Chapter 6

ESTIMATION OF THE LONG-RUN COVARIANCE MATRIX

An estimate of the long-run covariance matrix, $\mathbf{\Omega}$, is necessary to calculate asymptotic standard errors for the OLS and linear IV estimators presented in Chapter 5. Estimation of the long-run covariance matrix will be important for GMM estimators introduced later in Chapter 9 and many of the estimation and testing methods for nonstationary variables. Chapter 13 shows that an estimate of the long-run variance of a random variable is also useful in estimating the importance of the random walk component of some nonstationary random variables. This chapter discusses estimation methods for the long-run covariance matrix.

Let $\{\mathbf{u}_t : -\infty < t < \infty\}$ be a stationary and ergodic vector stochastic process with mean zero. We will discuss estimation methods of the long-run covariance matrix of $\mathbf{u}_t$:

$$\mathbf{\Omega} = \lim_{j\to\infty} \sum_{-j}^{j} E(\mathbf{u}_t \mathbf{u}'_{t-j}). \tag{6.1}$$

Depending on the application, we take different variables as $\mathbf{u}_t$. When $\mathbf{\Omega}$ is used for the calculation of the asymptotic standard errors for the OLS estimator, we take

$\mathbf{u}_t = \mathbf{x}_t(y_t - \mathbf{x}_t'\mathbf{b}_0)$. For the calculation of the asymptotic standard errors for the linear IV estimator, we take $\mathbf{u}_t = \mathbf{z}_t(y_t - \mathbf{x}_t'\mathbf{b}_0)$. Because $\mathbf{b}_0$ is unknown, the sample counterpart of $\mathbf{u}_t$, $\mathbf{z}_t(y_t - \mathbf{x}_t'\mathbf{b}_T)$, is used to estimate $\mathbf{\Omega}$ where $\mathbf{b}_T$ is a consistent estimator for $\mathbf{b}_0$. For the application in Chapter 13, $\mathbf{u}_t$ is a random variable such as the first difference of the log real GNP minus its expected value, and the first difference minus its estimated mean is used for the sample counterpart. Thus in many applications, $\mathbf{u}_t$ is unobservable and its sample counterpart is constructed from a consistent estimator for a parameter vector. When $\mathbf{\Omega}_T$ is a consistent estimator for $\mathbf{\Omega}$, $\mathbf{\Omega}_T^* = f(T)\mathbf{\Omega}_T$ is also a consistent estimator as long as $\lim_{T\to\infty} f(T) = 1$ for any real valued function $f(T)$. Therefore, we can consider various forms of $f(T)$ to improve small sample properties. If p parameters are estimated to compute the sample counterpart of $\mathbf{u}_t$, then $f(T) = \frac{T}{T-p}$ is a small sample degrees of freedom adjustment that is often used for each $\mathbf{\Omega}_T$ presented in this chapter.[1]

6.1 Serially Uncorrelated Variables

This section treats the case where $E(\mathbf{u}_t\mathbf{u}_{t-\tau}') = \mathbf{0}$ for $\tau \neq 0$. Many rational expectations models imply this property. In this case, $\mathbf{\Omega} = E(\mathbf{u}_t\mathbf{u}_t')$ can be estimated by $\frac{1}{T}\sum_{t=1}^{T}\mathbf{u}_t\mathbf{u}_t'$. For linear IV estimators, this is White's (1980) heteroskedasticity consistent estimator. In this case, $\mathbf{u}_t = \mathbf{z}_t(y_t - \mathbf{x}_t'\mathbf{b}_T)$ and $\frac{1}{T}\sum_{t=1}^{T}\mathbf{u}_t\mathbf{u}_t' = \frac{1}{T}\sum_{t=1}^{T}(y_t - \mathbf{x}_t'\mathbf{b}_T)^2\mathbf{z}_t\mathbf{z}_t'$.

