix, the condition of Le Cam's third lemma is satisfied. By that lemma, we conclude that $\boldsymbol{v}$ has a normal limiting distribution with mean $\boldsymbol{\tau}$ and covariance matrix V under the local alternative distributions $\{G_k\}$. □

Proof of Theorem 2. We first show that, under the $\{G_k\}$, the DELR statistic R_n is still approximated by the quadratic form on the RHS of (S2.4).

Under the $\{G_k\}$, we still have $-n^{-1}\partial^2 l_n(\boldsymbol{\theta}^*)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\intercal \to U$ and $\boldsymbol{v} = O_p(1)$. In addition, $\hat{\boldsymbol{\theta}}$ still admits the expansion

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) = U^{-1}\boldsymbol{v} + o_p(1) = O_p(1),$$

and hence it is root-n consistent for $\boldsymbol{\theta}^*$. Similarly, the constrained MELE $\tilde{\boldsymbol{\theta}}$ is also root-n consistent for $\boldsymbol{\theta}^*$ under the $\{G_k\}$. The root-n consistency of $\hat{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{\theta}}$ imply

$$R_n = \boldsymbol{\xi}^\intercal\{\Lambda^{-1} - J(J^\intercal\Lambda J)^{-1}J^\intercal\}\boldsymbol{\xi} + o_p(1)$$

when $q < md$, and $R_n = \boldsymbol{\xi}^\intercal\Lambda^{-1}\boldsymbol{\xi} + o_p(1)$ when $q = md$. The matrix in the quadratic form of the expansion of R_n is the same as that in (S2.4). What has changed is the distribution of $\boldsymbol{\xi} = (-U_{\boldsymbol{\beta\alpha}}U_{\boldsymbol{\alpha\alpha}}^{-1},\ I_{md})\boldsymbol{v}$.

By Lemma 3, under the local alternative $\{G_k\}$, $\boldsymbol{v}$ is asymptotically $N(\boldsymbol{\tau}, V)$. Hence $\boldsymbol{\xi}$ also has a normal limiting distribution. Since the asymptotic covariance matrix of $\boldsymbol{v}$ is the same as that under the $\{F_k\}$, the asymptotic covariance matrix of $\boldsymbol{\xi}$ is still Λ as we have shown in the proof of Theorem 1. The mean of the limiting distribution of $\boldsymbol{\xi}$ now is $\boldsymbol{\mu} = (-U_{\boldsymbol{\beta\alpha}}U_{\boldsymbol{\alpha\alpha}}^{-1},\ I_{md})\boldsymbol{\tau} = \Lambda\boldsymbol{\eta}$, where $\boldsymbol{\eta}$ is defined in Theorem 2 and the last equality is derived using (S1.5).

In the proof of Theorem 1, we have verified that the matrix

$$A = \Lambda^{1/2}\{\Lambda^{-1} - J(J^\intercal\Lambda J)^{-1}J^\intercal\}\Lambda^{1/2}$$

is idempotent with rank q. Hence, by Corollary 5.1.3a of Mathai (1992), the quadratic form in the above expansion of R_n, and hence R_n, has the claimed non-central chi-square limiting distribution.

In the last step we verify the condition for positiveness of the non-centrality parameter δ^2. When $q = md$, $\delta^2 = \boldsymbol{\eta}^\intercal \Lambda \boldsymbol{\eta} > 0$ because Λ is positive definite. When $q < md$, $\delta^2 = (\boldsymbol{\eta}^\intercal \Lambda^{1/2}) A (\Lambda^{1/2} \boldsymbol{\eta})$. We verified that A is an idempotent matrix. Hence, A is positive semidefinite and $\delta^2 \geq 0$. Moreover, $\delta^2 = 0$ if and only if $\Lambda^{1/2}\boldsymbol{\eta}$ is in the null space of A. The null space of A is the column space of $I - A = \Lambda^{1/2} J \Big(J^\intercal \Lambda J\Big)^{-1} J^\intercal \Lambda^{1/2}$, which is just the column space of $\Lambda^{1/2} J$. It is easily verified that $\Lambda^{1/2}\boldsymbol{\eta}$ is in the column space of $\Lambda^{1/2} J$ if and only if $\boldsymbol{\eta}$ is in the column space of J. Hence $\Lambda^{1/2}\boldsymbol{\eta}$ is in the null space of A and $\delta^2 = 0$ if and only if $\boldsymbol{\eta}$ is in the column space of J. □

