Proofs of theorems for the JRSS-B paper "Likelihood ratio tests in linear mixed models with one variance component".

Proof of Theorem 1. We partition first the vector of fixed effect parameters $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T|\boldsymbol{\beta}_2^T)^T$, where $\boldsymbol{\beta}_2 = \boldsymbol{\beta}_2^0$ are the known fixed effect parameters under the null hypothesis. We can also partition the matrix of fixed effects $\boldsymbol{X} = (\boldsymbol{X}_1|\boldsymbol{X}_2)$ corresponding to the partition of $\boldsymbol{\beta}$. Using the notations in the paper, the LRT statistic can be written as

$$\mathrm{LRT}_n = \sup_{\lambda \geq 0} \left\{ n \log \left(\boldsymbol{Y}^T P_0 \boldsymbol{Y} \right) - n \log (\boldsymbol{Y}^T P_\lambda^T V_\lambda^{-1} P_\lambda \boldsymbol{Y}) - \log |V_\lambda| \right\} + n \log \left(\frac{\boldsymbol{Y}^T S_1 \boldsymbol{Y}}{\boldsymbol{Y}^T P_0 \boldsymbol{Y}} \right) . \quad (1)$$

The first part of the right hand side of equation (1) corresponds to testing for the zero variance of random effects while the second part corresponds to testing for the fixed effects. One can easily show that $\log |V_\lambda| = \sum_{s=1}^K \log(1 + \lambda \xi_{s,n})$. Also, from Kuo, 1999 and Patterson and Thompson, 1971 there exists an $n \times (n-p)$ matrix W such that $W^T W = I_{n-p}$, $WW^T = P_0$, $W^T V_\lambda W = \mathrm{diag}\left\{(1 + \lambda \mu_{s,n})\right\}$, and

$$\boldsymbol{Y}^T P_\lambda^T V_\lambda^{-1} P_\lambda \boldsymbol{Y} = \boldsymbol{Y}^T W \mathrm{diag}\left\{ (1 + \lambda \mu_{s,n})^{-1} \right\} W^T \boldsymbol{Y} .$$

Denote by $\boldsymbol{w} = W^T \boldsymbol{Y} / \sigma_\epsilon$ and note that under the null hypothesis

$$E[\boldsymbol{w}] = (W^T X_1 \boldsymbol{\beta}_1 + W^T X_2 \boldsymbol{\beta}_2^0)/\sigma_\epsilon, \;\; \mathrm{Cov}[\boldsymbol{w}] = I_{n-p} .$$

We now show that $E[\boldsymbol{w}] = 0$. Denote by $A = W^T X$ and observe that $WA = P_0 X = 0$, and that $W^T W A = 0$. This shows that $A = 0$, that $W^T X_1 = 0$, and that $W^T X_2 = 0$. It now follows that $\boldsymbol{w} = (w_1, ..., w_{n-p})$ is an $n - p$ dimensional random vector with i.i.d. N(0,1) components. Putting all these together it follows that

$$L^{K,n}(\lambda) = -n \log \left\{ \sigma_\epsilon^2 \left(\sum_{s=1}^K \frac{w_s^2}{1 + \lambda \mu_{s,n}} + \sum_{s=K+1}^{n-p} w_s^2 \right) \right\} - \sum_{s=1}^K \log(1 + \lambda \xi_{s,n}) ,$$

where we used the fact that at most K eigenvalues $\mu_{s,n}$ and $\xi_{s,n}$ are not zero. In particular

$$L^{K,n}(0) = -n \log \left\{ \sigma_\epsilon^2 \left(\sum_{s=1}^{n-p} w_s^2 \right) \right\} ,$$

which is a standard result in regression analysis. Therefore we can write

$$L^{K,n}(\lambda) - L^{K,n}(0) = n \log \left\{ 1 + U_n(\lambda) \right\} ,$$

where $U_n(\lambda) = N_n(\lambda) / D_n(\lambda)$ and

$$N_n(\lambda) = \sum_{s=1}^K \frac{\lambda \mu_{s,n}}{1 + \lambda \mu_{s,n}} w_s^2, \;\; D_n(\lambda) = \sum_{s=1}^K \frac{w_s^2}{1 + \lambda \mu_{s,n}} + \sum_{s=K+1}^{n-p} w_s^2 .$$

