Presumptions: There are one or more treatment conditions, $W \geq 1$, and a control condition, $W = 0$. We see estimates of linear contrasts of the treatment condition(s), with the control; contrasts among the treatment conditions themselves are not of primary interest. There are one or more strata, within which propensity scores $\mathbf{Pr}(W = j|X)$ are taken to be uniform. Contrasts are to be combined across blocks B using given weights. Here we concern ourselves with estimation of variances and covariances of these contrasts, by combining the technical devices of conditioning on stratum totals

$$\sum_i \mathcal{I}\left[B_i = b\right] \mathcal{I}\left[W_i = c\right], \quad b, c \tag{1}$$

and the method of sandwich estimation.

There may be covariance adjustment, but in this case covariance parameters were estimated in a separate and prior fitting of the regression model. This needn't have been a linear model, but it's presumed to be an M-estimate or representable as such, and to satisfy smoothness and dimensionality conditions necessary for asymptotic linearity of the estimator. The sample used to estimate these covariance parameters may contain, overlap or be disjoint from the sample over which contrasts of treatment conditions are to be estimated.

Conjecture: clustering on one-many matched sets Call a block b a one-many matched set if $\sum_i \mathcal{I}\left[B_i = b\right] \mathcal{I}\left[W_i = 0\right] \geq 1$ while $\sum_{c>0} \sum_i \mathcal{I}\left[B_i = b\right] \mathcal{I}\left[W_i = c\right] = 0$. Under (1) and for given values of the covariance and contrast parameters parameters, the within-block sum of estimating functions, crossed with itself, unbiasedly estimates its expected value. (To do: state me non-tautologically.)

Conj: clustering on homogeneously full matched sets. Call a block b a homogeneous full match it is a full match in which no more than one treatment condition is represented. Under (1) and for given values of the covariance and contrast parameters parameters, the within-block sum of estimating functions, crossed with itself, unbiasedly estimates its expected value.

Conj: clustering on arbitrary full matched sets. Under (1) and for given values of the covariance and contrast parameters parameters, the within-block sum of estimating functions, crossed with itself, unbiasedly estimates its expected value.

Conj: strata with diverse representation of each represented condition (I'm not sure whether we can handle strata with a single member of one treatment condition and two members of another. Maybe it's easy, maybe it's not possible with the current scheme.)

Chain of estimating equations

A "detrending" step gives rise to parameter estimates $\hat{\theta}$, empirical solutions of a system of k estimating equations. The corresponding estimating function corresponds abstractly to a $k \times 1$ column vector $U(\theta)$ and concretely, post-estimation, to a $n \times k$ matrix $\mathbf{U}(\theta) = \{[U_i(\theta)]^t : i\}$, and $k \times 1$ column vector $\bar{U}(\theta) = n^{-1} \sum U_i(\theta)$. The purpose of this step is to set parameters involved in a residual transformation, $e_{\hat{\theta}}(\tilde{y}|r)$, which in turn figures in subsequent tests of treatment effect hypotheses.

Suppose $U(\theta)$ and θ to be partitionable as $[U_0(\theta_0)\,;\, U_1(\theta_0, \theta_1)\,;\, U_2(\theta_1, \alpha)]$ and $\theta = (\theta_0, \theta_1, \alpha)$, respectively, with neither θ_0 nor α figuring in the residual transformation, $e_\theta(\tilde{y}|r) \equiv e_{\theta_1}(\tilde{y}|r)$, and

$$U_2(\theta_1, \alpha) := e_\theta(\tilde{Y}|R) - \alpha.$$

(This structure would arise in robust MM-estimation of a linear detrender, which involves preliminary estimation of the scale parameter. Then θ_0 consists of this scale parameter and associated preliminary coefficients. Robust fitting ordinarily will not impose the constraint that $\sum_i e_{\hat{\theta}}(\tilde{y}_i|r_i) = 0$, except with specially selected $e_\theta(\cdot|\cdot)$; i.e. α may differ from 0.) Since $U_2(\hat{\theta}_1, \hat{\alpha}) = 0$, the definition of U_2 is equivalent to defining $\hat{\alpha}$ as $\overline{[e_{\hat{\theta}_1}(\tilde{Y}|R)]}$. We also have $\mathbf{U}(\theta) = [\mathbf{U}_0(\theta_0)\,;\, \mathbf{U}_1(\theta_0, \theta_1)\,;\, e_\theta(\tilde{\mathbf{Y}}|\mathbf{R}) - \alpha]$; $\bar{U} = [\bar{U}_0^t; \bar{U}_1^t; \bar{U}_2]^t$.