S4 Proof of Theorem 3

We first introduce a useful notation for Schur complements that will be frequently used in the subsequent proofs. Let matrix

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

be nonsingular. We write $M/A = D - CA^{-1}B$ and call it the *Schur complement of* M *with respect to its upper-left block* A. Also, we write $M/D = A - BD^{-1}C$ and call it the *Schur complement of* M *with respect to its lower-right block* D.

Recall that we defined two DELRT statistics $R_n^{(1)}$ and $R_n^{(2)}$ which are constructed using the samples from only the first $r+1$ populations $F_0, \cdots, F_r$, and the samples from all the populations, respectively. Let U be the information matrix based on all $m+1$ samples ($R_n^{(2)}$), and $\tilde{U}$ be that based on the first $r+1$ samples ($R_n^{(1)}$). Similar to the partition of U, we partition $\tilde{U}$ to $\tilde{U}_{\boldsymbol{\alpha\alpha}}$, $\tilde{U}_{\boldsymbol{\alpha\beta}}$, $\tilde{U}_{\boldsymbol{\beta\alpha}}$ and $\tilde{U}_{\boldsymbol{\beta\beta}}$, and similar to the definition $\Lambda = U/U_{\boldsymbol{\alpha\alpha}}$ given in Theorem 2, we define $\tilde{\Lambda} = \tilde{U}/\tilde{U}_{\boldsymbol{\alpha\alpha}}$. We also partition Λ as

$$\Lambda = \begin{pmatrix} \Lambda_a & \Lambda_b \\ \Lambda_b^T & \Lambda_c \end{pmatrix},$$

where Λ_a is the upper-left $rd \times rd$ block of Λ.

The null hypothesis of (4.1) under investigation contains a constraint $\boldsymbol{g}(\boldsymbol{\zeta}) = \boldsymbol{0}$ with $\boldsymbol{\zeta}^\intercal = (\boldsymbol{\beta}_1^\intercal, \ldots, \boldsymbol{\beta}_r^\intercal)$ related only to populations $F_0, \cdots, F_r$. As noted in the proof of Theorem 1, this null constraint is equivalent to $\boldsymbol{\zeta} = \mathcal{G}(\boldsymbol{\gamma})$ for some smooth function $\mathcal{G}$: $\mathbb{R}^{rd-q} \to \mathbb{R}^{rd}$ and parameter vector $\boldsymbol{\gamma}$. Denote the Jacobian of $\mathcal{G}$ evaluated at $\boldsymbol{\gamma}^*$ as J. By Theorem 2, under the $\{G_k\}$ defined by the local alternative model (4.2), $R_n^{(1)}$ and $R_n^{(2)}$ both have non-central chi-square limiting distributions of q degrees of freedom, but with different non-centrality parameters δ_1^2 and δ_2^2, respectively. We also know that for $R_n^{(1)}$,

$$\delta_1^2 = \rho\tilde{\boldsymbol{\eta}}^\intercal \left\{\tilde{\Lambda} - \tilde{\Lambda}J(J^\intercal\tilde{\Lambda}J)^{-1}J^\intercal\tilde{\Lambda}\right\}\tilde{\boldsymbol{\eta}},$$

where $\tilde{\boldsymbol{\eta}} = (\rho_1^{-1/2}\boldsymbol{c}_1^\intercal, \ldots, \rho_r^{-1/2}\boldsymbol{c}_r^\intercal)$. Moreover, under the same local alternative model, for $R_n^{(2)}$, we have $\boldsymbol{\eta}^\intercal = (\tilde{\boldsymbol{\eta}}^\intercal, 0_{m-r}^\intercal)$ and the corresponding Jacobian matrix of the null mapping is $J_2 = \mathrm{diag}(J, I_{(m-r)d})$. Thus

$$\delta_2^2 = \boldsymbol{\eta}^\intercal \left\{\Lambda - \Lambda J_2(J_2^\intercal\Lambda J_2)^{-1}J_2^\intercal\Lambda\right\}\boldsymbol{\eta}.$$

