Part 3c Asymptotic Results for MRR1 using PRESS* (and PRESS)**

As before, we will first find the optimal value of l minimizing PRESS*. But this time we'll illustrate that the rate of convergence to l^*, the optimal l from part 3a, is not as good as that of $\hat{l}^{*C}$, nor that of $\hat{l}^{P}$ as n -> ∞. In fact, we will show that the mixing parameter obtained using PRESS* may not really converge at all ($O_P(1)$). So the bulk of this section is quite theoretical since we will not be presenting the usual array of theorems on estimate convergence.

Recall from Mays (1995) that

$$\text{PRESS*}(l) = \frac{\sum_{i=1}^{n}(Y_i - (l\hat{f}^{(i)} + (1-l)\hat{g}^{(i)}))^2}{n - tr(H^{MRR1})}$$

$$= \frac{\sum_{i=1}^{n}(Y_i - (l\hat{f}^{(i)} + (1-l)\hat{g}^{(i)}))^2}{n - (l\,tr(H^{OLS}) + tr(H^{LLR}) - l\,tr(H^{LLR}))}$$

$$= \frac{\sum_{i=1}^{n}(Y_i - (l\hat{f}^{(i)} + (1-l)\hat{g}^{(i)}))^2}{n - l(tr(H^{OLS}) - tr(H^{LLR})) - tr(H^{LLR}))}$$

$$= \frac{\sum_{i=1}^{n}(Y_i - (l\hat{f}^{(i)} + (1-l)\hat{g}^{(i)}))^2}{n - lC - G} \text{ (say),}$$

where H^{MRR1} is the projection matrix for the MRR1 estimate consisting of H^{OLS}, the Ordinary Least Squares projection matrix, and H^{LLR}, the local linear regression projection matrix, and C is defined as $tr(H^{OLS})$-$tr(H^{LLR})$, while G is defined as $tr(H^{LLR})$.

Again we wish to minimize the last expression with respect to l. So $\frac{d}{dl}$PRESS*(l)

$$= \frac{\sum_{i=1}^{n}2(Y_i - (l\hat{f}^{(i)} + (1-l)\hat{g}^{(i)}))(\hat{g}^{(i)} - \hat{f}^{(i)})}{(n - lC - G)} + \frac{C\sum_{i=1}^{n}(Y_i - (l\hat{f}^{(i)} + (1-l)\hat{g}^{(i)}))^2}{(n - lC - G)^2}$$

=

$$\frac{(n-\lambda C-G)\sum_{i=1}^{n}2(Y_i-(\lambda\hat{f}^{(i)}+(1-\lambda)\hat{g}^{(i)}))(\hat{g}^{(i)}-\hat{f}^{(i)})}{(n-\lambda C-G)^2}+\frac{C\sum_{i=1}^{n}(Y_i-(\lambda\hat{f}^{(i)}+(1-\lambda)\hat{g}^{(i)}))^2}{(n-\lambda C-G)^2}$$

Setting this expression equal to zero yields

$$\hat{\lambda}^{P*} = \frac{\sum_{i=1}^{n}(Y_i-\hat{g}^{(i)})}{\sum_{i=1}^{n}(\hat{f}^{(i)}-\hat{g}^{(i)})} \text{ or } -\frac{\sum_{i=1}^{n}(Y_i-\hat{g}^{(i)})}{\sum_{i=1}^{n}(\hat{f}^{(i)}-\hat{g}^{(i)})}-2\frac{G-n}{C}. \qquad (3.C.1)$$

We must next investigate the behavior of the second derivative of PRESS*(λ).

$$\frac{d^2}{d\lambda^2}\text{PRESS*}(\lambda) = \frac{\sum_{i=1}^{n}((G-n)(\hat{g}^{(i)}-\hat{f}^{(i)})+C(\hat{g}^{(i)}-Y_i))^2}{(n-\lambda C-G)^3}.$$

Clearly for large n, the second derivative is positive when we utilize the first term of 3.C.1, and negative when we utilize the second term of 3.C.1. Hence, PRESS*(λ) is locally minimized at the first term and locally maximized at the second term of 3.C.1. The second tends to produce negative results in small samples because of the last fraction involving n. At any rate, the choice is immaterial asymptotically, for it is the first term in both cases that leads to asymptotic instability. This is illustrated next.

