# Time Series Analysis - Asymptotic for Dependent Data

ECON 722

Elena Pesavento — Emory University

Fall 2024

# Introduction

- So far we have discussed the main assumtions that our data needs to be stationary.
- Strict stationarity means that the distribution is constant over time.
- It does not mean, however, mean that the process has some sort of limited dependence, nor that there is an absence of periodic patterns.
- These restrictions are actually associated with the concepts of ergodicity and mixing, which we study next.

# Transformations of Stationary Processes

- The important properties of strict stationarity is that it is preserved by transformation. = transformations of strictly stationary processes are also strictly stationary.
- This includes transformations which include the full history of $\boldsymbol{y}_t$.

**Theorem**

If $y_t$ is strictly stationary and $x_t = \phi\left(y_t, y_{t-1}, y_{t-2}, \ldots\right) \in \mathbb{R}^q$ is a random vector, then $x_t$ is strictly stationary.

# Transformations of Stationary Processes

This Theorem extremely useful both for the study of stochastic processes which are constructed from underlying errors, and for the study of sample statistics such as linear regression estimators which are functions of sample averages of squares and cross-products of the original data.

As an example, it applies to the infinite-order moving average transformation

$$x_t = \sum_{j=0}^{\infty} a_j y_{t-j}$$

as long as the series converges almost surely. Sufficient conditions for this are that $\sup_t \mathbb{E}\,|y_t| < \infty$ and $\sum_{j=0}^{\infty} |a_j| < \infty$

# Ergodicity

- The assumption of stationarity is not sufficient for many purposes, as there are strictly stationary processes with no time series variation.
- We want a minimal sufficient assumption so that the law of large numbers will apply to the sample mean.
- It turns out that a sufficient condition is ergodicity. As it is a rather technical subject, we mention only a few highlights here. For a rigorous treatment see a standard textbook such as Walters (1982).
- If $\boldsymbol{y}_t$ is i.i.d., then it strictly stationary and ergodic.
- If $y_t$ is strictly stationary, ergodic, $\mathbb{E}\,|y_t| < \infty$, and $\sum_{j=0}^{\infty} |a_j| < \infty$ then $x_t = \sum_{j=0}^{\infty} a_j y_{t-j}$ is strictly stationary and ergodic.
- The conditions for ergodicity are hard. So we often assume it.

# Ergodicity

**Theorem**

If $y_t$ is strictly stationary, ergodic, $\mathbb{E}|y_t| < \infty$, and $\sum_{j=0}^{\infty} |a_j| < \infty$ then $x_t = \sum_{j=0}^{\infty} a_j y_{t-j}$ is strictly stationary and ergodic.

The next theorem is very useful as it shows that, on average, the autocovariances of a stationary and ergodic process converges to zero. This property will be important as it is sufficient for the weak law of large numbers.

**Theorem**

If $y_t$ is strictly stationary, ergodic, and $\mathbb{E}\left(y_t^2\right) < \infty$, then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{\ell=1}^{n} \operatorname{cov}\left(y_t, y_{t+\ell}\right) = 0.$$

# Ergodic Theorem

## Ergodic Theorem

If $y_t$ is strictly stationary, ergodic, and $\mathbb{E}\left\|\boldsymbol{y}_t\right\| < \infty$, then as $n \to \infty$,

$$\mathbb{E}\|\overline{\boldsymbol{y}} - \boldsymbol{\mu}\| \longrightarrow 0$$

and

$$\overline{\boldsymbol{y}} \xrightarrow{p} \mu$$

where $\boldsymbol{\mu} = \mathbb{E}\left(\boldsymbol{y}_t\right)$.

- The ergodic theorem shows that ergodicity is sufficient for consistent estimation.
- The moment condition $\mathbb{E}\left\|\boldsymbol{y}_t\right\| < \infty$ is the same as in the WLLN for i.i.d. samples.
- You can find a proof in Hansen Section 14.46 or Hamilton