In some cases, conditional homoskedasticity is assumed in the economic model, and an econometrician may wish to impose this property on the estimate for $\mathbf{\Omega} = E(e_t^2)E(\mathbf{z}_t\mathbf{z}_t')$. Then $\frac{1}{T}\sum_{t=1}^{T}(y_t - \mathbf{x}_t'\mathbf{b}_T)^2\frac{1}{T}\sum_{t=1}^{T}\mathbf{z}_t\mathbf{z}_t'$ with a small sample degree of

[1] Some other forms of small sample adjustments have been used (see, e.g., Ferson and Foerster, 1994).

freedom adjustment such as $\frac{T}{T-p}$ is used to estimate $\mathbf{\Omega}$.

6.2 Serially Correlated Variables

This section treats the case where the disturbance is serially correlated in the context of time series analysis.

6.2.1 Unknown Order of Serial Correlation

In many applications, the order of serial correlation is unknown. The estimators of the long-run covariance matrix in the presence of conditional heteroskedasticity and autocorrelation are called Heteroskedasticity and Autocorrelation Consistent (HAC) estimators.

Let $\mathbf{\Phi}(\tau) = E(\mathbf{u}_t\mathbf{u}'_{t-\tau})$. Many HAC estimators use the sample version of $\mathbf{\Phi}(\tau)$,

$$\mathbf{\Phi}_T(\tau) = \frac{1}{T}\sum_{t=\tau+1}^{T}\mathbf{u}_t\mathbf{u}'_{t-\tau} \quad \text{for} \quad 0 \le \tau \le T-1 \tag{6.2}$$

and $\mathbf{\Phi}_T(\tau) = \mathbf{\Phi}_T(-\tau)'$ for $\tau < 0$. Given the data of $\mathbf{u}_1, ..., \mathbf{u}_T$, $\mathbf{\Phi}_T(\tau)$ for a large lag τ cannot be estimated with many observations. For example, we have only one observation for $\mathbf{\Phi}_T(T-1)$. Hence it is natural to put much less weight on $\mathbf{\Phi}_T(\tau)$ with large τ than on $\mathbf{\Phi}_T(\tau)$ with small τ. The weights are described by a real valued function called a kernel function. The kernel HAC estimators for $\mathbf{\Omega}$ in the literature have the form

$$\mathbf{\Omega}_T = \sum_{\tau=-T+1}^{T-1} k(\frac{\tau}{S_T})\mathbf{\Phi}_T(\tau), \tag{6.3}$$

where $k(\cdot)$ is a real-valued kernel, and S_T is a band-width parameter.[2] Examples of

[2]These terminologies follow Andrews (1991), and are somewhat different from those in kernel estimations in other contexts.

kernels that have been used by econometricians include the following:

$$
\begin{aligned}
(6.4)\quad k(x) &= \begin{cases} 1 & \text{for } |x| \le 1 \\ 0 & \text{otherwise} \end{cases} && \text{Truncated kernel,} \\
k(x) &= \begin{cases} 1-|x| & \text{for } |x| \le 1 \\ 0 & \text{otherwise} \end{cases} && \text{Bartlett kernel,} \\
k(x) &= \begin{cases} 1-6x^2+6|x|^3 & \text{for } |x| \le \frac{1}{2} \\ 2(1-|x|)^3 & \text{for } \frac{1}{2} < |x| \le 1 \\ 0 & \text{otherwise} \end{cases} && \text{Parzen kernel,} \\
k(x) &= \frac{25}{12\pi^2 x^2}\left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos\left(\frac{6\pi x}{5}\right)\right) && \text{QS kernel.}
\end{aligned}
$$

The estimators of Hansen (1982) and White (1984, p.152) use the truncated kernel; the Newey and West (1987) estimator uses the Bartlett kernel; and the estimator of Gallant (1987, p.533) uses the Parzen kernel. The estimators corresponding to these kernels place zero weights on $\mathbf{\Phi}(\tau)$ for $\tau \ge S_T$, so that $S_T - 1$ is called the lag truncation number. Andrews (1991) advocates an estimator which uses the Quadratic Spectral (QS) kernel, which does not place zero weights on any $\mathbf{\Phi}(\tau)$ for $|\tau| \le T-1$.[3]

