LECTURE 8

LINEAR REGRESSION WITH WEAKLY DEPENDENT DATA

Consider the usual regression model $Y_t = X_t'\beta + U_t$, where $\beta \in R^k$ is unknown vector of parameters and the data consists of weakly dependent observations. The LS estimator of β is $\widehat{\beta}_n = \left(\sum_{t=1}^n X_t X_t'\right)^{-1} \sum_{t=1}^n X_t Y_t$. In this lecture, we discuss consistency, asymptotic normality and estimation of the asymptotic variance of $\widehat{\beta}_n$.

Consistency

We make the following assumptions.

(a) $\{(X_t', U_t)\}$ is a mixing sequence with ϕ of size $-r/(2r-1)$, $r \geq 1$, or α of size $-r/(r-1)$, $r > 1$.

(b) $EX_tU_t = 0$ for all t.

(c) $\sup_t E|X_{tj}|^{2r+\delta} < \Delta < \infty$ for some $\delta > 0$, and all $j = 1, \ldots, k$.

(d) $\sup_t E|U_t|^{2r+\delta} < \Delta < \infty$ for some $\delta > 0$.

(e) $M_n = n^{-1}\sum_{t=1}^n EX_tX_t'$ is uniformly positive definite for large n.

In order to show consistency of $\widehat{\beta}_n$, write, as usual,

$$\widehat{\beta}_n = \beta + \left(n^{-1}\sum_{t=1}^n X_t X_t'\right)^{-1} n^{-1}\sum_{t=1}^n X_t U_t.$$

Due to assumption (a) and Theorem 6 in Lecture 7, for $i, j = 1, \ldots, k$ we have that $\{X_{ti}X_{tj}\}$ and $\{X_{ti}U_t\}$ are mixing with ϕ of size $-r/(2r-1)$, $r \geq 1$, or α of size $-r/(r-1)$, $r > 1$. Next, set $\varepsilon = \delta/2$. By the Cauchy-Schwartz inequality, for all t,

$$E|X_{ti}X_{tj}|^{r+\varepsilon} \leq \sqrt{E|X_{ti}|^{2r+\delta} E|X_{tj}|^{2r+\delta}}.$$

Therefore, $\sup_t E|X_{ti}X_{tj}|^{r+\varepsilon} < \Delta < \infty$ for some $\varepsilon > 0$. Similarly, we can bound $\sup_t E|X_{ti}U_t|^{r+\varepsilon}$. Hence, by the SLLN,

$$n^{-1}\sum_{t=1}^n X_tX_t' - M_n \to_{a.s.} 0, \tag{1}$$

$$n^{-1}\sum_{t=1}^n X_tU_t \to_{a.s.} 0.$$

Next, as it was discussed in the proof of Corollary 2, Lecture 7, due to Assumption (e) $M_n^{-1} = O(1)$. Thus,

$$\begin{aligned}
\widehat{\beta}_n - \beta &= \left(\left(n^{-1}\sum_{t=1}^n X_tX_t' - M_n\right) + M_n\right)^{-1} n^{-1}\sum_{t=1}^n X_tU_t \\
&= \left(M_n^{-1}\left(n^{-1}\sum_{t=1}^n X_tX_t' - M_n\right) + I\right)^{-1} M_n^{-1} n^{-1}\sum_{t=1}^n X_tU_t \qquad (2)\\
&= \left(O(1)o_{a.s}(1) + I\right)^{-1} O(1)o_{a.s.}(1).
\end{aligned}$$

It follows that $\widehat{\beta}_n$ is a consistent estimator of β.

Equivalently, one can use the following extension of the Slutsky's theorem (Propositions 2.16 and 2.30 in White (1999)).

Lemma 1 *Let $g : R^k \to R^l$ be continuous on a compact set $C \subset R^k$. Suppose that $\{b_n\}$ is a sequence of random k-vectors, and $\{c_n\}$ is a sequence of k-vectors such that $b_n - c_n \to_{a.s.} 0$ ($b_n - c_n \to_p 0$), and, for all n sufficiently large, c_n is interior point to C uniformly in n. Then $g(b_n) - g(c_n) \to_{a.s.} 0$ ($g(b_n) - g(c_n) \to_p 0$).*

This result does not require that $\{c_n\}$ is a convergent sequence. The sequence $\{b_n\}$ does not necessary converge either, but, for large n, it follows the behavior of $\{c_n\}$. Since due to assumption (e), for n large enough M_n^{-1} exists, we have

$$\left(n^{-1}\sum_{t=1}^{n} X_tX_t'\right)^{-1} n^{-1}\sum_{t=1}^{n} X_tU_t = \left(n^{-1}\sum_{t=1}^{n} X_tX_t'\right)^{-1} n^{-1}\sum_{t=1}^{n} X_tU_t - M_n^{-1}0$$
$$\to_{a.s.} 0.$$

Remark. Assumption (b) above implies that the linear projection of Y_t against X_t is the same for all t's:

$$\begin{aligned} 0 = EX_tU_t &= EX_t\left(Y_t - X_t'\beta\right) \\ &= EX_tY_t - EX_tX_t'\beta, \end{aligned}$$

or

$$\beta = \left(EX_tX_t'\right)^{-1} EX_tY_t \text{ for all } t.$$

Now, suppose we replace Assumption (b) with the following weaker condition:

$$\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} EX_tU_t = 0. \tag{3}$$

This condition can be interpreted as X and U being uncorrelated only on average in the long-run (as opposed to exact period-by-period uncorrelatedness of Assumption (b)). We can still show that $\hat{\beta}_n \to_p \beta$, however, the meaning of β changes. Suppose that the limits of $n^{-1}\sum_{t=1}^{n} EX_tX_t'$ and $n^{-1}\sum_{t=1}^{n} EX_tY_t$ exist:

$$\begin{aligned} 0 = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} EX_tU_t &= \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} EX_t\left(Y_t - X_t'\beta\right) \\ &= \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} EX_tY_t - \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} EX_tX_t'\beta, \end{aligned}$$

or

$$\beta = \left(\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} EX_tX_t'\right)^{-1} \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} EX_tY_t.$$

