MURI Meeting

Phyllis Wan

Columbia University

April 15, 2016

Testing the goodness-of-fit of AR(p) model

$$X_t = \sum_{k=1}^{p} \phi_k X_{t-k} + Z_t$$

- $Z_t \overset{iid}{\sim} F$
- Causal
  - $1 - \sum_{k=1}^{p} \phi_k z^k = 0$ no roots inside the unit circle

Testing the goodness-of-fit of AR(p) model

$$X_t = \sum_{k=1}^{p} \phi_k X_{t-k} + Z_t$$

- Observations $(X_t)_{t=1,\ldots,n}$

Testing the goodness-of-fit of AR(p) model

$$X_t = \sum_{k=1}^{p} \hat{\phi}_k X_{t-k} + \hat{Z}_t$$

- Observations $(X_t)_{t=1,\ldots,n}$
- Parameter estimates $\hat{\phi}_k$ (e.g., least-square, Yule-Walker, etc.)
- Fitted residuals $(\hat{Z}_t)$

Proposal: Test the serial dependence of $(\hat{Z}_t)$

Measure of Dependence: Generalized Distance Covariance

$$T^2(X, Y; \mu) := \int |\varphi_{X,Y}(s,t) - \varphi_X(s)\varphi_Y(t)|^2 \, \mu(ds, dt)$$

- Random variables $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$,

$$X \perp\!\!\!\perp Y \iff \varphi_{X,Y}(s,t) = \varphi_X(s)\varphi_Y(t), \;\; \forall s, t$$

Measure of Dependence: Generalized Distance Covariance

$$T^2(X, Y; \mu) := \int |\varphi_{X,Y}(s,t) - \varphi_X(s)\varphi_Y(t)|^2 \, \mu(ds, dt)$$

- Random variables $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$,

$$X \perp\!\!\!\perp Y \iff \varphi_{X,Y}(s,t) = \varphi_X(s)\varphi_Y(t), \ \ \forall s, t$$

- Existence:
  - μ finite
  - μ infinite: $\int (1 \wedge |s|^\alpha)\,(1 \wedge |t|^\alpha)\,\mu(ds, dt) < \infty$ and $\mathbb{E}[|X|^\alpha + |Y|^\alpha + |XY|^\alpha] < \infty$

Measure of Dependence: Generalized Distance Covariance

$$T^2(X, Y; \mu) := \int \left|\varphi_{X,Y}(s,t) - \varphi_X(s)\varphi_Y(t)\right|^2 \mu(ds, dt)$$

- Random variables $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$,

$$X \perp\!\!\!\perp Y \iff \varphi_{X,Y}(s,t) = \varphi_X(s)\varphi_Y(t), \;\; \forall s, t$$

- Existence:
  - μ finite
  - μ infinite: $\int (1 \wedge |s|^\alpha)\,(1 \wedge |t|^\alpha)\,\mu(ds, dt) < \infty$ and $\mathbb{E}[|X|^\alpha + |Y|^\alpha + |XY|^\alpha] < \infty$
- Traditional distance covariance (Székely et al., 07):

$$\mu(ds, dt) = c|s|^{-2}|t|^{-2}$$

Measure of Dependence: Generalized Distance Covariance

$$T^2(X, Y; \mu) := \int |\varphi_{X,Y}(s,t) - \varphi_X(s)\varphi_Y(t)|^2 \, \mu(ds, dt)$$

- Random variables $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$,

$$X \perp\!\!\!\perp Y \iff \varphi_{X,Y}(s,t) = \varphi_X(s)\varphi_Y(t), \ \ \forall s, t$$

- Existence:
  - μ finite
  - μ infinite: $\int (1 \wedge |s|^\alpha)\,(1 \wedge |t|^\alpha)\,\mu(ds, dt) < \infty$ and $\mathbb{E}[|X|^\alpha + |Y|^\alpha + |XY|^\alpha] < \infty$
- Traditional distance covariance (Székely et al., 07):

$$\mu(ds, dt) = c|s|^{-2}|t|^{-2}$$

- Distance correlation:

$$R^2(X, Y; \mu) = \frac{T^2(X, Y; \mu)}{\sqrt{T^2(X, X; \mu)\, T^2(Y, Y; \mu)}}$$

Measure of Dependence: Generalized Distance Covariance

$$T^2(X,Y;\mu) := \int |\varphi_{X,Y}(s,t) - \varphi_X(s)\varphi_Y(t)|^2 \, \mu(ds,dt)$$