Recall that

$$\lambda^* = \frac{<\hat{f}-\hat{g},\theta-g>}{\|\hat{f}-\hat{g}\|^2} = \frac{\sum(\hat{f}-\hat{g})\theta-\sum(\hat{f}-\hat{g})\hat{g}}{\sum(\hat{f}-\hat{g})^2}.$$

So that

$$\lambda^*-\hat{\lambda}^{P*} = \lambda^*-\frac{\sum(Y-\hat{g}^{(i)})}{\sum(\hat{f}^{(i)}-\hat{g}^{(i)})} \qquad (3.C.2)$$

$$= \lambda^*-\frac{\sum(Y-\hat{g}^{(i)})}{\sum(\hat{f}^{(i)}-\hat{g}^{(i)})}\frac{\sum(\hat{f}^{(i)}-\hat{g}^{(i)})}{\sum(\hat{f}^{(i)}-\hat{g}^{(i)})}$$

$$= l^* - \frac{\sum_{i=1}^{n} (Y_i - \hat{g}^{(i)})(\hat{f}^{(i)} - \hat{g}^{(i)}) + \sum\sum_{i \neq j} (Y_i - \hat{g}^{(i)})(\hat{f}^{(j)} - \hat{g}^{(j)})}{\sum_{i=1}^{n} (\hat{f}^{(i)} - \hat{g}^{(i)})^2 + \sum\sum_{i \neq j} (\hat{f}^{(i)} - \hat{g}^{(i)})(\hat{f}^{(j)} - \hat{g}^{(j)})}$$

$$= l^* - \frac{a+c}{b+d} \text{ (say).}$$

Through some simple algebra we obtain the following important result.

$$\frac{a+c}{b+d}$$

$$= \frac{a}{b}\left(\frac{b}{b+d}\right) + \frac{c}{b+d}$$

$$= \frac{a}{b}\left(1 - \frac{d}{b+d}\right) + \frac{c}{b+d}$$

$$= \frac{a}{b} - \frac{ad}{b(b+d)} + \frac{c}{b+d}$$

$$= \frac{a}{b} - \frac{ad-bc}{b(b+d)}.$$

So that 3.C.2 becomes

$$l^* - \frac{\sum_{i=1}^{n} (Y_i - \hat{g}^{(i)})(\hat{f}^{(i)} - \hat{g}^{(i)})}{\sum_{i=1}^{n} (\hat{f}^{(i)} - \hat{g}^{(i)})^2} + \frac{ad-bc}{b(b+d)}$$

$$= l^* - \hat{l}^P + \frac{ad-bc}{b(b+d)}. \qquad \text{(3.C.3)}$$

Recall, from sections 3a and 3b, that the first part of 3.C.3 is asymptotically zero with various rates of stochastic convergence. We shall demonstrate that the right hand term,

$$\frac{ad-bc}{b(b+d)} = R2 \text{ (say),}$$

of 3.C.3 is $O_P(1)$, rendering $\hat{l}^{P*}$ asymptotically weak as a selection tool for the MRR1 procedure.

Note that

$$ad = (\sum_{i=1}^{n}(Y_i - \hat{g}^{(i)})(\hat{f}^{(i)} - \hat{g}^{(i)}) + \sum_{i=1}^{n}(\hat{f}^{(i)} - \hat{g}^{(i)})^2)\sum\sum_{i\neq j}(\hat{f}^{(i)} - \hat{g}^{(i)})(\hat{f}^{(j)} - \hat{g}^{(j)}),$$

while bc =

$$\sum_{i=1}^{n}(\hat{f}^{(i)} - \hat{g}^{(i)})^2(\sum\sum_{i\neq j}(Y_i - \hat{g}^{(i)})(\hat{f}^{(j)} - \hat{g}^{(j)}) + \sum\sum_{i\neq j}(\hat{f}^{(i)} - \hat{g}^{(i)})(\hat{f}^{(j)} - \hat{g}^{(j)})).$$

So that ad - bc =

$$\sum_{i=1}^{n}((Y_i - \hat{g}^{(i)})(\hat{f}^{(i)} - \hat{g}^{(i)}))\sum\sum_{i\neq j}((\hat{f}^{(i)} - \hat{g}^{(i)})(\hat{f}^{(j)} - \hat{g}^{(j)})) -$$

$$\sum_{i=1}^{n}((\hat{f}^{(i)} - \hat{g}^{(i)})^2)\sum\sum_{i\neq j}((Y_i - \hat{g}^{(i)})(\hat{f}^{(j)} - \hat{g}^{(j)}))$$