Thus, in this case β can be interpreted as the average long-run projection of Y_t against X_t. To show consistency under (3), first let

$$\begin{aligned} B_n &= \left(M_n^{-1}\left(n^{-1}\sum_{t=1}^{n} X_tX_t' - M_n\right) + I\right)^{-1} M_n^{-1} \\ &= O_{a.s.}(1). \end{aligned}$$

From (2) we have,

$$\begin{aligned} \hat{\beta}_n - \beta &= B_n n^{-1}\sum_{t=1}^{n} X_tU_t \\ &= B_n n^{-1}\sum_{t=1}^{n}\left(X_tU_t - EX_tU_t\right) + B_n n^{-1}\sum_{t=1}^{n} EX_tU_t \\ &= O_{a.s.}(1)\, o_{a.s.}(1) + O_{a.s.}(1)\, o(1), \end{aligned}$$

where $o_{a.s.}(1)$ in the first term is due to the SLLN and $o(1)$ in the second term is due to (3).

Asymptotic normality

For the asymptotic normality, we replace Assumption (a), (c) and (d) with

(a*) $\{(X_t', U_t)\}$ is a mixing sequence with ϕ of size $-r/(r-1)$, $r \geq 2$, or α of size $-2r/(r-2)$, $r > 2$.

(c*) $\sup_t EX_{tj}^{2r} < \Delta < \infty$ for all $j = 1, \ldots, k$.

(d*) $\sup_t EU_t^{2r} < \Delta < \infty$.

Note that (c*) and (d*) are stronger than (c) and (d) because in (a*) we require that $r \geq 2$ or $r > 2$. Assumption (a*) is stronger than what is required for the CLT, since we would need to insure that Ω_n is bounded (see equation (4) below). In addition, we make the following assumption.

(f) $\Omega_n = Var\left(n^{-1/2} \sum_{t=1}^n X_t U_t\right)$ is uniformly positive definite.

We will assume that condition (b), $EX_tU_t = 0$ for all t, holds. Define

$$\begin{aligned} V_n &= M_n^{-1} \Omega_n M_n^{-1} \\ &= \left(n^{-1} \sum_{t=1}^n EX_t X_t'\right)^{-1} Var\left(n^{-1/2} \sum_{t=1}^n X_t U_t\right) \left(n^{-1} \sum_{t=1}^n EX_t X_t'\right)^{-1}. \end{aligned}$$

Next, we will show that under the above assumptions,

$$V_n^{-1/2} n^{1/2} \left(\widehat{\beta}_n - \beta\right) \to_d N(0, I_k),$$

where

$$V_n^{-1/2} = \Omega_n^{-1/2} M_n,$$

the matrix square root of $M_n \Omega_n^{-1} M_n$. (This is Exercise 5.21 and Theorem 4.25 in White (1999). However, the moment conditions are stated differently, and the assumption on the mixing coefficients is stronger. The stronger mixing assumption is to ensure that $\Omega_n = O(1)$.)

First, consider

$$\begin{aligned} & V_n^{-1/2} n^{1/2} \left(\widehat{\beta}_n - \beta\right) - V_n^{-1/2} M_n^{-1} n^{-1/2} \sum_{t=1}^n X_t U_t \\ = \; & V_n^{-1/2} \left(\left(n^{-1} \sum_{t=1}^n X_t X_t'\right)^{-1} - M_n^{-1}\right) n^{-1/2} \sum_{t=1}^n X_t U_t \\ = \; & V_n^{-1/2} \left(\left(n^{-1} \sum_{t=1}^n X_t X_t'\right)^{-1} - M_n^{-1}\right) \Omega_n^{1/2} \Omega_n^{-1/2} n^{-1/2} \sum_{t=1}^n X_t U_t. \end{aligned} \tag{4}$$

Because of Assumptions (a*) and (d*), (1) holds, and due to Assumption (e) and Lemma 1 we have

$$\left(\left(n^{-1} \sum_{t=1}^n X_t X_t'\right)^{-1} - M_n^{-1}\right) = o_{a.s.}(1). \tag{5}$$

Next, by Assumptions (c*) and (d*) we have that

$$\begin{aligned} \sup_t E\left|X_{tj} U_t\right|^r &\leq \sqrt{\sup_t EX_{tj}^{2r} \sup_t EU_t^{2r}} \\ &< \Delta. \end{aligned}$$

Therefore, Assumptions (b), (f) and the vector CLT for weakly dependent processes (Corollary 2 in Lecture 7) imply that

$$\Omega_n^{-1/2} n^{-1/2} \sum_{t=1}^{n} X_t U_t \to_d N\left(0, I_k\right).$$

Hence,

$$\Omega_n^{-1/2} n^{-1/2} \sum_{t=1}^{n} X_t U_t = O_p(1). \tag{6}$$

(Let X_n be a sequence of scalars such that $X_n \to_d X$. This implies that for all large n and $\delta > 0$, $\left|P\left(|X_n| > \Delta_\delta\right) - P\left(|X| > \Delta_\delta\right)\right| < \delta$ provided that Δ_δ and $-\Delta_\delta$ are continuity points of the distribution of X. This in turn implies that $P\left(|X_n| > \Delta_\delta\right) < P\left(|X| > \Delta_\delta\right) + \delta$. However, $P\left(|X| > \Delta_\delta\right) < \delta$ if we choose Δ_δ large enough. Therefore, $P\left(|X_n| > \Delta_\delta\right) < 2\delta$ for all large n. This establishes that if $X_n \to_d X$ then $X_n = O_p(1)$. This is Lemma 4.5 in White (1999).)

