ORDINARY LEAST-SQUARES REGRESSION AND NON SPHERICAL DISTURBANCES

In cases where the structure of the dispersion matrix of the regression disturbances is known to depend on a small set of parameters, it will be possible to estimate the regression parameter β in the model $(y; X\beta, \sigma^2\Omega)$ via a method of feasible generalised least squares. The method uses an estimate Ω^* of the dispersion matrix of the disturbances within the formula $\beta^* = (X'\Omega^{*-1}X)^{-1}X'\Omega^{*-1}y$. In other cases, where there is no knowledge of the structure of the dispersion matrix, we may have to use the ordinary least-squares (OLS) estimator $\hat{\beta} = (X'X)^{-1}X'y$.

The OLS estimator will be unbiased and, subject to certain restrictions limiting the serial dependence of the disturbances, it will also be consistent. However, the dispersion matrix of the estimator will differ from that which obtains in the case of the OLS estimator of the classical model $(y; X\beta, \sigma^2 I)$, which is $D(\hat{\beta}) = \sigma^2(X'X)^{-1}$. In fact, the dispersion matrix of the OLS estimator of β in the model $(y; X\beta, \sigma^2\Omega)$ is given by

$$\begin{aligned} D(\hat{\beta}) &= (X'X)^{-1}X'D(y)X(X'X)^{-1} \\ &= (X'X)^{-1}\{\sigma^2 X'\Omega X\}(X'X)^{-1}, \end{aligned} \tag{1}$$

which is commonly referred to as the sandwich formula. Here, $D(y) = E(\varepsilon\varepsilon') = \sigma^2\Omega = \Sigma$ is a symmetric matrix of order T, which cannot be estimated on the basis of a sample of size T, unless there are sufficient restrictions on its structure. However, in order to implement the sandwich formula, it is required only to estimate of the matrix $W = \sigma^2 X'\Omega X = X'\Sigma X$, which is of the order k that corresponds to the number of explanatory variables in X.

For the purpose of deriving an asymptotic theory, we must consider the dispersion matrix of $\sqrt{T}\hat{\beta}$, which is

$$D(\sqrt{T}\hat{\beta}) = \left(\frac{X'X}{T}\right)^{-1}\left\{\frac{X'\Sigma X}{T}\right\}\left(\frac{X'X}{T}\right)^{-1}. \tag{2}$$

It is presumed that, as $T \to \infty$, $X'X/T$ tends to a limit that is a positive definite matrix with finite elements. The task is to demonstrate that, under certain weak assumptions, $W = X'\Sigma X/T$ will also tend to a finite limit.

To reveal the structure of this matrix, let us consider the elements of $W = [w_{ij}]$, $X = [x_{tj}]$, $X' = [x_{si}]'$ and $\Sigma = [\sigma_{st}]$. Then, there is

$$\begin{aligned} w_{ij} &= \sum_t \sum_s x_{is}\sigma_{st}x_{tj} \\ &= \sum_t \sum_s x_{is}E(\varepsilon_s\varepsilon_t)x_{tj}; \end{aligned} \tag{3}$$

and the matrix as a whole is given by $W = \sum_t \sum_s x'_{s\bullet} E(\varepsilon_s \varepsilon_t) x_{t\bullet}$, in a more summary notation that denotes the t-th *row* of X by $x_{t\bullet}$. For this to be estimable, some further restrictions are necessary. The restriction that removes the serial dependence from the disturbances, but which allows them to be heteroskedastic, sets

$$E(\varepsilon_s \varepsilon_t) = \begin{cases} \sigma_t^2, & \text{if } t = s, \\ 0, & \text{if } t \neq s. \end{cases} \tag{4}$$

Then, there is

$$w_{ij} = \sum_t \sigma_t^2 x_{it} x_{tj} \tag{5}$$

and

$$W = [w_{ij}] = \sum_t \sigma_t^2 x'_{t\bullet} x_{t\bullet}. \tag{6}$$

Here, $x'_{t\bullet} x_{t\bullet}; t = 1, \ldots, T$ is a sequence of matrices of rank 1, each formed as the outer product of a row of X and its column transpose. Each of the matrices is associated with an element of the diagonal matrix $\Sigma = \mathrm{diag}\{\sigma_1^2, \ldots, \sigma_T^2\}$. A visual analogy leads us to to describe (6) as the herringbone formula.

There are still as many parameters within the matrix $\Sigma = \mathrm{diag}\{\sigma_1^2, \ldots, \sigma_T^2\}$ as there are observations. Therefore, it cannot be estimated consistently. Nevertheless the product $T^{-1}W = T^{-1}X'\Sigma X$ can be estimated consistently via

$$\frac{1}{T}\hat{w}_{ij} = \frac{1}{T}\sum_t e_t^2 x_{it} x_{tj}, \tag{7}$$

which is obtained by replacing $\sigma_t^2 = E(\varepsilon_t^2)$ in (5) by the squared residual e_t^2. This is the heteroskedasticity-consistent estimator of White (1982).