# Proof of Ergodic Theorem for scalar when $\text{var}(y_t) = \sigma^2 < \infty$

$$\text{var}(\bar{y}) = \frac{1}{n^2} \sum_{t=1}^{n} \sum_{j=1}^{n} \gamma(t-j)$$

where $\gamma(\ell) = \text{cov}(x_t, x_{t+\ell})$. This is equals

$$\begin{aligned}\text{var}(\bar{y}) &= \frac{1}{n^2}\left(n\sigma^2 + 2(n-1)\gamma(1) + 2(n-2)\gamma(2) + \cdots + 2\gamma(n-1)\right) \\ &= \frac{\sigma^2}{n} + \frac{2}{n}\sum_{\ell=1}^{n}\left(1 - \frac{\ell}{n}\right)\gamma(\ell)\end{aligned}$$

$\text{var}(\bar{y}) = \sigma^2/n$ (same as for i.i.d. sampling) + weighted mean of the autocovariances. Under ergodicity, let $w_{n\ell} = 2\left(\ell/n^2\right)$,

$$\frac{2}{n}\sum_{\ell=1}^{n}\left(1 - \frac{\ell}{n}\right)\gamma(\ell) = \frac{2}{n^2}\sum_{\ell=1}^{n-1}\sum_{j=1}^{\ell}\gamma(j) = \sum_{\ell=1}^{n-1} w_{n\ell}\left(\frac{1}{\ell}\sum_{j=1}^{\ell}\gamma(j)\right) \longrightarrow 0$$

Thus $\text{var}(\bar{y}) = \sigma^2/n$ is $o(1)$ under ergodicity. Hence var $(\bar{y}) \to 0$. Markov's inequality establishes that $\bar{y} \xrightarrow{p} \mu$.

# MDS

Recall the definition of MDS

**Definition MDS**

The process $(e_t, \mathcal{F}_t)$ is a Martingale Difference Sequence (MDS) if $e_t$ is adapted to $\mathcal{F}_t$, $\mathbb{E}|e_t| < \infty$ and $\mathbb{E}(e_t \mid \mathcal{F}_{t-1}) = 0$.

- In words, a MDS $e_t$ is unforecastable in the mean.
- It is useful to notice that if we apply iterated expectations a MDS is mean zero.
- It is best to explicitly specify the information sets so there is no confusion.
- The term "martingale difference sequence" refers to the fact that the summed process $S_t = \sum_{j=1}^{t} e_j$ is a martingale, and $e_t$ is its first-difference. A martingale $S_t$ is defined as a process such that $\mathbb{E}(S_t \mid \mathcal{F}_{t-1}) = S_{t-1}$.
- If $e_t$ is i.i.d. and mean zero it is a MDS, but the reverse is not the case.

# CLT for Martingale Differences

We are interested in an asymptotic approximation for the distribution of standardized sample means such as

$$\boldsymbol{S}_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \boldsymbol{u}_t$$

where $\boldsymbol{u}_t$ is mean zero with variance $\mathbb{E}\left(\boldsymbol{u}_t \boldsymbol{u}_t'\right) = \boldsymbol{\Sigma} < \infty$. In this section we present a CLT for the case where $\boldsymbol{u}_t$ is a martingale difference sequence.

**Theorem: MDS CLT**

If $u_t$ is a strictly stationary and ergodic martingale difference sequence and $\mathbb{E}\left(\boldsymbol{u}_t \boldsymbol{u}_t'\right) = \boldsymbol{\Sigma} < \infty$, then as $n \to \infty$,

$$\boldsymbol{S}_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \boldsymbol{u}_t \xrightarrow{d} \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Sigma}).$$

# CLT for Martingale Differences

- The conditions for Theorem 14.15 are similar to the Lindeberg-Lévy CLT. The only difference is that the i.i.d. assumption has been replaced by the assumption of a strictly stationarity and ergodic MDS.
- It might be reasonable to conjecture that the CLT would hold under the broader assumption that $u_t$ is white noise. However, no such theory exists. At present, it is unknown if the MDS assumption can be weakened.
- We are not going to prove it. See Hansen for a sketch of the proof and references for where to find a full proof (not easy).
- For dependent data, we need a stronger restriction on the dependence between observations than ergodicity.

# Mixing

How can we measure *dependence*? We can measure the degree of dependence between two events A and B by the discrepancy

$$\alpha(A, B) = |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|.$$

This equals 0 when $A$ and $B$ are independent, and is positive otherwise. Now consider the two information sets ( $\sigma$-fields)

$$\begin{aligned} \mathcal{F}_{-\infty}^{t} &= \sigma\left(\ldots, \boldsymbol{y}_{t-1}, \boldsymbol{y}_{t}\right) \\ \mathcal{F}_{t}^{\infty} &= \sigma\left(\boldsymbol{y}_{t}, \boldsymbol{y}_{t+1}, \ldots\right). \end{aligned}$$

and separate them by $\ell$ periods that is is, take $\mathcal{F}_{-\infty}^{t-\ell}$ and $\mathcal{F}_{t}^{\infty}$.