We now focus on the second term in equation (1). Denote by $S_{X_1} = X_1 (X_1^T X_1)^{-1} X_1^T$ and by $S_X = X(X^T X)^{-1} X^T$. It is standard to show that

$$n \log \left(\frac{\boldsymbol{Y}^T S_1 \boldsymbol{Y}}{\boldsymbol{Y}^T P_0 \boldsymbol{Y}} \right) = n \log \left\{ \frac{(\boldsymbol{Y} - X_2 \boldsymbol{\beta}_2^0)^T (I_n - S_{X_1})(\boldsymbol{Y} - X_2 \boldsymbol{\beta}_2^0)}{\boldsymbol{Y}^T (I_n - S_X) \boldsymbol{Y}} \right\} .$$

Observe that $S_X X = X$, $S_X X_1 = X_1$, $S_X X_2 = X_2$, and $(I_n - S_X)X_2 = 0$. Hence $\boldsymbol{Y}^T(I_n - S_X)\boldsymbol{Y} = (\boldsymbol{Y} - X_2\boldsymbol{\beta}_2^0)^T(I_n - S_X)(\boldsymbol{Y} - X_2\boldsymbol{\beta}_2^0)$. Denoting by $\boldsymbol{V} = (\boldsymbol{Y} - X_2\boldsymbol{\beta}_2^0)/\sigma_\epsilon$ one obtains

$$n \log\left(\frac{\boldsymbol{Y}^T S_1 \boldsymbol{Y}}{\boldsymbol{Y}^T P_0 \boldsymbol{Y}}\right) = n \log\left\{1 + \frac{\boldsymbol{V}^T(S_X - S_{X_1})\boldsymbol{V}}{\boldsymbol{V}^T(I_n - S_X)\boldsymbol{V}}\right\} .$$

If S is an $n \times n$ idempotent, symmetric matrix of rank t, there exists an $n \times t$ matrix A so that $AA^T = S$ and $A^T A = I_t$. For the projection matrix $P_0 = I_n - S_X$ this matrix was denoted by $\boldsymbol{W}$. For the projection matrix $S_X - S_{X_1}$ of rank q let U be an $n \times q$ matrix so that $UU^T = S_X - S_{X_1}$ and $U^T U = I_q$. Because $W^T X_2 = 0$ it follows that $\boldsymbol{w} = W^T\boldsymbol{V}$. Define now $\boldsymbol{u} = U^T\boldsymbol{V}$ and note that under the null

$$E[\boldsymbol{u}] = \frac{U^T X_1 \boldsymbol{\beta}_1}{\sigma_\epsilon}, \;\; \text{Cov}[\boldsymbol{u}] = I_q .$$

Denoting by $B = U^T X_1$ it follows that $UB = (S_X - S_{X_1})X_1 = 0$. Hence $U^T UB = 0$ showing that $B = 0$ and $E[\boldsymbol{u}] = 0$. Also, note that $\text{Cov}(\boldsymbol{u}, \boldsymbol{w}) = U^T W$. If $C = U^T W$ then $UCW^T = (S_X - S_{X_1})P_0 = 0$. Therefore $U^T UCW^T W = 0$ or $C = 0$. Because the vector $(\boldsymbol{u}^T, \boldsymbol{w}^T)$ has a normal distribution, it follows that all entries are i.i.d. N(0,1) random variables. Denote $\boldsymbol{u} = (u_1, ..., u_q)^T$. We can now write

$$n \log\left(\frac{\boldsymbol{Y}^T S_1 \boldsymbol{Y}}{\boldsymbol{Y}^T P_0 \boldsymbol{Y}}\right) = n \log\left\{1 + \frac{\sum_{s=1}^q u_s^2}{\sum_{s=1}^{n-p} w_s^2}\right\} .$$