Our goal is to test $H : \mathbf{E}\{e_\theta(\tilde{Y}|R)|Z = 1\} = \mathbf{E}\{e_\theta(\tilde{Y}|R)|Z = 0\}$ The induced parameter α figures in a device for representing the test statistic $\overline{[e_\theta(\tilde{Y}|R)]}_{Z=1} - \overline{[e_\theta(\tilde{Y}|R)]}_{Z=0}$ in terms of yet another link in the estimating function chain. Define

$$V(\theta, \tau) = Z\{e_\theta(\tilde{Y}|R) - \alpha - \tau\},$$

so that $\bar{V}(\theta, \tau) = \bar{Z} \cdot \left\{ \overline{[e_\theta(\tilde{Y}|R)]}_{Z=1} - \alpha \right\}$. This definition of $\hat{\tau}$, as the solution of $\bar{V}(\hat{\theta}, \tau)$, is equivalent to

$$\begin{aligned}\hat{\tau} &= \overline{[e_\theta(\tilde{Y}|R)]}_{Z=1} - \overline{[e_\theta(\tilde{Y}|R)]} \\ &= (1 - \bar{Z})\left\{ \overline{[e_\theta(\tilde{Y}|R)]}_{Z=1} - \overline{[e_\theta(\tilde{Y}|R)]}_{Z=0} \right\}.\end{aligned}$$

So we can recover the variance of (this random multiple of) $\overline{[e_\theta(\tilde{Y}|R)]}_{Z=1} - \overline{[e_\theta(\tilde{Y}|R)]}_{Z=0}$ as the (τ, τ) component of the covariance of $(\hat{\theta}, \hat{\tau})$ — which we estimate in turn by analysis of the estimating equation stack

$$\begin{bmatrix} \bar{U}(\theta) \\ \bar{V}(\theta, \tau) \end{bmatrix} = 0.$$

To accomplish this analysis we repeatedly apply formulas for covariances of chains of estimating equations.

A_{UU}, B_{UU} and $\widehat{\mathbf{V}}(\hat{\theta})$

Applied to U alone, the sandwich formula (Stefanski and Boos, 2002) alleges that $\mathbf{V}[\hat{\theta}] \approx n^{-1} A_{UU}^{-1} B_{UU} A_{UU}^{-t}$, where

$$A_{UU} = \nabla_\theta \bar{U} = \nabla_\theta \begin{bmatrix} \bar{U}_0(\theta_0) \\ \bar{U}_1(\theta_0, \theta_1) \\ \bar{U}_2(\theta_0, \theta_1, \alpha) \end{bmatrix} = \begin{bmatrix} A_{00} & 0 & 0 \\ A_{10} & A_{11} & 0 \\ A_{20} & A_{21} & A_{22} \end{bmatrix}$$

and

$$B_{UU} = \begin{bmatrix} B_{00} & B_{10}^t & B_{20}^t \\ B_{10} & B_{11} & B_{21}^t \\ B_{20} & B_{21} & B_{22} \end{bmatrix}.$$

To simplify matrix algebra, define

$$\begin{aligned} A_{[01][01]} &= \nabla_{\theta_0,\theta_1} \begin{bmatrix} \bar{U}_0(\theta_0) \\ \bar{U}_1(\theta_0, \theta_1) \end{bmatrix} = \begin{bmatrix} A_{00} & 0 \\ A_{10} & A_{11} \end{bmatrix}, \\ A_{2[01]} &= \nabla_{\theta_0,\theta_1} \bar{U}_2 = [A_{20}\,;\, A_{21}] = [0\,;\, \nabla_{\theta_1} \overline{[e_\theta(\tilde{Y}|R)]}]; \\ B_{[01][01]} &= \begin{bmatrix} B_{00} & B_{10}^t \\ B_{10} & B_{11} \end{bmatrix}, \\ B_{2[01]} &= \mathbf{U}_2^t [\mathbf{U}_0\,;\, \mathbf{U}_1]. \end{aligned}$$

This allows us to write

$$A_{UU} = \begin{bmatrix} A_{[01][01]} & 0 \\ A_{2[01]} & -1 \end{bmatrix}$$

where $A_{2[01]} = \nabla_{\theta_0,\theta_1} \overline{[e_\theta(\tilde{Y}|R)]}$.

