Sample Splitting for Assessing Goodness-of-Fit in Time Series

Richard A. Davis
Columbia University

Leon Fernandes
Columbia University

ASA Georgia Winter Lecture

December 7, 2023

The Story—illustrated with AR(10)

AR(10) model:

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_{10} X_{t-10} + Z_t, \qquad \{Z_t\} \sim IID(0, \sigma^2)$$

$$X_t = \boldsymbol{\phi}^\top \boldsymbol{X}_{t-1} + Z_t,$$

where

$$\boldsymbol{\phi} = (\phi_1, \dots, \phi_{10})^\top \quad \textit{and} \quad \boldsymbol{X}_{t-1} = (X_{t-1}, \dots, X_{t-10})^\top$$

Estimation: Based on data, $X_1, \dots, X_n$, estimate $\boldsymbol{\phi}$ using least squares, i.e.,

$$\widehat{\boldsymbol{\phi}} = \hat{\Gamma}_{10}^{-1} n^{-1} \sum_{j=11}^{n} X_t \boldsymbol{X}_{t-1},$$

where $\hat{\Gamma}_{10} = n^{-1} \sum_{j=11}^{n} \boldsymbol{X}_{t-1} \boldsymbol{X}_{t-1}^\top$ is the sample covariance matrix.

Now,

$$\sqrt{n}\left(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0\right) \xrightarrow{d} Q \sim N(\mathbf{0}, \sigma^2 \Gamma_{10}^{-1})$$

The Story—illustrated with AR(10)

AR(10) model: $X_t = \boldsymbol{\phi}_0^\top \boldsymbol{X}_{t-1} + Z_t, \quad \{Z_t\} \sim IID(0, \sigma^2)$

Define

$$Z_t(\boldsymbol{\phi}) = X_t - \boldsymbol{\phi}^\top \boldsymbol{X}_{t-1}; \quad Z_t(\boldsymbol{\phi}_0) = X_t - \boldsymbol{\phi}_0^\top \boldsymbol{X}_{t-1} = Z_t.$$

Residuals:

$$\widehat{Z}_t = Z_t(\widehat{\boldsymbol{\phi}}) = X_t - \widehat{\boldsymbol{\phi}}^\top \boldsymbol{X}_{t-1} = Z_t - (\widehat{\boldsymbol{\phi}}^\top - \boldsymbol{\phi}_0^\top) \boldsymbol{X}_{t-1}$$

Remarks:

- $\widehat{Z}_t$'s are not independent (not even uncorrelated)
- $\widehat{Z}_t = Z_t + O_P(\frac{1}{\sqrt{n}}) \boldsymbol{X}_{t-1}$

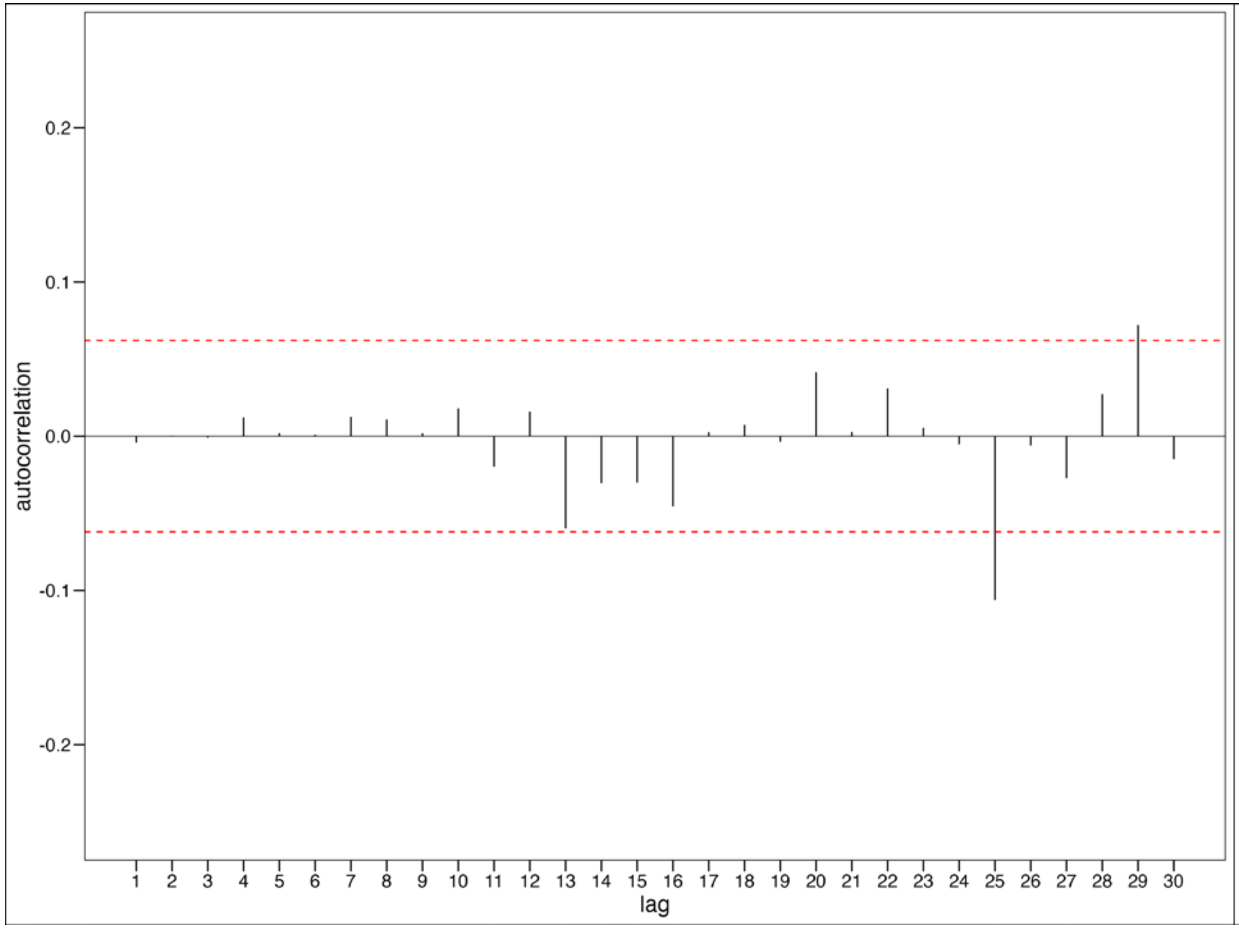
Goodness of fit test: compute ACF of $\widehat{Z}_t$;