Given Assumption (a*), Lemma 3 on page 9 of Lecture 7 implies that

$$\Omega_n = O(1). \tag{7}$$

Assumptions (c*) and (f) together imply that

$$V_n^{-1/2} = O(1). \tag{8}$$

The result in (8) holds because $V_n^{-1/2} = \Omega_n^{-1/2} M_n$, $\Omega_n^{-1} = O\left(1\right)$ because Ω_n is uniformly positive definite by Assumption (f); $M_n = O\left(1\right)$ because for its element i, j we have

$$\begin{aligned}
\left|[M_n]_{ij}\right| &= \left|n^{-1} \sum_{t=1}^{n} E X_{ti} X_{tj}\right| \\
&\leq n^{-1} \sum_{t=1}^{n} \left|E X_{ti} X_{tj}\right| \\
&\leq n^{-1} \sum_{t=1}^{n} \left(E X_{ti}^2\right)^{\frac{1}{2}} \left(E X_{tj}^2\right)^{\frac{1}{2}} \\
&\leq n^{-1} \sum_{t=1}^{n} \left(E\left|X_{ti}\right|^{2r}\right)^{\frac{1}{2r}} \left(E\left|X_{tj}\right|^{2r}\right)^{\frac{1}{2r}} \\
&\leq \left(\sup_t E\left|X_{ti}\right|^{2r}\right)^{\frac{1}{2r}} \left(\sup_t E\left|X_{tj}\right|^{2r}\right)^{\frac{1}{2r}} \\
&< \Delta^{\frac{1}{r}} < \infty,
\end{aligned}$$

where the inequality in the third line is by Cauchy-Schwartz, and the inequality in the fourth line is by the norm inequality since $r \geq 2$ (see Davidson (1994) page 138).

Now, it follows from (4) and (5)-(8) that

$$V_n^{-1/2} n^{1/2} \left(\widehat{\beta}_n - \beta\right) - V_n^{-1/2} M_n^{-1} n^{-1/2} \sum_{t=1}^{n} X_t U_t \to_p 0.$$

Consequently, the asymptotic distribution of

$$V_n^{-1/2} n^{1/2} \left(\widehat{\beta}_n - \beta\right)$$

is the same as that of

$$V_n^{-1/2} M_n^{-1} n^{-1/2} \sum_{t=1}^{n} X_t U_t.$$

Lastly,

$$
\begin{aligned}
& V_n^{-1/2} M_n^{-1} n^{-1/2} \sum_{t=1}^{n} X_t U_t \\
& = \Omega_n^{-1/2} n^{-1/2} \sum_{t=1}^{n} X_t U_t \\
& \rightarrow_d N(0, I_k).
\end{aligned}
$$

Remark. We can also show asymptotic normality when the condition $EX_tU_t = 0$ does not hold for every period t. In that case, we replace the average uncorrelatedness in (3) with a somewhat stronger condition:

$$
n^{-1/2} \sum_{t=1}^{n} EX_t U_t = o(1).
$$

Now, write:

$$
\begin{aligned}
& V_n^{-1/2} n^{1/2} \left(\hat{\beta}_n - \beta - M_n^{-1} n^{-1} \sum_{t=1}^{n} EX_t U_t \right) \\
= \; & V_n^{-1/2} \left(n^{-1} \sum_{t=1}^{n} X_t X_t' \right)^{-1} n^{-1/2} \sum_{t=1}^{n} X_t U_t - V_n^{-1/2} M_n^{-1} n^{-1/2} \sum_{t=1}^{n} EX_t U_t \\
= \; & V_n^{-1/2} \left(n^{-1} \sum_{t=1}^{n} X_t X_t' \right)^{-1} n^{-1/2} \sum_{t=1}^{n} (X_t U_t - EX_t U_t) \\
& + V_n^{-1/2} \left(\left(n^{-1} \sum_{t=1}^{n} X_t X_t' \right)^{-1} - M_n^{-1} \right) n^{-1/2} \sum_{t=1}^{n} EX_t U_t \\
= \; & V_n^{-1/2} \left(n^{-1} \sum_{t=1}^{n} X_t X_t' \right)^{-1} n^{-1/2} \sum_{t=1}^{n} (X_t U_t - EX_t U_t) \\
& + V_n^{-1/2} \left(\left(n^{-1} \sum_{t=1}^{n} X_t X_t' \right)^{-1} - \left(n^{-1} \sum_{t=1}^{n} EX_t X_t' \right)^{-1} \right) n^{-1/2} \sum_{t=1}^{n} EX_t U_t.
\end{aligned}
$$

The second term in the above expression is asymptotically negligible. For the first term, by the same argument as in (4), one can show that

$$
V_n^{-1/2} \left(\left(n^{-1} \sum_{t=1}^{n} X_t X_t' \right)^{-1} - M_n^{-1} \right) n^{-1/2} \sum_{t=1}^{n} (X_t U_t - EX_t U_t) = o_p(1).
$$

Since

$$
V_n^{-1/2} M_n^{-1} n^{-1/2} \sum_{t=1}^{n} (X_t U_t - EX_t U_t) \rightarrow_d N(0, I_k),
$$

we therefore obtain that

$$
V_n^{-1/2} n^{1/2} \left(\hat{\beta}_n - \beta - M_n^{-1} n^{-1} \sum_{t=1}^{n} EX_t U_t \right) \rightarrow_d N(0, I_k).
$$

However,

$$
V_n^{-1/2} n^{1/2} \left(\hat{\beta}_n - \beta - M_n^{-1} n^{-1} \sum_{t=1}^{n} EX_t U_t \right) = V_n^{-1/2} n^{1/2} (\hat{\beta}_n - \beta) - \Omega_n^{-1/2} n^{-1/2} \sum_{t=1}^{n} EX_t U_t
$$

$$= V_n^{-1/2} n^{1/2} (\hat{\beta}_n - \beta) + o(1).$$

We now conclude that $V_n^{-1/2} n^{1/2}(\hat{\beta}_n - \beta) \to_d N(0, I_k)$.
Note that to avoid asymptotic bias, it is insufficient to assume that the errors and regressors are uncorrelated on average ($n^{-1}\sum_{t=1}^n EX_tU_t = o(1)$) as in that case one can still have non-negligible asymptotic bias. For example, if $\Omega_n^{-1/2} n^{-1/2} \sum_{t=1}^n EX_tU_t \to \delta \neq 0$ for some $\delta \in R^k$, then $V_n^{-1/2} n^{1/2}(\hat{\beta}_n - \beta) \to N(\delta, I_k)$.

