Ulrich Müller
Economics 513
Fall 2011

Notes on Consistent Long-Run Variance Estimation

Importance of the 'Long-Run Variance'

- As we have already seen, the 'Long-Run Variance' crops up everywhere in time series econometrics: when conducting GMM type inference with autocorrelated data, but also in nonstandard problems such as testing for parameter stability, and more generally when FCLT-type approximations are invoked. It is thus important to understand the rationale for alternative estimators, and empirical results often hinge crucially on a specific estimator for the long-run variance, ie. the results change quite radically when another estimator is employed.

Discrete Fourier Transform

- Because of this intimate link between the spectral density and the long-run variance, most methods of estimating the long-run variance are most easily understood from a spectral analysis point of view. A very instructive way of putting this literature into perspective is by considering the properties of the Discrete Fourier Transform (DFT) of any mean-zero second-order stationary series with absolutely summable autocovariance function.

  The DFT of the series $\{y_t\}_{t=1}^T$ are the T variables

$$\begin{aligned}\xi_l &= T^{-1/2}\sum_{t=1}^{T} e^{-\mathbf{i}t\lambda_l} y_t \\ &= T^{-1/2}\sum_{t=1}^{T}\cos(\lambda_l t)y_t - \mathbf{i}T^{-1/2}\sum_{t=1}^{T}\sin(\lambda_l t)y_t\end{aligned}$$

  where $\lambda_l = 2\pi l/T$ for all $l \in F_T \subset \mathbb{Z}$ such that $-\pi < \lambda_l \leq \pi$. Note that ξ_l is complex even if $\{y_t\}$ is real. But with $\{y_t\}$ real, $\xi_{-l} = \overline{\xi_l}$, so that the DFT generates T independent real quantities Z_l. In the following, let T be odd. The T independent quantities are then given by Z_l, $l \in F_T = \{-T/2+1/2, \cdots, T/2-1/2\}$, where

$$Z_l = \begin{cases} \sqrt{2}T^{-1/2}\sum_{t=1}^{T}\cos(2\pi lt/T)y_t \text{ for } l<0 \\ T^{-1/2}\sum_{t=1}^{T} y_t \;\; \text{ for } l=0 \\ \sqrt{2}T^{-1/2}\sum_{t=1}^{T}\sin(2\pi lt/T)y_t \text{ for } l>0 \end{cases}$$

- Before stating what is interesting about this transformation, recall that any random vector $v \sim (0, \Sigma)$ can be orthogonalized by a suitable orthonormal matrix Q (i.e. $Q'Q = I$), that is there exists a orthonormal matrix Q such that $Q'v \sim (0, D)$, where D is diagonal. In other words, with the columns of Q describing the weights, there exist T weighted averages of the original data $Q'v$ that are uncorrelated. [One choice for Q is the matrix P whose columns are the eigenvectors of Σ, i.e. $\Sigma = P\Lambda P'$ (i.e. $P'\Sigma P = \Lambda$), where the columns of P are the eigenvectors of Σ and Λ is the diagonal matrix of the eigenvalues of Σ, since then $P'v \sim (0, \Lambda)$.] Being able to orthogonalize data in this way would often be very handy, since it is much easier to work with uncorrelated random variables than with correlated ones. Unfortunately, it seems, that in order to be able to orthogonalize the vector v, i.e. construct a valid orthonormal Q, one needs to know the variance-covariance matrix Σ.

  Now the amazing property of the DFT is that the T weighted averages of $\{y_t\}_{t=1}^T$ corresponding to the T (real) random variables $\{Z_l\}_{l \in F_T}$ asymptotically orthogonalize *any* mean-zero second-order stationary series $\{y_t\}$ with absolutely summable autocovariance function! This is a truly astounding result: Any pair of elements in $\{Z_l\}_{l \in F_T}$ are asymptotically uncorrelated, so that orthogonalization is possible despite the fact that Σ is unknown. The variances of Z_l turn out to be closely related to the spectral density f of y_t.