We reinsert the denominator and consider the left term of $R2$ given by

$$\frac{\sum_{i=1}^{n}((Y_i - \hat{g}^{(i)})(\hat{f}^{(i)} - \hat{g}^{(i)}))\sum\sum_{i\neq j}((\hat{f}^{(i)} - \hat{g}^{(i)})(\hat{f}^{(j)} - \hat{g}^{(j)}))}{\sum_{i=1}^{n}((\hat{f}^{(i)} - \hat{g}^{(i)})^2)(\sum_{i=1}^{n}((\hat{f}^{(i)} - \hat{g}^{(i)})^2) + \sum\sum_{i\neq j}((\hat{f}^{(i)} - \hat{g}^{(i)})(\hat{f}^{(j)} - \hat{g}^{(j)})))}$$

=

$$\frac{(\sum_{i=1}^{n}((Y_i - \hat{g}^{(i)})(\hat{f}^{(i)} - \hat{g}^{(i)}))/n)(\sum\sum_{i\neq j}((\hat{f}^{(i)} - \hat{g}^{(i)})(\hat{f}^{(j)} - \hat{g}^{(j)}))/n^2)}{(\sum_{i=1}^{n}((\hat{f}^{(i)} - \hat{g}^{(i)})^2)/n)((\sum_{i=1}^{n}((\hat{f}^{(i)} - \hat{g}^{(i)})^2) + \sum\sum_{i\neq j}((\hat{f}^{(i)} - \hat{g}^{(i)})(\hat{f}^{(j)} - \hat{g}^{(j)})))/n^2)}$$

=

$$\frac{(\sum_{i=1}^{n}((Y_i - \hat{g}^{(i)})(\hat{f}^{(i)} - \hat{g}^{(i)}))/n)(\sum\sum_{i\neq j}((\hat{f}^{(i)} - \hat{g}^{(i)})(\hat{f}^{(j)} - \hat{g}^{(j)}))/n^2)}{(\sum_{i=1}^{n}((\hat{f}^{(i)} - \hat{g}^{(i)})^2)/n)(O_P(n^{-1}) + \sum\sum_{i\neq j}((\hat{f}^{(i)} - \hat{g}^{(i)})(\hat{f}^{(j)} - \hat{g}^{(j)}))/n^2)}$$

$$= \frac{(\sum_{i=1}^{n}((Y_i - \hat{g}^{(i)})(\hat{f}^{(i)} - \hat{g}^{(i)}))/n)(T1)}{(\sum_{i=1}^{n}((\hat{f}^{(i)} - \hat{g}^{(i)})^2)/n)(O_P(n^{-1}) + T1)} \text{ (say).} \qquad (3.C.4)$$

The right term of $R2$ is

$$\frac{\sum_{i=1}^{n}((\hat{f}^{(i)} - \hat{g}^{(i)})^2)\sum\sum_{i\neq j}((Y_i - \hat{g}^{(i)})(\hat{f}^{(j)} - \hat{g}^{(j)}))}{\sum_{i=1}^{n}((\hat{f}^{(i)} - \hat{g}^{(i)})^2)(\sum_{i=1}^{n}((\hat{f}^{(i)} - \hat{g}^{(i)})^2) + \sum\sum_{i\neq j}((\hat{f}^{(i)} - \hat{g}^{(i)})(\hat{f}^{(j)} - \hat{g}^{(j)})))}$$

=

$$\frac{(\sum\sum_{i\neq j}((Y_i-\hat{g}^{(i)})(\hat{f}^{(j)}-\hat{g}^{(j)}))/n^2)}{((\sum_{i=1}^{n}((\hat{f}^{(i)}-\hat{g}^{(i)})^2)+\sum\sum_{i\neq j}((\hat{f}^{(i)}-\hat{g}^{(i)})(\hat{f}^{(j)}-\hat{g}^{(j)})))/n^2)}$$

$$= \frac{(\sum\sum_{i\neq j}((Y_i-\hat{g}^{(i)})(\hat{f}^{(j)}-\hat{g}^{(j)}))/n^2)}{((O_P(n^{-1})+\sum\sum_{i\neq j}((\hat{f}^{(i)}-\hat{g}^{(i)})(\hat{f}^{(j)}-\hat{g}^{(j)})))/n^2)} \quad (3.C.5)$$

Recall a maneuver from section 3b, in which the ratio $\frac{a}{c+b}=\frac{a}{b}-\frac{a}{b}(\frac{c}{b})$ provided that b dominates c. We will use this result after showing that the b term does, in fact, dominate the c term in question.