Estimation of asymptotic variance matrix

Consistent estimation of V_n is required for hypothesis testing. Due to (1), we have a natural estimator for M_n, $n^{-1}\sum_t X_tX_t'$. Suppose that there exists $\widehat{\Omega}_n$ such that

$$\Omega_n - \widehat{\Omega}_n \to_p 0,$$

and, therefore, for n sufficiently large, $\widehat{\Omega}_n$ is positive definite. We say that such $\widehat{\Omega}_n$ is a consistent estimator of Ω_n. Note that Ω_n allows for general form of heteroskedasticity and autocorrelation for $\{X_tU_t\}$. A consistent estimator of Ω_n is referred as heteroskedasticity and autocorrelation consistent (HAC). Define

$$\widehat{V}_n = \left(n^{-1} \sum_t X_tX_t' \right)^{-1} \widehat{\Omega}_n \left(n^{-1} \sum_t X_tX_t' \right)^{-1}.$$

It follows from Lemma 1 that

$$V_n - \widehat{V}_n \to_p 0.$$

Further, set

$$\widehat{V}_n^{-1/2} = \widehat{\Omega}_n^{-1/2} \left(n^{-1} \sum_t X_tX_t' \right).$$

Then,

$$\begin{aligned} & \widehat{V}_n^{-1/2} n^{1/2} \left(\widehat{\beta}_n - \beta \right) - V_n^{-1/2} n^{1/2} \left(\widehat{\beta}_n - \beta \right) \\ = & \left(\widehat{V}_n^{-1/2} - V_n^{-1/2} \right) V_n^{1/2} \left(V_n^{-1/2} n^{1/2} \left(\widehat{\beta}_n - \beta \right) \right) \\ = & o_p(1) O(1) O_p(1) \\ = & o_p(1), \end{aligned}$$

where $V_n = O(1)$, since both Ω_n and M_n^{-1} are $O(1)$ as it was discussed above. Therefore,

$$\widehat{V}_n^{-1/2} n^{1/2} \left(\widehat{\beta}_n - \beta \right) \to_d N\left(0, I_k\right).$$

As usual, inference about β can be based on a Wald statistic. Under $H_0 : \beta = \beta_0$,

$$\begin{aligned} W_n &= n \left(\widehat{\beta}_n - \beta_0 \right)' \widehat{V}_n^{-1} \left(\widehat{\beta}_n - \beta_0 \right) \\ &\to_d \chi_k^2. \end{aligned}$$

Next, we consider estimation of Ω_n:

$$\begin{aligned} \Omega_n &= \Omega_n(0) + \sum_{j=1}^{n-1} \left(\Omega_n(j) + \Omega_n(j)' \right), \text{ where} \\ \Omega_n(j) &= n^{-1} \sum_{t=j+1}^{n} E\left(X_tU_t \right) \left(X_{t-j}U_{t-j} \right)'. \end{aligned}$$

A consistent estimator was suggested, for example, by Newey and West (1987).

We will proceed in two steps. First, we will suggest a consistent but infeasible estimator that uses the true disturbances U_t. In the second step, we will argue that the estimator is still consistent when U_t is replaced by $\widehat{U}_t = Y_t - X_t'\widehat{\beta}_n$. For $j = 0, 1, \ldots,$ let

$$\widetilde{\Omega}_n(j) = n^{-1} \sum_{t=j+1}^{n} (X_t U_t)(X_{t-j} U_{t-j})'.$$

The mixing and moments conditions that we made imply that for all (fixed) j's:

$$\Omega_n(j) - \widetilde{\Omega}_n(j) \to_p 0.$$

Therefore, one may consider the following estimator of Ω_n:

$$\widetilde{\Omega}_n(0) + \sum_{j=1}^{n-1} \left(\widetilde{\Omega}_n(j) + \widetilde{\Omega}_n(j)' \right).$$

However, it turns out that such an estimator is not consistent. The reason for that is that the number of estimated autocovariances $\widetilde{\Omega}_n(j)$ grows too fast with n. A solution is to allow the number of autocovariances to grow with n, but at a slower rate. Let

$$\widetilde{\Omega}_n = \widetilde{\Omega}_n(0) + \sum_{j=1}^{m_n} w(j, m_n) \left(\widetilde{\Omega}_n(j) + \widetilde{\Omega}_n(j)' \right).$$

In the above expression, $w(j, m_n)$ represents the weight assigned to autocovariance term j. The weight ensure that the resulting estimator is positive semi-definite. If $w(j, m_n) = 1$, the resulting matrix does not have to be positive semi-definite. An example of weights that yield positive semi-definite $\widetilde{\Omega}_n$ is

$$w(j, m) = 1 - \frac{j}{m+1}$$

for $1 \leq j \leq m$, and zero otherwise. Such weights are called the Bartlett weights or kernel. The number of autocovariance elements m_n is called the bandwidth or lag truncation parameter, since the autocovariances with $j > m_n$ receive zero weights.

Assume the following assumptions.

(c)** $\sup_t E|X_{tj}|^{4r+\delta} < \Delta < \infty$ for some $\delta > 0$ and all $j = 1, \ldots, k$.

(d)** $\sup_t E|U_t|^{4r+\delta} < \Delta < \infty$.

(g) $|w(j, m_n)| \leq C < \infty$ for all j.

(h) $\lim_{m_n \to \infty} w(j, m_n) = 1$ for each j.

(i) $m_n \to \infty$, and $m_n / n^{1/4} \to 0$.