Let A denote the upper-left $rd \times rd$ block of $\Lambda - \Lambda J_2(J_2^\intercal\Lambda J_2)^{-1}J_2^\intercal\Lambda$. Since $\boldsymbol{\eta}$ consists of $\tilde{\boldsymbol{\eta}}$ and a zero vector, we have

$$\delta_2^2 = \tilde{\boldsymbol{\eta}}^\intercal A\tilde{\boldsymbol{\eta}},$$

The upper-left block of Λ is Λ_a. By the quadratic form decomposition formula of Lemma 2, the upper-left block of $\Lambda J_2(J_2^\intercal\Lambda J_2)^{-1}J_2^\intercal\Lambda$ is found to be

$$(\Lambda/\Lambda_c)J(J^\intercal(\Lambda/\Lambda_c)J)^{-1}J^\intercal(\Lambda/\Lambda_c) + \Lambda_b\Lambda_c^{-1}\lambda_b^T.$$

Hence, the expression of δ_2^2 becomes

$$\begin{aligned}
\delta_2^2 &= \tilde{\boldsymbol{\eta}}^\intercal A \tilde{\boldsymbol{\eta}} \\
&= \tilde{\boldsymbol{\eta}}^\intercal \left\{ \Lambda_a - \Lambda_b \Lambda_c^{-1} \lambda_b^T - (\Lambda/\Lambda_c) J (J^\intercal (\Lambda/\Lambda_c) J)^{-1} J^\intercal (\Lambda/\Lambda_c) \right\} \tilde{\boldsymbol{\eta}} \\
&= \tilde{\boldsymbol{\eta}}^\intercal \left\{ (\Lambda/\Lambda_c) - (\Lambda/\Lambda_c) J (J^\intercal (\Lambda/\Lambda_c) J)^{-1} J^\intercal (\Lambda/\Lambda_c) \right\} \tilde{\boldsymbol{\eta}}.
\end{aligned}$$

Therefore, to show the claimed result $\delta_2^2 \geq \delta_1^2$, it suffices to show that

$$(\Lambda/\Lambda_c) - (\Lambda/\Lambda_c) J (J^\intercal (\Lambda/\Lambda_c) J)^{-1} J^\intercal (\Lambda/\Lambda_c) \geq \rho \left\{ \tilde{\Lambda} - \tilde{\Lambda} J (J^\intercal \tilde{\Lambda} J)^{-1} J^\intercal \tilde{\Lambda} \right\}. \quad \text{(S4.1)}$$

We prove this equality in the sequel.

Recall that we defined $\boldsymbol{\theta}_k^\intercal = (\alpha_k, \boldsymbol{\beta}_k^\intercal)$. Denote the information matrix with respect to $(\boldsymbol{\theta}_1^\intercal, \ldots, \boldsymbol{\theta}_r^\intercal)^\intercal$ under the DRM based on the first $r+1$ samples as U_1, and that with respect to $(\boldsymbol{\theta}_1^\intercal, \ldots, \boldsymbol{\theta}_m^\intercal)^\intercal$ under the DRM based on all $m+1$ samples as U_2. Let $U_{2,c}$ be the lower-right $(m-r)(d+1) \times (m-r)(d+1)$ block of U_2. Let $\rho = \lim_{n\to\infty} (\sum_{k=0}^r n_k)/n$.

Lemma 4. *Adopt the conditions of Theorem 1. We have:*

(i) $U_2/U_{2,c} \geq \rho U_1$. *That is,* $U_2/U_{2,c} - \rho U_1$ *is positive semidefinite.*

(ii) $\Lambda/\Lambda_c \geq \rho \tilde{\Lambda}$.