- Assume $\mu(ds,dt) = \mu_1(ds) \times \mu_2(dt)$ symmetric about the origin
- Let $\tilde{\nu}(s) = \begin{cases} \int_{\mathbb{R}^d} e^{i\langle s,x\rangle} \, \nu(dx) & , \quad \nu \text{ finite} \\ \int_{\mathbb{R}^d} (1-\cos\langle s,x\rangle) \, \nu(dx) & , \quad \nu \text{ infinite} \end{cases}$

$$\begin{aligned} T(X,Y;\mu) \quad = \quad & \mathbb{E}[\tilde{\mu}_1(X-X')\,\tilde{\mu}_2(Y-Y')] \\ & + \mathbb{E}[\tilde{\mu}_1(X-X')]\,\mathbb{E}[\tilde{\mu}_2(Y-Y')] \\ & - 2\,\mathbb{E}[\tilde{\mu}_1(X-X')\tilde{\mu}_2(Y-Y'')]\,. \end{aligned}$$

  - where X', Y', Y'' are iid copies of X, Y, Y respectively

Measure of Dependence: Generalized Distance Covariance

$$T_n^2(\mathbf{X},\mathbf{Y};\mu) := \int \left|\varphi^n_{\mathbf{X},\mathbf{Y}}(s,t) - \varphi^n_{\mathbf{X}}(s)\varphi^n_{\mathbf{Y}}(t)\right|^2 \mu(ds,dt)$$

- Assume $\mu(ds,dt) = \mu_1(ds)\times\mu_2(dt)$ symmetric about the origin
- Let $\tilde{\nu}(s) = \begin{cases} \int_{\mathbb{R}^d} e^{i\langle s,x\rangle}\,\nu(dx) & , \quad \nu \text{ finite} \\ \int_{\mathbb{R}^d} (1-\cos\langle s,x\rangle)\,\nu(dx) & , \quad \nu \text{ infinite} \end{cases}$

$$\begin{aligned} T_n(X,Y;\mu) &= \frac{1}{n^2}\sum_{i,j=1}^n \tilde{\mu}_1(X_i - X_j)\,\tilde{\mu}_2(Y_i - Y_j) \\ &\quad + \frac{1}{n^4}\sum_{i,j,k,l=1}^n \tilde{\mu}_1(X_i - X_j)\tilde{\mu}_2(Y_k - Y_l) \\ &\quad - \frac{2}{n^3}\sum_{i,j,k=1}^n \tilde{\mu}_1(X_i - X_j)\tilde{\mu}_2(Y_i - Y_k). \end{aligned}$$

Measure of Dependence: Generalized Distance Covariance

$$T_n^2(\mathbf{X}, \mathbf{Y}; \mu) := \int \left|\varphi^n_{\mathbf{X},\mathbf{Y}}(s,t) - \varphi^n_{\mathbf{X}}(s)\varphi^n_{\mathbf{Y}}(t)\right|^2 \mu(ds, dt)$$

- Consistency under ergodicity
- Asymptoticity under certain α-mixing condition

Auto-Distance Covariance Function

Let $\mathbf{Z}_1 = (Z_1, \ldots, Z_{n-h})$, $\mathbf{Z}_{h+1} = (Z_{h+1}, \ldots, Z_n)$, then