- Theorem: Let y_t be mean-zero second-order covariance stationary process with absolutely summable autocovariances $\gamma(k) = E[y_t y_{t-k}]$ and spectral density function $f(\cdot)$. Then

$$\max_{l,j \in F_T, l \neq j} |E[Z_l Z_j]| \to 0$$

  Furthermore, for any sequence of integers $l_T \in F_T$ such that $\eta_T = 2\pi l_T / T \to \eta \in [-\pi, \pi]$, $E[Z_{l_T}^2] \to 2\pi f(\eta)$.

Sketch of proof

- Note that for mean zero complex random variables $A = A_r + \mathbf{i}A_i$ and $B = B_r + \mathbf{i}B_i$,

$$\begin{aligned}
E[AB^*] &= E[(A_r + \mathbf{i}A_i)(B_r - \mathbf{i}B_i)] \\
&= E[A_r B_r + A_i B_i + \mathbf{i}(A_i B_r - A_r B_i)] \\
&= E[A_r B_r] + E[A_i B_i] + \mathbf{i}(E[A_i B_r] - E[A_r B_i]) \\
E[AB] &= E[A_r B_r] - E[A_i B_i] + \mathbf{i}(E[A_i B_r] + E[A_r B_i])
\end{aligned}$$

  If it can be shown that $E[AB^*] = E[AB] = 0$, then this implies uncorrelatedness of all possible combinations of products of the real and imaginary parts of A, and B, i.e.

$E[A_rB_r] = E[A_iB_i] = E[A_iB_r] = E[A_rB_i] = 0$. Furthermore, if $E[A^2] = 0$, then the real and imaginary part of A are uncorrelated, since the imaginary part of $E[A^2] = 0$ is $2E[A_iA_r]$.

- Because of these equalities, it suffices for the first claim of the theorem to show that the modulus of $E[A_lA_j^*] = E[A_lA_{-j}] \to 0$ uniformly for $l \neq j$, $l, j \in F_T$ (including the case $j = -l$).

- Now

$$
\begin{aligned}
TE[A_lA_j^*] &= E\left[\left(\sum_{t=1}^{T} u_t e^{-\mathbf{i}\lambda_l t}\right)\left(\sum_{t=1}^{T} u_t e^{\mathbf{i}\lambda_j t}\right)\right] \\
&= \begin{pmatrix} e^{-\mathbf{i}\lambda_l} \\ e^{-2\mathbf{i}\lambda_l} \\ e^{-3\mathbf{i}\lambda_l} \\ \\ e^{-T\mathbf{i}\lambda_l} \end{pmatrix}' \begin{pmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \gamma(T-1) \\ \gamma(1) & \gamma(0) & \gamma(1) & \gamma(T-2) \\ \gamma(2) & \gamma(1) & \gamma(0) & \gamma(T-3) \\ \\ \gamma(T-1) & \gamma(T-2) & \gamma(T-3) & \gamma(0) \end{pmatrix} \begin{pmatrix} e^{\mathbf{i}\lambda_j} \\ e^{2\mathbf{i}\lambda_j} \\ e^{3\mathbf{i}\lambda_j} \\ \\ e^{T\mathbf{i}\lambda_j} \end{pmatrix} \\
&= \gamma(0)\sum_{t=1}^{T} e^{-\mathbf{i}(\lambda_l-\lambda_j)t} + \sum_{k=1}^{T-1}\gamma(k)\sum_{t=1}^{T-k} e^{-\mathbf{i}(\lambda_l-\lambda_j)t-\mathbf{i}\lambda_l k} + \sum_{k=1}^{T-1}\gamma(k)\sum_{t=1}^{T-k} e^{-\mathbf{i}(\lambda_l-\lambda_j)t+\mathbf{i}\lambda_j k} \\
&= \gamma(0)\sum_{t=1}^{T} e^{-\mathbf{i}(\lambda_l-\lambda_j)t} + \sum_{k=1}^{T-1}\gamma(k)e^{-\mathbf{i}\lambda_l k}\sum_{t=1}^{T-k} e^{-\mathbf{i}(\lambda_l-\lambda_j)t} + \sum_{k=1}^{T-1}\gamma(k)e^{\mathbf{i}\lambda_j k}\sum_{t=1}^{T-k} e^{-\mathbf{i}(\lambda_l-\lambda_j)t}
\end{aligned}
$$

