

Orthogonal/Proxy sampling in time series

Suhasini Subba Rao
Texas A&M University

NBER-NSF Time Series Conference, Vienna 2015

Background

- In time series we often want to estimate parameters or test a hypothesis under a minimal number of modelling assumptions.
- However, the ‘cost’ of weak modelling assumptions usually mean that inference will depend on nuisance parameters which need to estimated.
- The objective here is to estimate these nuisance parameters.
- The main set of assumptions we use is that $\{X_t\}$ is a univariate stationary time series whose fourth moment exists (and sufficient mixing-type assumptions). We suppose we observe $\{X_t\}_{t=1}^T$.
 - It is clear how to generalised to the multivariate set-up.

Motivation

- The sample autocovariance is $\tilde{c}_T(j) = \frac{1}{T} \sum_{t=1}^{T-|j|} (X_t - \bar{X})(X_{t+|j|} - \bar{X})$.

$$\begin{aligned} & T \text{cov}[\tilde{c}_T(j_1), \tilde{c}_T(j_2)] \\ &= \sum_{k=-\infty}^{\infty} c(k)c(k + j_1 - j_2) + \sum_{k=-\infty}^{\infty} c(k - j_1)c(k + j_2) + \\ & \underbrace{\sum_{k=-\infty}^{\infty} \kappa_4(j_1, k, k + j_2)}_{\text{Fourth order cumulant}} + O(T^{-1}), \end{aligned}$$

- The sample covariance can be rewritten in terms of a weighted periodogram, $I_T(\omega_k) = |J_T(\omega_k)|^2$ where $J_T(\omega_k) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^T X_t e^{it\omega_k}$ and $\omega_k = \frac{2\pi k}{T}$ are the so called Fourier frequencies.

- In the next few examples we consider statistics which are formulated in terms of the weighted periodogram

$$A_T(\phi_j) = \frac{1}{T} \sum_{k=1}^T \phi_j(\omega_k) |J_T(\omega_k)|^2.$$

Different $\phi_j(\cdot)$ lead to estimators of different quantities:

- The Whittle likelihood, $\phi(\omega) = \nabla_\theta f_\theta^{-1}(\omega)$ (where f_θ denotes the spectral density function).
- The Autocovariance $\phi_j(\omega) = e^{ij\omega}$.
- Density estimation $\phi(\omega) = \frac{1}{b} W\left(\frac{u-\omega}{b}\right)$.
- Goodness of fit tests: $\phi_j(\omega) = e^{ij\omega} f_\theta(\omega)^{-1}$

In general, any quadratic form $T^{-1} \sum_{t,\tau=1}^T h_{t-\tau} X_t X_\tau$ can be approx. written in the above form (under sufficient conditions).

Examples with their variance

The Whittle likelihood involves the variance

$$\frac{2}{2\pi} \int_0^{2\pi} [\nabla f_\theta(\omega)] [\nabla f_\theta(\omega)]' d\omega + \\ \underbrace{\frac{1}{(2\pi)^2} \int_0^{2\pi} \int_0^{2\pi} \nabla f_\theta(\omega_1) [\nabla f_\theta(\omega_2)]' f_4(\omega_1, -\omega_1, \omega_2) d\omega_1 d\omega_2}_{\text{Under linearity} = 0}.$$

The spectral density estimator $\widehat{f}_T(\omega_j) = \frac{1}{bT} \sum_k W\left(\frac{\omega_j - \omega_k}{b}\right) |J_T(\omega_k)|^2$:

$$bT \text{var}[\widehat{f}_T(\omega_j)] = \underbrace{\frac{1}{2\pi b} \int_0^{2\pi} W\left(\frac{\omega_j - \omega}{b}\right) f(\omega)^2 d\omega}_{\approx f(\omega_j)^2} + \\ \underbrace{\frac{1}{(2\pi)^2 b} \int_0^{2\pi} \int_0^{2\pi} W\left(\frac{\omega_j - \omega_1}{b}\right) W\left(\frac{\omega_j - \omega_2}{b}\right) f_4(\omega_1, -\omega_1, \omega_2) d\omega_1 d\omega_2}_{=O(b)} + O\left(\frac{1}{bT}\right).$$

- Spectral based tests, where null hypothesis is written in the form $\psi(f(\omega), \omega) = 0$ for all $\omega \in [0, 2\pi]$. We then use the test statistic

$$S_T = \frac{1}{T} \sum_{k=1}^T \left\| \psi \left(\hat{f}_T(\omega_k), \omega_k \right) \right\|_2^2$$

In the general case S_T has mean and variance that involves a function of the spectral density function and lower order cumulant terms (see Taniguchi and Kakizawa (2000) and Eichler (2007)).

What to do?

- To estimate the variance and sampling distribution of various statistics of time series. Several different methods have been developed.
- **Self-normalisation** Lobato (2001) and Shao (2009), (2010).
- **Bootstrap methods**
 - **Frequency domain bootstrap** See Hurvich and Zeger (1987) and Franke and Härdle (1992)).
Dahlhaus and Janas (1996) generalize the methods to so called ratio statistics, of linearity of the time series
 - **Residual bootstrap** Mainly designed for linear processes (but can also be used to obtain second order properties) (see, Kreiss (1996) and Kreiss, Paparoditis and Politis (2011)).

- **Block bootstrap** Not restricted to linear time series. The Block bootstrap can be applied to many statistics (see Künsch (1989) and Lahiri (2003) for a review). However it depends on the choice of block length (tuning parameter). Recently the **fixed-b bootstrap** has been developed which avoids the influence of bandwidth (see, Lahiri (2001), Keifer and Vogelsang (2005), Hashimzade and Vogelsang (2007), Shao and Politis (2013)).
- Note: this list is far from incomplete.

Overview

- All the methods described above work well in various situations. However some can be computationally cumbersome.
- Here we describe a ‘quick, dirty and lazy’ method (it takes $O(T \log T)$ computing operations) for estimating the nuisance and distributions for a broad class of statistics.
- We use this method:
 - Estimate the variance of an estimator and quantify its uncertainty.
 - Estimate the distribution function of test statistics under the null.