To demonstrate the consistency, we note that, if $\hat{\beta} \to \beta$ as $T \to \infty$, then $e_t^2 \to \varepsilon_t^2$. Therefore, it is sufficient to consider the limiting behaviour of

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T} \varepsilon_t^2 x_{it} x_{tj} &= \frac{1}{T}\sum_{t=1}^{T} (\sigma_t^2 + \nu_t) x_{it} x_{tj} \\ &= \frac{1}{T}\sum_{t=1}^{T} \sigma_t^2 x_{it} x_{tj} + \frac{1}{T}\sum_{t=1}^{T} \nu_t x_{it} x_{tj}. \end{aligned} \tag{8}$$

In the second term on the LHS, there is a random variable ν_t, representing the deviation of ε_t^2 from its expected value $E(\varepsilon_t^2) = \sigma_t^2$, which has $E(\nu_t) = 0$ and which is independent of the elements x_{it} and x_{tj}. We can expect the second term to converge to zero. Then, since $e_t^2 \to \varepsilon_t^2$, it follows that $T^{-1}\hat{w}_{ij} = T^{-1}\sum_t e_t^2 x_{it} x_{tj}$ converges to $T^{-1}w_{ij}$.

The restriction that eliminates the serial dependence of the disturbances is much stronger than it need be. Given that the matrix W is of a constant order k, whereas the sample size T may grow indefinitely, there is hope of estimating W consistently in circumstances where the disturbances are generated by a stationary stochastic process. Then, the matrix $\Sigma = [\sigma_{ts}] = [\sigma_{|t-s|}]$ has a Toeplitz form with elements that are repeated along each of the N.W.–S.W. diagonals.

Consider the writing the matrix $T^{-1}W = T^{-1}X'E(\varepsilon\varepsilon)X$ as

$$\begin{aligned}\frac{1}{T}W &= \sum_{s=1}^{T}\sum_{t=1}^{T}\frac{1}{T}x'_{s\bullet}E(\varepsilon_s\varepsilon_t)x_{t\bullet}\\ &= \sum_{j=1-T}^{T-1}\Big\{\frac{1}{T}\sum_{t=j+1}^{T}x'_{[t-j]\bullet}E(\varepsilon_{t-j}\varepsilon_t)x_{t\bullet}\Big\}.\end{aligned} \tag{9}$$

Whereas in (6) there were T rank-1 matrices of the form $\sigma_t^2 x'_{t\bullet}x_{t\bullet}$, associated with the non-zero elements of the diagonal matrix $\Sigma = \mathrm{diag}\{\sigma_1^2,\ldots,\sigma_T^2\}$, there are now T^2 rank-1 matrices of the form $\sigma_{ts}x'_{s\bullet}x_{t\bullet}$, each of which is associated with an element of a positive definite matrix Σ, which is subject only to the restrictions of symmetry.

In the first expression of (9), the summations over s and t run thoughout the rows and columns of Σ. The second expression is derived by defining $j = t - s$ and setting $s = t - j$. Each value of j indexes one of the diagonals, beginning in the bottom left corner, where $s = T$, $t = 1$ and $j = t - s = 1 - T$, and rising through the principal diagonal to the top right corner where $j = T - 1$.

The dispersion matrix of a stationary stochastic process has constant values along these diagonals. Therefore, the second expression is appropriate to cases where both the data and the disturbances are generated by stationary processes.

Under such circumstances, the second equality of (9) can be written as

$$\frac{1}{T}W = \sum_{j=1-T}^{T-1}\Gamma_j = \Gamma_0 + \sum_{j=1}^{T-1}(\Gamma_j + \Gamma'_j), \tag{10}$$

where Γ_j is the expression within the braces. The second equality reflects the symmetry of W. The empirical counterpart of Γ_j is

$$G_j = \frac{1}{T}\sum_{t=j+1}^{T}x'_{[t-j]\bullet}e_{t-j}e_t x_{t\bullet}. \tag{11}$$

If the number j is small in comparison with T, then we can expect G_j to be an adequate estimate of Γ_j. Moreover, for a fixed j, we can expect $G_j \to \Gamma_j$ as $T \to \infty$.

Replacing Γ_j by G_j in (10) for all j results in the matrix $T^{-1}X'ee'X$, which does not constitute a viable estimator. The difficulty lies in the estimates G_j when j is close to T. In that case, the estimate will comprise a limited amount of information from $T - j$ sample points. Various recourses for avoiding the problem are available. The simplest of these is to limit the range of the index j so that its absolute value does not exceed some threshold value p. Then, we obtain the estimator of Hansen (1982), which is

$$W_H = \sum_{j=p}^{p} G_j = G_0 + \sum_{j=1}^{p} (G_j + G'_j). \tag{12}$$

An alternative estimator, which is due to Newey and West (1987), applies a gradual discount to the matrices G_j as j increases. It takes the form of

$$W_N = G_0 + \sum_{j=1}^{p} \left(1 - \frac{j}{p+1}\right)(G_j + G'_j). \tag{13}$$

References

Hansen, L.P., (1982), Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica,* 50, 1029–1054.

Newey, W.K. and K.D. West, (1987), A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica,* 55, 703–708.

White, H., (1980), A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, *Econometrica,* 48, 817–838.