Online Appendix to: Mallows criterion for heteroskedastic linear regressions with many regressors

Stanislav Anatolyev
CERGE-EI and NES

Proof of Theorem 1

First we will show that $\left|GC_p(q) - GC_p^0(q)\right| = o_P(n)$ uniformly in q. For brevity, we will denote $P(q)$ by P, and $M(\bar{q})$ by M. We will heavily use the following properties of projection matrices: $P_{ii} \leq 1$, $M_{ii} \leq 1$ and $\sum_{j=1}^n M_{ij}M_{jk} = M_{ik}$ for any $i, k \in \{1, ..., n\}$, and $\sum_{i=1}^n \sum_{j=1}^n |M_{ij}| \leq \sum_{i=1}^n n^{1/2} \big(\sum_{j=1}^n M_{ij}^2 \big)^{1/2} = n^{1/2} \sum_{i=1}^n M_{ii}^{1/2} \leq n^{3/2}$.

We need to bound

$$\sum_{i=1}^n P_{ii}\hat{\sigma}_i^2 - \sum_{i=1}^n P_{ii}\sigma_i^2 = a_{1,n} + a_{2,n} + a_{3,n} + a_{4,n} + a_{5,n} + a_{6,n} + a_{7,n},$$

where

$$\begin{aligned}
a_{1,n} &= \sum_{i=1}^n P_{ii}\left(e_i^2 - \sigma_i^2\right), \ a_{2,n} = \sum_{i=1}^n \sum_{j \neq i} \frac{P_{ii}}{M_{ii}} M_{ij} e_i e_j, \\
a_{3,n} &= \sum_{i=1}^n \frac{P_{ii}}{M_{ii}} \sum_{j=1}^n M_{ij} g_j e_i, \ a_{4,n} = \sum_{i=1}^n \frac{P_{ii}}{M_{ii}} (g_i - \bar{g}_n) \sum_{j=1}^n M_{ij} e_j, \\
a_{5,n} &= -\bar{e}_n \sum_{i=1}^n \frac{P_{ii}}{M_{ii}} \sum_{j=1}^n M_{ij} g_j, \ a_{6,n} = -\bar{e}_n \sum_{i=1}^n \frac{P_{ii}}{M_{ii}} \sum_{j=1}^n M_{ij} e_j, \\
a_{7,n} &= \sum_{i=1}^n \frac{P_{ii}}{M_{ii}} (g_i - \bar{g}_n) \sum_{j=1}^n M_{ij} g_i,
\end{aligned}$$

in case if estimates (6) are used, and some of the terms are missing if estimates (5) are used. All $a_{1,n}$, $a_{2,n}$, $a_{3,n}$, $a_{4,n}$ and $a_{5,n}$ have zero mean and variances

$$
\begin{aligned}
var\left(a_{1,n}\right) &= \sum_{i=1}^{n} P_{ii}^{2} var\left(e_i^2-\sigma_i^2\right) \le \sum_{i=1}^{n} P_{ii}^{2} E\left[e_i^4\right] = O_P\left(n\right), \\
var\left(a_{2,n}\right) &= 2\sum_{i=1}^{n}\sum_{j\neq i}\frac{P_{ii}^{2}M_{ij}^{2}}{M_{ii}^{2}}\sigma_i^2\sigma_j^2 \le 2\underline{C}_M^{-2}\overline{C}_\sigma^2\sum_{i=1}^{n}\sum_{j\neq i}M_{ij}^2 \le O_P\left(n\right), \\
var\left(a_{3,n}\right) &= \sum_{i=1}^{n}\frac{P_{ii}^{2}}{M_{ii}^{2}}\left(Mg\right)_i^2\sigma_i^2 \le \underline{C}_M^{-2}\overline{C}_\sigma\sum_{i=1}^{n}\left(Mg\right)_i^2 = \underline{C}_M^{-2}\overline{C}_\sigma\overline{C}_{Mg,n} \le o_P\left(n\right),
\end{aligned}
$$

$$
\begin{aligned}
var\left(a_{4,n}\right) &= \sum_{j=1}^{n}\left(\sum_{i=1}^{n}\frac{P_{ii}}{M_{ii}}M_{ij}\left(g_i-\bar{g}_n\right)\right)^2\sigma_j^2 \le \overline{C}_\sigma\sum_{j=1}^{n}\left(\sum_{i=1}^{n}\frac{P_{ii}}{M_{ii}}M_{ij}\left(g_i-\bar{g}_n\right)\right)^2 \\
&= \overline{C}_\sigma\sum_{i=1}^{n}\frac{P_{ii}}{M_{ii}}\left(g_i-\bar{g}_n\right)\sum_{k=1}^{n}\frac{P_{kk}}{M_{kk}}\left(g_k-\bar{g}_n\right)\sum_{j=1}^{n}M_{ij}M_{kj} \\
&\le \overline{C}_\sigma\underline{C}_M^{-2}\sum_{i=1}^{n}\sum_{k=1}^{n}\left|M_{ik}\left(g_i-\bar{g}_n\right)\left(g_k-\bar{g}_n\right)\right| \\
&\le \overline{C}_\sigma\underline{C}_M^{-2}\Big(\sum_{i=1}^{n}\sum_{k=1}^{n}M_{ik}^2\Big)^{1/2}\left(\sum_{i=1}^{n}\sum_{k=1}^{n}\left|\left(g_i-\bar{g}_n\right)\left(g_k-\bar{g}_n\right)\right|^2\right)^{1/2} \\
&= \overline{C}_\sigma\underline{C}_M^{-2}\Big(\sum_{i=1}^{n}M_{ii}\Big)^{1/2}\sum_{i=1}^{n}\left(g_i-\bar{g}_n\right)^2 \le 2\overline{C}_\sigma\underline{C}_M^{-2}n^{1/2}\left(g'g\right) = O_P\left(n^{3/2}\right),
\end{aligned}
$$