$$\hat{\rho}_{\hat{Z}}(h) = \frac{\sum_{t=11}^{n-h} \hat{Z}_t \hat{Z}_{t+h}}{\sum_{t=11}^{n} \hat{Z}_t^2}$$

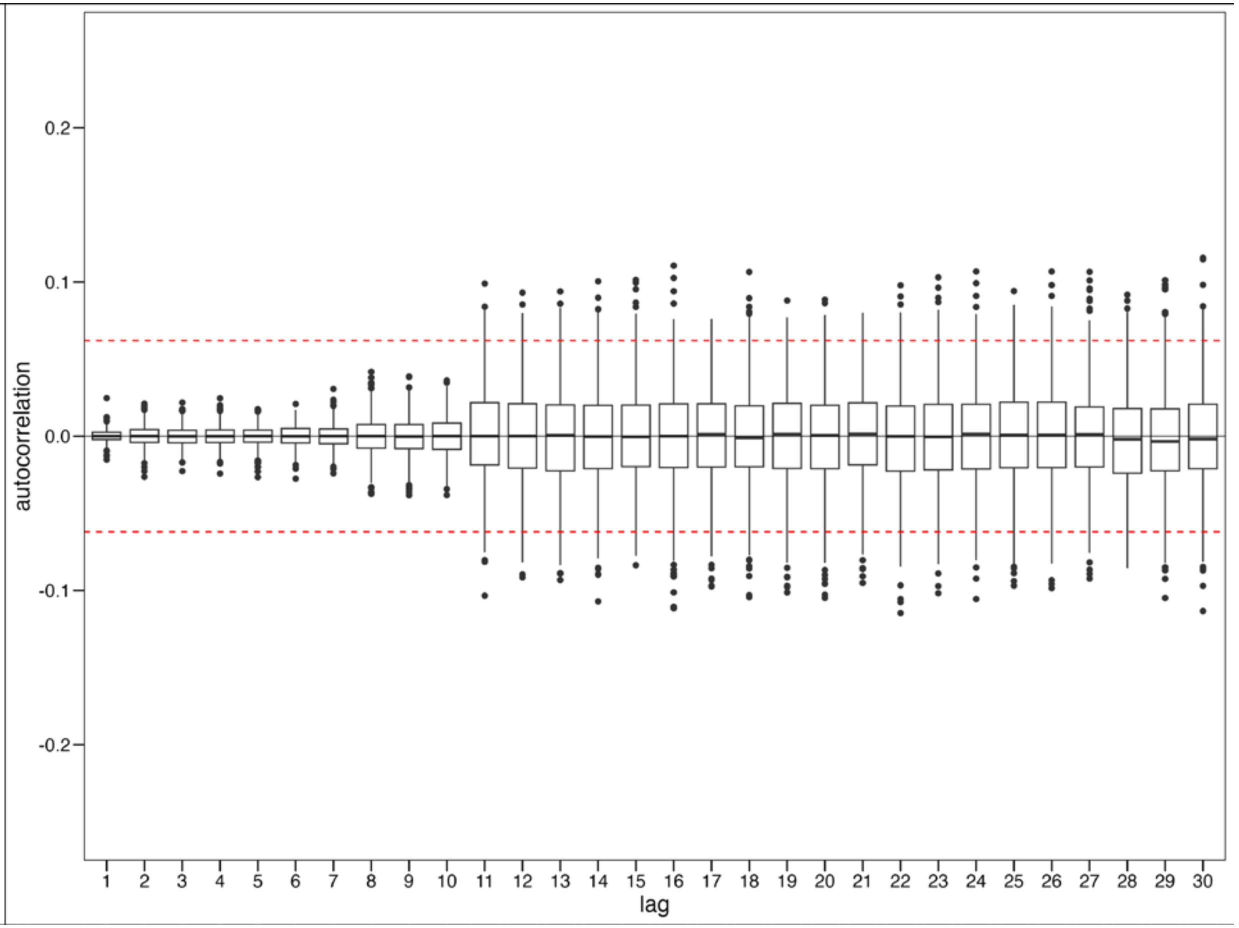
and compare with $\pm \frac{1.96}{\sqrt{n}}$.