Proof of Theorem 2. We continue to use notations from the proof of theorem 1. For $R(x) = \log(1+x) - x$, $\lim_{x\to 0} R(x)/x = 0$ and $\lim_{x\to 0} R(x)/x^2 = -1/2$. Using the Taylor expansion around 0, $\log(1+x) = x + R(x)$ and taking into account that $\sum_{s=K+1}^{n-p} w_s^2/n$ converges almost surely to 1 one obtains

$$n \log\left\{1 + \frac{\sum_{s=1}^q u_s^2}{\sum_{s=1}^{n-p} w_s^2}\right\} = \sum_{s=1}^q u_s^2 + V_n ,$$

where V_n converges almost surely to 0. Denoting by $W_q = \sum_{s=1}^q u_s^2$ one obtains $\text{LRT}_n = \sup_{\lambda \geq 0} \text{LRT}_n(\lambda)$ where

$$\text{LRT}_n(\lambda) = n \log\{1 + U_n(\lambda)\} - \sum_{s=1}^K \log(1 + \lambda\xi_{s,n}) + W_q + V_n ,$$

where $U_n(\lambda)$ is independent of W_q, and V_n converges almost surely to 0. Denote now by $f_n(d) = n \log\{1 + U_n(n^{-\alpha}d)\} - \sum_{s=1}^K \log(1 + dn^{-\alpha}\xi_{s,n}) + W_q$ and we will show that

$$\sup_{d\geq 0} f_n(d) \Rightarrow \sup_{d\geq 0} \text{LRT}_\infty(d) + W_q .$$

This proof consists of two steps

1 Prove that $f_n(\cdot)$ converges weakly to $\text{LRT}_\infty(\cdot) + W_q$ on the space $C[0,\infty)$ of continuous functions with support $[0,\infty)$.

2 Prove that a Continuous Mapping Theorem type result holds for the $\sup_{d\geq 0} f_n(d)$.

We show the weak convergence for any $\mathcal{C}[0,M]$. Denote $f(d) = \mathrm{LRT}_\infty(d) + W_q$, $\eta_{s,n} = n^{-\alpha}\mu_{s,n}$, $\zeta_{s,n} = n^{-\alpha}\xi_{s,n}$. Note that $\lim_{n\to\infty}\eta_{s,n} = \mu_s$ and $\lim_{n\to\infty}\zeta_{s,n} = \xi_s$. We first establish the finite dimensional convergence of $f_n(d)$ to $f(d)$ and then we prove that $f_n(d)$ is a tight sequence in $\mathcal{C}[0,M]$.

To show finite dimensional convergence it is sufficient to show that for a fixed d the convergence is almost sure. Note that

$$n\log\left\{1 + U_n(n^{-\alpha}d)\right\} = nU_n(n^{-\alpha}d) + nR(n^{-\alpha}d)\ ,$$

where $R(\cdot)$, $U_n(\cdot)$, $N_n(\cdot)$ and $D_n(\cdot)$ were defined earlier. It follows immediately that almost surely

$$\lim_{n\to\infty} nU_n(n^{-\alpha}d) = \sum_{s=1}^{K}\frac{d\mu_s}{1+d\mu_s}w_s^2\ ,$$

Because $nR\{U_n(n^{-\alpha}d)\} = \{nU_n(n^{-\alpha}d)\}\left\{R(U_n(n^{-\alpha}d))/U_n(n^{-\alpha}d)\right\}$, it follows that $nR(U_n(n^{-\alpha}d))$ converges to zero almost surely ($\lim_{x\to} R(x)/x = 0$). Note that $\lim_{n\to\infty}\sum_{s=1}^{K}\log(1+d\zeta_{s,n}) = \sum_{s=1}^{K}\log(1+d\xi_s)$ for every fixed d. We proved that, for every fixed d, $f_n(d)$ converges almost surely to $\mathrm{LRT}_\infty(d) + W_q$.