Lemma 5. *Let A be a $s \times s$ positive definite matrix and B be a $s \times s$ positive semidefinite matrix. Also let X and Y be $s \times t$ matrices, and suppose the* *column space of Y is contained in that of B. Then*

$$(X+Y)^\intercal(A+B)^{-1}(X+Y) \le X^\intercal A^{-1}X + Y^\intercal B^\dagger Y$$

where $B^\dagger$ is the Moore-Penrose pseudoinverse of B.

The proofs of the above two lemmas are given after the proof of Theorem 3.

Proof of Theorem 3. We prove equality (S4.1). Define $M = \Lambda/\Lambda_c - \rho\tilde{\Lambda}$. In Lemma 5, let $A = \rho J^\intercal \tilde{\Lambda} J$, $B = J^\intercal M J$, $X = \rho J^\intercal \tilde{\Lambda}$, $Y = J^\intercal M$. Then $A + B = J^\intercal(\Lambda/\Lambda_c)J$ and $X + Y = J^\intercal(\Lambda/\Lambda_c)$. Matrix A is positive definite because $\tilde{\Lambda}$ is positive definite and J is of full rank. B is positive semidefinite because M is positive semidefinite by Lemma 4 (ii). Moreover, it is easily seen that the column space of Y is the same as that of B. Hence the conditions of Lemma 5 are satisfied, and we have

$$(\Lambda/\Lambda_c)J(J^\intercal(\Lambda/\Lambda_c)J)^{-1}J^\intercal(\Lambda/\Lambda_c) \le \rho\tilde{\Lambda}J(J^\intercal\tilde{\Lambda}J)^{-1}J^\intercal\tilde{\Lambda} + MJ(J^\intercal MJ)^\dagger J^\intercal M.$$

The above inequality and $\Lambda/\Lambda_c = \rho\tilde{\Lambda} + M$ imply that

$$\begin{aligned}&(\Lambda/\Lambda_c) - (\Lambda/\Lambda_c)J(J^\intercal(\Lambda/\Lambda_c)J)^{-1}J^\intercal(\Lambda/\Lambda_c)\\ \ge&\rho\{\tilde{\Lambda} - \tilde{\Lambda}J(J^\intercal\tilde{\Lambda}J)^{-1}J^\intercal\tilde{\Lambda}\} + \{M - MJ(J^\intercal MJ)^\dagger J^\intercal M\}.\end{aligned}$$

The term $M - MJ(J^\intercal MJ)^\dagger J^\intercal M$ is positive semidefinite because

$$M - MJ(J^\intercal MJ)^\dagger J^\intercal M = M^{1/2}\{I - M^{1/2}J(J^\intercal MJ)^\dagger J^\intercal M^{1/2}\}M^{1/2},$$

and $I - M^{1/2}J(J^\intercal MJ)^\dagger J^\intercal M^{1/2}$ is easily verified to be idempotent, hence positive semidefinite. Therefore inequality (S4.1) holds and the claimed result is true. □

Proof of Lemma 4 (i). We prove the result for $m = r+1$, namely $R_n^{(1)}$ uses all sample except for the last one. The general result is true by mathematical induction.

Let $U_{2,a}$ be the upper-left $r(d+1) \times r(d+1)$ block, and $U_{2,b}$ be the upper-right $r(d+1) \times (m-r)(d+1)$ block, of U_2. Note that $U_2/U_{2,c} = U_{2,a} - U_{2,b}U_{2,c}^{-1}U_{2,b}^\intercal$, so to show the claimed result of $U_2/U_{2,c} \geq \rho\tilde{U_1}$, it suffices to show that

$$(U_{2,a} - \rho U_1) - U_{2,b}U_{2,c}^{-1}U_{2,b}^\intercal$$

is positive semidefinite. Notice that the above matrix is the Shur complement of

$$D = \begin{pmatrix} U_{2,a} - \rho U_1 & U_{2,b} \\ U_{2,b}^\intercal & U_{2,c} \end{pmatrix} = U_2 - \mathrm{diag}(\rho U_1,\, 0). \tag{S4.2}$$

By standard matrix theory, the positive semidefiniteness is implied by that of D.