One important problem is how to choose the bandwidth parameter S_T. Andrews (1991) provides formulas for the optimal choice of the bandwidth parameter, S_T^*, for a variety of kernels. The S_T^* is optimal in the sense of minimizing the MSE for a given positive semidefinite matrix $\mathbf{W}$:[4]

$$
(6.5)\qquad S_T^* = \begin{cases} 1.1447(\alpha(1)T)^{\frac{1}{3}} & \text{Bartlett kernel} \\ 2.6614(\alpha(2)T)^{\frac{1}{5}} & \text{Parzen kernel} \\ 1.3221(\alpha(2)T)^{\frac{1}{5}} & \text{QS kernel,} \end{cases}
$$

[3]Hansen (1992) relaxes an assumption made by these authors to show the consistency of the kernel estimators.

[4]To be exact, the optimal bandwidth parameter minimizes the asymptotic truncated MSE. See Andrews (1991).

and

$$\alpha(q) = \frac{2(\mathrm{vec}\mathbf{f}^{(q)})'\mathbf{W}\mathrm{vec}\mathbf{f}^{(q)}}{\mathrm{tr}\mathbf{W}(\mathbf{I}+\mathbf{K}_{pp})\mathbf{f}^{(0)}\otimes\mathbf{f}^{(0)}}, \qquad (6.6)$$

$$\mathbf{f}^{(q)} = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}|j|^q\mathbf{\Phi}(\tau),$$

where $\mathbf{W}$ is a $p^2\times p^2$ weight matrix and $\mathbf{K}_{pp}$ is the $p^2\times p^2$ commutation matrix that transforms vec($\mathbf{A}$) into vec($\mathbf{A}'$), i.e., $\mathbf{K}_{pp}=\sum_{i=1}^{p}\sum_{j=1}^{p}\mathbf{e}_i\mathbf{e}_j'\otimes\mathbf{e}_j\mathbf{e}_i'$, where $\mathbf{e}_i$ is the i-th elementary p-vector. Here $\mathbf{f}^{(0)}$ is the spectral density at frequency zero, and the long-run covariance matrix $\mathbf{\Omega}$ is equal to $2\pi\mathbf{f}^{(0)}$. Unfortunately, these formulas include the unknown parameters we wish to estimate. This outcome presents a serious circular problem.

Andrews proposes automatic bandwidth estimators in which these unknown parameters are estimated from the data by a parameterized model. His method involves two steps. The first step is to parameterize the model to estimate the law of motion of $\mathbf{u}_t$. For example, we can use an AR(1) model for each element of $\mathbf{u}_t$ or a VAR(1) model for $\mathbf{u}_t$. The second step is to calculate the parameters for the optimal bandwidth parameter from the estimated law of motion. For example, we calculate the unknown parameters by assuming that the estimated AR(1) model is true. In his Monte Carlo simulations, Andrew uses an AR(1) parameterization for each term of the disturbance, which seems to work well in the models he considers. Newey and West (1994) propose an alternative method based on truncated sums of the sample autocovariances; this method avoids the use of any parametric model.

Another issue is the choice of the kernel. One serious problem with the truncated kernel is that the corresponding estimator is not guaranteed to be positive semidefinite. Andrews (1991) shows that the QS kernel is an optimal kernel in the sense that it minimizes the asymptotic MSE among the kernel estimators that are guaranteed to be positive semidefinite. His Monte Carlo simulations show that the QS kernel and the Parzen kernel work better than the Bartlett kernel in most of the models he considers. He also finds that even the estimators based on the QS kernel and the Parzen kernel are not satisfactory in the sense that the standard errors calculated from these estimators are not accurate in small samples when the amount of autocorrelation is large.

