14.384 Time Series Analysis, Fall 2007
Professor Anna Mikusheva
Paul Schrimpf, scribe
September 13, 2007
revised September 16, 2009

Lecture 3

More HAC and Intro to Spectrum

HAC

Continuing with our setup from last time:

- $\{z_t\}$ stationary
- $\gamma_k = \text{cov}(z_t, z_{t+k})$
- $\mathcal{J} = \sum_{k=-\infty}^{\infty} \gamma_k$

we want to estimate $\mathcal{J}$. Last time, we first considered just summing all the sample covariances. This was inconsistent because high order covariances will always be noisily estimated. Then we considered a truncated sum of covariances with bandwidths slowly increasing to infinity. This may be consistent (under certain conditions), but could lead to a not positive definite $\hat{\mathcal{J}}$. To fix this, we considered the kernel estimator:

$$\hat{\mathcal{J}} = \sum_{-S_T}^{S_T} k_T(j)\hat{\gamma}_j, \quad \hat{\gamma}_j = \frac{1}{T}\sum_{t=1}^{T-j} z_t z_{t+j}$$

We want to choose S_T and $k_T()$ such that (1) $\hat{\mathcal{J}}$ is consistent and (2) $\hat{\mathcal{J}}$ is always positive definite.

Consistency

Theorem 1. *Assume that:*

- $\sum_{-\infty}^{\infty} |\gamma_j| < \infty$
- $k_T(j) \to 1$ *as* $T \to \infty$ *and* $|k_T(j)| < C$ $\forall j$ *and some constant* C
- *for all* j *a sequence* $\xi_{t,j} = z_t z_{t+j} - \gamma_j$ *is stationary and* $\sup_j \sum_k |cov(\xi_{t,j}, \xi_{t+k,j})| < C$ *for some constant* C *(limited dependence)*
- $S_T \to \infty$ *and* $\frac{S_T^3}{T} \to 0$

then $\hat{\mathcal{J}} = \sum_{-S_T}^{S_T} k_T(j)\hat{\gamma}_j$ *is a consistent estimator of* $\mathcal{J}$

Proof. This is an informal "proof" that sketches the ideas, but isn't completely rigorous.

$$\hat{\mathcal{J}} - \mathcal{J} = -\sum_{|j|>S_T} \gamma_j + \sum_{j=-S_T}^{S_T} (k_T(j) - 1)\gamma_j + \sum_{j=-S_T}^{S_T} k_T(j)(\hat{\gamma}_j - \gamma_j) \tag{1}$$

We can interprete these three terms as follows;

1. $\sum_{|j|>S_T} \gamma_j$ is truncation error

2. $\sum_{j=-S_T}^{S_T}(k_T(j)-1)\gamma_j$ is error from using the kernel

3. $\sum_{j=-S_T}^{S_T} k_T(j)(\hat{\gamma}_j-\gamma_j)$ is error from estimating the covariances

Terms 1 and 2 are non-stochastic. They represent bias. The third term is stochastic; it is responsible for uncertainty. We want to show that each of these terms goes to zero. We also will put special attention to a bias-variance tradeoff arising in this problem.

1. Disappears as long as $S_T \to \infty$, since we assumed $\sum_{-\infty}^{\infty}|\gamma_j| < \infty$.

2. $\sum_{j=-S_T}^{S_T}(k_T(j)-1)\gamma_j \le \sum_{j=-S_T}^{S_T}|k_T(j)-1||\gamma_j|$ This will converge to zero as long as $k_T(j)\to 1$ as $T\to\infty$ (we use here $\sum_{-\infty}^{\infty}|\gamma_j|<\infty$ and $|k_T(j)|<C\ \forall j$).

3. Notice that for the first term we wanted S_T big enough to eliminate it. Here, we'll want S_T to be small enough.

   First, note that $\hat{\gamma}_j \equiv \frac{1}{T}\sum_{k=1}^{T-j} z_k z_{k+j}$ is not unbiased. $E\hat{\gamma}_j = \frac{T-j}{T}\gamma_j = \tilde{\gamma}_j$. However, it's clear that this bias will disappear as $T\to\infty$. Formally,

$$\sum_{j=-S_T}^{S_T} k_T(j)(\hat{\gamma}_j-\gamma_j) = \sum_{j=-S_T}^{S_T} k_T(j)(\hat{\gamma}_j-\tilde{\gamma}_j) + \sum_{j=-S_T}^{S_T} k_T(j)(\tilde{\gamma}_j-\gamma_j),$$

   and the last term goes to zero for exactly the same reasons as 2.