Also

$$B_{UU} = \begin{bmatrix} B_{[01][01]} & n^{-1} \mathbf{U}_{[01]}^t U_2 \\ n^{-1} U_2^t \mathbf{U}_{[01]} & \hat{\sigma}^2_{e_\theta(\tilde{Y}|R)} \end{bmatrix},$$

$\hat{\sigma}^2_{e_\theta(\tilde{Y}|R)} = n^{-1} U_2' U_2$. In virtue of $A_{[01][01]}$ and A_{UU} both having 0 upper-right submatrices,

$$\begin{aligned} A_{[01][01]}^{-1} &= \begin{bmatrix} A_{00}^{-1} & 0 \\ A_{11}^{-1} A_{10} A_{00}^{-1} & A_{11}^{-1} \end{bmatrix} \text{ and} \\ A_{UU}^{-1} &= \begin{bmatrix} A_{[01][01]}^{-1} & 0 \\ \{\nabla_{\theta_0,\theta_1} \overline{[e_\theta(\tilde{Y}|R)]}\} A_{[01][01]}^{-1} & -1 \end{bmatrix}. \end{aligned}$$

Putting A_{UU}^{-1} and B_{UU} together gives sandwich estimates of $\mathbf{V}(\hat{\theta}) = \mathrm{Cov}([\hat{\theta}_0, \hat{\theta}_1, \hat{\alpha}])$. Knowing that $\hat{\theta}_0$ won't directly contribute to the test statistic, one might restrict attention to B_{UU} and the submatrix of A_{UU}^{-1} consisting of its lowermost rows, the gradients of $U_1(\theta_0, \theta_1)$ and $U_2(\theta_0, \theta_1, \alpha)$. If we calculate and store $A_{11}^{-1} A_{10} A_{00}^{-1}$ and A_{11}^{-1}, we won't also need to store A_{00}^{-1} in order to estimate $\mathrm{Cov}([\hat{\theta}_1, \hat{\alpha}])$.

Ultimately we're interested in $\mathbf{V}\hat{\tau}$, not $\mathrm{Cov}(\hat{\theta})$; $\mathrm{Cov}([\hat{\theta}; \hat{\tau}])$ is more relevant. We'll see that for computing the $\mathbf{V}(\hat{\tau})$ component of that covariance we won't need A_{00}^{-1}, either, provided that we have $A_{11}^{-1} A_{10} A_{00}^{-1}$ and A_{11}^{-1}.

But it's useful to note the quantities that a sandwich estimation routine will have to have computed, directly or indirectly, en route to providing sandwich estimates of $\mathbf{V}(\hat{\theta}_1)$: $A_{11}^{-1} A_{10} A_{00}^{-1}$ and A_{11}^{-1}, if not necessarily the other part of $A_{[01][01]}^{-1}$ (ie A_{00}^{-1}).

1 A_{VU}, A_{VV}, B_{UV}, B_{VV} and $\mathrm{Cov}(\hat{\tau})$

The sandwich formula also says $\mathrm{Cov}(\hat{\theta}, \hat{\tau}) \approx n^{-1} A^{-1} B A^{-t}$, where

$$\begin{aligned} A = \quad & \nabla_{\theta,\tau} \begin{pmatrix} \bar{U} \\ \bar{V} \end{pmatrix} = \left[\begin{array}{c|c} A_{UU} & 0 \\ \hline A_{VU} & A_{VV} \end{array}\right] = \left[\begin{array}{c|c} A_{UU} & 0 \\ \hline \nabla_\theta \bar{V} & (\partial/\partial\tau)\bar{V} \end{array}\right] \\ = \quad & \left[\begin{array}{c|c} A_{UU} & 0 \\ \hline A_{VU} & -\bar{Z} \end{array}\right], \end{aligned}$$

where

$$\begin{aligned} A_{VU} &= \nabla_\theta \bar{V} = \nabla_{\theta_0,\theta_1,\alpha} \bar{Z} \{ \overline{[e_\theta(\tilde{Y}|R)]}_{Z=1} - \alpha - \tau \} \\ &= \bar{Z} \cdot [0\,;\, \nabla_{\theta_1} \{ \overline{[e_\theta(\tilde{Y}|R)]}_{Z=1} \}\,;\, -1], \end{aligned}$$