$$T_n^2(h) := T_n^2(\mathbf{Z}_1, \mathbf{Z}_{h+1}; \mu)$$

If $Z_t \overset{iid}{\sim} F$, then

$$n\, T_n^2(h) \xrightarrow{d} \int |G_F(s,t)|^2 \, \mu(ds, dt)$$

- G_F is a Gaussian field dependent on F

Example: Kilkenny wind speed time series

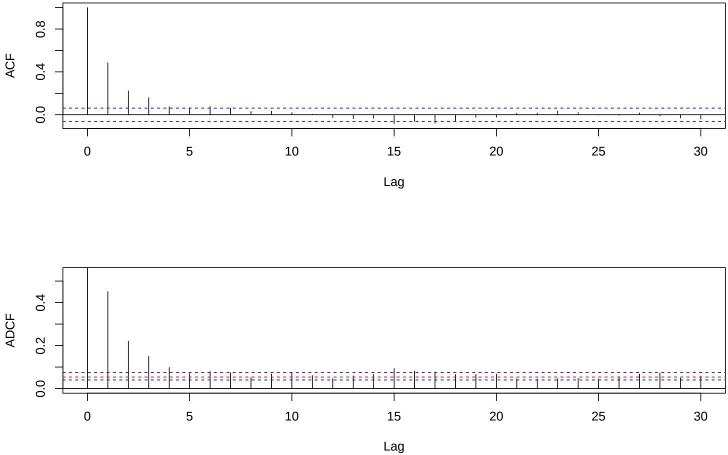

Figure : ACF and auto-distance correlation function (ADCF) of Kilkenny daily wind speed time series from 1/1/61 - 9/27/63

Testing the goodness-of-fit of AR(p) model

$$X_t = \sum_{k=1}^{p} \hat{\phi}_k X_{t-k} + \hat{Z}_t$$

- Observations $(X_t)_{t=1,\ldots,n}$
- Parameter estimates $\hat{\phi}_k$
- Fitted residuals $(\hat{Z}_t)$

Proposal: Test the serial dependence of $(\hat{Z}_t)$

Statistic of interest: $\tilde{T}_n^2(h) := T_n^2(\hat{\mathbf{Z}}_1, \hat{\mathbf{Z}}_{h+1}; \mu)$

Auto-distance covariance of AR(p) residuals

$$X_t = \sum_{k=1}^{p} \hat{\phi}_k X_{t-k} + \hat{Z}_t$$

Statistic of interest: $\tilde{T}_n^2(h) := T_n^2(\hat{\mathbf{Z}}_1, \hat{\mathbf{Z}}_{h+1}; \mu)$

Theorem

- *Assume that*

$$\int (s^2 \wedge t^2 \wedge (st)^2)\, \mu(ds, dt) < \infty$$

- *If* $\mathbb{E}Z^2 < \infty$, *then*

$$n\, \tilde{T}_n^2(h) \xrightarrow{d} \int \left| G_F(s,t) + \xi_h(s,t) \right|^2 \mu(ds, dt)$$

where

$$\xi_h(s,t) = t\varphi_Z(t)\, \varphi_Z'(s) \Psi_h^T \mathbf{Q}$$

  - $\Psi_h = (\psi_{h-j})_{j=1,\ldots,p}$ *where* ψ_j *are the coefficients in the causal representation* $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$
  - $\mathbf{Q}$ *is the limit distribution of* $\sqrt{n}(\hat{\phi} - \phi)$

Example: Auto-distance covariance of AR(10) residuals: $Z_t \sim N(0,1)$

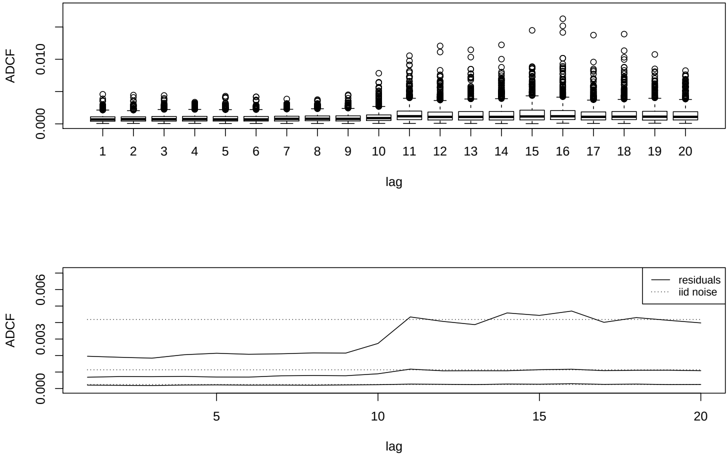

Figure : Left panel: empirical box plots of $\tilde{T}_n^2$; Right panel: empirical 5%, 50%, 95% quantiles of $\tilde{T}_n^2$ and T_n^2