Since the kernel HAC estimators do not seem satisfactory in many cases, Andrews and Monahan (1992) propose an estimator based on VAR prewhitening. The intuition behind this proposition is that the kernel HAC estimators only take care of the MA components of $\mathbf{u}_t$ and cannot handle the AR components well in small samples. The first step in the VAR prewhitening method is to run a VAR of the form

$$\mathbf{u}_t = \mathbf{A}_1\mathbf{u}_{t-1} + \mathbf{A}_2\mathbf{u}_{t-2} + \cdots + \mathbf{A}_p\mathbf{u}_{t-p} + \mathbf{v}_t. \tag{6.7}$$

Note that the model (6.7) need not be a true model in any sense. The estimated VAR is used to form an estimate $\mathbf{v}_t$ and a kernel HAC estimator is applied to the estimated $\mathbf{v}_t$ to estimate the long-run variance of $\mathbf{v}_t$, $\mathbf{\Omega}^*_T$. The estimator based on the QS kernel with the automatic bandwidth parameter can be used to find $\mathbf{v}_t$ for example. Then the sample counterpart of the formula

$$\mathbf{\Omega} = [\mathbf{I} - \sum_{\tau=1}^{p} \mathbf{A}_\tau]^{-1}\mathbf{\Omega}^*[\mathbf{I} - \sum_{\tau=1}^{p} \mathbf{A}'_\tau]^{-1} \tag{6.8}$$

is used to form an estimate of $\mathbf{\Omega}$. Andrews and Monahan use the VAR of order one in their Monte Carlo simulations. Their results suggest that the prewhitened kernel HAC estimator performs better than the non-prewhitened kernel HAC estimators for the purpose of calculating the standard errors of estimators.[5]

In a recent paper, den Haan and Levin (1996) propose a HAC estimator based on (6.7) without using any kernel estimation, which is called the Vector Autoregression Heteroskedasticity and Autocorrelation Consistent (VARHAC) estimator. This estimator has an advantage over any estimator that involves kernel estimation in that the circular problem associated with estimating the optimal bandwidth parameter can be avoided. For the VARHAC estimator, a usual method is to choose the order of AR such as the AIC is applied to (6.7). Then the sample counterpart of (6.8) with $\mathbf{\Omega}^* = E(\mathbf{v}_t\mathbf{v}_t')$ is used to estimate $\mathbf{\Omega}$. Their Monte Carlo evidence indicates that the VARHAC estimator performs better than the non-prewhitened and prewhitened kernel estimators in many cases. On the other hand, Cochrane (1988) basically argues that kernel estimators are better than VARHAC estimators for his purpose of estimating the random walk component as discussed in Chapter 13. Thus, it seems necessary to compare VARHAC estimators with other estimators in different contexts for various applications.

In sum, existing Monte Carlo evidence for estimation of $\mathbf{\Omega}$ recommends VAR prewhitening and either the QS or Parzen kernel estimator together with Andrews' (1991) automatic bandwidth parameter if a kernel HAC estimator is to be utilized. Though the QS kernel estimator may be preferred to the Parzen kernel estimator because of its asymptotic optimality, it takes more time to calculate the QS kernel estimators than the Parzen kernel estimators. This difference may be important when estimation is repeated many times. The VARHAC estimator of den Haan and Levin (1996) seems to have important advantages over estimators involving kernel

[5] Park and Ogaki's (1991) Monte Carlo simulations suggest that the VAR prewhitening improves estimators of $\mathbf{\Omega}$ in the context of cointegrating regressions.

estimation, even though it is a relatively new method, and has more Monte Carlo evidence for various applications.

6.2.2 Known Order of Serial Correlation

In some applications, the order of serial correlation is known in the sense that the economic model implies a particular order. Assume that the order of serial correlation is known to be s.