Assumption (g) says that the weight are uniformly bounded in j. Assumption (h) implies that asymptotically all autocovariances receive the unit weight. Assumption (i) says that the number of autocovariances used increases with the sample size, however, at the rate slower than n. The rate of growth $n^{1/4}$ is not optimal. Andrews (1991) establishes consistency for $m_n = o(n)$. Thus, it is sufficient that the lag truncation parameter grows at the rate just slower than n. However, he shows that the optimal rates for various weighting schemes are typically slower than $n^{1/2}$. For example, for the Bartlett kernel it is $n^{1/3}$, which is faster than $n^{1/4}$ assumed here.

Lemma 2 *Under Assumptions (a*), (b), (c**), (d**), (g)-(i),* $\widetilde{\Omega}_n - \Omega_n \to_p 0$.

Proof. Note that $\widetilde{\Omega}_n - \Omega_n \to_p 0$ if and only if $c'\widetilde{\Omega}_n c - c'\Omega_n c \to_p 0$ for all $c \in R^k$. Let

$$h_t = c' X_t U_t.$$

Then $\{h_t\}$ is mixing of the same size as in (a*), $Eh_t = 0$. Further,

$$\begin{aligned} c'\Omega_n c &= n^{-1}\sum_{t=1}^{n} Eh_t^2 + 2\sum_{j=1}^{n-1} n^{-1}\sum_{t=j+1}^{n} Eh_t h_{t-j}, \\ c'\widetilde{\Omega}_n c &= n^{-1}\sum_{t=1}^{n} h_t^2 + 2\sum_{j=1}^{m_n} w(j,m_n)\, n^{-1}\sum_{t=j+1}^{n} h_t h_{t-j}, \end{aligned}$$

and, therefore,

$$c'\widetilde{\Omega}_n c - c'\Omega_n c = R_{n,0} + 2R_{n,1} + 2R_{n,2} - 2R_{n,3},$$

where

$$\begin{aligned} R_{n,0} &= n^{-1}\sum_{t=1}^{n}\left(h_t^2 - Eh_t^2\right), \\ R_{n,1} &= \sum_{j=1}^{m_n} w(j,m_n)\, n^{-1}\sum_{t=j+1}^{n}\left(h_t h_{t-j} - Eh_t h_{t-j}\right), \\ R_{n,2} &= \sum_{j=1}^{m_n}\left(w(j,m_n) - 1\right) n^{-1}\sum_{t=j+1}^{n} Eh_t h_{t-j}, \\ R_{n,3} &= \sum_{j=m_n+1}^{n-1} n^{-1}\sum_{t=j+1}^{n} Eh_t h_{t-j}. \end{aligned}$$

We need to show that each one of the four terms converge to zero in the appropriate sense.

First, since $r \geq 2$,

$$\begin{aligned} \sup_t E\left|h_t^2\right|^{r+\delta/4} &\leq \sqrt{\sup_t E\left|U_t\right|^{4r+\delta}\sup_t E\left|c'X_t\right|^{4r+\delta}} \\ &\leq \sqrt{\sup_t E\left|U_t\right|^{4r+\delta}\left(\sum_{j=1}^{k}\left|c_j\right|\left(\sup_t E\left|X_{tj}\right|^{4r+\delta}\right)^{1/(4r+\delta)}\right)^{4r+\delta}} \qquad (9) \\ &\leq \Delta\left(\sum_{j=1}^{k}\left|c_j\right|\right)^{2r+\delta/2} < \infty. \end{aligned}$$

Therefore,

$$\begin{aligned} R_{n,0} &= n^{-1}\sum_{t=1}^{n}\left(h_t^2 - Eh_t^2\right) \\ &\to_{a.s.} 0, \end{aligned}$$

as it follows from the SLLN for weakly dependent processes (Theorem 7 in Lecture 7).

By the same argument as in the proof of Lemma 3 in Lecture 7, Assumption (a*) implies that there exists some constants K and $\varepsilon > 0$, such that

$$n^{-1}\sum_{t=j+1}^{n} E\left|h_t h_{t-j}\right| \leq K j^{-1-\varepsilon} \text{ for some } \varepsilon > 0. \qquad (10)$$

Hence,

$$\begin{aligned}|R_{n,3}| &\leq K \sum_{j=m_n+1}^{n-1} j^{-1-\varepsilon} \\ &\to 0,\end{aligned}$$

since $m_n \to \infty$. To show that, one can approximate the sum by an integral as follows. Consider a series $\{a_j : j \geq m\}$ such that $a_j \geq 0$ for all j's and $a_{j+1} < a_j$. Recall that, if we can find a continuous function $f(x)$ such that $f(j) = a_j$, then

$$\sum_{j=m+1}^{\infty} a_j \leq \int_m^{\infty} f(x)\, dx \leq \sum_{j=m}^{\infty} a_j,$$

where $\sum_{j=m}^{\infty} a_j = \sum_{j=m+1}^{\infty} a_{j-1}(j-(j-1))$ is the upper Riemann sum, and $\sum_{j=m+1}^{\infty} a_j = \sum_{j=m+1}^{\infty} a_j(j-(j-1))$ is the lower Riemann sum. Hence,

$$\begin{aligned}\sum_{j=m_n+1}^{n-1} \frac{1}{j^{1+\varepsilon}} &< \int_{m_m}^{\infty} \frac{dx}{x^{1+\varepsilon}} \\ &= \frac{1}{\varepsilon m_n^{\varepsilon}} \\ &\to 0.\end{aligned}$$

Next, from (10) we have

$$|R_{n,2}| \leq K \sum_{j=1}^{m_n} |w(j, m_n) - 1|\, j^{-1-\varepsilon}.$$

Since $|w(j, m_n) - 1| \leq C + 1$ for all j, and $\lim_{n\to\infty} \sum_{j=1}^{m_n} j^{-1-\varepsilon}$ is finite, by the dominated convergence theorem we can bring the limit inside the sum, so that due to Assumption (h),

$$\begin{aligned}\lim_{n\to\infty} |R_{n,2}| &\leq K \sum_{j=1}^{\infty} \lim_{n\to\infty} |w(j, m_n) - 1|\, j^{-1-\varepsilon} \\ &= 0.\end{aligned}$$