For $\lambda_j \neq \lambda_l$

$$
\sum_{t=1}^{T} e^{-\mathbf{i}(\lambda_l-\lambda_j)t} = e^{-\mathbf{i}(\lambda_l-\lambda_j)}\frac{1-e^{-\mathbf{i}(\lambda_l-\lambda_j)T}}{1-e^{-\mathbf{i}(\lambda_l-\lambda_j)}} = 0
$$

since

$$
1 - e^{-\mathbf{i}(\lambda_l-\lambda_j)T} = 1 - e^{-\mathbf{i}2\pi(l-j)} = 0.
$$

and series expansion is valid because $e^{-\mathbf{i}(\lambda_l-\lambda_j)} = e^{\mathbf{i}2\pi(j-l)/T} \neq 1$ for $-T < l-j < T$.

- Furthermore

$$
\sum_{t=1}^{T-k} e^{-\mathbf{i}(\lambda_l-\lambda_j)t} = \sum_{t=1}^{T} e^{-\mathbf{i}(\lambda_l-\lambda_j)t} - \sum_{t=T-k+1}^{T} e^{-\mathbf{i}(\lambda_l-\lambda_j)t} = -\sum_{t=T-k+1}^{T} e^{-\mathbf{i}(\lambda_l-\lambda_j)t}
$$

so that

$$
\left|\sum_{t=1}^{T-k} e^{-\mathbf{i}(\lambda_l-\lambda_j)t}\right| \leq \left|\sum_{t=T-k+1}^{T} e^{-\mathbf{i}(\lambda_l-\lambda_j)t}\right| \leq k
$$

Note that this inequality does not depend on λ_l and λ_j, so that, for any $\lambda_l = 2\pi l/T \neq \lambda_j = 2\pi j/T$,

$$|T^{-1}E\left[\sum_{t=1}^{T} u_t e^{\mathbf{i}\lambda_l t} \sum_{s=1}^{T} u_s e^{-\mathbf{i}\lambda_j s}\right]| \leq T^{-1} \sum_{k=-T+1}^{T-1} |k\gamma(k)| \to 0 \tag{1}$$

from the absolute summability of the autocovariances. Because the right-hand side of the inequality (1) is independent of l, j, the inequality is also true for the choice of l and j that makes the left-hand side as large as possible.

- For the second claim, note that for $l_T \neq 0$

$$\begin{aligned} E[A_{l_T} A_{l_T}^*] &= \gamma(0) + \sum_{k=1}^{T-1} \frac{T-k}{T}\gamma(k)e^{-\mathbf{i}\eta_T k} + \sum_{k=1}^{T-1} \frac{T-k}{T}\gamma(k)e^{\mathbf{i}\eta_T k} \\ &= \sum_{k=-T+1}^{T-1} \frac{T-|k|}{T}\gamma(k)e^{-\mathbf{i}\eta_T k} \to \sum_{k=-\infty}^{\infty} e^{-\mathbf{i}\eta k}\gamma(k) = 2\pi f_u(\eta) \end{aligned}$$

since $\sum_{k=-T+1}^{T-1} \frac{T-|k|}{T}\gamma(k)e^{-\mathbf{i}\lambda k} \to 2\pi f_u(\lambda)$

- Furthermore, note that $E[A_{l_T} A_{l_T}^*] = E[A_{l_T,r}^2] + E[A_{l_T,i}^2]$, and since $E[A_{l_T} A_{l_T}] = E[A_{l_T,r}^2] - E[A_{l_T,i}^2] + 2iE[A_{l_T,i}A_{l_T,r}] \to 0$ for $l_T \neq 0$ and $l_T \neq T/2$, (so that $E[A_{l_T,r}^2] - E[A_{l_T,i}^2] \to 0$), it follows that $E[A_{l_T,r}^2] \to 2\pi f_u(\eta)/2$ and $E[A_{l_T,i}^2] \to 2\pi f_u(\eta)/2$. This 'sharing of the variance' between the real and imaginary part is the reason for the scaling by $\sqrt{2}$ in the definition of $\{Z_l\}$, which ensures that $E[Z_{l_T}^2] \to 2\pi f_u(\eta)$. The special case $Z_0 = A_0 = T^{-1/2}\sum_{t=1}^{T} u_t$ is immediate.