In this case, there exist the zero restrictions on the autocovariances that $\mathbf{\Phi}(\tau) = \mathbf{0}$ for $|\tau| > s$. Imposing these zero restrictions on the estimator of $\mathbf{\Omega}$ leads to a more efficient estimator.[6] Since $\mathbf{\Omega} = \sum_{\tau=-s}^{s} \mathbf{\Phi}(\tau)$ in this case, a natural estimator is

$$\mathbf{\Omega}_T = \sum_{\tau=-s}^{s} \mathbf{\Phi}_T(\tau), \tag{6.9}$$

which is the truncated kernel estimator.

Hansen and Hodrick (1980) study a multi-period forecasting model that leads to $s \geq 1$. They use (6.9) with conditional homoskedasticity imposed (as discussed at the end of Section 6.1 above). Their method of calculating the standard errors for linear regressions is known as Hansen-Hodrick correction.

A possible problem with the estimator (6.9) is that $\mathbf{\Omega}_T$ is not guaranteed to be positive semidefinite if $s \geq 1$. In applications, researchers often encounter cases where $\mathbf{\Omega}_T$ is invertible but is not positive semidefinite. If this happens, $\mathbf{\Omega}_T$ should not be used to form the optimal GMM estimator (e.g., Newey and West, 1987). There exist at least two ways to handle this problem. One way is to use Eichenbaum, Hansen, and Singleton's (1988) modified Durbin method. The first step of this method is

[6] In some applications, the order of serial correlation may be different for different terms of $\mathbf{u}_t$. The econometrician may wish to impose these restrictions.

to estimate the VAR (6.7) for a large p by solving the Yule Walker equations. The second step is to estimate an MA(s) representation

$$\mathbf{u}_t = \mathbf{B}_1\mathbf{v}_{t-1} + ... + \mathbf{B}_s\mathbf{v}_{t-s} + \mathbf{e}_t, \tag{6.10}$$

by regressing the estimated $\mathbf{u}_t$ on estimated lagged $\mathbf{v}_t$'s. Then the sample counterpart of

$$\mathbf{\Omega} = (\mathbf{I} + \mathbf{B}_1 + ... + \mathbf{B}_s)E(\mathbf{e}_t\mathbf{e}_t')(\mathbf{I} + \mathbf{B}_1 + ... + \mathbf{B}_s)' \tag{6.11}$$

is used to form an estimate of $\mathbf{\Omega}$ that imposes the zero restrictions. This method is not reliable when the number of elements in $\mathbf{u}_t$ is large relative to the sample size because too many parameters in (6.7) need to be estimated. The number of elements in $\mathbf{u}_t$ need to be kept as small as possible when using this method.

Another method uses one of the kernel HAC estimators (or VAR prewhitened kernel estimators if s is large) that is guaranteed to be positive semidefinite. When employing this method, the zero restrictions should *not* be imposed even though $\mathbf{\Phi}(\tau)$ is known to be zero for $|\tau| > s$. In order to illustrate this method in a simple example, consider the case where $s = 1$ and Newey and West's (1987) Bartlett kernel estimator is used. Then

$$\mathbf{\Omega}_T = \sum_{\tau=-\ell}^{\ell} \frac{S_T - |\tau|}{S_T}\mathbf{\Phi}_T(\tau), \tag{6.12}$$

where $\ell = S_T - 1$ is the lag truncation number. If $\ell = 1$ is used to impose the zero restrictions, then $\mathbf{\Omega}_T$ converges to $\mathbf{\Phi}(0) + \frac{1}{2}\mathbf{\Phi}(1) + \frac{1}{2}\mathbf{\Phi}(-1)$, which is not equal to $\mathbf{\Omega} = \mathbf{\Phi}(0) + \mathbf{\Phi}(1) + \mathbf{\Phi}(-1)$. Thus ℓ must increase as T increases to obtain a consistent estimator. On the other hand, if $\ell > 1$ is used and the zero restrictions are imposed by setting $\mathbf{\Phi}_T(\tau)$ in (6.6) equal to zero for $|\tau| > 1$, then the resulting estimator is no longer guaranteed to be positive semidefinite.