Motivating the approach: Toy example I

- Suppose $\{X_t\}_{t=1}^T$ are iid $N(\mu, \sigma^2)$. Let $\bar{X} = \frac{1}{T} \sum_{t=1}^T X_t$ and $s^2 = \frac{1}{T-1} \sum_{t=1}^T (X_t - \bar{X})^2$. It is well known that
$$\left(\sqrt{T} \frac{(\bar{X} - \mu)}{\sigma}, (T-1) \frac{s^2}{\sigma^2} \right) \sim (Z, \chi_{T-1}^2) \quad \text{which gives the t-distribution.}$$

- **Reminder of PROOF** $(T-1) \frac{s^2}{\sigma^2} \sim \chi_{T-1}^2$

Define the vector $\mathbf{e}_0 = T^{-1/2}(1, 1, \dots, 1)$, and $(T-1)$ vectors $\{\mathbf{e}_j; j = 1, \dots, T-1\}$ such that $\{\mathbf{e}_j; j = 0, \dots, T-1\}$ forms an orthonormal basis of \mathbb{R}^T . Let $\mathbf{X} = (X_1, \dots, X_T)$. The properties of $a_T(j) = \langle \mathbf{e}_j, \mathbf{X} \rangle$ are:

$$(1) \quad \sqrt{T} \bar{X} = \langle \mathbf{e}_0, \mathbf{X} \rangle.$$

(2) For $1 \leq j \leq T - 1$, $\mathbf{E}[\langle \mathbf{e}_j, \mathbf{X} \rangle] = 0$ and for all j

$$\text{cov}(\langle \mathbf{e}_{j_1}, \mathbf{X} \rangle, \langle \mathbf{e}_{j_2}, \mathbf{X} \rangle) = \begin{cases} \sigma^2 & j_1 = j_2 \\ 0 & j_1 \neq j_2 \end{cases}$$

Thus we have the decomposition

$$\left(\sqrt{T}(\bar{X} - \mu), \langle \mathbf{e}_1, \mathbf{X} \rangle, \dots, \langle \mathbf{e}_{T-1}, \mathbf{X} \rangle \right) \sim N(0, \sigma^2 I_T).$$

(3) $\{\langle \mathbf{e}_j, \mathbf{X} \rangle; j = 1, \dots, T - 1\}$ has zero mean and same variance as $\sqrt{T}\bar{X}$ and can be used to estimate the variance with

$$\hat{\sigma}^2 = \frac{1}{T-1} \sum_{j=1}^{T-1} \langle \mathbf{e}_j, \mathbf{X} \rangle^2 = s^2 \text{ (using projection arguments).}$$

- By the uncorrelatedness of $\langle \mathbf{e}_j, \mathbf{X} \rangle$ it is immediately clear that $\sum_{j=1}^{T-1} \langle \mathbf{e}_j, \mathbf{X} \rangle^2 = (T - 1)s^2 \sim \chi_{T-1}^2$.
- Returning to $\frac{1}{T-1} \sum_{j=1}^{T-1} \langle \mathbf{e}_j, \mathbf{X} \rangle^2$, we can consider $\{\langle \mathbf{e}_j, \mathbf{X} \rangle; j = 1, \dots, T - 1\}$ as the orthogonal/proxy sample associated with $\sqrt{T}\bar{X}$, meaning that (1) it contains **no** information about the mean, μ (2) since it has the same variance as $\sqrt{T}\bar{X}$, we can say it has similar properties to the *centralized* sample mean $\sqrt{T}(\bar{X} - \mu)$, and thus can be used to estimate the variance.

Toy example II: sample mean again!

- Suppose we relax the condition on $\{X_t\}$ and assume it is a second order stationary time series, ie. $\mathbf{E}[X_t] = \mu$ and $\text{cov}(X_t, X_{t+s}) = c(s)$.
- Let $\sqrt{T}\bar{X} = \langle \mathbf{X}, \mathbf{e}_0 \rangle$. However, unlike the iid case, using *any* set of orthonormal vectors which are orthogonal to \mathbf{e}_0 will not yield a suitable orthogonal sample.

Their mean will be zero, **but** the variance will not be the same, which renders the approach useless.

- Nevertheless, using some well known results about stationary time series does yield a suitable orthogonal basis.

The Discrete Fourier Transform

- We recall that the Discrete Fourier transform (DFT) is a linear transformation of the data:

$$J_T(\omega_k) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^T X_t \exp(it\omega_k) = \frac{1}{\sqrt{2\pi}} \langle \mathbf{e}_j, \mathbf{X} \rangle \quad k = 0, \dots, T/2,$$

where $\mathbf{e}_k = (1, e^{i\omega_k}, e^{2i\omega_k}, \dots, e^{i(T-1)\omega_k})$ and

- $\sqrt{2\pi} J_T(0) = \sqrt{2\pi} \langle \mathbf{e}_0, \mathbf{X} \rangle = \sqrt{T} \bar{X}$.
- For $1 \leq k \leq T/2$ $\mathbf{E}[J_T(\omega_k)] = 0$. Let f denote the spectral density then

$$\text{cov} [J_T(\omega_k), J_T(\omega_{k+r})] = \begin{cases} f(\omega_k) + O(T^{-1}) & r = 0 \\ O(T^{-1}) & r \neq 0. \end{cases}$$

- Therefore the DFT almost orthogonalizes the (stationary) time series and for frequencies which are close together the variance of the DFT is almost the same.
- Returning to the sample mean, for fixed M and large T , we have

$$\left(\sqrt{\frac{T}{2\pi}}(\bar{X} - \mu), \mathbf{J}_T(\omega_1), \dots, \mathbf{J}_T(\omega_M) \right) \sim N(0, f(0)I_{2M+1})$$

where $\mathbf{J}_T(\omega_k) = \sqrt{2}(\Re J_T(\omega_k), \Im J_T(\omega_k))$ and $f(0) = \frac{1}{2\pi} \sum_{s \in \mathbb{Z}} c(s)$.