To show that $f_n(d)$ form a tight sequence it is sufficient to show that for every ϵ and η strictly positive, there exist $\delta = \delta(\epsilon,\eta)$, $0<\delta<1$ and $n_0 = n_0(\epsilon,\delta)$ such that for $n \geq n_0$

$$\frac{1}{\delta}P\left\{\sup_{t\leq u\leq t+\delta}|f_n(u) - f_n(t)| \geq \epsilon\right\} \leq \eta\ .$$

Observe first that

$$|f_n(u) - f_n(t)| \leq n\log\left\{\frac{D_n(n^{-\alpha}t)}{D_n(n^{-\alpha}u)}\right\} + \sum_{s=1}^{K}\log\frac{1+u\zeta_{s,n}}{1+t\zeta_{s,n}}\ ,$$

and because $\log(1+x) < x$ for every $x>0$ we obtain the following inequalities

$$\log\left\{\frac{D_n(n^{-\alpha}t)}{D_n(n^{-\alpha}u)}\right\} \leq \frac{D_n(n^{-\alpha}t) - D_n(n^{-\alpha}u)}{D_n(n^{-\alpha}u)} \leq (s-t)C\frac{\sum_{s=1}^{K}w_s^2}{\sum_{s=K+1}^{n-p}w_s^2}\ ,$$

where $C>0$ is a constant so that $n\zeta_{s,n}/(n-p-K) \leq C$ for every s and n. It follows that the following inequality holds

$$n\log\left\{\frac{D_n(n^{-\alpha}t)}{D_n(n^{-\alpha}u)}\right\} \leq (u-t)CKF_{K,n}\ ,$$

where $F_{K,n}$ is a random variable with an F distribution with $(K, n-p-K)$ degrees of freedom. Similarly

$$\sum_{s=1}^{K}\log\left(\frac{1+u\zeta_{s,n}}{1+t\zeta_{s,n}}\right) \leq (u-t)CK\ .$$

We conclude that

$$P\left\{\sup_{t\leq u\leq t+\delta}|f_n(u) - f_n(t)| \geq \epsilon\right\} \leq P\left\{F_{K,n} \geq \frac{\epsilon}{CK\delta} - 1\right\}\ ,$$

and it is sufficient to find δ, n_0 so that for every $n \geq n_0$ the c.d.f. $H_{K,n}$ of $F_{K,n}$ satisfies

$$1 - H_{K,n}\left(\frac{\epsilon}{CK\delta} - 1\right) \leq \eta\delta \, . \tag{2}$$

If H_K is the c.d.f. of a χ^2 distribution with K degrees of freedom then, for every x, $\lim_{n\to\infty} H_{K,n}(x) = H_K(Kx)$. Because (using for example l'Hospital rule and the pdf of a χ^2 distribution with K degrees of freedom)

$$\lim_{\delta \downarrow 0} \left\{1 - H_K\left(\frac{\epsilon}{C\delta} - K\right)\right\} \Big/ \left\{\frac{\eta\delta}{2}\right\} = 0 \, ,$$

one can find $\delta = \delta(\epsilon, \eta)$, $\delta < 1$, with $\frac{\epsilon}{C\delta} - K > 0$ so that $1 - H_K\left(\frac{\epsilon}{C\delta} - K\right) \leq \frac{\eta\delta}{2}$. Because of the convergence of $H_{K,n}$ to H_K, one can find $n_0 = n_0(\epsilon, \eta)$ so that for $n \geq n_0$ the following inequality holds

$$\left| H_{K,n}\left(\frac{\epsilon}{C\delta} - K\right) - H_K\left(\frac{\epsilon}{C\delta} - K\right) \right| \leq \frac{\eta\delta}{2} \, ,$$

which finishes the proof of the inequality in equation (2). We conclude that $f_n(d)$ converges weakly to $f(d)$ on $\mathcal{C}[0, M]$ for each M, and therefore on $\mathcal{C}[0, \infty)$.