and

$$B = \left[\begin{array}{c|c} B_{UU} & B_{VU}^t \\ \hline B_{VU} & B_{VV} \end{array}\right] = \begin{bmatrix} B_{UU} & n^{-1}\mathbf{U}^t\mathbf{V} \\ n^{-1}\mathbf{V}^t\mathbf{U} & \bar{Z}\hat{\sigma}^2_{e_\theta(\tilde{Y}|R)|Z=1} \end{bmatrix},$$

$\hat{\sigma}^2_{e_\theta(\tilde{Y}|R)|Z=1} = \mathbf{V}^t\mathbf{V}/(\mathbf{Z}'\mathbf{Z})$.

The lower-left entry of $A^{-1} B A^{-t}$, ie. $n\widehat{\mathbf{V}}(\hat{\tau})$, depends only on B and A^{-1}'s bottom row.

Applying inversion formulas for blocked matrices with upper-right 0's as above,

$$A^{-1} = \left[\begin{array}{c|c} A_{UU}^{-1} & 0 \\ \hline \bar{Z}^{-1} A_{VU} A_{UU}^{-1} & -\bar{Z}^{-1} \end{array}\right]. \tag{2}$$

The lower-left submatrix (left part of bottom row) is expressible as

$$\begin{aligned}
\bar{Z}^{-1}A_{VU}A_{UU}^{-1} &= \left[\{\nabla_{\theta_0,\theta_1}\overline{[e_\theta(\tilde{Y}|R)]}_{Z=1}\}\,;\,-1\right]\left[\begin{array}{cc} A_{[01][01]}^{-1} & 0 \\ \{\nabla_{\theta_0,\theta_1}\overline{[e_\theta(\tilde{Y}|R)]}\}A_{[01][01]}^{-1} & -1 \end{array}\right] \\
&= \left[\{\nabla_{\theta_0,\theta_1}\overline{[e_\theta(\tilde{Y}|R)]}_{Z=1}\}A_{[01][01]}^{-1} - \{\nabla_{\theta_0,\theta_1}\overline{[e_\theta(\tilde{Y}|R)]}\}A_{[01][01]}^{-1}\,;\,1\right] \\
&= \left[(\nabla_{\theta_0,\theta_1}\{\overline{[e_\theta(\tilde{Y}|R)]}_{Z=1} - \overline{[e_\theta(\tilde{Y}|R)]}\})A_{[01][01]}^{-1}\,;\,1\right] \\
&= \left[\left[0\,;\,\nabla_{\theta_1}\{\overline{[e_\theta(\tilde{Y}|R)]}_{Z=1} - \overline{[e_\theta(\tilde{Y}|R)]}\}\right]A_{[01][01]}^{-1}\,;\,1\right],
\end{aligned}$$

the final equality following from $e_\theta(\cdot|\cdot) \equiv e_{\theta_1}(\cdot\ \cdot)$. So the bottom row as a whole is

$$\left[\left[0\,;\,\nabla_{\theta_1}\{\overline{[e_\theta(\tilde{Y}|R)]}_{Z=1} - \overline{[e_\theta(\tilde{Y}|R)]}\}\right]A_{[01][01]}^{-1} \qquad ;\,1\,;\,-\bar{Z}^{-1}\right]$$

$$\left[\left[0\,;\,\nabla_{\theta_1}\{\overline{[e_\theta(\tilde{Y}|R)]}_{Z=1} - \overline{[e_\theta(\tilde{Y}|R)]}\}\right]\left[\begin{array}{cc} A_{00}^{-1} & 0 \\ A_{11}^{-1}A_{10}A_{00}^{-1} & A_{11}^{-1} \end{array}\right] \qquad ;\,1\,;\,-\bar{Z}^{-1}\right]$$

$$\left[\nabla_{\theta_1}\{\overline{[e_\theta(\tilde{Y}|R)]}_{Z=1} - \overline{[e_{\theta_1}(\tilde{Y}|R)]}\}A_{11}^{-1}A_{10}A_{00}^{-1}\,;\,\nabla_{\theta_1}\{\overline{[e_\theta(\tilde{Y}|R)]}_{Z=1} - \overline{[e_\theta(\tilde{Y}|R)]}\}A_{11}^{-1} \quad ;\,1\,;\,-\bar{Z}^{-1}\right]$$