- Hence $\{\mathbf{J}_T(\omega_k); 1 \leq k \leq M\}$ (where $M \ll T$) can be considered as the orthogonal sample associated with \bar{X} .

- To estimate the variance of the sample mean, $\sum_{s \in \mathbb{Z}} c(s)$ we use

$$\hat{\sigma}^2 = \frac{2\pi}{M} \sum_{k=1}^M |J_T(\omega_k)|^2 \text{ and } \frac{\sqrt{T}(\bar{X} - \mu)}{\sqrt{\frac{2\pi}{M} \sum_{k=1}^M |J_T(\omega_k)|^2}} \xrightarrow{\mathcal{D}} t_{2M}$$

- This is a local average of the periodogram and is simply one of the well known methods of estimating the long run variance.
- **Objective** The above can be generalized to ‘mean-type’ statistics such as the sample covariance, however, our aim is to define the orthogonal samples for a broader class of statistics.

Variance estimation for quadratic forms

- We generalize the above approach to the integrated periodogram estimators

$$A_T(\phi_j) = \frac{1}{T} \sum_{k=1}^T \phi_j(\omega_k) |J_T(\omega_k)|^2 \left(\approx T^{-1} \sum_{t,\tau=1}^T h_{t-\tau} X_t X_\tau \right)$$

which estimates the functional

$$A(\phi_j) = \frac{1}{2\pi} \int_0^{2\pi} \phi_j(\omega) f(\omega) d\omega$$

- $\text{var}[\sqrt{T} A_T(\phi_j)] = \nu_j + O(T^{-1})$ and $\text{cov}[\sqrt{T} A_T(\phi_{j_1}), \sqrt{T} A_T(\phi_{j_2})] = \nu_{j_1 j_2} + O(T^{-1})$

where (using Rosenblatt-Brillinger results) we have

$$\begin{aligned}\nu_j &= \frac{1}{2\pi} \int_0^{2\pi} f(\omega)^2 \left(|\phi_j(\omega)|^2 + \phi_j(\omega) \overline{\phi_j(-\omega)} \right) d\omega + \\ &\quad \frac{1}{(2\pi)^2} \int_0^{2\pi} \int_0^{2\pi} \phi_j(\omega_1) \overline{\phi_j(\omega_2)} f_4(\omega_1, -\omega_1, \omega_2) d\omega_1 d\omega_2 \\ \nu_{j_1 j_2} &= \frac{1}{2\pi} \int_0^{2\pi} f(\omega)^2 \left(\phi_{j_1}(\omega) \overline{\phi_{j_2}(\omega)} + \phi_{j_1}(\omega) \overline{\phi_{j_2}(-\omega)} \right) d\omega + \\ &\quad \frac{1}{(2\pi)^2} \int_0^{2\pi} \int_0^{2\pi} \phi_{j_1}(\omega_1) \overline{\phi_{j_2}(\omega_2)} f_4(\omega_1, -\omega_1, \omega_2) d\omega_1 d\omega_2,\end{aligned}$$

(for the spectral density case we multiple ν by b).

- **Objective** Find an orthogonal sample associated with $A_T(\phi)$ to estimate ν_j and $\nu_{j_1 j_2}$ (we sometimes suppress the notation j).

The ingredients for the orthogonal sample

- To construct the orthogonal sample associated with $A_T(\phi)$ it should have mean zero and variance close to $\text{var}[A_T(\phi)]$. $A_T(\phi)$ is a weighted average of the periodogram $|J_T(\omega_k)|^2 = \langle J_T(\omega_k), J_T(\omega_k) \rangle$.
 - Using the periodogram $|J_T(\omega_k)|^2 = \langle J_T(\omega_k), J_T(\omega_k) \rangle$ we shift one-side by $r J_T(\omega_k) \overline{J_T(\omega_{k+r})} = \langle J_T(\omega_k), J_T(\omega_{k+r}) \rangle$ (this is the bi-periodogram as defined in Goodman (1965));
- (1) $|J_T(\omega_k)|^2$ estimates the spectral density whereas its shift $J_T(\omega_k) \overline{J_T(\omega_{k+r})}$ estimates zero.
- (2) The sequence $\{|J_T(\omega_k)|^2, \sqrt{2}\Re J_T(\omega_k) \overline{J_T(\omega_{k+r})}, \sqrt{2}\Im J_T(\omega_k) \overline{J_T(\omega_{k+r})}; r = 1, \dots, M\}$ is near uncorrelated with approximately the same variance.

The orthogonal sample associated with $A_T(\phi)$

- Using the same argument, given the estimator

$$A_T(\phi) = \frac{1}{T} \sum_{k=1}^T \phi(\omega_k) |J_T(\omega_k)|^2$$

we replace $|J_T(\omega_k)|^2$ with its shift $J_T(\omega_k) \overline{J_T(\omega_{k+r})}$ and define

$$A_T(\phi; r) = \frac{1}{T} \sum_{k=1}^T \phi(\omega_k) J_T(\omega_k) \overline{J_T(\omega_{k+r})}.$$

- To simplify notation, we split $A_T(\phi; r)$ into its real and imaginary parts, and define the vector $\mathbf{A}_T(\phi; r) = \sqrt{2} (\Re A_T(\phi; r), \Im A_T(\phi; r))$

Properties of the orthogonal sample

Lemma Let us suppose that $\{X_t\}$ is a short memory stationary time series with finite fourth moment:

The mean

$$\mathbf{E}[A_T(\phi; r)] = \begin{cases} \frac{1}{2\pi} \int_0^{2\pi} \phi(\omega) f(\omega) d\omega + O(T^{-1}) & r = 0 \text{ (statistic of interest)} \\ O(T^{-1}) & 0 < r < T/2 \end{cases}$$

The variance

$$T \text{cov}[\mathbf{A}_T(\phi; r_1), \mathbf{A}_T(\phi; r_2)] = \begin{cases} \nu I_2 + O([1 + |r|]T^{-1}) & 0 < r_1 = r_2 (= r) \\ O(T^{-1}) & 0 < r_1 \neq r_2 \leq T/2 \end{cases}$$