We now show D is positive semidefinite. We first give useful algebraic expressions for U_2 and ρU_1. Notice that $(\boldsymbol{\theta}_1^\intercal, \ldots, \boldsymbol{\theta}_m^\intercal)$ is just permuted $\boldsymbol{\theta}^\intercal = (\boldsymbol{\alpha}^\intercal, \boldsymbol{\beta}^\intercal)$, the information matrix (3.7) of which helps us to obtain algebraic expressions for U_1 and U_2. Recall $\boldsymbol{Q}(x) = (1,\, \boldsymbol{q}^\intercal(x))^\intercal$. For $R_n^{(2)}$, we get

$$U_2 = \mathrm{E}_0\Big\{H(\boldsymbol{\theta}^*, x) \otimes \{\boldsymbol{Q}(x)\boldsymbol{Q}^\intercal(x)\}\Big\}.$$

For $R_n^{(1)}$, we find

$$\rho U_1 = \mathrm{E}_0\Big\{H_r(\boldsymbol{\theta}^*, x) \otimes \{\boldsymbol{Q}(x)\boldsymbol{Q}^\intercal(x)\}\Big\},$$

where $H_r(\boldsymbol{\theta}, x)$ is the H matrix defined in (3.6) based on the first $r+1$ samples. Substituting the above expressions of U_2 and ρU_1 into the expression (S4.2) of D, we get

$$D = \rho_m \mathrm{E}_0\Big\{\{\boldsymbol{w}(x)\boldsymbol{w}^\intercal(x)\} \otimes \{\boldsymbol{Q}(x)\boldsymbol{Q}^\intercal(x)\}\Big\},$$

with

$$\boldsymbol{w}(x) = \sqrt{\varphi_m(\boldsymbol{\theta}^*,\, x)}\,\Big(\boldsymbol{h}_r^\intercal(\boldsymbol{\theta}^*,\, x),\, s_r(\boldsymbol{\theta}^*,\, x)\Big)^\intercal \Big/ \sqrt{s(\boldsymbol{\theta}^*,\, x)s_r(\boldsymbol{\theta}^*,\, x)},$$

where $\boldsymbol{h}_r(\boldsymbol{\theta},\, x)$ and $s_r(\boldsymbol{\theta},\, x)$ are the $\boldsymbol{h}$ vector and s defined in (3.6) based on the first $r+1$ samples, respectively. Since D is the expectation of the Kronecker product of two squares of vectors, it is positive semidefinite. This completes the proof. □

To prove Lemma 4 (ii), partition $U_{\boldsymbol{\alpha\alpha}}$, $U_{\boldsymbol{\alpha\beta}}$ and $U_{\boldsymbol{\beta\beta}}$ as follows:

$$U_{\boldsymbol{\alpha\alpha}} = \begin{pmatrix} U_{\boldsymbol{\alpha\alpha},a} & U_{\boldsymbol{\alpha\alpha},b} \\ U_{\boldsymbol{\alpha\alpha},b}^\intercal & U_{\boldsymbol{\alpha\alpha},c} \end{pmatrix},\ U_{\boldsymbol{\alpha\beta}} = \begin{pmatrix} U_{\boldsymbol{\alpha\beta},a} & U_{\boldsymbol{\alpha\beta},b} \\ U_{\boldsymbol{\alpha\beta},c} & U_{\boldsymbol{\alpha\beta},d} \end{pmatrix},\ U_{\boldsymbol{\beta\beta}} = \begin{pmatrix} U_{\boldsymbol{\beta\beta},a} & U_{\boldsymbol{\beta\beta},b} \\ U_{\boldsymbol{\beta\beta},b}^\intercal & U_{\boldsymbol{\beta\beta},c} \end{pmatrix},$$

where $U_{\boldsymbol{\alpha\alpha},a}$, $U_{\boldsymbol{\alpha\beta},a}$ and $U_{\boldsymbol{\beta\beta},a}$ are the corrsponding upper-left $r \times r$, $r \times rd$ and $rd \times rd$ blocks.

