Notes on the Huber-Eicker-White Procedure for Obtaining Consistent Estimates of OLS Standard Errors under Unrestricted Heteroskedasticity

Let the regression model be specified as

$$y = X\beta + \varepsilon,$$

where

- $E(\varepsilon|X) = 0$
- $E(\varepsilon\varepsilon'|X) = \sigma^2 \begin{bmatrix} \psi_{11} & 0 & \cdots & 0 \\ 0 & \psi_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \psi_{nn} \end{bmatrix} = \sigma^2\Psi,$

  where the ψ_{ii} are unknown constants.

Let X be $n \times k$, and let the rank of X be equal to k. The model then consists of $k+n$ unknown parameters, the k element coefficient vector β and the n conditional variance parameters in the Ψ matrix.

We know that the OLS estimator of β, $\hat{\beta} = (X'X)^{-1}X'y$, remains unbiased and consistent in this case. We also know that the covariance matrix of $\hat{\beta}$ is given by

$$V(\hat{\beta}|X) = \sigma^2(X'X)^{-1}X'\Psi X(X'X)^{-1}. \tag{0.1}$$

Now we know the following. Let z be some unkown parameter, and let $\tilde{z}$ be some random variable for which plim $\tilde{z} = z$. Let $g(z)$ be a known, "smooth" function of z - differentiability is more than enough. Then

$$\text{plim}\, g(\tilde{z}) = g(z).$$

This type of result is used repeatedly throughout the course.

If we let $\Sigma = \sigma^2\Psi$, then if we had a consistent estimator of Σ, call it $\hat{\Sigma}$, we could consistently estimate the covariance matrix $V(\hat{\beta}|X, \Sigma)$ by $V(\hat{\beta}|X, \hat{\Sigma})$ since (1) we assume $\text{plim}(\hat{\Sigma}) = \Sigma$) and (2) $V(\hat{\beta}|X, \Sigma)$ is a known, differentiable function of Σ. For reasons stated in class however, with n pieces of information (and only one observation per individual), we can never hope to consistently estimate

n individual specific variances. This is a problem of incidental, or nuisance parameters, in which the dimension of the parameter space grows with sample size. Thus we cannot consistently estimate $\hat{\Sigma}$ and all appears lost.

This case is not so hopeless after all, as was recognized by the various authors cited in the title. They recognized that to consistently estimate [0.1] didn't require a consistent estimator for Σ, but rather only for $X'\Psi X$, which is after all a $k \times k$ matrix the size of which doesn't increase in n. Note that if Σ is known, the the estimate of the asymptotic covariance matrix of $\hat{\beta}$ would be

$$\hat{V}_n(\hat{\beta}|X) = \frac{1}{n}\left(\frac{X_n'X_n}{n}\right)^{-1}\frac{1}{n}X'\Sigma X\left(\frac{1}{n}X'X\right)^{-1}.$$

To consistently estimate this quantity, what is required is a consistent estimator $Q^* = \text{plim}(Q_n^*)$, where

$$\begin{aligned} Q_n^* &= \frac{1}{n}X_n'\Sigma_n X_n \\ &= \frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 x_i' x_i, \end{aligned}$$

where x_i is the i^{th} row of X [and so is of dimension $1 \times k$. Say that the true distrubances could be observed [i.e., we knew β]. Then each term in the above summation could be rewritten so that

$$\sigma_i^2 x_i' x_i = E[\varepsilon_i^2 x_i' x_i | x_i]$$

Under mild conditions on the behavior of the x_i, a law of large numbers [LLN] argument can be constructed to show that

$$\text{plim}\,\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 x_i' x_i = \text{plim}\,\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 x_i' x_i.$$

The final step involves replacing the unknown disturbances with consistent estimates of them. Since the OLS estimator remains consistent in the case of unrestricted heterogeneity, $\text{plim}\,\hat{\beta} = \beta$, which implies that the OLS $e_i = y - x_i\hat{\beta}$ will converge to $\varepsilon_i = y_i - x_i\beta$, and by the same token, $\text{plim}(e_i^2) = \varepsilon_i^2$. Then

$$Q^* = \text{plim}\,\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 x_i' x_i$$

$$
\begin{aligned}
&= \operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i^2 x_i' x_i \\
&= \operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} e_i^2 x_i' x_i .
\end{aligned}
$$

All this implies the following. In sufficiently large samples, the covariance matrix of $\hat{\beta}$ is well-approximated by

$$
\hat{V}_n(\hat{\beta}|X) \approx (X_n' X_n)^{-1} \sum_{i=1}^{n} e_i^2 x_i' x_i (X_n' X_n)^{-1}. \tag{0.2}
$$

To compute this quantity, recognize that we have to first obtain the OLS residuals e. Thus, first estimate $\hat{\beta}$, obtain the residual vector, and to conserve memory, run through a DO LOOP in which the summation $\sum_{i=1}^{n} e_i^2 x_i' x_i$ is formed. For example:

```
cum_mat = zeros(k,k);
i=1;
do until i gt n;
   cum_mat = cum_mat + e[i]^2 * x[i,.]'x[i,.];
   i=i+1;
endo;
```

This will produce the matrix of the quadratic form given in [0.2]. The rest should be straightforward.

This is a very useful result, and in the absence of any strong reasons to suspect that the homogeneity assumption is appropriate, standard errors computed in this manner should always be used [in cross-sectional types of analysis where an independence assumption is appropriate]. While standard errors computed under this assumption could be found to be larger or smaller than those computed under a homoskedasticity assumption, in standard practice you should expect to see the HEW standard errors to be a bit larger (say on the order of 10 percent).