It remains to show that $R_{n,1} \to_p 0$. Let

$$Z_{j,t} = h_t h_{t-j} - E h_t h_{t-j}. \tag{11}$$

Using the fact that $|w(j, m_n)| \leq C$, we obtain

$$\begin{aligned}&P(|R_{n,1}| > \varepsilon) \\ = \;& P\left(\left|\sum_{j=1}^{m_n} w(j, m_n)\, n^{-1} \sum_{t=j+1}^{n} Z_{j,t}\right| > \varepsilon\right) \\ \leq \;& P\left(\sum_{j=1}^{m_n} |w(j, m_n)|\, n^{-1} \left|\sum_{t=j+1}^{n} Z_{j,t}\right| > \varepsilon\right) \\ \leq \;& P\left(\sum_{j=1}^{m_n} n^{-1} \left|\sum_{t=j+1}^{n} Z_{j,t}\right| > \frac{\varepsilon}{C}\right) \\ \leq \;& P\left(n^{-1}\left|\sum_{t=2}^{n} Z_{1,t}\right| > \frac{\varepsilon}{Cm_n} \text{ or } n^{-1}\left|\sum_{t=3}^{n} Z_{2,t}\right| > \frac{\varepsilon}{Cm_n} \text{ or } \ldots \text{ or } n^{-1}\left|\sum_{t=m_n+1}^{n} Z_{m_n,t}\right| > \frac{\varepsilon}{Cm_n}\right)\end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{j=1}^{m_n} P\left(n^{-1}\left|\sum_{t=j+1}^{n} Z_{j,t}\right| > \frac{\varepsilon}{Cm_n}\right) \\
&\leq \frac{m_n^2 C^2}{n^2\varepsilon^2}\sum_{j=1}^{m_n} E\left|\sum_{t=j+1}^{n} Z_{j,t}\right|^2, \qquad (12)
\end{aligned}
$$

where the last inequality is by Chebyshev's. Next, we will show that there exists a constant Δ^* such that

$$
E\left|\sum_{t=j+1}^{n} Z_{j,t}\right|^2 \leq \Delta^* n\,(j+2)\,. \qquad (13)
$$

We will show that the result holds in the case of uniform mixing. For the strong mixing case, the proof is identical. Write

$$
\begin{aligned}
E\left|\sum_{t=j+1}^{n} Z_{j,t}\right|^2 &= \sum_{t=j+1}^{n} EZ_{j,t}^2 + 2\sum_{l=1}^{n-j-1}\sum_{t=j+1+l}^{n} EZ_{j,t}Z_{j,t-l} \\
&\leq \sum_{t=j+1}^{n} EZ_{j,t}^2 + 2\sum_{l=1}^{n-j-1}\sum_{t=j+1+l}^{n} \left|EZ_{j,t}Z_{j,t-l}\right| \\
&= \sum_{t=j+1}^{n} EZ_{j,t}^2 + 2\sum_{l=1}^{j}\sum_{t=j+1+l}^{n} \left|EZ_{j,t}Z_{j,t-l}\right| + 2\sum_{l=j+1}^{n-j-1}\sum_{t=j+1+l}^{n} \left|EZ_{j,t}Z_{j,t-l}\right|. \qquad (14)
\end{aligned}
$$

For the first term in (14), $\sup_t EZ_{j,t}^2$ is finite if $\sup_t Eh_t^4$ is finite. Applying the Cauchy-Schwartz and Minkowski's inequalities, as in (9), we obtain

$$
\begin{aligned}
\sup_t Eh_t^4 &\leq \sqrt{\sup_t EU_t^8 \sup_t E\left|c'X_t\right|^8} \\
&\leq \sqrt{\sup_t EU_t^8\left(\sum_{j=1}^{k}|c_j|\left(\sup_t E\left|X_{tj}\right|^8\right)^{1/8}\right)^8} \\
&\leq \Delta' < \infty,
\end{aligned}
$$

for some constant Δ'. The last inequality holds due to assumptions (c**) and (d**), and since $r \geq 2$. Next, by Cauchy-Schwartz, $\left|EZ_{j,t}Z_{j,t-l}\right| \leq \sup_t EZ_{j,t}^2$. Hence, for the first two terms in (14), we obtain that

$$
\begin{aligned}
&\sum_{t=j+1}^{n} EZ_{j,t}^2 + 2\sum_{l=1}^{j}\sum_{t=j+1+l}^{n} \left|EZ_{j,t}Z_{j,t-l}\right| \\
&\leq (n-j)\sup_t EZ_{j,t}^2 + 2\sum_{l=1}^{j}(n-j-l)\sup_t EZ_{j,t}^2 \\
&\leq (n-j)\sup_t EZ_{j,t}^2 + 2\,(n-3/2j-1/2)\,j\sup_t EZ_{j,t}^2 \\
&\leq 2\,(n-j)\,(j+1)\sup_t EZ_{j,t}^2 \\
&\leq 2n\,(j+1)\sup_t EZ_{j,t}^2 \qquad (15)
\end{aligned}
$$

For the third term in (14), a different strategy has to be used. If we choose again to use Cauchy-Schwartz and replace $EZ_{j,t}^2$ with $\sup_t EZ_{j,t}^2$, we will obtain $\sum_{l=j+1}^{n-j-1}\sum_{t=j+1+l}^{n}\left|EZ_{j,t}Z_{j,t-l}\right| \leq \sup_t EZ_{j,t}^2 \sum_{l=j+1}^{n-j-1}(n-$

$j-l$) which behaves as n^2 for large n and all j's fixed we. If we substitute it back into (12), n^2 will cancel out in the numerator and denominator, and the bound on $P(|R_{n,1}| > \varepsilon)$ will be uninformative since the numerator in (12) grows with m_n.