# CLT for Martingale Differences

We can measure the degree of dependence between the information sets by taking all events in each, and then taking the largest discrepancy. This is

$$\alpha(\ell) = \sup_{A \in \mathcal{F}_{-\infty}^{t-\ell}, B \in \mathcal{F}_t^{\infty}} \alpha(A, B).$$

The constants $\alpha(\ell)$ are known as the **mixing coefficients**.

- We say that $\boldsymbol{y}_t$ is strong mixing if $\alpha(\ell) \to 0$ as $\ell \to \infty$..
- This means that as the time separation increases between the information sets, the degree of dependence decreases, eventually reaching independence.
- It can be shown that a mixing process is ergodic.

# Strong Mixing

- For applications, mixing is often useful when we can characterize the rate at which the coefficients $\alpha(\ell)$ decline to zero.
- There are two types of conditions which are seen in asymptotic theory: rates and summation.
  - Rate conditions take the form $\alpha(\ell) = O\left(\ell^{-r}\right)$ or $\alpha(\ell) = o\left(\ell^{-r}\right)$.
  - Summation conditions take the form $\sum_{\ell=0}^{\infty} \alpha(\ell)^r < \infty$ or $\sum_{\ell=0}^{\infty} \ell^s \alpha(\ell)^r < \infty$.
- There are alternative measures of dependence and many have been proposed. Strong mixing is one of the weakest and often not sufficient. For example $\beta - \textit{mixing}$ (which implies strong mixing). For now we will just assume strong-mixing unless you have super technical work.

# Mixing

Mixing is useful as it is preserved by transformations.

**Theorem**

If $y_t$ has mixing coefficients $\alpha_y(\ell)$ and $x_t = \phi\left(\boldsymbol{y}_t, \boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \ldots, \boldsymbol{y}_{t-q}\right)$ then $\boldsymbol{x}_t$ has mixing coefficients $\alpha_x(\ell) \leq \alpha_y(\ell - q)$ (for $\ell \geq q$). The coefficients $\alpha_x(m)$ satisfy the same summation and rate conditions as $\alpha_y(\ell)$.

A limitation of the above result is that it is confined to a finite number of lags, unlike the transformation results for stationarity and ergodicity. Mixing can be a useful tool because of the following inequalities.

# Inequalities for Mixing processes

**Theorem**

Suppose that $x_{t-\ell}$ and $z_t$ are random variables which are $\mathcal{F}_{-\infty}^{t-\ell}$ and $\mathcal{F}_t^{\infty}$ measurable, respectively. 1. If $|x_t| \leq C_1$ and $|z_t| \leq C_2$ then

$$|\text{cov}\,(x_{t-\ell}, z_t)| \leq 4C_1C_2\alpha(\ell).$$

2. If $\mathbb{E}\,|x_t|^r < \infty$ and $\mathbb{E}\,|z_t|^q < \infty$ for $1/r + 1/q < 1$ then

$$|\text{cov}\,(x_{t-\ell}, z_t)| \leq 8\left(\mathbb{E}\,|x_t|^r\right)^{1/r}\left(\mathbb{E}\,|z_t|^q\right)^{1/q}\alpha(\ell)^{1-1/r-1/q}.$$

3. If $\mathbb{E}\,(y_t) = 0$ and $\mathbb{E}\,|y_t|^r < \infty$ for $r \geq 1$ then

$$\mathbb{E}\left|\mathbb{E}\left(y_t \mid \mathcal{F}_{-\infty}^{t-\ell}\right)\right| \leq 6\left(\mathbb{E}\,|y_t|^r\right)^{1/r}\alpha(\ell)^{1-1/r}.$$

# i.i.d and Mixing

Finally, it follows directly that

**Theorem**

If $y_t$is i.i.d. then it is strong mixing and ergodic.

Now we are ready to derive a CLS for correlated observations/dependent processes.