and the lower-right entry of $n^{-1}A^{-1}BA^{-t}$ is

$$n^{-1}\left[\nabla_{\theta_1}\{\overline{[e_\theta(\tilde{Y}|R)]}_{Z=1} - \overline{[e_{\theta_1}(\tilde{Y}|R)]}\}[A_{11}^{-1}A_{10}A_{00}^{-1}\,;\,A_{11}^{-1}]\,;\,1\,;\,-\bar{Z}^{-1}\right]$$
$$\cdot B\left[\begin{array}{c} A_{00}^{-t}A_{10}^{t}A_{11}^{-t}\nabla_{\theta_1}^{t}\{\overline{[e_\theta(\tilde{Y}|R)]}_{Z=1} - \overline{[e_{\theta_1}(\tilde{Y}|R)]}\} \\ A_{11}^{-t}\nabla_{\theta_1}^{t}\{\overline{[e_\theta(\tilde{Y}|R)]}_{Z=1} - \overline{[e_{\theta_1}(\tilde{Y}|R)]}\} \\ 1 \\ -\bar{Z}^{-1} \end{array}\right],$$

with

$$B = \left[\begin{array}{cccc} B_{[01][01]} & & n^{-1}\mathbf{U}_0^t\mathbf{U}_2 & n^{-1}\mathbf{U}_0^t\mathbf{V} \\ & & n^{-1}\mathbf{U}_1^t\mathbf{U}_2 & n^{-1}\mathbf{U}_1^t\mathbf{V} \\ n^{-1}\mathbf{U}_2^t\mathbf{U}_0 & n^{-1}\mathbf{U}_2^t\mathbf{U}_1 & n^{-1}\mathbf{U}_2^t\mathbf{U}_2 & n^{-1}\mathbf{U}_2^t\mathbf{V} \\ n^{-1}\mathbf{V}^t\mathbf{U}_0 & n^{-1}\mathbf{V}^t\mathbf{U}_1 & n^{-1}\mathbf{V}^t\mathbf{U}_2 & \bar{Z}\hat{\sigma}^2_{e_\theta(\tilde{Y}|R)|Z=1} \end{array}\right].$$

To summarize what this means for computation:

- What's needed from the 0 and 1 fits in terms of A matrices (bread) is the same as what a fitting or covariance estimation routine has to carry forward from stage 0 in order to build a proper sandwich estimate of stage 1 coefficients, $\mathbf{V}(\theta_1)$. For this one needs only the lower, U_1 submatrix of $A_{[01][01]}^{-1}$.
- To complete A matrix calcs we also require a means of computing $\nabla_{\theta_1}[e_\theta(y|r)]$.
- In terms of B matrices (meat), one appears to need everything from stages 0 and 1: not only $B_{[01][01]}$ but also $\mathbf{U}_{[01]}$ (from which $B_{[01][01]}$ can be regenerated).

Simplifications when detrending parameters are estimated via M-S

The B-matrix requirements are meaningfully reduced in the important special case that stages 0 and 1 comprise an S/M estimator chain, a.k.a. MM estimators Yohai (1987): one gets away with passing forward $\mathbf{U}_\sigma$, the stage-0 estimating function for the scale parameter, but not other parts of remainder of $\mathbf{U}_0$. . This is noted by Croux, Dhaene, and Hoorelbeke (2004), whose development is the basis for sandwich estimates of variance for robust regression as implemented in matlab, stata and R.…

An upshot is that one could define a $(k+1)\times k$ "bread matrix" as the value of `bread.lmrob()` and an $n \times (k+1)$ matrix $[\mathbf{U}_\sigma\ \mathbf{U}_1]$ as the value of `estfun.lmrob()`.

References

Croux, C., Dhaene, G., and Hoorelbeke, D. (2004), "Robust standard errors for robust estimators," Tech. rep., KU Leuven, Faculty of Economics and Applied Economics: Department of Economics.

Stefanski, L. and Boos, D. (2002), "The calculus of M-estimation," *American Statistician*, 56, 29–38.

Yohai, V. J. (1987), "High breakdown-point and high efficiency robust estimates for regression," *The Annals of Statistics*, 642–656.