The Story (ACF of $\hat{Z}_t$)

$\hat{\rho}_{\hat{Z}}(h)$: one realization; $\hat{Z}_t = \hat{Z}_t(\hat{\boldsymbol{\phi}})$



$\hat{\rho}_{\hat{Z}}(h)$: Boxplots 1000 reps



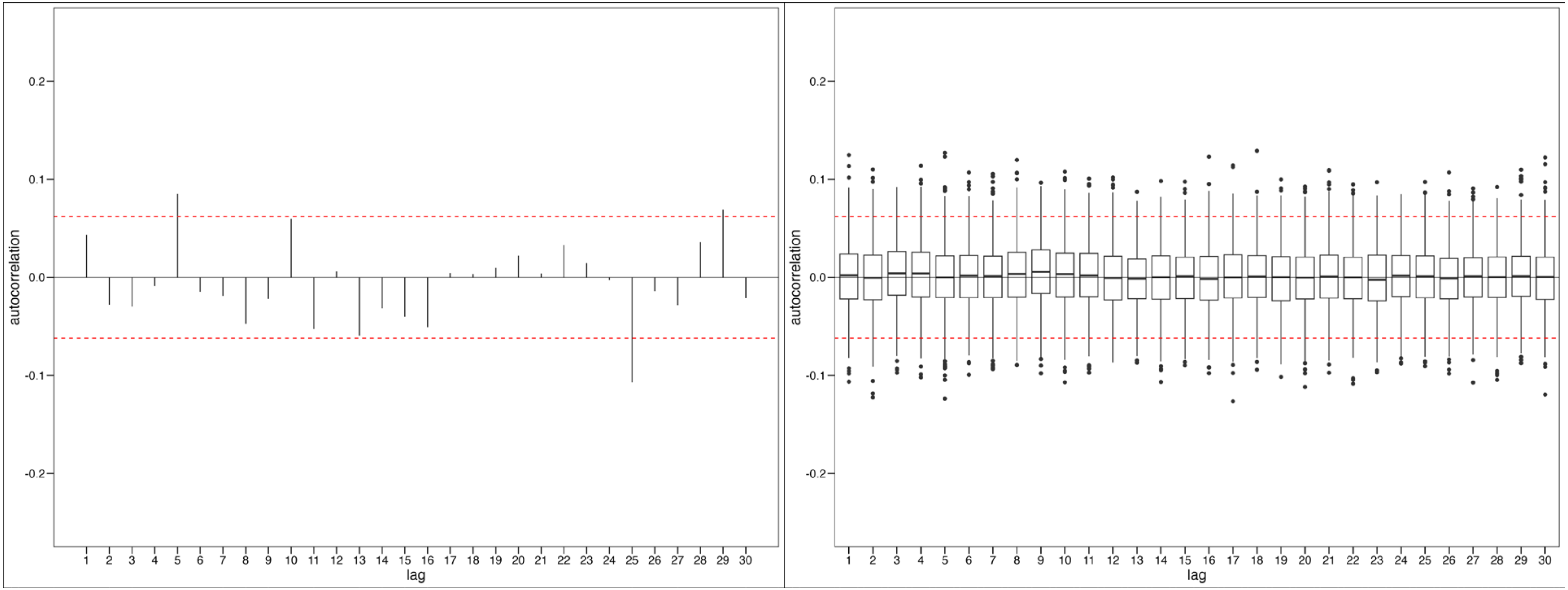
Box-Pierce (1970) gives limit behavior of $\hat{\rho}_{\hat{Z}}(h)$, which depends on the AR coefficients.