We will use therefore a different approach based on the mixing inequalities (Lemma 2 on page 11 of Lecture 7) for the third term in (14). Let ϕ_{Z_j} and ϕ_h be the uniform mixing coefficients of $Z_{j,t}$ and h_t respectively. The mixing inequalities can be used because $l > j$ in this case, and as in proof of Theorem 8 on page 8 of Lecture 7,

$$\phi_{Z_j}(l) \leq \phi_h(l-j).$$

(The mixing inequalities cannot be used for the second term in (14) because for it $l \leq j$.) By Lemma 2 in Lecture 7,

$$\begin{aligned} |EZ_{j,t}Z_{j,t-l}| &\leq 2\phi_{Z_j}(l)^{1-1/r}\left(EZ_{j,t-l}^2\right)^{1/2}\left(E|Z_{j,t}|^r\right)^{1/r} \\ &\leq 2\phi_h(l-j)^{1-1/r}\left(\sup_t EZ_{j,t}^2\right)^{1/2}\left(\sup_t E|Z_{j,t}|^r\right)^{1/r}. \end{aligned}$$

As we argued above, $\sup_t EZ_{j,t}^2 < \infty$. By the definition of $Z_{j,t}$ and Minkowski's inequality,

$$\begin{aligned} \left(E|Z_{j,t}|^r\right)^{1/r} &= \left(E|h_t h_{t-j} - Eh_t h_{t-j}|^r\right)^{1/r} \\ &\leq \left(E|h_t h_{t-j}|^r\right)^{1/r} + \left(|Eh_t h_{t-j}|^r\right)^{1/r}, \end{aligned}$$

and since $r \geq 2$, $\sup_t E|Z_{j,t}|^r$ is finite if $\sup_t E|h_t h_{t-j}|^r < \infty$ which can be shown by the same argument as in (9). Thus, we have

$$\begin{aligned} &\sum_{l=j+1}^{n-j-1}\sum_{t=j+1+l}^{n}|EZ_{j,t}Z_{j,t-l}| \\ \leq\ & 2\left(\sup_t EZ_{j,t}^2\right)^{1/2}\left(\sup_t E|Z_{j,t}|^r\right)^{1/r}(n-j-l)\sum_{l=j+1}^{n-j-1}\phi_h(l-j)^{1-1/r} \\ \leq\ & 2\left(\sup_t EZ_{j,t}^2\right)^{1/2}\left(\sup_t E|Z_{j,t}|^r\right)^{1/r} n\sum_{l=1}^{\infty}\phi_h(l)^{1-1/r}. \end{aligned} \tag{16}$$

The conditions on mixing coefficients (a*) are sufficient to ensure that

$$\sum_{l=1}^{\infty}\phi_h(l)^{1-1/r} < \infty,$$

as it was previously discussed in the proof of Lemma 3 in Lecture 7. The results in equations (14), (15), and (16) together with the norm inequality $\left(EZ_{j,t}^2\right)^{1/2} \leq \left(E|Z_{j,t}|^r\right)^{1/r}$ imply that the result in (13) holds:

$$\begin{aligned} E\left|\sum_{t=j+1}^{n} Z_{j,t}\right|^2 &\leq 2n(j+1)\sup_t EZ_{j,t}^2 + 2\left(\sup_t EZ_{j,t}^2\right)^{1/2}\left(\sup_t E|Z_{j,t}|^r\right)^{1/r} n\sum_{l=1}^{\infty}\phi_h(l)^{1-1/r} \\ &= 2n(j+1)\Delta^{2/r} + 2\Delta^{2/r} n\sum_{l=1}^{\infty}\phi_h(l)^{1-1/r} \\ &\leq 2\Delta^{2/r}\left(1 + \sum_{l=1}^{\infty}\phi_h(l)^{1-1/r}\right) n(j+2). \end{aligned}$$

Lastly, from (12) and (13),

$$P(|R_{n,1}| > \varepsilon) \leq \frac{m_n^2 C^2}{n^2\varepsilon^2}\Delta^* n\sum_{j=1}^{m_n}(j+2)$$

$$
\begin{aligned}
&\leq \frac{C^2\Delta^*}{2\varepsilon^2}\frac{m_n^3(m_n+5)}{n}\\
&\to 0,
\end{aligned}
$$

since, by assumption (i), $m_n = o\left(n^{1/4}\right)$. ■

Now, we construct a feasible HAC estimator of Ω_n. Let

$$\widehat{U}_t = Y_t - X_t'\widehat{\beta}_n.$$

Define

$$
\begin{aligned}
\widehat{\Omega}_n &= \widehat{\Omega}_n(0) + \sum_{j=1}^{m_n} w(j, m_n)\left(\widehat{\Omega}_n(j) + \widehat{\Omega}_n(j)'\right),\\
\widehat{\Omega}_n(j) &= n^{-1}\sum_{t=j+1}^{n}\left(X_t\widehat{U}_t\right)\left(X_{t-j}\widehat{U}_{t-j}\right)',
\end{aligned}
$$

Theorem 3 *Under assumptions Under Assumptions (a*), (b), (c**), (d**), (g)-(i),* $\widehat{\Omega}_n - \Omega_n \to_p 0$.