- The same arguments also show that the weights in the definition of $\{Z_l\}_{l \in F_T}$ are orthogonal to each other and have length one. In other words, for T odd, the $T \times T$ matrix

$$W = T^{-1/2} \begin{pmatrix} \sqrt{2}\sin(\lambda_{T/2-1/2}) & \cdots & \sqrt{2}\sin(\lambda_1) & 1 & \sqrt{2}\cos(\lambda_1) & \cdots & \sqrt{2}\cos(\lambda_{T/2-1/2}) \\ \sqrt{2}\sin(2\lambda_{T/2-1/2}) & \cdots & \sqrt{2}\sin(2\lambda_1) & 1 & \sqrt{2}\cos(2\lambda_1) & \cdots & \sqrt{2}\cos(2\lambda_{T/2-1/2}) \\ \sqrt{2}\sin(3\lambda_{T/2-1/2}) & \cdots & \sqrt{2}\sin(3\lambda_1) & 1 & \sqrt{2}\cos(3\lambda_1) & \cdots & \sqrt{2}\cos(3\lambda_{T/2-1/2}) \\ \\ \sqrt{2}\sin(T\lambda_{T/2-1/2}) & \cdots & \sqrt{2}\sin(T\lambda_1) & 1 & \sqrt{2}\cos(T\lambda_1) & \cdots & \sqrt{2}\cos(T\lambda_{T/2-1/2}) \end{pmatrix}$$

satisfies (check!)

$$W'W = I.$$

Discrete Fourier Transform and Parametric Long-Run Variance Estimators

- The message of the Theorem is that asymptotically, we can treat $\{Z_l\}_{l\in F_T}$ as a sequence of uncorrelated random variables $Z_l \sim (0, 2\pi f_u(\lambda_l))$. Let Σ be the covariance matrix of $u = (u_1, \cdots, u_T)'$. Then

$$W'u \sim (0, W'\Sigma W) \overset{a}{\sim} (0, D)$$

where $D = 2\pi \operatorname{diag}(f_u(\lambda_{[T/2-1/2]}), \cdots, f_u(\lambda_1), f_u(0), f_u(\lambda_1), \cdots, f_u(\lambda_{[T/2]}))$, and '$\overset{a}{\sim}$' stands for 'is distributed approximately' (recall that $f_u(\cdot)$ is an even function, ie $f_u(x) = f_u(-x)$). The problem of dependence/autocorrelation in the original sequence $\{u_t\}$ has hence been transformed into the problem of heteroskedasticity/non-constant variance. What makes this possible is that second-order stationarity and absolutely summability of the autocovariances imply that the covariance matrix of $u = (u_1, \cdots, u_T)'$ is band-diagonal, and asymptotically, most of the action takes place very close to the diagonal. This is enough to enable the orthogonalization independent of the exact form of the dependence. Note that the sequence $\gamma(k)$ is fully recoverable from $f_u(\cdot)$ (and vice versa): One can think of the DFT as an operator that maps summable sequences $\gamma(k)$ into continuous functions $f_u(\cdot)$.

- The *periodogram* merges the information in Z_l and Z_{-l} about the Fourier frequency ω_l into a single number: For $-\pi < \lambda_l < \pi$, $l \in F_T$, $l \neq 0$ the periodogram $I : [-\pi, \pi] \mapsto \mathbb{R}_0^+$ is defined as

$$\begin{aligned} I_T(\lambda_l) &= \frac{Z_l^2 + Z_{-l}^2}{2} \\ &= \frac{\left(\sqrt{2}\sum \cos(\lambda_l t)u_t\right)^2 + \left(\sqrt{2}\sum \sin(\lambda_l t)u_t\right)^2}{2} \\ &= \left|\sum_{t=1}^{T} e^{-\mathbf{i}\lambda_l t} u_t\right|^2 \\ &= (A_l^r)^2 + (A_l^i)^2 = A_l A_l^* = A_l A_{-l} \end{aligned}$$