The Story (ACF of $\hat{Z}_t(\hat{\boldsymbol{\phi}}_{f_n})$)

Sample splitting: Estimate $\boldsymbol{\phi}$ using the first $f_n = \frac{n}{2}$ of the observations. Call estimate $\hat{\boldsymbol{\phi}}_{f_n}$ and compute $\hat{Z}_t(\hat{\boldsymbol{\phi}}_{f_n})$

$\hat{\rho}_{\hat{Z}}(h)$: one realization; $\hat{Z}_t = \hat{Z}_t(\hat{\boldsymbol{\phi}}_{f_n})$

$\hat{\rho}_{\hat{Z}}(h)$: Boxplots 1000 reps



Sample-splitting: deeper dive with AR(1)

Motivation: Comes from Pfister, Bühlmann, Schölkopf, and Peters (2018) who used sample splitting for examining residuals using permutation procedures from a *regression model.*

Goal: Extend PBSP's idea to time series models (independence of data was assumed in their paper).

Unfortunately, this idea does not work in time series!

Deeper dive with AR(1): $X_t = \phi X_{t-1} + Z_t, \ \{Z_t\} \sim IID(0, \sigma^2)$.

Sample-splitting: Split data $X_1, \ldots, X_n$ into two possibly overlapping sets

$\{X_1, \ldots, X_{f_n}\}$ -- first f_n set of obs: used to estimate $\phi \Rightarrow \hat{\phi}$.

$\{X_{n-l_n+1}, \ldots, X_n\}$ -- last l_n set of obs: used to compute residuals $\hat{Z}_t = Z_t - (\hat{\phi} - \phi) X_{t-1}, t = n - l_n + 1, \ldots, n$

Since

$$\hat{\gamma}_{l_n}^{\hat{Z}}(0) = \frac{1}{l_n} \sum_{j=n-l_n+1}^{n} \hat{Z}_j^2 \xrightarrow{P} \sigma^2,$$

we focus on the sample covariances.

Sample-splitting: deeper dive with AR(1)

$\{X_1, \ldots, X_{f_n}\}$ -- first f_n set of obs: used to estimate $\phi \Rightarrow \hat{\phi}$.

$\{X_{n-l_n+1}, \ldots, X_n\}$ -- last l_n set of obs: used to compute residuals $\hat{Z}_t(\hat{\phi}) = Z_t - (\hat{\phi} - \phi)X_{t-1}, t = n - l_n + 1, \ldots, n$

Sample autocovariance:

$$\sqrt{l_n}\,\hat{\gamma}_{l_n}^{\hat{Z}}(h) = \sqrt{l_n}\sum_{t=n-l_n+1}^{n} \hat{Z}_t\hat{Z}_{t+h}$$

$$= \sqrt{l_n}\,\hat{\gamma}_{l_n}^{Z}(h) - \phi^{h-1}(1-\phi^2)\frac{\sqrt{l_n}}{f_n}\sum_{j=2}^{f_n} X_{j-1}Z_j + o_P(1).$$

Joint CLT:

$$\left(\sqrt{l_n}\,\hat{\gamma}_{l_n}^{Z}(h), \frac{\sqrt{l_n}}{f_n}\sum_{j=2}^{f_n} X_{j-1}Z_j\right) \xrightarrow{d} N\left(\mathbf{0}, \sigma^4\begin{pmatrix} 1 & k_{ov}\phi^{h-1} \\ k_{ov}\phi^{h-1} & k_{ra}(1-\phi^2)^{-1}\end{pmatrix}\right),$$

where $k_{ra} = \lim_n l_n/f_n$ (limiting ratio) and $k_{ov} = \lim_n \max\{0, f_n + l_n - n\}/f_n$ (limiting overlap).

Theorem. $\sqrt{l_n}\,\hat{\rho}_{l_n}^{\hat{Z}}(h) \xrightarrow{d} N(0, \sigma_\gamma^2)$;