Under sufficient mixing/moment conditions we have

$$\sqrt{T} [(A_T(\phi) - A(\phi)), \mathbf{A}_T(\phi; 1), \dots, \mathbf{A}_T(\phi; M)] \xrightarrow{\mathcal{D}} N(0, vI_{2M+1})$$

Estimating the variance ν

- Using the orthogonal sample we estimate the variance with:

$$\widehat{\nu}_M = \frac{T}{M} \sum_{r=1}^M |A_T(\phi; r)|^2.$$

- **Lemma** Suppose that $\{X_t\}$ is a stationary time series with finite 8th moment (and satisfies certain mixing conditions). Then $\widehat{\nu}_M/\nu \xrightarrow{\mathcal{D}} \frac{1}{2M}\chi_{2M}^2$ and

$$\mathbf{E} (\widehat{\nu}_M - \nu)^2 = O \left(\frac{M^2}{T^2} + \frac{1}{M} \right).$$

Application

- Using these results, we have

$$\frac{\sqrt{T}[A_T(\phi) - A(\phi)]}{\sqrt{\widehat{\nu}_M}} \xrightarrow{\mathcal{D}} \frac{Z}{\sqrt{\frac{1}{2M}\chi_{2M}^2}} \sim t_{2M},$$

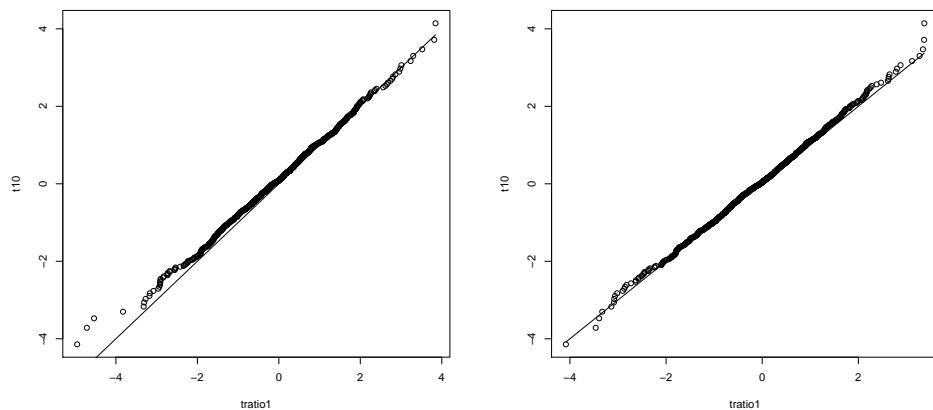
- In general, for several estimators $\{A_T(\phi_j)\}$ together

$$T(A_T - A)' \widehat{\Sigma}_M^{-1} (A_T - A) \xrightarrow{\mathcal{D}} T_{p,2M}^2$$

where $A_T = (A_T(\phi_1), \dots, A_T(\phi_p))$ is the estimator of interest (for example the sample autocovariance at various lags), $\widehat{\Sigma}_M$ is the analogous variance estimator (of Σ_M) based on the orthogonal sample and T^2 denotes Hotelling's T-squared distribution.

Example 1: iid normal (sanity check)

For sample sizes $T = 100$ and 200 , we estimated the sample autocovariance and estimated the sample variance using the orthogonal sample $\{A_T(e^{i\cdot}; r); r = 1, \dots, 5\}$



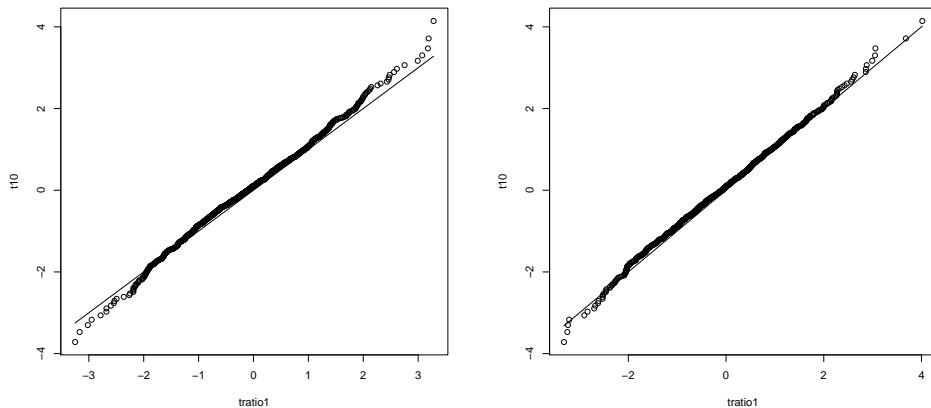
$$t_{10} = \frac{A_T(e^{i\cdot})}{\sqrt{\frac{1}{5} \sum_{r=1}^5 |A_T(e^{i\cdot}; r)|^2}}.$$

This is a QQplot against the t-distribution with 10df using $T = 100$ and 200 over 1000 replications.