The c term is clearly $\sum_{i=1}^{n}((\hat{f}^{(i)}-\hat{g}^{(i)})^2)/n^2 = O_P(n^{-1})$. So in order to show that the b term ($T1$) dominates, we need to demonstrate that $T1$ has a rate of convergence slower than $O_P(n^{-1})$. We need the following lemmas.

Lemma 3.c.1: Assuming conditions A1-A6 hold…

$$T1 = \sum\sum_{i\neq j}((\hat{f}^{(i)}-\hat{g}^{(i)})(\hat{f}^{(j)}-\hat{g}^{(j)}))/n^2 = O_P(g_n^2)$$

Lemma 3.c.2: Assuming conditions A1-A6 hold…

$$\frac{\sum e_i}{n} = O_P(n^{-.5}).$$

Some comments concerning the proofs (found in appendix 3c) are in order. Note that we did not invoke the distance measure d_n in A3.C.1. This is probably too conservative since the optimal parametric function f will automatically yield a sum of zero in the term $\sum(\hat{f}-q)/n$. Instead we use only the rate of convergence of the parametric estimate to f, which is $O_P(n^{-.5})$.

The same is not necessarily true for non-parametric functions, so $T1$ can converge only as fast as the nonparametric estimate converges to q.

The first lemma gives us some idea of the convergence rate of $T1$, at least in a maximal sense. Taking the result of Lemma 3.c.2 and combining that with Burman and Chaudhuri (1992) equations 6.11 and 6.19 and the Cauchy-Schwartz inequality confirms that $T1$ is at least $O_P(n^{-1})$. This is important in the remaining developments. We may proceed with finding the asymptotic convergence rate of $R2$.

Rewriting 3.C.4 we have $\dfrac{(\sum_{i=1}^{n}((Y_i - \hat{g}^{(i)})(\hat{f}^{(i)} - \hat{g}^{(i)}))/n)(T1)}{(\sum_{i=1}^{n}((\hat{f}^{(i)} - \hat{g}^{(i)})^2)/n)(O_P(n^{-1}) + T1)}$

=

$$\frac{(\sum_{i=1}^{n}((Y_i - \hat{g}^{(i)})(\hat{f}^{(i)} - \hat{g}^{(i)}))/n)}{(\sum_{i=1}^{n}((\hat{f}^{(i)} - \hat{g}^{(i)})^2)/n)} - \frac{(\sum_{i=1}^{n}((Y_i - \hat{g}^{(i)})(\hat{f}^{(i)} - \hat{g}^{(i)}))/n)}{(\sum_{i=1}^{n}((\hat{f}^{(i)} - \hat{g}^{(i)})^2)/n)}\left(\frac{O_P(n^{-1})}{T1}\right) \quad \text{(3.C.6).}$$

Similarly, 3.C.5 becomes $\dfrac{(\sum\sum_{i \neq j}((Y_i - \hat{g}^{(i)})(\hat{f}^{(j)} - \hat{g}^{(j)}))/n^2)}{((O_P(n^{-1}) + \sum\sum_{i \neq j}((\hat{f}^{(i)} - \hat{g}^{(i)})(\hat{f}^{(j)} - \hat{g}^{(j)})))/n^2)}$

=

$$\frac{(\sum\sum_{i \neq j}((Y_i - \hat{g}^{(i)})(\hat{f}^{(j)} - \hat{g}^{(j)}))/n^2)}{(\sum\sum_{i \neq j}((\hat{f}^{(i)} - \hat{g}^{(i)})(\hat{f}^{(j)} - \hat{g}^{(j)}))/n^2)} - \frac{(\sum\sum_{i \neq j}((Y_i - \hat{g}^{(i)})(\hat{f}^{(j)} - \hat{g}^{(j)}))/n^2)}{(\sum\sum_{i \neq j}((\hat{f}^{(i)} - \hat{g}^{(i)})(\hat{f}^{(j)} - \hat{g}^{(j)}))/n^2)}\left(\frac{O_P(n^{-1})}{T1}\right)$$