where the last two lines are the definition of $I_T(\cdot)$ also for $l = 0$ or $l = T/2$ for T even. <<Authors differ about the definition of $I(\lambda)$ at frequencies that are not Fourier frequencies. Some define it to be a piecewise constant function that takes on the value of the 'nearest' Fourier frequency, whereas others define it to be $I_T(\lambda) = \left|\sum_{t=1}^{T} e^{-\mathbf{i}\lambda t} u_t\right|^2$ for all $-\pi \leq \lambda \leq \pi$.>>

- With the help of the approximation $Z_l \sim (0, 2\pi f_u(\lambda_l))$, it becomes easier to understand what is really going on with parametric estimators of the long-run variance. Consider the special case of an estimator based on an AR model $u_t = \phi_1 u_{t-1} + \cdots + \phi_p u_{t-p} + \varepsilon_t$

with $\varepsilon_t \sim iid(0, \sigma^2)$, so that the estimator $\hat{\omega}_{AR}$ is

$$\begin{aligned} \hat{\omega}^2_{AR} &= 2\pi \widehat{f_{AR}(0)} \\ &= \frac{\hat{\sigma}^2}{(1 - \hat{\phi}_1 \cdots - \hat{\phi}_p)^2} \end{aligned}$$

The easiest way to obtain estimators of $\phi = (\phi_1, \cdots, \phi_p)$ is to estimate them by OLS

$$\begin{aligned} \hat{\phi}_{OLS} &= \arg \min_{\phi=(\phi_1,\cdots,\phi_p)} \sum_{t=p+1}^{T} (u_t - \phi_1 u_{t-1} \cdots - \phi_p u_{t-p})^2 \\ &= \arg\min_{\phi} u'\Sigma(\phi)^{-1} u \\ \hat{\sigma}^2_{OLS} &= T^{-1} \sum_{t=p+1}^{T} (u_t - \hat{\phi}_1 u_{t-1} \cdots - \hat{\phi}_p u_{t-p})^2 = T^{-1} u'\Sigma(\hat{\phi}_{OLS})^{-1} u \end{aligned}$$

where $u = (u_1, \cdots, u_T)'$ and, except for the first p rows and p columns, $\Sigma(\phi)^{-1}$ is the inverse of the variance-covariance matrix of T observations of a stationary AR(p) with parameters ϕ and innovation variance one. <<The OLS estimator $\hat{\phi}_{OLS}$ corresponds to the ML estimator for Gaussian disturbances, except for the fact that the contribution of $(u_1, \cdots, u_p)$ to the likelihood is ignored. As written, $\Sigma(\phi)^{-1}$ is singular, but we will ignore this in our approximations... See Hamilton, Chapter 5, for more details on this and related points.>>

As long as all roots of the lag polynomial are outside the unit circle, an AR(p) process is second-order stationary with absolutely summable autocovariances. With the above approximation, we hence find

$$\begin{aligned} u'\Sigma(\phi)^{-1}u &= u'WW'\Sigma(\phi)^{-1}WW'u \\ &= u'W(W'\Sigma(\phi)W)^{-1}W'u \\ &\simeq Z'D(\phi)^{-1}Z \\ &= \sum_{l\in F_T} \frac{Z_l^2}{f_{AR}(\lambda_l;\phi,1)} \\ &= \sum_{l\in F_T} \frac{Z_l^2 + Z_{-l}^2}{2f_{AR}(\lambda_l;\phi,1)} = \sum_{l\in F_T} \frac{I(\lambda_l)}{f_{AR}(\lambda_l;\phi,1)} \end{aligned}$$

where

$$D(\phi) = 2\pi \operatorname{diag}(f_{AR}(\lambda_{[T/2-1/2]};\phi,1), \cdots, f_{AR}(\lambda_1;\phi,1), f_{AR}(0;\phi,1), f_{AR}(\lambda_1;\phi,1), \cdots, f_{AR}(\lambda_{[T/2]};\phi,1))$$