Example: Auto-distance covariance of AR(10) residuals: $Z_t \sim N(0,1)$

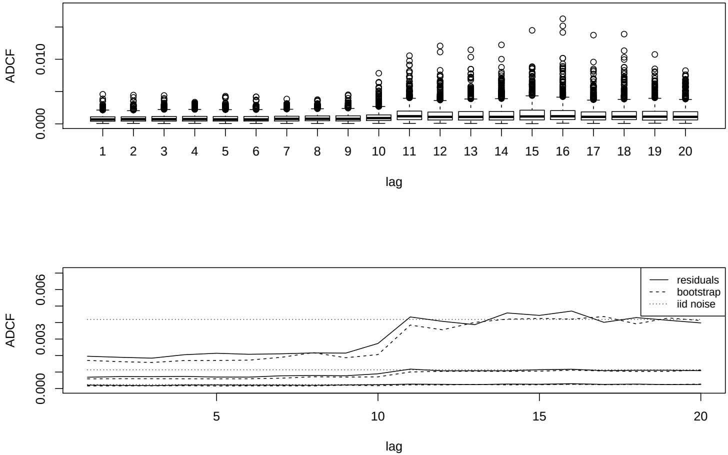

Figure : Left panel: empirical box plots of $\tilde{T}_n^2$; Right panel: empirical 5%, 50%, 95% quantiles of $\tilde{T}_n^2$, from simulations and from bootstrapping, and that of T_n^2

Example: Auto-distance covariance of AR(10) residuals

$$X_t = \sum_{k=1}^{p} \hat{\phi}_k X_{t-k} + \hat{Z}_t$$

Statistic of interest: $\tilde{T}_n^2(h) := T_n^2(\hat{\mathbf{Z}}_1, \hat{\mathbf{Z}}_{h+1}; \mu)$

Theorem

- *Assume that*

$$\int (s^2 \wedge t^2 \wedge (st)^2)\, \mu(ds, dt) < \infty$$

- *If Z is in the domain of attraction of a stable distribution of index $\alpha \in (0, 2)$, then*

$$n\, \tilde{T}_n^2(h) \xrightarrow{d} \int |G_F(s, t)|^2\, \mu(ds, dt).$$

Example: Auto-distance covariance of AR(10) residuals: $Z_t \sim t(1.5)$

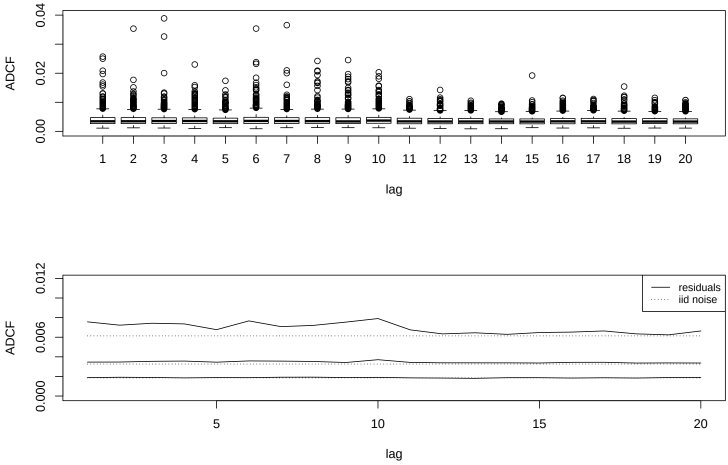

Figure : Left panel: empirical box plots of $\tilde{T}_n^2$; Right panel: empirical 5%, 50%, 95% quantiles of $\tilde{T}_n^2$ and T_n^2

Example: Kilkenny wind speed time series

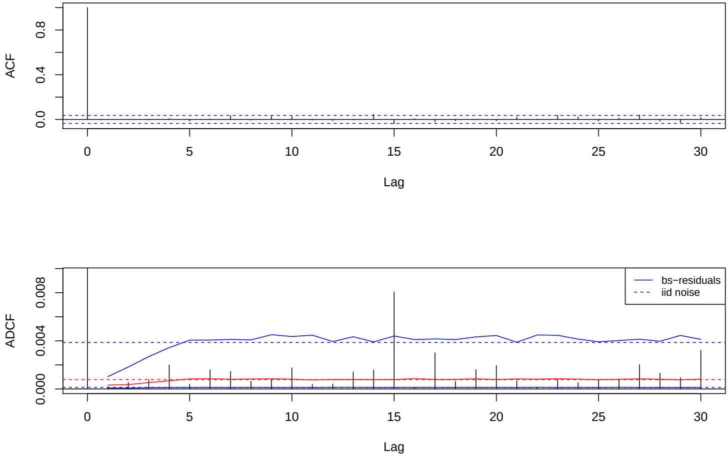

Figure : ACF and auto-distance correlation function (ADCF) of Kilkenny daily wind speed time series after AR(3) fitting.