# CLT for Correlated observations

Recall

$$S_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} u_t$$

We saw that, in the scalar case

$$\text{var}\,(S_n) = \sigma^2 + 2 \sum_{\ell=1}^{n} \left(1 - \frac{\ell}{n}\right) \gamma(\ell)$$

where $\sigma^2 = \text{var}\,(u_t)$ and $\gamma(\ell) = \text{cov}\,(u_t, u_{t-\ell})$. Since $\gamma(-\ell) = \gamma(\ell)$ this can be written as

$$\text{var}\,(S_n) = \sum_{\ell=-n}^{n} \left(1 - \frac{|\ell|}{n}\right) \gamma(\ell)$$

# CLT for Correlated observations

In the vector case define the variance

$$\boldsymbol{\Sigma} = \mathbb{E}\left(\boldsymbol{u}_t \boldsymbol{u}_t'\right)$$

and the matrix covariance

$$\Gamma(\ell) = \mathbb{E}\left(u_t u_{t-\ell}'\right)$$

which satisfies $\boldsymbol{\Gamma}(-\ell) = \boldsymbol{\Gamma}(\ell)'$. We obtain by a calculation analogous to (14.14)

$$\begin{aligned}\operatorname{var}\left(\boldsymbol{S}_n\right) &= \boldsymbol{\Sigma} + \sum_{\ell=1}^{n}\left(1 - \frac{\ell}{n}\right)\left(\boldsymbol{\Gamma}(\ell) + \boldsymbol{\Gamma}(\ell)'\right) \\ &= \sum_{\ell=-n}^{n}\left(1 - \frac{|\ell|}{n}\right)\boldsymbol{\Gamma}(\ell)\end{aligned}$$

# CLT for Correlated observations

A necessary condition for $\boldsymbol{S}_n$ to converge to a normal distribution is that the variance (14.15) converges to a limit. Indeed,

$$\sum_{\ell=1}^{n}\left(1-\frac{\ell}{n}\right)\boldsymbol{\Gamma}(\ell)=\frac{1}{n}\sum_{\ell=1}^{n-1}\sum_{j=1}^{\ell}\boldsymbol{\Gamma}(j)\longrightarrow\sum_{\ell=0}^{\infty}\boldsymbol{\Gamma}(\ell)$$

where the convergence holds if the limit is convergent. A necessary condition for this to hold is that the covariances $\boldsymbol{\Gamma}(\ell)$ decline to zero as $\ell\to\infty$, which is stronger than ergodicity. A sufficient condition is that the covariances are absolutely summable, which can see that $\text{var}\,(\boldsymbol{S}_n)$ converges if $\mathbb{E}\,\|\boldsymbol{u}_t\|^r<\infty$ and $\sum_{\ell=0}^{\infty}\alpha(\ell)^{1-2/r}<\infty$.

$$\text{var}\,(\boldsymbol{S}_n)\longrightarrow\sum_{\ell=-\infty}^{\infty}\boldsymbol{\Gamma}(\ell)\overset{\text{def}}{=}\Omega.$$

It turns out that these conditions are sufficient for the CLT.

# CLT for Martingale Differences

## Theorem: CLT for Dependent Processes

Theorem 14.19 If $\boldsymbol{u}_t$ is strictly stationary with mixing coefficients $\alpha(\ell)$ $\mathbb{E}(\boldsymbol{u}_t) = \mathbf{0}$, for some $r > 2, \mathbb{E}\|\boldsymbol{u}_t\|^r < \infty$ and $\sum_{\ell=1}^{\infty} \alpha(\ell)^{1-2/r} < \infty$, then (14.17) is convergent, and

$$\boldsymbol{S}_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \boldsymbol{u}_t \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{\Omega}).$$

- The theorem requires $r > 2$ finite moments which is stronger than the MDS CLT.
- The summability condition on the mixing coefficients is considerably stronger than ergodicity.
- There is a trade off involving the choice of $r$. A larger $r$ means more moments are required finite, but a slower decay in the coefficients $\alpha(\ell)$ is allowed. Smaller $r$ is less restrictive regarding moments, but requires a faster decay rate in the mixing coefficients.