In this chapter, we focused on consistent estimators for the long-run covariance matrix. Recently, some researchers have pointed out that we may not need to have consistent estimators for some purposes such as computing standard errors for regression estimators or computing Wald tests. For example, small sample properties of Wald tests computed form inconsistent estimates of the long-run covariance matrix may be better for some data generating processes. See Kiefer, Vogelsang, and Bunzel (2000), Kiefer and Vogelsang (2002a,b), and Müller (2004).

Exercises

6.1 *(The Multi-Period Forecasting Model)* Suppose that I_t is an information set generated by $\{\mathbf{Y}_t, \mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, \cdots\}$, where $\mathbf{Y}_t$ is a stationary and ergodic vector stochastic process. Economic agents are assumed to use current and past $\mathbf{Y}_t$ to generate their information set. Let X_t be a stationary and ergodic random variable in the information set I_t with $E(|X_t|^2) < \infty$. We consider a 3-period forecast of X_t, $E(X_{t+3}|\mathrm{I}_t)$, and the forecast error, $e_t = X_{t+3} - E(X_{t+3}|\mathrm{I}_t)$.

(**a**) Give an expression for the long-run variance of e_t. Which methods do you suggest to use in order to estimate the long-run variance?

(**b**) Let $\mathbf{Z}_t$ be a random vector with finite second moments in the information set I_t. Define $\mathbf{f}_t = \mathbf{Z}_t e_t$ Give an expression for the long covariance matrix of $\mathbf{f}_t$. Which methods do you suggest to use in order to estimate the long-run variance?

References

Andrews, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59(3), 817–858.

Andrews, D. W. K., and J. C. Monahan (1992): "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica*, 60(4), 953–966.

Cochrane, J. H. (1988): "How Big is the Random Walk in GNP?," *Journal of Political Economy*, 96(5), 893–920.

den Haan, W. J., and A. Levin (1996): "Inference From Parametric And Non-Parametric Covariance Matrix Estimation Procedures," NBER Technical Working Paper No. 195.

Eichenbaum, M., L. P. Hansen, and K. J. Singleton (1988): "A Time Series Analysis of Representative Agent Models of Consumption and Leisure Choice under Uncertainty," *Quarterly Journal of Economics*, 103, 51–78.

Ferson, W. E., and S. R. Foerster (1994): "Finite Sample Properties of the Generalized Methods of Moments in Tests of Conditional Asset Pricing Models," *Journal of Financial Economics*, 36, 29–55.

Gallant, A. R. (1987): *Nonlinear Statistical Models*. John Wiley and Sons, New York.

Hansen, B. E. (1992): "Consistent Covariance Matrix Estimation for Dependent Heterogeneous Processes," *Econometrica*, 60(4), 967–972.

Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50(4), 1029–1054.

Hansen, L. P., and R. J. Hodrick (1980): "Forward Exchange Rates as Optimal Predictors of Future Spot Rates," *Journal of Political Economy*, 88, 829–853.

Kiefer, N. M., and T. J. Vogelsang (2002a): "Heteroskedasticity-Autocorrelation Robust Standard Errors Using the Bartlett Kernel Without Truncation," *Econometrica*, 70, 2093–2095.

——— (2002b): "Heteroskedasticity-Autocorrelation Robust Testing Using Bandwidth Equal to Sample Size," *Econometric Theory*, 18, 1350–1366.

Kiefer, N. M., T. J. Vogelsang, and H. Bunzel (2000): "Simple Robust Testing of Regression Hypotheses," *Econometrica*, 68(3), 695–714.

Müller, U. K. (2004): "A Theory of Robust Long-Run Variance Estimation," Manuscript, Princeton University.

Newey, W. K., and K. D. West (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55(3), 703–708.

——— (1994): "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, 61, 631–653.

Park, J. Y., and M. Ogaki (1991): "VAR Prewhitening to Estimate Short-Run Dynamics: On Improved Method of Inference in Cointegrated Models," RCER Working Paper No. 281.

White, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48(4), 817–838.

——— (1984): *Asymptotic Theory For Econometricians*. Academic Press, New York.