We also introduce an important property of the Schur complement. Let

$$M = \begin{pmatrix} \underset{s\times s}{A} & \underset{s\times t}{B} \\ \underset{t\times s}{C} & \underset{t\times t}{D} \end{pmatrix} \quad \text{and} \quad D = \begin{pmatrix} \underset{u\times u}{E} & \underset{u\times v}{F} \\ \underset{v\times u}{G} & \underset{v\times v}{H} \end{pmatrix},$$

where $u + v = t$. Suppose M, A and D are nonsingular. By Theorem 1.4 of Zhang (2005), the lower-right $u \times u$ block of M/H is just D/H, and

$$M/D = (M/H)/(D/H). \tag{S4.3}$$

The above equality is known as the *quotient formula*. Similar quotient formula holds for M/A.

Proof of Lemma 4 (ii). We first give an algebraic expression for Λ/Λ_c. Recall the definition $\Lambda = U_{\boldsymbol{\beta\beta}} - U_{\boldsymbol{\beta\alpha}} U_{\boldsymbol{\alpha\alpha}}^{-1} U_{\boldsymbol{\alpha\beta}}$, so

$$\Lambda = \Psi / U_{\boldsymbol{\alpha\alpha}},$$

where

$$\Psi = \begin{pmatrix} U_{\boldsymbol{\beta\beta}} & U_{\boldsymbol{\beta\alpha}} \\ U_{\boldsymbol{\alpha\beta}} & U_{\boldsymbol{\alpha\alpha}} \end{pmatrix}.$$

Let Ψ_1 be the lower-right $\{(m-r)d + m\} \times \{(m-r)d + m\}$ block of Ψ. Then Λ_c, the lower-right $(m-r)d \times (m-r)d$ block of $\Lambda = \Psi/U_{\boldsymbol{\alpha\alpha}}$, satisfies

$$\Lambda_c = \Psi_1 / U_{\boldsymbol{\alpha\alpha}}.$$

Therefore

$$\Lambda/\Lambda_c = (\Psi/U_{\boldsymbol{\alpha\alpha}})/(\Psi_1/U_{\boldsymbol{\alpha\alpha}}) = \Psi/\Psi_1,$$

where the second equality above is by quotient formula (S4.3).

It is easily seen that $\Psi/\Psi_1 = \Omega/\Omega_1$, where

$$\Omega = \left(\begin{array}{cc:cc} U_{\boldsymbol{\beta\beta},a} & U_{\boldsymbol{\beta\alpha},a} & U_{\boldsymbol{\beta\beta},b} & U_{\boldsymbol{\beta\alpha},b} \\ U_{\boldsymbol{\alpha\beta},a} & U_{\boldsymbol{\alpha\alpha},a} & U_{\boldsymbol{\alpha\beta},b} & U_{\boldsymbol{\alpha\alpha},b} \\ \hdashline U_{\boldsymbol{\beta\beta},b}^{\intercal} & U_{\boldsymbol{\beta\alpha},c} & U_{\boldsymbol{\beta\beta},c} & U_{\boldsymbol{\beta\alpha},d} \\ U_{\boldsymbol{\alpha\beta},c} & U_{\boldsymbol{\alpha\alpha},b}^{\intercal} & U_{\boldsymbol{\alpha\beta},d} & U_{\boldsymbol{\alpha\alpha},c} \end{array}\right)$$

and Ω_1 is the lower-right block of Ω with the same size as that of Ψ_1. Thus we get

$$\Lambda/\Lambda_c = \Psi/\Psi_1 = \Omega/\Omega_1.$$

Let Ω_2 be the lower-right $(m-r)(d+1) \times (m-r)(d+1)$ block of Ω_1. Matrix Ω_1/Ω_2 is just the lower-right $r \times r$ block of Ω/Ω_2, and $\Omega/\Omega_1 = (\Omega/\Omega_2)/(\Omega_1/\Omega_2)$ by quotient formula (S4.3). Hence, we finally get

$$\Lambda/\Lambda_c = \Omega/\Omega_1 = (\Omega/\Omega_2)/(\Omega_1/\Omega_2).$$

The above identity implies that our cliam of $\Lambda/\Lambda_c \geq \rho\tilde{\Lambda}$ is equivalent to