# CLT for MA

For an MA, $x_t = c(L)e_t$, we have $\sum_{j=1}^{\infty} |c_j| < \infty$ implies $\sum_{-\infty}^{\infty} |\gamma_j| < \infty$

$$\gamma_k = \sum_{j=0}^{\infty} c_j c_{j+k}$$

$$\sum_{k=0}^{\infty} |\gamma_k| = \sum_{k=0}^{\infty} \left| \sum_{j=0}^{\infty} c_j c_{j+k} \right| \leq \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} |c_j| \, |c_{j+k}|$$

$$\leq \sum_{l=0}^{\infty} \sum_{j=0}^{\infty} |c_j| \, |c_l| = \left( \sum_{j=0}^{\infty} |c_j| \right)^2 < \infty$$

$$\text{Var}\left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i \right) = \gamma_0 + 2 \sum_{k=1}^{n} \gamma_k \left( 1 - \frac{k}{n} \right) \to \gamma_0 + 2 \sum_{k=1}^{\infty} \gamma_k = \mathcal{J}$$

$\mathcal{J}$ is called the long-run variance and is a correct scale measure.

# CLT for $MA(\infty)$

Let $y_t = \mu + \sum_{j=0}^{\infty} c_j e_{t-j}$, where $e_t$ is independent white noise and $\sum_{j=0}^{\infty} |c_j| < \infty$, then

$$\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T} y_t - \mu\right) \Rightarrow N(0, \mathcal{J})$$

# LR Variance for $AR(1)$ and ARMA

AR(1) $y_t = \rho y_{t-1} + e_t$

$$\gamma_k = \frac{\sigma^2 \rho^k}{1-\rho^2}$$

$$\mathcal{J} = \gamma_0 + 2\sum_{k=1}^{\infty} \gamma_k = \frac{\sigma^2}{1-\rho^2}\left(1 + 2\sum_{k=1}^{\infty} \rho^k\right) = \frac{\sigma^2}{(1-\rho)^2} = \gamma(1)$$

If $a(L)y_t = b(L)e_t$, then

$$\gamma(\xi) = \sigma^2 \frac{b(\xi)b\left(\xi^{-1}\right)}{a(\xi)a\left(\xi^{-1}\right)}$$

So

$$\mathcal{J} = \left(\frac{b(1)}{a(1)}\right)^2 \sigma^2$$

# LR Variance for VARMA

If $\{y_t\}$ is a vector, then let $\Gamma_k = \text{cov}(y_t, y_{t+k})$ and $\mathcal{J} = \sum_{-\infty}^{\infty} \Gamma_k$. The only thing that's different from the scalar case is that $\Gamma_k \neq \Gamma_{-k}$. Instead, $\Gamma_k = \Gamma'_{-k}$. All the formulas above also hold, except in matrix notation. For example, for a VARMA,

$$\mathcal{J} = A^{-1} B \Sigma B' A^{-1'}$$

Do you recognize this?

# OLS Review (skip?)

- Suppose $y_t = x_t\beta + u_t$.
- In cross-section $x_t$ is always independent from $u_s$ if $s \neq t$ due to iid assumption, so the exclusion restriction is formulated as $E(u_t \mid x_t) = 0$.
- In time series, however, we have to describe the dependence between error terms and all regressors.
- **Definition.** $x_t$ is weakly exogenous if $E(u_t \mid x_t, x_{t-1}, \ldots) = 0$
- **Definition.** $x_t$ is strictly exogenous if $E\left(u_t \mid \{x_t\}_{t=-\infty}^{\infty}\right) = 0$
- Usually, strict exogeneity is too strong an assumption, it is difficult to find a good empirical example for it.
- The weak exogeneity is much more functional (and we will mainly assume it).OLS estimator: $\hat{\beta} = (X'X)^{-1}(X'y)$

What is the asymptotic distribution?

# OLS Review (skip?)

$$\sqrt{T}(\hat{\beta} - \beta) = \left(\frac{1}{T}X'X\right)^{-1}\left(\frac{1}{\sqrt{T}}X'u\right)$$
$$= \left(\frac{1}{T}\sum_t x_t x_t'\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_t x_t u_t\right)$$

Appropriate assumptions will gives us a LLN for $\left(\frac{1}{T}\sum_t x_t x_t'\right) \to M$. Assume also condition for for $z_t = x_t u_t$ CLT are satisfied .

$$\left(\frac{1}{\sqrt{T}}\sum_t x_t u_t\right) \Rightarrow N(0, \mathcal{J}),$$

which means that

$$\sqrt{T}(\hat{\beta} - \beta) \Rightarrow N\left(0, M^{-1}\mathcal{J}M^{-1}\right)$$

The only thing that is different from usual is the $\mathcal{J}$. $\mathcal{J} = \sum_{-\infty}^{\infty} \gamma_j$ (where $\gamma_j$ are the autocovariances of $z_t = x_t u_t$) is called the long-run variance. Thus the need for HAC standard errors

# LR VCV matrix estimation

Now you can see what estimation of the LR variance is important (HAC and any asymptotic involving dependent data). How do we estimate the LR variance?