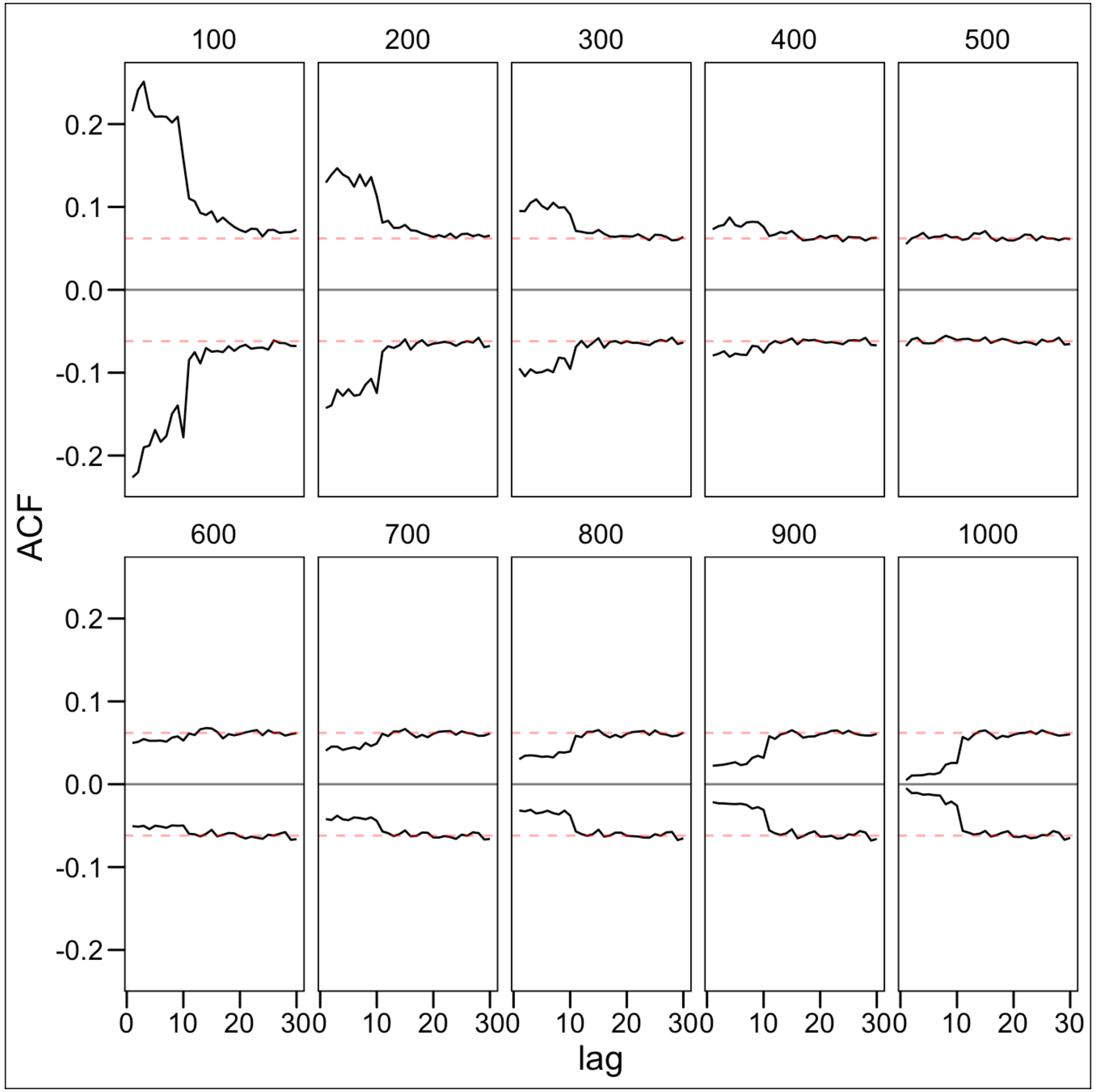
$$\sigma_\gamma^2 = \left(1 + (k_{ra} - 2k_{ov})\phi^{2(h-1)}(1-\phi^2)\right)$$

$$= 1 \quad if \quad k_{ra} - 2k_{ov} = 0 \quad e.g., l_n = n, f_n = n/2$$

$l_n = n, f_n = n \implies k_{ra} - 2k_{ov} = -1 \implies$ Box and Pierce (1970)

The Story (ACF of $\hat{Z}_t(\widehat{\boldsymbol{\phi}}_{f_n})$)

Sample splitting: First $f_n = k * 100, k = 1, \ldots, 10$ of the observations to estimate $\widehat{\boldsymbol{\phi}}$; $l_n = n$. Laplace noise.



$$k_{ra} = \lim_n l_n/f_n = \frac{10}{k}; \quad k_{ov} = 1$$

$$k_{ra} - 2k_{ov} \begin{cases} > 0, & k < 5 \\ = 0, & k = 5 \\ < 0, & k > 5 \end{cases}$$

$$\sigma_\gamma^2 \begin{cases} > 1, & k < 5 \\ = 1, & k = 5 \\ < 1, & k > 5 \end{cases}$$

More General Models

Same results hold for more general models: Assume $\{X_t\}$ follows the causal and invertible model

$$X_j = g\left(Z_{-\infty:j}; \boldsymbol{\beta}_0\right), \qquad \{Z_t\} \sim IID(0, \sigma^2)$$

$$Z_j = h\left(X_{-\infty:j}; \boldsymbol{\beta}_0\right),$$

Now set,

$$Z_j(\boldsymbol{\beta}) = h\left(X_{-\infty:j}; \boldsymbol{\beta}\right).$$

As before set

$$\hat{Z}_j\left(\widehat{\boldsymbol{\beta}}\right) = h\left(X_{-\infty:j}; \widehat{\boldsymbol{\beta}}\right), \qquad j = n - l_n + 1, \dots, n$$

where $\widehat{\boldsymbol{\beta}}$ is an estimate based on $X_1, \dots, X_{f_n}$.