Proof. It is sufficient to show that $\widehat{\Omega}_n - \widetilde{\Omega}_n \to_p 0$.

$$
\begin{aligned}
&\widehat{\Omega}_n - \widetilde{\Omega}_n\\
= \; & n^{-1}\sum_{t=1}^{n}\widehat{U}_t^2 X_t X_t' + \sum_{j=1}^{m_n} w(j, m_n)\, n^{-1}\sum_{t=j+1}^{n}\widehat{U}_t\widehat{U}_{t-j}\left(X_t X_{t-j}' + X_{t-j}X_t'\right)\\
&-n^{-1}\sum_{t=1}^{n}U_t^2 X_t X_t' - \sum_{j=1}^{m_n} w(j, m_n)\, n^{-1}\sum_{t=j+1}^{n}U_t U_{t-j}\left(X_t X_{t-j}' + X_{t-j}X_t'\right).
\end{aligned}
$$

Since

$$\widehat{U}_t = U_t - X_t'\left(\widehat{\beta}_n - \beta\right),$$

We have

$$\widehat{\Omega}_n - \widetilde{\Omega}_n = -2B_{n,1} + B_{n,2} - B_{n,3} - B_{n,4} + B_{n,5},$$

where

$$
\begin{aligned}
B_{n,1} &= n^{-1}\sum_{t=1}^{n}\left(\left(\widehat{\beta}_n - \beta\right)' X_t U_t\right) X_t X_t',\\
B_{n,2} &= n^{-1}\sum_{t=1}^{n}\left(\left(\widehat{\beta}_n - \beta\right)' X_t\right)^2 X_t X_t',\\
B_{n,3} &= \sum_{j=1}^{m_n} w(j, m_n)\, n^{-1}\sum_{t=j+1}^{n}\left(\left(\widehat{\beta}_n - \beta\right)' X_t U_{t-j}\right)\left(X_t X_{t-j}' + X_{t-j}X_t'\right),\\
B_{n,4} &= \sum_{j=1}^{m_n} w(j, m_n)\, n^{-1}\sum_{t=j+1}^{n}\left(\left(\widehat{\beta}_n - \beta\right)' X_{t-j} U_t\right)\left(X_t X_{t-j}' + X_{t-j}X_t'\right),\\
B_{n,5} &= \sum_{j=1}^{m_n} w(j, m_n)\, n^{-1}\sum_{t=j+1}^{n}\left(\left(\widehat{\beta}_n - \beta\right)' X_t\right)\left(\left(\widehat{\beta}_n - \beta\right)' X_{t-j}\right)\left(X_t X_{t-j}' + X_{t-j}X_t'\right).
\end{aligned}
$$

By the same argument as on page 4 of Lecture 1,

$$\|B_{n,1}\| \leq \left\|\widehat{\beta}_n - \beta\right\| n^{-1}\sum_{t=1}^{n}|U_t|\,\|X_t\|^3.$$

Further, by Holder's inequality with $p = 4$, $q = 4/3$

$$E\left(|U_t|\,\|X_t\|^3\right)^{r+\delta/4} \leq \left(E\,|U_t|^{4r+\delta}\right)^{1/4}\left(E\,\|X_t\|^{4r+\delta}\right)^{3/4},$$

and bounded uniformly in t. Therefore, by the SLLN, $n^{-1}\sum_{t=1}^{n}|U_t|\,\|X_t\|^3 = O_{a.s.}(1)$. Hence, since $\widehat{\beta}_n$ is strongly consistent, $B_{n,1} = o_{a.s.}(1)$. Similarly,

$$\begin{aligned}\|B_{n,2}\| &\leq \left\|\widehat{\beta}_n - \beta\right\|^2 n^{-1}\sum_{t=1}^{n}\|X_t\|^4 \\ &= o_{a.s.}(1)O_{a.s.}(1).\end{aligned}$$

Next, consider $B_{n,3}$. By the triangle inequality and since the weights $w(j,m)$ are non-negative,

$$\|B_{n,3}\| \leq 2\|\widehat{\beta}_n - \beta\|\sum_{j=1}^{m_n} w\left(j, m_n\right) n^{-1}\sum_{t=j+1}^{n}|U_{t-j}|\|X_t\|^2\|X_{t-j}\|. \tag{17}$$

By the same argument as for $R_{n,1}$ in the proof of Lemma 2, we can show that

$$\sum_{j=1}^{m_n} w\left(j, m_n\right) n^{-1}\sum_{t=j+1}^{n}\left(|U_{t-j}|\|X_t\|^2\|X_{t-j}\| - E\left[|U_{t-j}|\|X_t\|^2\|X_{t-j}\|\right]\right) = o_p(1). \tag{18}$$

This can be achieved by re-defining $Z_{j,t}$ in (11) as

$$Z_{j,t} = |U_{t-j}|\|X_t\|^2\|X_{t-j}\| - E\left[|U_{t-j}|\|X_t\|^2\|X_{t-j}\|\right]. \tag{19}$$

Note that similarly to the definition of $Z_{j,t}$ in (11), the expression in (19) also involves a cross-moment of order 4, and is ϕ- or α-mixing of the same size. Therefore, the rest of the argument used in showing that $|R_{n,1}| \to_p 0$ goes through without any essential changes. Moreover, given the assumed mixing and moment conditions,

$$\sum_{j=1}^{m_n} w\left(j, m_n\right) n^{-1}\sum_{t=j+1}^{n} E\left[|U_{t-j}|\|X_t\|^2\|X_{t-j}\|\right] = O(1). \tag{20}$$

The results in (17), (18), and (20) together with consistency of $\hat{\beta}_n$ imply that

$$B_{n,3} = o_p(1).$$

Using the same argument as for $B_{n,3}$, we can show that $B_{n,4}$ and $B_{n,5}$ are $o_p(1)$. This can be done since

$$\begin{aligned}\|B_{n,4}\| &\leq 2\|\widehat{\beta}_n - \beta\|\sum_{j=1}^{m_n} w\left(j, m_n\right) n^{-1}\sum_{t=j+1}^{n}|U_{t-j}|\|X_t\|\|X_{t-j}\|^2, \\ \|B_{n,5}\| &\leq 2\|\widehat{\beta}_n - \beta\|^2\sum_{j=1}^{m_n} w\left(j, m_n\right) n^{-1}\sum_{t=j+1}^{n}\|X_t\|^2\|X_{t-j}\|^2.\end{aligned}$$

The rest of the prove for $B_{n,4}$ and $B_{n,5}$ is essentially the same as that for $B_{n,3}$ (or $R_{n,1}$). ■