- Parametric Approach
- Non parametric Approach

# Parametric LR Variance Estimation

Assume $z_t$ is $\mathrm{AR(p)}$ :

$$z_t = a_1 z_{t-1} + \ldots + a_p z_{t-p} + e_t$$

then $\mathcal{J} = \frac{\sigma^2}{a(1)^2}$, where $a(L) = 1 - a_1 L - \ldots a_p L^p$. We can proceed in the following way: run OLS regression of $z_t$ on $z_{t-1}, \ldots, z_{t-p}$, get $\hat{a}_1, \ldots, \hat{a}_p$ and $\hat{\sigma}^2$, then use $\hat{a}(L) = 1 - \hat{a}_1 L - .. - \hat{a}_p L^p$ to construct $\hat{\mathcal{J}}$,

$$\hat{\mathcal{J}} = \frac{\hat{\sigma}^2}{\hat{a}(1)^2}.$$

Two important practical questions:

- What $p$ should we use? - model selection criteria, BIC (Bayesian informaiton criteria)
- What if $z_t$ is not $\mathrm{AR(p)}$ ? (this is still an open question)
- Den Haan and Levin (1997) showed that if $z_t$ is AR(p), then the convergence of the parameteric estimator is faster than the kernel estimator described below.

# Non-Parametric LR Variance Estimation - Naive Approach

$\mathcal{J}$ is the sum of all auto-covariance. We can estimate $T-1$ of these, but not all. What if we just use the ones we can estimate, i.e.

$$\tilde{\mathcal{J}} = \sum_{k=T-1}^{T-1} \hat{\gamma}_k, \hat{\gamma}_k = \frac{1}{T} \sum_{j=1}^{T-k} z_j z_{j+k}$$

It turns out that this is a very bad idea.

$$\tilde{\mathcal{J}} = \sum_{k=-(T-1)}^{T-1} \hat{\gamma}_k = \frac{1}{T} \sum_{k=-(T-1)}^{T-1} \sum_{j=1}^{T-k} z_j z_{j+k} = \frac{1}{T} \left( \sum_{t=1}^{T} z_t \right)^2$$

$$= \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} z_t \right)^2 \Rightarrow N(0, \mathcal{J})^2$$

so $\tilde{\mathcal{J}}$ is not consistent; it converges to a distribution instead of a point. The problem is that we're summing too many imprecisely estimated covariances. So, the noise does not die out.

# Non-Parametric LR Variance Estimation - Truncated Sum

What if we don't use all the covariances?

$$\tilde{\mathcal{J}}_2 = \sum_{k=-S_T}^{S_T} \hat{\gamma}_k$$

where $S_T < T$ and $S_T \to \infty$ as $T \to \infty$, but more slowly.

- Notice that due to truncation there will be a finite sample bias.
- As $S_T$ will increase the bias due to truncation should be smaller and smaller.
- But we don't want to increase $S_T$ too fast for the reason stated above (we don't want to sum up noises)
- Even if we find a $S_T$ in such a way that this estimator is consistent, we might still face another bad small sample property as the estimate of long run variance may be negative

# Non-Parametric LR Variance Estimation - Weighted Truncated Sum

The renewed suggestion is to create a weighted sum of sample auto-covariances with weights guaranteeing positive-definiteness:

$$\hat{J} = \sum_{j=-S_T}^{S_T} k_T(j)\hat{\gamma}_j$$

Remark 16. $k_T()$ is called a kernel.

- We need conditions on $S_T$ and $k_T()$ to give us consistency and positive-definiteness.
- $S_T$ should increase $S_T \to \infty$ as $T \to \infty$, but but not too fast.
- $k_T()$ needs to be such that it guarantees positive-definiteness by down-weighting high lag covariances.
- Also need $k_T() \to 1$ for consistency.