Up to the approximation, the OLS estimator of ϕ can hence be thought of as being determined by finding the shape of the spectral density that most closely matches the periodogram, where the match is measured by $\sum_{l \in F_T} \frac{I(\lambda_l)}{f_{AR}(\lambda_l;\phi,1)}$. For an AR(1), for instance,

$$\hat{\phi}_{OLS,1} \simeq \arg\min_{\phi_1} \sum_{l \in F_T} I(\lambda_l) 2\pi(1 + \phi_1^2 - 2\phi_1 \cos\lambda)$$

The estimator of σ^2 is then approximately

$$\begin{aligned} \hat{\sigma}^2_{OLS} &= T^{-1} u' \Sigma(\hat{\phi}_{OLS})^{-1} u \\ &\simeq T^{-1} \sum_{l \in F_T} \frac{Z_l^2}{f_{AR}(\lambda_l; \hat{\phi}_{OLS}, 1)} \end{aligned}$$

The minimizing value of ϕ, and hence the implied long-run variance estimator $\hat{\omega}^2_{AR} = 2\pi\hat{\sigma}^2 f_{AR}(0;\phi,1)$, is determined by the whole sequence $\{Z_l\}$. This is good if the functional form is correct, because in this case, matching $f_{AR}(\lambda;\phi,1)$ to Z_l^2 for all l extracts valuable information about ϕ (and hence ω^2). But if the functional form is wrong, the estimated spectral density $\hat{\sigma}^2 f_{AR}(\lambda;\hat{\phi},1)$ might be driven to differ substantially from the true spectral density at $\lambda = 0$, simply because $\hat{\sigma}^2 f_{AR}(\lambda;\hat{\phi},1)$ is the best overall match to the true spectral density, but not a particularly accurate one for values of λ close to zero.

Nonparametric Long-Run Variance Estimators

- One can avoid this dependence on functional forms in several ways. One way is to allow a lot of flexibility in the functional form by letting the order of the AR process be big. In fact, one can show that any spectral density can be arbitrarily well approximated by an AR(p) with sufficiently large p. By letting p be an increasing function of T, it is hence possible to eliminate the bias in $\hat{\omega}^2_{AR}$ that stems from a mistaken functional form. At the same time, p must not grow too fast, or the estimators of ϕ become too imprecise. The math to show that all the little epsilons are actually small is quite a mess and we don't pursue this.

- There is an easier, and also more natural approach to generate estimators of the long-run variance that do not hinge on getting a functional form right. Simply note that for any $l_T \in F_T$ which satisfies $l_T \to \infty$ and $l_T = o(T)$, $\eta_T = 2\pi l_T / T \to 0$, so that the Theorem states that for such a l_T

  $$E[Z^2_{l_T}] \to 2\pi f_u(0) = \omega^2$$

  Since we can choose l_T to be increasing with T, asymptotically there are an infinite number of Z_l's that have asymptotic variance ω^2, so a natural way to estimate ω^2 is to take some kind of average of those.

- Let $K(s) : \mathbb{R} \mapsto [0;1]$ be a continuous function that satisfies

$$\begin{aligned} K(s) &> 0 \\ K(-s) &= K(s) \\ \int_{-\infty}^{\infty} K(s) ds &= 1 \\ \int_{-\infty}^{\infty} K(s) s^2 ds &< \infty \end{aligned} \tag{2}$$

Let b_T be a sequence of real values that satisfies $b_T \to \infty$, but $b_T/T \to 0$. Let $\Delta = 2\pi/T$. The *weighted periodogram* of ω^2 with *bandwidth* b_T and *spectral window* (or *kernel*) K is defined as

$$\begin{aligned} \hat{\omega}^2_{WP} &= b_T \Delta \sum_{l \in F_T} K(b_T \lambda_l) Z_l^2 \\ &= b_T \Delta \sum_{l \in F_T} K(b_T \lambda_l) \frac{(Z_l^2 + Z_{-l}^2)}{2} \\ &= b_T \Delta \sum_{l \in F_T} K(b_T \lambda_l) I_T(\lambda_l) \end{aligned}$$