   As for the first summand let's consider $\xi_{t,j} = z_t z_{t+j} - \gamma_j$, so $\hat{\gamma}_j - \tilde{\gamma}_j = \frac{1}{T}\sum_{\tau=1}^{T-j}\xi_{\tau,j}$. We need to assess the speed with which the sum of $\xi_{t,j}$ converges to zero in probability.

$$\begin{aligned} E(\hat{\gamma}_j-\tilde{\gamma}_j)^2 =& \frac{1}{T^2}\sum_{k=1}^{T-j}\sum_{t=1}^{T-j}\operatorname{cov}(\xi_{k,j},\xi_{t,j}) \\ \le& \frac{1}{T^2}\sum_{k=1}^{T-j}\sum_{t=1}^{T-j}|\operatorname{cov}(\xi_{k,j},\xi_{t,j})| \end{aligned}$$

   We need an assumption to guarantee that the covariances of ξ disappear. The assumption that $\xi_{t,j}$ are sationary for all j and $\sup_j \sum_k |\operatorname{cov}(\xi_{t,j},\xi_{t+k,j})| < C$ for some constant C implies that

$$\frac{1}{T^2}\sum_{k=1}^{T-j}\sum_{t=1}^{T-j}|\operatorname{cov}(\xi_{k,j},\xi_{t,j})| \le \frac{C}{T}$$

   By Chebyshev's inequality we have:

$$P(|\hat{\gamma}_j-\tilde{\gamma}_j|>\epsilon) \le \frac{E(\hat{\gamma}_j-\tilde{\gamma}_j)^2}{\epsilon^2} \le \frac{C}{\epsilon^2 T}$$

   This characterize the accuracy with which we estimate each covariance. Now we need to assess how many auto-covariances we can estimate well simultaneously:

$$\begin{aligned} P(\sum_{-S_T}^{S_T}|\hat{\gamma}_j-\tilde{\gamma}_j|>\epsilon) \le& \sum_{-S_T}^{S_T} P(|\hat{\gamma}_j-\tilde{\gamma}_j| > \frac{\epsilon}{2S_T+1}) \\ \le& \sum_{-S_T}^{S_T}\frac{E(\hat{\gamma}_j-\gamma_j)^2}{\epsilon^2}(2S_T+1)^2 \\ \le& \sum_{-S_T}^{S_T}\frac{C}{T}(2S_T+1)^2 \approx C_1\frac{S_T^3}{T} \end{aligned}$$

   so, it is enough to assume $\frac{S_T^3}{T}\to 0$ as $T\to\infty$ to make the last term in (1) go to 0 in probability.

□

We used the assumption that $S_T^3/T \to 0$. In fact, we could do better than that, but it requires a little bit more work.

Positive Definiteness

Now we will address the question of positive definiteness of the estimate of long-run variance. It is easiest to characterize positive definiteness using the Fourier transformation. To do this I will refresh your memory about complex numbers. We also will need this material for the spectrum.

Complex Numbers

Definition 2. $i = \sqrt{-1}$

We write complex numbers in rectangular as $a+bi$, where a and b are real. We add and multiply complex numbers as:

$$\begin{aligned}(a+bi)+(c+di) =& a+c+(b+d)i \\ (a+bi)(c+di) =& ac-bd+(bc+ad)i\end{aligned}$$

We can also write complex numbers in polar form as

$$r(\cos\phi + i\sin\phi) = re^{i\phi}$$

which simplifies multiplication

$$r_1 e^{i\phi_1} r_2 e^{i\phi_2} = r_1 r_2 e^{i(\phi_1+\phi_2)}$$

Fourier Transfrom

Definition 3. Given a series $\{x_t\}_{-\infty}^{\infty}$, the *Fourier transform* is $d_x(\omega) = \sum_{t=-\infty}^{\infty} x_t e^{-i\omega t}$

Definition 4. The *inverse Fourier transform* of $d_x(\omega)$ is $x_t = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\omega t} d_x(\omega) d\omega$

As the names suggest, these operations are the inverse of one another.

Proof.

$$\begin{aligned}\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\omega t} d_x(\omega) d\omega =& \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\omega t}\sum_s e^{-i\omega s} x_s d\omega \\ =& \sum_s x_s \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\omega(t-s)} d\omega \\ =& x_t\end{aligned}$$

where the last line froms from the fact that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\omega k} d\omega = \begin{cases} 1 & k=0 \\ 0 & k\neq 0\end{cases}$$

□

Now we are ready to characterize kernels that would lead us to non-negative estimates of $\mathcal{J}$. Assume $k_T(j)$ is an inverse Fourier transform of $K_T(\omega)$, *i.e.*

$$k_T(j) = \frac{1}{2\pi}\int_{-\pi}^{\pi} K_T(\omega) e^{-i\omega j} d\omega$$

Lemma 5. *$\hat{\mathcal{J}}$ is non-negative with probability 1 if and only if $K_T(\omega) \geq 0$ and $K_T(\omega) = K_T(-\omega)$*

Proof. Let us introduce Fourier transform of z_t: $d_z(\omega) = \sqrt{\frac{1}{T}} \sum_{t=1}^{T} z_t e^{-i\omega t}$, and what's called a periodiogram $I(\omega) = \frac{1}{2\pi} d_z(\omega) \overline{d_z(\omega)}'$. We wish to show that

$$\widehat{\mathcal{J}} = \int_{-\pi}^{\pi} K_T(\omega) I(\omega) d\omega \tag{2}$$