We want to show now that $\sup_{d \geq 0} f_n(d) \Rightarrow \sup_{d \geq 0} f(d)$. First we find a random variable $T_{K,n}$ so that

$$\sup_{d \geq 0} f_n(d) = \max_{d \in [0, T_{K,n}]} f_n(d) \, .$$

Note first that $f_n(0) = W_q$ for every n. Also, using again the inequality $\log(1 + x) \leq x$ for $x \geq 0$ it is easy to prove that

$$f_n(d) \leq n \frac{\sum_{s=1}^{K} w_s^2}{\sum_{s=K+1}^{n-p} w_s^2} - K \log(1 + dm) + W_q \, .$$

where $m > 0$ is chosen so that $\zeta_{s,n} \geq m$ for all s and n. Hence

$$f_n(d) \leq \frac{nK}{n - p - K} F_{K,n} - K \log(1 + dm) + W_q \, .$$

Denote by

$$T_{K,n} = \frac{1}{m}\left\{\exp\left(\frac{n}{n - p - K} F_{K,n}\right) - 1\right\}$$

and observe that for $d > T_{K,n}$ we have $f_n(d) < W_q$ which shows that $T_{K,n}$ has the desired property. Observe now that for every fixed $M > 0$ and for every $t \geq 0$

$$\text{pr}\left\{\sup_{d \geq 0} f_n(d) \leq t\right\} \leq \text{pr}\left\{\max_{d \in [0, M]} f_n(d) \leq t\right\} \, .$$

Taking $\limsup$ for $n \to \infty$ one obtains

$$\limsup_{n\to\infty} \text{ pr}\left\{\sup_{d \geq 0} f_n(d) \leq t\right\} \leq \limsup_{n\to\infty} \text{ pr}\left\{\max_{d \in [0, M]} f_n(d) \leq t\right\} \, .$$

Because $f_n(d) \Rightarrow f(d)$ on $C[0, M]$ and max is a continuous function on $C[0, M]$ one can apply the Continuous Mapping Theorem for the right hand side of the inequality and we obtain

$$\lim_{n\to\infty} \text{ pr}\left\{\max_{d \in [0, M]} f_n(d) \leq t\right\} = \text{pr}\left\{\max_{d \in [0, M]} f(d) \leq t\right\} \, . \tag{3}$$

Using that for any two events A and B, $P(A \cap B) \geq P(A) - P(B^C)$ we obtain the following sequence of relations

$$\begin{aligned}
\text{pr}\left\{\sup_{d\geq 0} f_n(d) \leq t\right\} &\geq \text{pr}\left\{\sup_{d\geq 0} f_n(d) \leq t, T_{K,n} < M\right\} = \\
&= \text{pr}\left\{\max_{d\in[0,M]} f_n(d) \leq t, T_{K,n} < M\right\} \\
&\geq \text{pr}\left\{\max_{d\in[0,M]} f_n(d) \leq t\right\} - \text{pr}\left(T_{K,n} \geq M\right)
\end{aligned}$$