Note that the weights will (approximately) sum to one, since

$$b_T \Delta \sum_{l \in F_T} K(b_T \lambda_l) \to \int_{-\infty}^{\infty} K(s) ds = 1$$

The advantage of defining estimator in this way is that they are for sure nonnegative. It is possible to rewrite this estimator in terms of a 'weighted sum of covariances estimator', since

$$\begin{aligned} \hat{\omega}^2_{SW} &= b_T \Delta \sum_{l \in F_T} K(b_T \lambda_l) \frac{(Z_l^2 + Z_{-l}^2)}{2} \\ &= b_T \Delta \sum_{l \in F_T} K(b_T \lambda_l) \left(T^{-1/2} \sum_{t=1}^{T} e^{-\mathbf{i} \lambda_l t} u_t T^{-1/2} \sum_{t=1}^{T} e^{\mathbf{i} \lambda_l t} u_t \right) \\ &= b_T \Delta \sum_{k=-T+1}^{T-1} \left(\sum_{l \in F_T} K(b_T \lambda_l) e^{-\mathbf{i} k \lambda_l} \right) T^{-1} \sum_{t=k+1}^{T} u_t u_{t-k} \\ &= \sum_{k=-T+1}^{T-1} \left(b_T \Delta \sum_{l \in F_T} K(b_T \lambda_l) e^{-\mathbf{i} k \lambda_l} \right) T^{-1} \sum_{t=k+1}^{T} u_t u_{t-k} \\ &= \sum_{k=-T+1}^{T} w_T(k) \hat{\gamma}(k) \end{aligned}$$

where

$$w_T(k) = b_T \Delta \sum_{l \in F_T} K(b_T \lambda_l) e^{-\mathbf{i} k \lambda_l}$$

Note that if $k_T = cb_T + o(b_T)$, then

$$\begin{aligned} w_T(k_T) &= b_T \Delta \sum_{l \in F_T} K(b_T \lambda_l) e^{-\mathbf{i}(cb_T + o(b_T))\lambda_l} \\ &\to \int_{-\infty}^{\infty} K(s) e^{-\mathbf{i}sc} ds \equiv H(c) \end{aligned}$$

i.e. $H(\cdot)$ is the Fourier Transform of $K(\cdot)$ (with the inverse transform $K(c) = \frac{1}{2\pi} \int_{-\infty}^{\infty} H(s) e^{-\mathbf{i}sc} ds$), so that

$$\sum_{k=-T+1}^{T} w_T(k) \hat{\gamma}(k) \simeq \sum_{k=-T+1}^{T} H(\frac{k}{b_T}) \hat{\gamma}(k)$$

The function $H(\cdot)$ is sometimes called *spectral lag window.*

- Now let's find b_T and $K(s)$ that are in some sense good. For that, we will assume that $Z_l \sim \mathcal{N}(0, 2\pi f_u(\lambda_l))$. The assumption of Gaussianity seems convenient but arbitrary, but in fact it is not. The reason is simply that Central Limit Theorem type arguments are applicable, since the Z_l's are weighted averages of $\{u_t\}$, so that $Z_{l_T} \Rightarrow \mathcal{N}(0, 2\pi f_u(\eta))$.

- Now the bias of $\hat{\omega}^2_{WP}$ is given by

$$E[\hat{\omega}^2_{WP} - \omega^2] = b_T \Delta \sum_{l \in F_T} K(b_T \lambda_l) 2\pi (f_u(\lambda_l) - f_u(0))$$