Example 2

Non-causal, uncorrelated, linear time series $\{X_t\}_{t=1}^{200}$

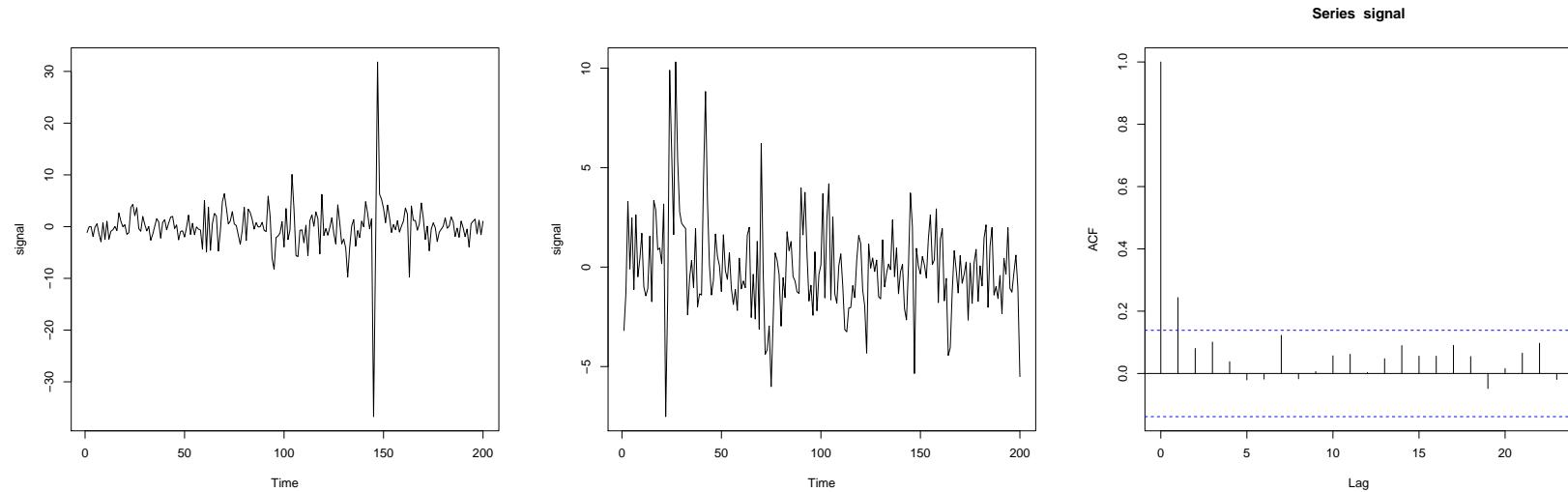
$$X_t = \sum_{j=0}^{\infty} 0.6^j t_{5,t-j} - \frac{0.6}{1 - 0.6^2} t_{5,t+1} \quad \{t_{5,j}\}_j \text{ iid t-dist. with 5df.}$$



This is a QQplot against the t-distribution with 10df using $T = 100$ and 200 over 1000 replications.

$$t_{10} = \frac{A_T(e^{i\cdot})}{\sqrt{\frac{1}{5} \sum_{r=1}^5 |A_T(e^{i\cdot}; r)|^2}}.$$

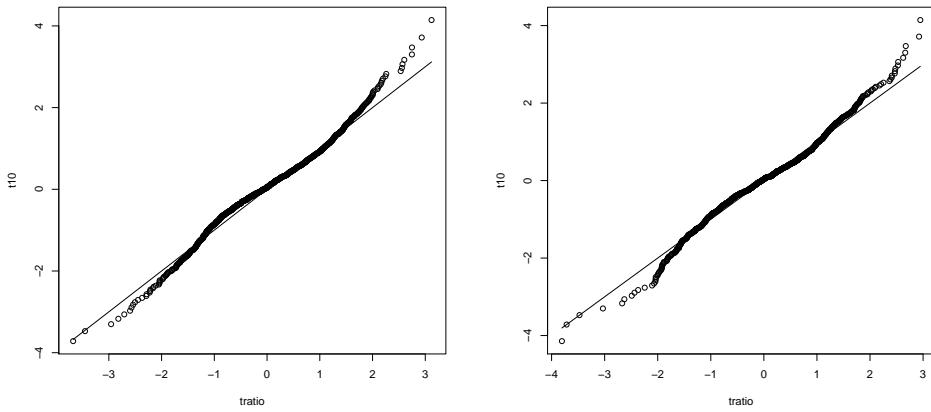
Example 3



Realisations and empirical ACF plots from a dependent but uncorrelated process:

$$X_t = \sum_{j=0}^{\infty} 0.6^j U_{t-j} - \frac{0.6}{1 - 0.6^2} U_{t+1} \quad \text{where}$$

$$U_t = \sigma_t Z_t \quad \sigma_t^2 = 1 + 0.7 U_{t-1}^2 \quad \text{ARCH model.}$$



$$t_{10} = \frac{A_T(e^{i\cdot})}{\sqrt{\frac{1}{5} \sum_{r=1}^5 |A_T(e^{i\cdot}; r)|^2}}.$$

- This is a QQplot against the t-distribution with 10df using $T = 100$ and 200 over 1000 replications.
- Note the fourth moment of $\{X_t\}$ does not exist (hence the variance of the estimator is not finite).

Extensions to spectral based tests

- Using the above construction, the orthogonal sample associated with the spectral density estimator

$$\widehat{f}(\omega_j) = \frac{1}{bT} \sum_k W\left(\frac{\omega_{j-k}}{b}\right) |J_T(\omega_k)|^2 \rightarrow \widehat{f}(\omega_j; r) = \frac{1}{bT} \sum_k W\left(\frac{\omega_{j-k}}{b}\right) J_T(\omega_k) \overline{J_T(\omega_{k+r})}.$$

- **Example** Testing equality of two spectral densities

$$S_T = \frac{1}{T} \sum_{j=1}^T |\widehat{f}_1(\omega_j) - \widehat{f}_2(\omega_j)|^2.$$

Replacing in S_T , $\widehat{f}_1(\omega)$ and $\widehat{f}_2(\omega)$ with $\widehat{f}_1(\omega; r)$ and $\widehat{f}_2(\omega; r)$ gives rise to its orthogonal sample $\{S_T(r)\}$, which can be used to estimate the mean and variance under the null.

Application to testing

Testing the hypothesis $H_0 : A(\phi_j) = 0$ for $1 \leq j \leq L$ against the alternative at least one is non-zero

Examples: Testing for zero correlation $c(j) = 0$, goodness of fit.

- One possibility is to use the T-square statistic, which under the null is

$$T(A_T(\phi_1), \dots, A_T(\phi_L)) \widehat{\Sigma}_M^{-1} (A_T(\phi_1), \dots, A_T(\phi_L))' \xrightarrow{\mathcal{D}} T_{p,2L}^2,$$

and $\widehat{\Sigma}_M$ is the sample variance covariance matrix estimated using the corresponding orthogonal sample.

- However, this does require the inversion of the matrix $\widehat{\Sigma}_M$.