Indeed

$$\int_{-\pi}^{\pi} K_T(\omega) I(\omega) d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(K_T(\omega) \frac{1}{T} \sum_{t=1}^{T} z_t e^{-i\omega t} \sum_{s=1}^{T} z_s' e^{+i\omega s} \right) d\omega =$$

$$= \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} z_t z_s' \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(K_T(\omega) e^{-i\omega(t-s)} \right) d\omega = \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} z_t z_s k_T(t-s) = \widehat{\mathcal{J}}.$$

But what we can notice is that if K is symmetric, then k_T is real. If in addition K is positive, then (2) implies that $\widehat{\mathcal{J}}$ is positive. The inverse is also true, in a sense that if any realization of $\{z_t\}_{t=1}^{T}$ is possible, then to guarantee that $\widehat{\mathcal{J}}$ is positive we need positiveness of K.

□

Last comments on the estimation of long-run variance.

Below are some examples of good kernels (that satisfy the positiveness condition).

Definition 6. *Bartlett kernel $k_T(j) = k(j/S_T)$ where* $k(x) = \begin{cases} 1 - |x| & x \in [0,1] \\ 0 & \text{otherwise} \end{cases}$

Newey-West (1987) (this is one of the most cited papers in economics) suggested to use Bartlett kernel.

Definition 7. *Parsen kernel* $k(x) = \begin{cases} 1 - 6x^2 - 6|x|^3 & 0 \leq x \leq 1/2 \\ 2(1-|x|)^3 & 1/2 < x \leq 1 \\ 0 & \text{otherwise} \end{cases}$

Bandwidth How do we choose S_T? The general idea here is that we are facing a bias-variance trade-off. Namely, the bigger S_T reduces the cut-off bias, however, it increases the number of estimated covariances used(and hence the variance of the estimate). One idea of how to assess the quality of estimate is to minimize the mean squared error (MSE)

$$MSE(\widehat{\mathcal{J}}) = E(\widehat{\mathcal{J}} - \mathcal{J})^2 = bias(\widehat{\mathcal{J}})^2 + var(\widehat{\mathcal{J}})$$

Andrews (1991) did this minimization. His findings are: $S_T = \text{const}T^C$, where $C = \begin{cases} \frac{1}{3} & \text{for Newey-West} \\ \frac{1}{5} & \text{for Parsen} \end{cases}$. Andrews also gives a formula for const. This constant depends on properties of the process and is quite complicated.

Keifer & Vogelsang Keifer & Vogelsang (2002) consider setting $S_T = T - 1$. This gives $\hat{\mathcal{J}}$ inconsistent estimate(as we've seen on the last lecture). However, $\hat{\mathcal{J}}$ usually isn't what we care about. We care about testing $\hat{\beta}$, say by looking at the t statistic. We can use $\hat{\mathcal{J}}$ with $S_T = T - 1$ to compute $t = \frac{\hat{\beta}}{se(\hat{\beta})}$, which will converge to some (non-normal) distribution. What is important here that this distribution does not contain any nuisance parameters (even though it depends on the kernel used). Why it is important: we can simulate this distribution (bootstrap)and perform testing. The motivation for doing this is that Newey-West often works poorly in small samples due to bias.

Spectrum

Spectral density

Definition 8. *Spectral density* is the Fourier transform of covariances

$$S(\omega) = \sum_{j=-\infty}^{\infty} e^{-i\omega j}\gamma_j = \gamma(e^{-i\omega})$$

(where $\gamma()$ is the covariance function from lecture 1)

Example 9. ARMA: $a(L)y_t = b(L)e_t$

$$\begin{aligned}\gamma(\xi) =& \sigma^2 \frac{b(\xi)b(\xi^{-1})}{a(\xi)a(\xi^{-1})}\\ S(\omega) =& \sigma^2 \frac{b(e^{-i\omega})b(e^{i\omega})}{a(e^{-i\omega})a(e^{i\omega})} = \sigma^2 \frac{|b(e^{i\omega})|^2}{|a(e^{i\omega})|^2}\end{aligned}$$

Why use the spectrum?

- Easier to write down and work with than covariances
- Spectral decomposition

Filtering

We have a series $\{x_t\}$

Definition 10. We *filter* the series by $B(L)$ to produce $y_t = B(L)x_t$, where $B(L)$ might include both positive and negative powers of L.

The spectral density of y is related to the spectral density of x by a simple equation

$$S_y(\omega) = S_x(\omega)|B(e^{i\omega})|^2$$

Note that we can also write the covariances of y in terms of the covariances of x, but the relationship is not so simple

$$\begin{aligned}\gamma_k^y =& \mathrm{cov}(y_t, y_{t+k})\\ =& \mathrm{cov}(\sum_j b_j x_{t-j}, \sum_l b_l x_{t+k-l})\\ =& \sum_{j,l} b_j b_l \gamma^x_{k-l+j}\end{aligned}$$