(3.C.7)

From the comments following the first two lemmas, we may assert that the asymptotic properties of $R2$ can be found by investigating the following combination of 3.C.6 and 3.C.7 given by

$$\frac{(\sum_{i=1}^{n}((Y_i - \hat{g}^{(i)})(\hat{f}^{(i)} - \hat{g}^{(i)}))/n)}{(\sum_{i=1}^{n}(\hat{f}^{(i)} - \hat{g}^{(i)})^2 /n)} - \frac{(\sum\sum_{i \neq j}((Y_i - \hat{g}^{(i)})(\hat{f}^{(j)} - \hat{g}^{(j)}))/n^2)}{(\sum\sum_{i \neq j}((\hat{f}^{(i)} - \hat{g}^{(i)})(\hat{f}^{(j)} - \hat{g}^{(j)}))/n^2)} \quad \text{(3.C.8)}$$

To analyze the parts of 3.C.8 we will need the following lemmas.

Lemma 3.c.3: Assuming conditions A1-A6 hold…

$$\frac{(\sum_{i=1}^{n}((Y_i - \hat{g}^{(i)})(\hat{f}^{(i)} - \hat{g}^{(i)})) / n)}{(\sum_{i=1}^{n}(\hat{f}^{(i)} - \hat{g}^{(i)})^2 / n)} = \begin{cases} O_P(g_n),\ if\ \lim_{n\to\infty} d_n \neq 0 \\ O_P(1),\quad if\ d_n = 0 \end{cases}.$$

The reader will recognize this last term as $\hat{l}^P$, the mixing parameter estimate obtained from the PRESS statistic. Indeed, the result of Lemma 3.c.3 coincides with the result of section 3b. The proof for this lemma is found in appendix 3c.

Lemma 3.c.4: Assuming conditions A1-A6 hold…

$$\frac{\sum\sum_{i \neq j}((Y_i - \hat{g}^{(i)})(\hat{f}^{(j)} - \hat{g}^{(j)}))}{n^2} = O_P(g_n^2)$$

Lemma 3.c.4 is necessary for the analysis of the upper right portion of 3.C.8, while Lemma 3.c.5 is for the entire right side of the same term. Both proofs are found in appendix 3c.

Lemma 3.c.5: Assuming conditions A1-A6 hold…

$$\frac{(\sum\sum_{i \neq j}((Y_i - \hat{g}^{(i)})(\hat{f}^{(j)} - \hat{g}^{(j)})) / n^2)}{(\sum\sum_{i \neq j}((\hat{f}^{(i)} - \hat{g}^{(i)})(\hat{f}^{(j)} - \hat{g}^{(j)})) / n^2)} = O_P(1).$$

Finally we can conclude from Lemmas 3.c.3 and 3.c.5 that term 3.C.8 (and consequently $R2$ and terms 3.C.2 and 3.C.3) is $O_P(1)$ if $\lim_{n\to\infty} d_n \neq 0$. If $d_n = 0$, the asymptotic rate of convergence is unknown, since 3.C.8 is $O_P(1)$ - $O_P(1)$ which is not necessarily zero. In fact, in most cases this difference is $O_P(1)$ as well, although we'll not pursue that issue here. The fact remains that

$$l^* - \hat{l}^{P*} = \begin{cases} O_P(1), & if\ d_n \neq 0 \\ O_P(1) - O_P(1), & if\ d_n = 0 \end{cases} \qquad (3.C.9)$$

Conclusion and Thoughts on PRESS**

We have demonstrated the asymptotic instability of the PRESS* statistic. No doubt, this is due mathematically to the existence of two roots in 3.C.1 which represent a minimum and a maximum (most likely neither of which is global). This is an artifact of the "correction factor" (the trace of the MRR1 hat matrix) in the denominator. Evidently, PRESS** will find the same asymptotic result, since this statistic involves two "correction factors" in the denominator, the first of which is identical to the aforementioned trace (see section 2c). This is not to say that these techniques are not useful. On the contrary Mays (1995) has demonstrated their usefulness in the small sample setting. For now, however, we shall turn our attention to dealing with the asymptotic results of the MRR2 estimate.