Theorem (CLT). Under additional conditions on g and h and the form of the estimate $\widehat{\boldsymbol{\beta}}$, we have with

$f_n = \frac{n}{2}, l_n = n$

$$\sqrt{n}\, \hat{\rho}_n^{\hat{Z}}(h) \xrightarrow{d} N0,1)$$

and are asymptotically independent at different lags h.

Remark: Theorem covers ARMA and GARCH processes (see later for GARCH(1,1)).

Auto Distance Correlation Function (ADCF)

Distance covariance: The distance covariance between two random p −vectors, X and Y, is defined as

$$T(X,Y) = \int_{\mathbb{R}^p\times\mathbb{R}^p} |\varphi_{X,Y}(s,t) - \varphi_X(s)\varphi_Y(t)|^2\, \mu(ds,dt)$$

where $d\mu$ is some measure on $\mathbb{R}^p \times \mathbb{R}^p$.

Common choice for μ (Székely et al. (2007)):

$$d\mu(s,t) = \frac{dsdt}{c_p c_q |s|^{p+\alpha}|t|^{p+\alpha}}$$

In this case,

$$T(X,Y) = \mathbb{E}\left|X-\dot{X}\right|^{\alpha}\left|Y-\dot{Y}\right|^{\alpha} + \mathbb{E}\left|X-\dot{X}\right|^{\alpha}\mathbb{E}\left|Y-\dot{Y}\right|^{\alpha} - 2\mathbb{E}\left|X-\dot{X}\right|^{\alpha}\left|Y-\ddot{Y}\right|^{\alpha},$$

where $(\dot{X}, \dot{Y})$, $(\ddot{X}, \ddot{Y})$ are iid copies of (X,Y).

More generally if $\mu = \mu_1 \times \mu_2$ is a product measure with symmetric μ_1 and μ_2

$$T(X,Y) = \mathbb{E}\left[\hat{\mu}_1(X-\dot{X})\hat{\mu}_2(Y-\dot{Y})\right] + \mathbb{E}\left[\hat{\mu}_1(X-\dot{X})\right]\mathbb{E}\left[\hat{\mu}_1(Y-\dot{Y})\right] - 2\mathbb{E}\left[\hat{\mu}_1\left(X-\dot{X}\right)\hat{\mu}_2\left(Y-\ddot{Y}\right)\right],$$

where $\hat{\mu}_i$ is the Fourier transform of $\mu_i, i = 1,2$.

Auto Distance Correlation Function (ADCF)

$$T(X,Y\ ) = \int_{\mathbb{R}^p \times \mathbb{R}^p} |\varphi_{X,Y}(s,t) - \varphi_X(s)\varphi_Y(t)|^2\, \mu(ds,dt)$$

Estimation of $T(X,Y\ )$:

$$\hat{T}(Y,Z\ ) := \int_{\mathbb{R}^p \times \mathbb{R}^p} |\hat{\varphi}_{X,Y}(s,t) - \hat{\varphi}_X(s)\hat{\varphi}_Y(t)|^2\, d\mu(s)\ d\mu(t)$$

$\hat{\varphi}_{X,Y}(s,t) = \frac{1}{n}\sum_{j=1}^{n} \mathrm{e}^{\mathrm{i}sX_j + \mathrm{i}tY_j}$ is the empirical *joint* characteristic function.

ADCF: For observations $X_1, \ldots, X_n$ from a stationary time series the empirical ADCF at lag h is

$$\hat{R}_n^X(h) = \frac{\hat{T}_n^X(h)}{\hat{T}_n^X(0)} \quad \left(= \frac{\hat{T}(X_t, X_{t+h})}{\hat{T}(X_t, X_t)} \right)$$

Auto Distance Correlation Function (ADCF)

Davis, Matsui, Mikosch, & Wan (2018) show: under suitable conditions on μ (either finite measure or infinite with moment condition on the rvs),

$$n\hat{R}_n^Z(h) = n\frac{\hat{T}_n^Z(h)}{\hat{T}_n^Z(0)} \xrightarrow{d} \int_{\mathbb{R}^2} |G_h(s,t)|^2 \mu(ds,dt)\,/T^Z(0)$$

where $G_h(s,t)$ is a complex-valued Gaussian process.