A second order Taylor expansion of $f_u(\cdot)$ yields

$$f_u(\lambda_l) \simeq f_u(0) + f'_u(0)\lambda_l + \tfrac{1}{2} f''_u(0) \lambda_l^2$$

since $f_u(\cdot)$ is even, $f'_u(0) = 0$ and we find

$$\begin{aligned} E[\hat{\omega}^2_{WP} - \omega^2] &\approx \pi b_T \Delta \sum_{l \in F_T} K(b_T \lambda_l) f''_u(0) \lambda_l^2 \\ &= b_T^{-2} \pi f''_u(0) \Delta b_T \sum_{l \in F_T} K(b_T \lambda_l)(\lambda_l b_T)^2 \end{aligned}$$

so that

$$b_T^2 E[\hat{\omega}^2_{WP} - \omega^2] \to \pi f''_u(0) \int_{-\infty}^{\infty} K(s) s^2 ds$$

and the variance of $\hat{\omega}^2_{WP}$ is given by

$$\begin{aligned} V[\hat{\omega}^2_{WP}] &= (b_T\Delta)^2 \sum_{l\in F_T} K(b_T\lambda_l)^2 2(2\pi f_u(\lambda_l))^2 \\ &= \frac{b_T}{T} 16\pi^3 (b_T\Delta) \sum_{l\in F_T} K(b_T\lambda_l)^2 f_u(\lambda_l)^2 \end{aligned}$$

so that

$$\frac{T}{b_T} V[\hat{\omega}^2_{WP}] \to 16\pi^3 f_u(0)^2 \int_{-\infty}^{\infty} K(s)^2 ds$$

- In order to minimize the MSE, we need the square of the bias to be of the same order of magnitude than the variance. For this we need b_T^4 to be of the same order of magnitude than T/b_T, which requires $b_T = aT^{1/5}$. For this choice of bandwidth, the MSE is given by

$$\begin{aligned} T^{4/5} MSE[\hat{\omega}^2_{WP}] &= T^{4/5} E[\hat{\omega}^2 - \omega^2]^2 + T^{4/5} V[\hat{\omega}^2_{WP}] \\ &\to a^{-4}\pi^2 f_u''(0)^2 \left[\int_{-\infty}^{\infty} K(s)s^2 ds\right]^2 + 8\pi^2 f_u(0)^2 a 2\pi \int_{-\infty}^{\infty} K(s)^2 ds \end{aligned}$$

  In order to minimize the asymptotic MSE, one hence needs to choose a as solving

$$-4a^{-5} f_u''(0)^2 \left[\int_{-\infty}^{\infty} K(s)s^2 ds\right]^2 + 8 f_u(0)^2 2\pi \int_{-\infty}^{\infty} K(s)^2 ds = 0$$

  or

$$a = \left(\frac{f_u''(0)^2 \left[\int K(s)s^2 ds\right]^2}{2 f_u(0)^2 2\pi \int K(s)^2 ds}\right)^{1/5}$$

  Andrews (1991) suggests using estimators from parametric models (for instance, from an AR(1)) to obtain estimates for the ratio $f''(0)/f(0)^2$ (which is a highly nonlinear function of the estimated AR(1) parameter and the estimated innovation variance). The point is that one does not need great accuracy in the choice of the bandwidth in order to still obtain reasonable long-run variance estimators $\hat{\omega}^2_{WP}$, so the AR(1) model suffices to be a very rough approximation.

- There is also the question of how to optimally pick $K(\cdot)$, which from the above computations leads to the question which $K(\cdot)$ minimizes

$$\left[\int_{-\infty}^{\infty} K(s)^2 ds\right]^2 \int_{-\infty}^{\infty} K(s)s^2 ds$$

  (just plug in the solution for a in the minimal MSE) of all kernels that satisfy the requirements (2) above. It can be shown by variational methods that the best kernel is the *Quadratic Spectral Kernel*

$$K_{QS}(x) = \begin{cases} \frac{5}{8\pi}\left(1 - \frac{25x^2}{36\pi^2}\right) \text{ for } |x| < 6\pi/5 \\ 0 \text{ else} \end{cases}$$

The corresponding spectral lag window $H(\cdot)$ is

$$H_{QS}(x) = \frac{25}{12\pi^2 x^2}\left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\right)$$

- The focus on the mean square error is mathematically convenient, but not obviously of a natural objective function, as it is unclear how the MSE relates to the approximation accuracy of replacing ω with $\hat{\omega}_{WP}$ in, say, the denominator of a t-statistic, which is what we really care about.