Taking the $\liminf$ when $n \to \infty$ in the first and last expressions we obtain

$$\liminf_{n\to\infty} \text{ pr}\left\{\sup_{d\geq 0} f_n(d) \leq t\right\} \geq \text{pr}\left\{\sup_{d\in[0,M]} f(d) \leq t\right\} - \text{pr}(T_K \geq M) ,$$

where we used equation (3) and $T_K = \left\{\exp\left(\sum_{s=1}^K w_s^2/K\right) - 1\right\}\Big/ m$. Consider now a sequence of integers $M \to \infty$. Then $\lim_{M\to\infty} \text{pr}(T_K \geq M) = 0$. Therefore if we prove that

$$\lim_{M\to\infty} \text{ pr}\left\{\max_{d\in[0,M]} f(d) \leq t\right\} = \text{pr}\left\{\sup_{d\geq 0} f(d) \leq t\right\} , \tag{4}$$

then it follows that $\lim_{n\to\infty} \text{pr}\left\{\sup_{d\geq 0} f_n(d) \leq t\right\}$ exists and

$$\lim_{n\to\infty} \text{pr}\left\{\sup_{d\geq 0} f_n(d) \leq t\right\} = \text{pr}\left\{\sup_{d\geq 0} f(d) \leq t\right\} ,$$

proving that

$$\sup_{d\geq 0} f_n(d) \Rightarrow \sup_{d\geq 0} f(d) .$$

Denote by $A_M = \left\{\max_{t\in[0,M]} f(d) \leq t\right\}$. Then $A_M \supset A_{M+1}$ and $\lim_{M\to\infty} \text{pr}(A_M) = \text{pr}\left(\bigcap_{M\geq 1} A_M\right)$. But $\bigcap_{M\geq 1} A_M = \left\{\sup_{t\geq 0} f(d) \leq t\right\}$ which ends the proof of equation (4).

Observe now that

$$\text{LRT}_n = \sup_{\lambda\geq 0} \text{LRT}_n(\lambda) = \sup_{d\geq 0} f_n(d) + V_n .$$

Because V_n converges almost surely to 0 it follows that

$$\text{LRT}_n \Rightarrow \sup_{d\geq 0} f(d) = \sup_{d\geq 0} \text{LRT}_\infty(d) + W_q .$$

To end the proof one only needs to show that $\sup_{d\geq 0} \text{LRT}_\infty(d)$ and W_q are independent. But, for any fixed $d \geq 0$, $\text{LRT}_\infty(d)$ is independent of W_q. Because $\text{LRT}_\infty(d)$ is continuous in d then

$$\sup_{d\geq 0} \text{LRT}_\infty(d) = \sup_{d\in Q\cap[0,\infty)} \text{LRT}_\infty(d) = \sup_{i\geq 1} \text{LRT}_\infty(d_i) ,$$

where $(d_i)_{i\geq 1}$ is an enumeration of $Q \cap [0, \infty)$. Let x, $w \geq 0$, $A_M = \bigcap_{i=1}^M \{\text{LRT}_\infty(d_i) \leq x\} \bigcap (W_q < w)$. Note that $A_M \supset A_{M+1}$ and if $A_\infty = \bigcap_{M=1}^\infty A_M$ then $\text{pr}(A_\infty) = \lim_{M\to\infty} \text{pr}(A_M)$. But

$$A_\infty = \left\{\sup_{i\geq 1} \text{LRT}_\infty(d_i) \leq x\right\} \bigcap (W_q < w) ,$$

and $\text{pr}(A_M) = \text{pr}\left[\bigcap_{i=1}^{M}\{\text{LRT}_\infty(d_i) \le x\}\right]\text{pr}(W_q \le w)$ due to the independence of W_q and the vector $\{\text{LRT}_\infty(d_1), \ldots, \text{LRT}_\infty(d_M)\}$ for every M. Using a similar procedure it is easy to show that

$$\lim_{M\to\infty} \text{pr}\left[\bigcap_{i=1}^{M}\{\text{LRT}_\infty(d_i) \le x\}\right] = \text{pr}\left\{\sup_{i\ge 1}\text{LRT}_\infty(d_i) \le x\right\} .$$

Therefore we proved that for every x, $w \ge 0$

$$\text{pr}\left[\left\{\sup_{d\ge 0} LRT_\infty(d) \le x\right\}\bigcap(W_q < w)\right] = \text{pr}\left\{\sup_{d\ge 0}\text{LRT}_\infty(d) \le x\right\}\text{pr}(W_q < w) ,$$

showing that $\sup_{d\ge 0}\text{LRT}_\infty(d)$ and W_q are independent.