They also show that for ***residuals*** from an AR model:

$$n\hat{R}_n^{\hat{Z}}(h) = n\frac{\hat{T}_n^{\hat{Z}}(h)}{\hat{T}_n^{\hat{Z}}(0)} \xrightarrow{d} \int_{\mathbb{R}^2} |G_h(s,t) + \xi_h(s,t)|^2 \mu(ds,dt)\,/T^Z(0)$$

where $(G_h(s,t), \xi_h(s,t))$ is a bivariate Gaussian random field.

See also Wan and Davis (2022) for a similar result for more general linear and nonlinear TS models.

ADCF and Sample Splitting

Recall: $\{X_1, \dots, X_{f_n}\}$ -- first f_n set of obs: used to estimate $\boldsymbol{\beta} \Rightarrow \widehat{\boldsymbol{\beta}}_{f_n}$

$\{X_{n-l_n+1}, \dots, X_n\}$ -- last l_n set of obs to compute residuals based on $\widehat{\boldsymbol{\beta}}_{f_n}$:

Then

$$l_n \hat{R}_{l_n}^{\hat{Z}}(h) = l_n \frac{\hat{T}_{l_n}^{\hat{Z}}(h)}{\hat{T}_{l_n}^{\hat{Z}}(0)} \xrightarrow{d} \int_{\mathbb{R}^2} |G_h(s,t) + \xi_h(s,t)|^2 \mu(ds,dt) \,/T^Z(0)$$

where

$$\xi_h(s,t) = \sqrt{k_{ra}}\nu_h(s,t)Q$$

Here: $k_{ra} = \lim_n l_n/f_n$;

- $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\right) = \frac{1}{\sqrt{n}} \sum_{t=1}^n m(Z_{-\infty:t}; \boldsymbol{\beta}_0) + o_p(1) \xrightarrow{d} Q, \quad Q \sim N(\mathbf{0}, \Sigma),$
- $\nu_h(s,t) = it\mathbb{E}\left((e^{isZ_0} - \varphi_Z(s))e^{itZ_h} L_h(\boldsymbol{\beta}_0)\right); L_h(\boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\beta}} h\left(X_{-\infty:j}; \boldsymbol{\beta}\right) = \frac{\partial}{\partial \boldsymbol{\beta}} Z_j(\boldsymbol{\beta})$
- $\mathrm{Cov}(G_h(s,t), Q) = \frac{k_{ov}}{\sqrt{k_{ra}}} \mu_h(s,t)$
- $\mu_h(s,t) = \mathbb{E}\left((e^{isZ_0} - \varphi_Z(s))e^{itZ_h} m(Z_{-\infty:h}; \boldsymbol{\beta}_0)\right)$

ADCF and Sample Splitting

$$l_n \hat{R}_{l_n}^{\hat{Z}}(h) = l_n \frac{\hat{T}_{l_n}^{\hat{Z}}(h)}{\hat{T}_{l_n}^{\hat{Z}}(0)} \xrightarrow{d} \int_{\mathbb{R}^2} |G_h(s,t) + \xi_h(s,t)|^2 \mu(ds,dt) \,/ T^Z(0)$$

Question: When is

$$\int_{\mathbb{R}^2} |G_h(s,t) + \xi_h(s,t)|^2 \mu(ds,dt) \begin{pmatrix} \preccurlyeq_{st} \\ =_{st} \\ \succcurlyeq_{st} \end{pmatrix} \int_{\mathbb{R}^2} |G_h(s,t)|^2 \mu(ds,dt)\,?$$

Observation:

$\preccurlyeq_{st}$ $\Longrightarrow$ asymptotic quantiles for $\hat{R}_{l_n}^{\hat{Z}}(h)$ ***smaller*** than those for $\hat{R}_{l_n}^{Z}(h)$

$\succcurlyeq_{st}$ $\Longrightarrow$ asymptotic quantiles for $\hat{R}_{l_n}^{\hat{Z}}(h)$ ***larger*** than those for $\hat{R}_{l_n}^{Z}(h)$

ADCF and Sample Splitting

$$l_n \hat{R}_{l_n}^{\hat{Z}}(h) = l_n \frac{\hat{T}_{l_n}^{\hat{Z}}(h)}{\hat{T}_{l_n}^{\hat{Z}}(0)} \xrightarrow{d} \int_{\mathbb{R}^2} |G_h(s,t) + \xi_h(s,t)|^2 \mu(ds,dt) \,/T^Z(0)$$

Question: When is

$$\int_{\mathbb{R}^2} |G_h(s,t) + \xi_h(s,t)|^2 \mu(ds,dt) \begin{pmatrix} \preccurlyeq_{st} \\ =_{st} \\ \succcurlyeq_{st} \end{pmatrix} \int_{\mathbb{R}^2} |G_h(s,t)|^2 \mu(ds,dt)\,?$$