- Another possibility is simply to use the ‘classical’ L_2 statistic

$$Q_T = T \sum_{j=1}^L |A_T(\phi_j)|^2$$

- Unless under the null $\{A_T(\phi_j)\}$ are asymptotically uncorrelated with the same variance (for example testing for uncorrelatedness but assuming independence), even under the null Q_T will have a $\mathbf{Z}'_L \Sigma_M \mathbf{Z}_L$ (nonstandard chi-square distribution with L -df).
- We estimate the finite sample distribution of Q_T under the null using the orthogonal sample.

- We recall that the orthogonal sample associated with $A_T(\phi_j)$ is

$$\mathbf{A}_T(\phi_j; r) = \{\sqrt{2}\Re A_T(\phi_j; r), \sqrt{2}\Im A_T(\phi_j; r); r = 1, \dots, M\},$$

under the null ($A(\phi_j) = 0$) they have approximately the same sampling properties.

- Therefore we can easily construct the orthogonal sample associated $Q_T = T \sum_{j=1}^L |A_T(\phi_j)|^2$ by simply replacing $A_T(\phi_j)$ with $\mathbf{A}_T(\phi_j; r)$:

$$Q_{T,R}(r) = T \sum_{j=1}^L |\sqrt{2}\Re A_T(e^{ij\cdot}; r)|^2, \quad Q_{T,I}(r) = T \sum_{j=1}^L |\sqrt{2}\Im A_T(e^{ij\cdot}; r)|^2.$$

Estimating the distribution of Q_T

- The above result means that under the null hypothesis of no correlation (but not necessarily independence) Q_T and $\{Q_{T,R}(r), Q_{T,I}(r); 1 \leq r \leq M\}$ asymptotically have equivalent mean, variance and higher order cumulants, but are uncorrelated.
- Using this result we estimate the distribution function of Q_T with the empirical distribution

$$\widehat{F}_{M,T}(x) = \frac{1}{2M} \left(\sum_{r=1}^M [I(Q_{T,R}(r) \leq x) + I(Q_{T,I}(r) \leq x)] \right).$$

- We reject the null at the $\alpha\%$ -level if the test statistic is such that $1 - \widehat{F}_{M,T}(Q_T) < \alpha\%$.

Selecting the number in the orthogonal sample, M

- All the procedures described above require us to select the number of elements, M , in the orthogonal sample.
- In order to select M we need to understand how well $A_T(\phi; r)$ can model the variance of $A_T(\phi) = A_T(\phi; 0)$. It can be shown that $\text{var}[\sqrt{T}A_T(\phi; r)] = \nu(\omega_r) + O(T^{-1})$, where $\nu : [0, 2\pi] \rightarrow \mathbb{R}$ is a smooth function defined as

$$\begin{aligned}\nu(\omega_r) &= \frac{1}{2\pi} \int_0^{2\pi} f(\omega) f(\omega + \omega_r) \left(|\phi(\omega)|^2 + \phi(\omega) \overline{\phi(-\omega - \omega_r)} \right) d\omega + \\ &\quad \frac{1}{(2\pi)^2} \int_0^{2\pi} \int_0^{2\pi} \phi(\omega_1) \overline{\phi(\omega_2)} f_4(\omega_1, -\omega_1 - \omega_r, \omega_2) d\omega_1 d\omega_2.\end{aligned}$$

- Since $\text{var}[\sqrt{T}A_T(\phi)] = \nu(0) + O(T^{-1})$, the closer ω_r is to zero, the closer the variance of $A_T(\phi; r)$ and $A_T(\phi)$.

- Recalling that $\text{var}[\sqrt{T}A_T(\phi; r)] \approx \nu(\omega_r)$ and using the near uncorrelatedness of $\{|\sqrt{T}A_T(\phi; r)|^2; 1 \leq r \leq T/2\}$ implies that

$$\left(\frac{|\sqrt{T}A_T(\phi; r)|^2}{\nu(\omega_r)} - 1 \right) \quad 1 \leq r < T/2,$$

is an almost uncorrelated sequence with mean zero and variance one.

- Returning to $\nu(\omega_r)$, we recall that we estimate ν using the local average $\widehat{\nu}_M = \frac{1}{M} \sum_{k=1}^M |A_T(\phi; k)|^2$. Thus the analogous estimator of $\nu(\omega_r)$ is

$$\widehat{\nu}_M(\omega_r) = \frac{1}{M} \sum_{k=1}^M |\sqrt{T}A_T(\phi; (r+k))|^2.$$

Just as in spectral density estimation, if M is ‘too small’ $\text{var}[\widehat{\nu}_M(\omega_r)]$ will be large, if M is ‘too large’ the bias $[\widehat{\nu}_M(\omega_r)]$ will be large.

Defining the cross validation criterion

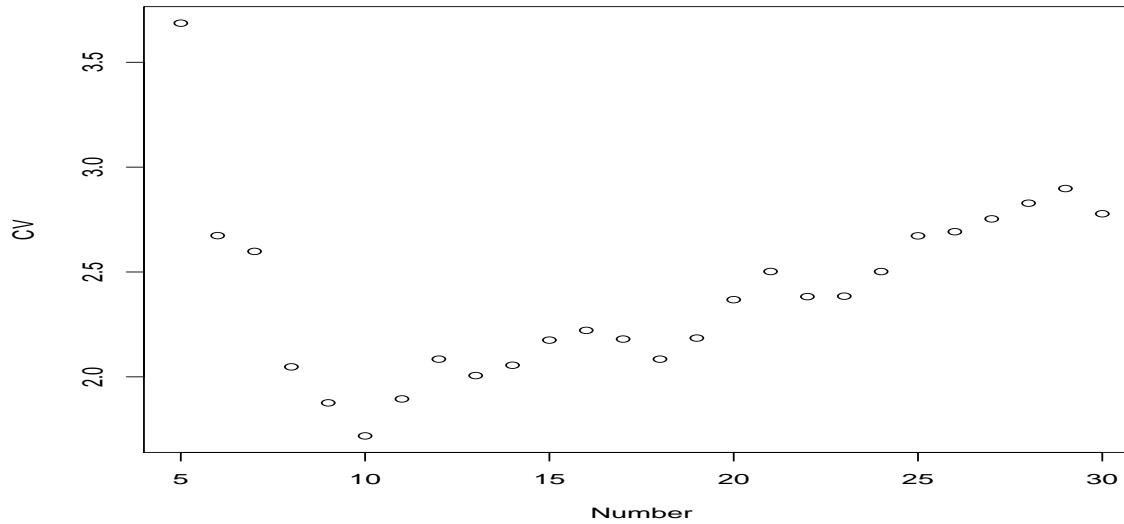
- We use this trade-off to define the cross-validation criterion

$$\mathcal{C}_T(M) = \sum_{r=1}^{T/P} \left(\frac{|\sqrt{T}A_T(\phi; r)|^2}{\widehat{\nu}_M(\omega_r)} - 1 \right)^2$$

Note we don't define the CV-criterion over all $T/2$ but only frequencies close to zero, as our objective is to estimate $\nu(\omega)$ at zero.