Proof of Theorem 3. Suppose that $\lambda = \lambda_0$ is the true value of the ratio $\sigma_b^2/\sigma_\epsilon^2$. The restricted profile log-likelihood function is

$$2l^{K,n}(\lambda) = -\log|V_\lambda| - \log|X^T V_\lambda^{-1} X| - (n-p)\log(Y^T P_\lambda^T V_\lambda^{-1} P_\lambda Y) ,$$

where $P_\lambda = I_n - X(X^T V_\lambda^{-1} X)^{-1} X^T V_\lambda^{-1}$. From Kuo (1999) and Patterson and Thompson (1971) there exists an $n \times (n-p)$ matrix W such that $W^T W = I_{n-p}$, $WW^T = P_0$, $W^T V_\lambda W = \text{diag}\{(1+\lambda\tau_{s,n})\}$, and

$$Y^T P_\lambda^T V_\lambda^{-1} P_\lambda Y = Y^T W \text{diag}\left\{(1+\lambda\tau_{s,n})^{-1}\right\} W^T Y , \tag{5}$$

where $\tau_{s,n}$, $s = 1, \ldots, n-p$ are the eigenvalues of $W^T Z\Sigma Z^T W$. If

$$w = \text{diag}\left\{(1+\lambda_0\tau_{s,n})^{-1/2}\right\} W^T Y/\sigma_\epsilon$$

then w is a normal vector with mean zero and covariance matrix I_{n-p}. Indeed, if $A = W^T X$ then $WA = P_0 X = 0$ and $W^T W A = A = 0$. Because $E(W^T Y) = W^T X\beta = 0$ it follows that $E(w) = 0$. Moreover,

$$\text{Cov}(w) = \text{diag}\left\{(1+\lambda_0\tau_{s,n})^{-1/2}\right\}\left(W^T V_{\lambda_0} W\right)\text{diag}\left\{(1+\lambda_0\tau_{s,n})^{-1/2}\right\} = I_{n-p} ,$$

since $W^T V_{\lambda_0} W = \text{diag}\{(1+\lambda_0\tau_{s,n})\}$. Because Y is a normal vector, the entries w_i of the vector w are i.i.d. $N(0,1)$ random variables. Replacing these results back in equation (5) we obtain

$$Y^T P_\lambda^T V_\lambda^{-1} P_\lambda Y = \sigma_\epsilon^2 w^T \text{diag}\left\{\left(\frac{1+\lambda_0\tau_{s,n}}{1+\lambda\tau_{s,n}}\right)\right\} w .$$

Note that at most K eigenvalues of $W^T Z\Sigma Z^T W$ are non-zero and these eigenvalues are, in fact, the eigenvalues of $\Sigma^{1/2} Z^T P_0 Z \Sigma^{1/2}$. Therefore $\tau_{s,n} = \mu_{s,n}$ for $s = 1, \ldots, K$ and $\tau_{s,n} = 0$ for $s = K+1, \ldots, n-p$ showing that

$$Y^T P_\lambda^T V_\lambda^{-1} P_\lambda Y = \sigma_\epsilon^2\left(\sum_{s=1}^{K}\frac{1+\lambda_0\mu_{s,n}}{1+\lambda\mu_{s,n}} w_s^2 + \sum_{s=K+1}^{n-p} w_s^2\right) ,$$

where w_s are i.i.d. $N(0,1)$ random variables. Using a result from Kuo, 1999 it follows immediately that

$$\log|V_\lambda| + \log|X^T V_\lambda^{-1} X| = \log(|\boldsymbol{X}^T \boldsymbol{X}|) + \sum_{s=1}^{K} \log(1 + \lambda \mu_{s,n}) \, .$$

The result of the theorem now follows.