$$(\Omega/\Omega_2)/(\Omega_1/\Omega_2) \geq \rho\tilde{\Lambda}.$$

Further notice that $\tilde{\Lambda} = \check{U}/\tilde{U}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}$, where

$$\check{U} = \begin{pmatrix} \tilde{U}_{\boldsymbol{\beta}\boldsymbol{\beta}} & \tilde{U}_{\boldsymbol{\beta}\boldsymbol{\alpha}} \\ \tilde{U}_{\boldsymbol{\alpha}\boldsymbol{\beta}} & \tilde{U}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} \end{pmatrix},$$

so, the above inequality is equivalent to

$$(\Omega/\Omega_2)/(\Omega_1/\Omega_2) \geq \rho(\check{U}/\tilde{U}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}). \tag{S4.4}$$

In the last step, we prove the above inequality (S4.4). By standard matrix theory, if matrices M and N are both positive definite and $M \geq N$, then the corresponding Schur complements satisfy the same inequality. Note that both Ω/Ω_2 and $\check{U}$ are positive definite, so to show (S4.4), it is enough to show that

$$\Omega/\Omega_2 \geq \rho\check{U}.$$

Note that parameter $\boldsymbol{\phi}^\intercal = (\boldsymbol{\beta}_1^\intercal, \ldots, \boldsymbol{\beta}_r^\intercal, \alpha_1, \ldots, \alpha_r, \boldsymbol{\beta}_{r+1}^\intercal, \ldots, \boldsymbol{\beta}_m^\intercal, \alpha_{r+1}$ $, \ldots, \alpha_m)$ is just permuted $(\boldsymbol{\theta}_1^\intercal, \ldots, \boldsymbol{\theta}_m^\intercal)$, so the conculsion of Lemma 4 (i) also applies to the information matrix with respect to $\boldsymbol{\phi}$. The information matrix with respect to $\boldsymbol{\phi}$ for $R_n^{(2)}$ is just Ω, and its lower-right $(m-r)(d+1) \times (m-r)(d+1)$ block is Ω_2. For $R_n^{(1)}$, the infromation matrix is just $\check{U}$. Thus by Lemma 4 (i), we have $\Omega/\Omega_2 \geq \rho\check{U}$. The proof is complete. □

Proof of Lemma 5. Notice that

$$\begin{pmatrix} A+B & X+Y \\ (X+Y)^\intercal & X^\intercal A^{-1}X + Y^\intercal B^\dagger Y \end{pmatrix} = \begin{pmatrix} A & X \\ X^\intercal & X^\intercal A^{-1}X \end{pmatrix} + \begin{pmatrix} B & Y \\ Y^\intercal & Y^\intercal B^\dagger Y \end{pmatrix}.$$

The first matrix on the RHS is positive semidefinite by Theorem 1.12 of Zhang (2005), and since Y is in the column space of B, the second matrix on the RHS is also positive semidefinite by Theorem 1.20 of Zhang (2005). Therefore the matrix on the left hand side (LHS) is positive semidefinite. Also note that $A+B$ is positive definite. Hence the Schur complement of the LHS with respect to its upper-left block $A+B$,

$$X^\intercal A^{-1}X + Y^\intercal B^\dagger Y - (X+Y)^\intercal (A+B)^{-1}(X+Y),$$

must also be positive semidefinite. The claimed result then follows. □

References

Chen, J. and Y. Liu (2013). Quantile and quantile-function estimations under density ratio model. *The Annals of Statistics* **41**, 1669–1692.

Harville, D. A. (2008). *Matrix Algebra From a Statistician's Perspective.* Springer, New York.

Mathai, A. M. (1992). *Quadratic Forms in Random Variables: Theory and Applications.* Marcel Dekker, New York.

van der Vaart, A. W. (2000). *Asymptotic Statistics.* Cambridge University Press, Cambridge.

Zhang, B. (2002). Assessing goodness-of-fit of generalized logit models based on case-control data. *Journal of Multivariate Analysis* **82**, 17–38.

Zhang, F. (Ed.) (2005). *The Schur complement and its applications*. Springer, New York.

Zorich, V. A. (2004). *Mathematical Analysis II*. Springer-Verlag, Berlin.