- Let \mathcal{S} is the set of candidate M (in the simulations we use $\mathcal{S} = \{9, \dots, 30\}$) and $\widehat{M} = \arg \min_{M \in \mathcal{S}} \mathcal{C}_T(M)$.
- Based on this we use $\{A_T(\phi; r); 1 \leq r \leq \widehat{M}\}$ in the orthogonal sample.

Example



- Estimate the autocovariance of the AR(2) model $X_t = 1.5X_{t-1} - 0.75X_{t-2} + \varepsilon_t$, where $T = 200$ using $A_T(e^{ij\cdot})$.

The above gives the cross-validation criterion for the number of orthogonal samples required to estimate the variance of the sample covariance at lag one.

Simulations: Testing for uncorrelatedness

- We use as the test statistic the sample covariance at the first five lags:

$$Q_T = T \sum_{j=1}^5 |A_T(e^{ij\cdot})|^2$$

- To test for statistical significance Q_T we use
 1. The proposed orthogonal sampling method to obtain the critical value at the 5% and 10% level. The number of elements in the orthogonal sample were evaluated using the cross-validation procedure.
 2. The classical Box-Pierce test using a chi-squared distribution with 5df.

3. The robust Box-Pierce test, where the variance of Q_T is obtained under the assumption that the time series are martingale differences. The assumption of martingale differences (ie. $\mathbf{E}[X_t | X_{t-1}, \dots, X_1] = 0$) is weaker than iid, but stronger than uncorrelated.
4. Block bootstrap: The time series is blocked into blocks of length B . These blocks are drawn at random with replacement to produce a bootstrap time series of the same length. The sample autocovariance is evaluated using the bootstrap time series $\{A_T^*(e^{ij\cdot})\}_{j=1}^L$ and the bootstrap test statistic constructed

$$Q_T^* = T \sum_{j=1}^5 |A_T^*(e^{ij\cdot}) - A_T(e^{ij\cdot})|^2.$$

This is done 1000 times. The empirical distribution of $\{Q_T^*\}$ is used as an estimator of the distribution of Q_T under the null.
 For a given data set it is not clear what is the optimal block size, so we use the block sizes $B = 5, 10$ and 20 .

Under the null

Model	Proxy Q_T		Regular \tilde{Q}_T		Robust Q_T^*		Block Bootstrap					
	5%	10%	5%	10%	5%	10%	B=5		B=10		B=20	
							5%	10%	5%	10%	5%	10%
Normal	6.52	11.1	4.1	8.34	5.42	10.18	0.0	0.1	1.14	4.36	5.14	11.94
t_5	6.34	11.42	4.08	8.46	4.92	10.5	0.0	0.08	0.96	4.12	4.74	10.70
$X_{3,t} = Z_t Z_{t-1}$	5.02	9.44	10.66	16.64	4.38	9.52	0.14	0.66	1.2	4.68	4.14	11.14
$X_{4,t}$	0.86	1.82	3.5	4.86	1.1	2.32	0.06	0.3	0.22	1.0	0.64	2.26
$X_{5,t}$	4.26	8.14	23.56	31.6	5.4	9.98	0.14	1.06	1.22	4.96	3.54	11.00
$X_{6,t}$	3.16	6.42	17.64	24.22	4.20	8.38	0.08	0.4	0.6	2.92	2.08	7.78
$X_{7,t}$	5.1	10.46	13.22	20.46	6.88	12.8	0.16	0.86	1.18	5	4.56	11.88
$X_{8,t}$	4.46	8.36	8.2	13.18	3.5	8.38	0.04	0.58	0.74	4	3.1	10.48

Table 1: Test for uncorrelatedness, under the null hypothesis, $T = 100$ over 5000 replications.

Under the null

Model	Proxy Q_T		Regular \tilde{Q}_T		Robust Q_T^*		Block Bootstrap					
	5%	10%	5%	10%	5%	10%	B=5		B=10		B=20	
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
Normal	5.9	11.1	4.56	9.44	4.74	9.86	0.08	0.44	1.48	4.22	3.7	8.78
t_5	6.1	10.82	4.8	9.58	4.92	9.98	0.04	0.34	1.24	4.18	3.3	8.66
$X_{3,t} = Z_t Z_{t-1}$	5.00	9.82	15.26	22.26	4.9	9.16	0.7	2.52	2.5	6.56	3.6	9.16
$X_{4,t}$	1.0	1.86	30.52	38.56	5.6	11.16	1.56	4.48	2.72	7.94	3.1	10.02
$X_{5,t}$	3.76	7.06	49.86	58.76	4.82	9.34	1.06	4.10	2.08	6.9	2.52	7.96
$X_{6,t}$	2.88	6.22	42.48	50.46	3.64	7.24	0.64	2.4	1.24	4.68	1.5	6.62
$X_{7,t}$	4.48	8.88	20.38	28.32	6.02	11.46	0.78	2.58	1.52	5.04	2.78	7.5
$X_{8,t}$	5.28	9.46	15.08	20.46	4.12	8.54	1.28	3.82	3.04	7.36	4.34	9.92

Table 2: Test for uncorrelatedness, under the null hypothesis, $T = 500$ over 5000 replications.