and

$$
\begin{aligned}
var\left(a_{5,n}\right) &= \frac{\sum_{i=1}^{n}\sigma_i^2}{n^2}\left(\sum_{i=1}^{n}\frac{P_{ii}}{M_{ii}}\sum_{j=1}^{n}M_{ij}g_j\right)^2 \le \frac{\overline{C}_\sigma}{n}\left(\sum_{i=1}^{n}\frac{P_{ii}}{M_{ii}}\left(Mg\right)_i\right)^2 \\
&\le n^{-1}\overline{C}_\sigma\underline{C}_M^{-2}n\sum_{i=1}^{n}\left(Mg\right)_i^2 = \overline{C}_\sigma\underline{C}_M^{-2}\overline{C}_{Mg,n} \le O_P\left(n^{\delta_{Mg}}\right).
\end{aligned}
$$

Next, the first term in $a_{6,n}$ is $-\bar{e}_n = O_P\left(n^{-1/2}\right)$, and the second term has mean zero and variance

$$
\begin{aligned}
var\left(\sum_{i=1}^{n}\frac{P_{ii}}{M_{ii}}\sum_{j=1}^{n}M_{ij}e_j\right) &= \sum_{j=1}^{n}\left(\sum_{i=1}^{n}\frac{P_{ii}}{M_{ii}}M_{ij}\right)^2\sigma_j^2 \le \overline{C}_\sigma\sum_{i=1}^{n}\frac{P_{ii}}{M_{ii}}\sum_{k=1}^{n}\frac{P_{kk}}{M_{kk}}\sum_{j=1}^{n}M_{ij}M_{jk} \\
&\le \overline{C}_\sigma\underline{C}_M^{-2}\sum_{i=1}^{n}\sum_{k=1}^{n}\left|M_{ik}\right| \le \overline{C}_\sigma\underline{C}_M^{-2}n^{3/2} = O_P\left(n^{3/2}\right),
\end{aligned}
$$

so $a_{6,n} = O_P\left(n^{1/4}\right)$. Finally, the omitted variables bias term is bounded:

$$\begin{aligned} |a_{7,n}| &= \left|\sum_{i=1}^{n} M_{ii}^{-1} P_{ii}\left(g_i - \bar{g}_n\right)\left(Mg\right)_i\right| \le \left(\sum_{i=1}^{n} M_{ii}^{-2} P_{ii}^2 \left(\left(g_i - \bar{g}_n\right)\right)^2\right)^{1/2} \left(\sum_{i=1}^{n} \left(Mg\right)_i^2\right)^{1/2} \\ &\le \underline{C}_M^{-1}\left(g'g\right)^{1/2}\left(g'Mg\right)^{1/2} \le O_P\left(n^{\delta_{Mg}/2+1/2}\right). \end{aligned}$$

In total,

$$\begin{aligned} \left|GC_p\left(q\right) - GC_p^0\left(q\right)\right| &= 2\left|\sum_{i=1}^{n} P_{ii}\left(q\right)\hat{\sigma}_i^2 - \sum_{i=1}^{n} P_{ii}\left(q\right)\sigma_i^2\right| \\ &\le O_P\left(n^{1/2} + n^{3/4} + n^{\delta_{Mg}/2} + n^{1/4} + n^{\delta_{Mg}/2+1/2}\right) \\ &= O_P\left(n^{\max\{3/4,\, \delta_{Mg}/2+1/2\}}\right) = o_P\left(n\right) \end{aligned}$$

uniformly in $q \in Q$. The objective functions for selection of $\hat{q}$ and $\hat{q}^0 \equiv \arg\min_{q\in Q} GC_p^0\left(q\right)$ differ by no more than $o_P\left(n\right)$, while the objective function itself is of order n. Hence, $R\left(\hat{q}\right) - R\left(\hat{q}^0\right) = o_P(1)\, R\left(\hat{q}^0\right) \le o_P(1)\left(g'g + \max_{q\in Q} \dim(z_i(q))\right) = o_P(n)$. Now, because $\min_{q\in Q} R\left(q\right) \ge \min_{q\in Q}\left\{g'M\left(q\right)g + \underline{C}_\sigma \dim(z_i(q))\right\} > \min\{1, \underline{C}_\sigma\}\,\underline{C}_R n$, we have

$$\frac{R\left(\hat{q}\right) - R\left(\hat{q}^0\right)}{\min_{q\in Q} R\left(q\right)} < \frac{o_P(n)}{\min\{1, \underline{C}_\sigma\}\,\underline{C}_R n} = o_P\left(1\right).$$

By Theorem 2.1* of Andrews (1991), as Assumption 1 validates his condition (A.2*) with $m = 1$, and finiteness of Q validates his condition (A.3) while his condition (A.1) is ensured by $M(q)$ being projection matrices, the asymptotic optimality of $GC_p^0\left(q\right)$ holds:

$$\frac{R\left(\hat{q}^0\right)}{\min_{q\in Q} R\left(q\right)} \xrightarrow{P} 1.$$

Combining the pieces, we arrive at

$$\frac{R\left(\hat{q}\right)}{\min_{q\in Q} R\left(q\right)} = \frac{R\left(\hat{q}^0\right)}{\min_{q\in Q} R\left(q\right)} + \frac{R\left(\hat{q}\right) - R\left(\hat{q}^0\right)}{\min_{q\in Q} R\left(q\right)} \xrightarrow{P} 1 + o_P\left(1\right) = 1.$$

□