Under the alternative

Model	Proxy Q_T		Regular \tilde{Q}_T		Robust Q_T^*		Block Bootstrap					
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
$Y_{1,t}$	27.06	38.88	29.52	40.84	28.60	40.50	2.62	9.4	14.12	29.34	25.74	43.22
$Y_{2,t}$	12.68	20.58	21.78	30.6	11.08	18.68	0.5	3.3	3.94	13.84	9.96	24.02
$Y_{3,t}$	55.7	68.6	71.94	79.58	61.8	72.32	18.36	41.46	41.62	65.14	52.86	74.30

Table 3: Test for uncorrelatedness, under the alternative hypothesis, $T = 100$ over 5000 replications.

Model	Proxy Q_T		Regular \tilde{Q}_T		Robust Q_T^*		Block Bootstrap					
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
$Y_{1,2}$	94.86	97.44	95.94	97.84	95.24	97.60	82.30	91.74	92.96	96.48	93.76	97.28
$Y_{2,t}$	49.50	60.58	69.60	77.24	37	49.90	29.92	45.30	41.48	58.28	42.2	60.98
$Y_{3,t}$	98.86	99.36	99.94	99.98	98.5	99.16	99.2	99.68	98.98	99.82	97.94	99.82

Table 4: Test for uncorrelatedness under the alternative hypothesis, $T = 500$ over 5000 replications.

Goodness of fit tests

- Let us suppose f is the spectral density of the observed time series, our objective is to test $H_0 : f(\omega) = g(\omega; \theta)$ for all $\omega \in [0, 2\pi]$ against $H_A : f(\omega) \neq g(\omega; \theta)$ for some $\omega \in [0, 2\pi]$. The test statistics

$$G_T = T \sum_{j=1}^L |A_T(e^{ij\cdot} g(\cdot; \theta)^{-1})|^2.$$

where $\frac{1}{T} \sum_{k=1}^T \frac{\exp(i j \omega_k)}{g(\omega_k; \theta)} |J_T(\omega_k)|^2$.

- To obtain the critical values of the test statistic we use
 - Orthogonal samples
 - Block bootstrap with $B = 5, 10, 20, 30$ and 40 .

Under the null

Model	Proxy Q_T	Block Bootstrap									
		B=5		B=10		B=20		B=30		B=40	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
$X_{0.6,t}^G$	2.32	4.94	0.00	0.00	0.18	0.96	2.18	7.06	4.66	11.24	8.64
$X_{0.6,t}^\chi$	2.24	4.4	0	0.02	0.04	0.26	0.94	3.48	2.24	6.82	4.94
$X_{0.9,t}^\chi$	0.96	1.74	0	0	0	0.02	0.04	0.26	0.2	0.96	0.56

Table 5: Goodness of fit test, under the null, $T=100$ over 5000 replications

Model	Proxy Q_T	Block Bootstrap									
		B=5		B=10		B=20		B=30		B=40	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
$X_{0.6,t}^G$	5.24	10.28	0	0	0.08	0.52	1.54	3.92	2.48	6	3.32
$X_{0.6,t}^\chi$	5.2	9.78	0	0	0.02	0.16	0.52	1.84	0.04	0.14	2.12
$X_{0.9,t}^\chi$	2.24	4.84	0	0	0	0	0	0	0.04	0.62	0.04

Table 6: Goodness of fit test, under the null, $T=500$ over 5000 replications

Under the alternative, $T = 100$ and 500

Model	Null	Proxy Q_T	Block Bootstrap										
			B=5		B=10		B=20		B=30		B=40		
			5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	
$X_{0.6,t}^G$	$\sigma = 1, \phi = 0.3$	60.4	71.24	1.98	11.32	35.42	55.12	53.24	70.78	61.2	76.5	68.62	80.34
$X_{0.6,t}^X$	$\sigma = \sqrt{2}, \phi = 0.3$	64.32	76.12	1.3	8.68	33.2	56.12	54.24	74.58	62.4	80.42	71.06	84.14
$X_{0.6,t}^G$	$\sigma = 1, \phi = 0.45$	15.1	22.28	0	0	1.94	7.24	10.94	22.06	15.9	28.62	23.14	36.36
$X_{0.6,t}^X$	$\sigma = \sqrt{2}, \phi = 0.45$	13.98	22.38	0	0	0.82	3.96	6.42	16.78	10.72	23.6	17.9	32.74
$X_{0.9,t}^X$	$\sigma = \sqrt{2}, \phi = 0.7$	41.14	51.82	0	0	3.2	10.28	17.54	36	25.54	45.58	36.24	56.44

Model	Null	Proxy Q_T	Block Bootstrap										
			B=5		B=10		B=20		B=30		B=40		
			5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	
$X_{0.6,t}^G$	$\sigma = 1, \phi = 0.3$	99.98	100	98.88	99.72	100	100	100	100	100	100	100	100
$X_{0.6,t}^X$	$\sigma = \sqrt{2}, \phi = 0.3$	99.98	100	99.2	99.92	100	100	100	100	99.92	100	99.92	100
$X_{0.6,t}^G$	$\sigma = 1, \phi = 0.45$	86.90	92.52	0.18	1.54	48.88	65.62	77.36	86.7	81	89.93	82.58	90.72
$X_{0.6,t}^X$	$\sigma = \sqrt{2}, \phi = 0.45$	89.92	94.22	0.06	0.56	42.16	61.88	75.60	87.44	81.94	90.8	84.24	92.26
$X_{0.9,t}^G$	$\sigma = \sqrt{2}, \phi = 0.7$	99.96	99.98	0	0	94.46	98.12	99.86	99.96	99.9	99.98	99.92	99.98

Concluding remarks

- The orthogonal sample associated with an estimator is a computationally very fast method for estimating the uncertainty in the estimator and also estimating the uncertainty in the uncertainty estimator.
- With Soutir Bandyopadhyay and Carsten Jentsch we have used this method to test for stationarity for stochastic processes.
- In several cases the finite sample distribution of many estimators, $A_T(\phi)$ tend to be skewed, in such cases it may be advisable to use the power-transformation method (described in Chen and Deo (2002) to normalise the estimator).