Thoughts: Choose sample splits so that

$$G_h(s,t) + \xi_h(s,t) \triangleq G_h(s,t) \quad \text{for all } s,t.$$

Using the results on previous slide,

$$\mathrm{Cov}\big(G_h(s_1,t_1) + \xi_h(s_1,t_1), G_h(s_2,t_2) + \xi_h(s_2,t_2)\big) =$$

$$\mathrm{Cov}\big(G_h(s_1,t_1), G_h(s_2,t_2)\big) + \left[k_{ov} \nu_h(s_1,t_1)\overline{\mu_h(s_2,t_2)} + k_{ov}\overline{\nu_h(s_2,t_2)}\mu_h(s_1,t_1) + k_{ra}\nu_h(s_1,t_1)\mathrm{Var(Q)}\overline{\nu_h(s_2,t_2)}\right]$$

Need term in brackets [ ] to be zero.

ADCF and Sample Splitting: ARMA case

$\mathrm{Cov}\big(G_h(s_1,t_1)+\xi_h(s_1,t_1), G_h(s_2,t_2)+\xi_h(s_2,t_2)\big) =$

$\mathrm{Cov}\big(G_h(s_1,t_1), G_h(s_2,t_2)\big) + \big[k_{ov}v_h(s_1,t_1)\overline{\mu_h(s_2,t_2)} + k_{ov}\overline{v_h(s_2,t_2)}\mu_h(s_1,t_1) + k_{ra}v_h(s_1,t_1)\mathrm{Var}(\mathrm{Q})\overline{v_h(s_2,t_2)}\big]$

In the case of an **ARMA process**, this reduces to

$\mathrm{Cov}\big(G_h(s_1,t_1)+\xi_h(s_1,t_1), G_h(s_2,t_2)+\xi_h(s_2,t_2)\big)$

$$= Cov\big(G_h(s_1,t_1), G_h(s_2,t_2)\big) + \left[k_{ov}\left(\frac{\varphi_Z'(t_1)}{\sigma^2 t_1\varphi_Z(t_1)} + \frac{\varphi_Z'(t_2)}{\sigma^2 t_2\varphi_Z(t_2)}\right) + k_{ra}\right] v_h(s_1,t_1)\mathrm{Var}(\mathrm{Q})\overline{v_h(s_2,t_2)}$$

Remarks: [term in brackets]

(1) If $Z \sim N(0,\sigma^2)$, then $\frac{\varphi_Z'(t)}{\sigma^2 t\varphi_Z(t)} = -1$ for all t $\Longrightarrow$ term in **brackets** is $-2k_{ov} + k_{ra} = 0$ if $k_{ov} = \frac{k_{ra}}{2}$. Thus,

$\mathrm{Cov}\big(G_h(s_1,t_1)+\xi_h(s_1,t_1), G_h(s_2,t_2)+\xi_h(s_2,t_2)\big) = Cov\big(G_h(s_1,t_1), G_h(s_2,t_2)\big)$ and

$$l_n\hat{R}_{l_n}^{\hat{Z}}(h) = l_n\frac{\hat{T}_{l_n}^{\hat{Z}}(h)}{\hat{T}_{l_n}^{\hat{Z}}(0)} \xrightarrow{d} \int_{\mathbb{R}^2}|G_h(s,t)+\xi_h(s,t)|^2\mu(ds,dt)\,/T^Z(0) \triangleq \int_{\mathbb{R}^2}|G_h(s,t)|^2\mu(ds,dt)\,/T^Z(0)$$

ADCF and Sample Splitting: ARMA case

$$\mathrm{Cov}\big(G_h(s_1,t_1) + \xi_h(s_1,t_1), G_h(s_2,t_2) + \xi_h(s_2,t_2)\big)$$

$$= \mathrm{Cov}\big(G_h(s_1,t_1), G_h(s_2,t_2)\big) + \left[k_{ov}\left(\frac{\varphi_Z'(t_1)}{\sigma^2 t_1 \varphi_Z(t_1)} + \frac{\varphi_Z'(t_2)}{\sigma^2 t_2 \varphi_Z(t_2)}\right) + k_{ra}\right] v_h(s_1,t_1)\mathrm{Var}(\mathrm{Q})\overline{v_h(s_2,t_2)}$$

Remarks:

(2) If Z is Laplace or student's t with finite variance, then

$$\frac{\varphi_Z'(t)}{\sigma^2 t \varphi_Z(t)} \geq -1 \text{ for all t,}$$ (proof in paper; uses special functions)

$\Longrightarrow$ term in **brackets** is **positive** whenever $k_{ra} \geq 2k_{ov}$. Thus,

$\mathrm{Cov}\big(G_h(s_1,t_1) + \xi_h(s_1,t_1), G_h(s_2,t_2) + \xi_h(s_2,t_2)\big) \succcurlyeq \mathrm{Cov}\big(G_h(s_1,t_1), G_h(s_2,t_2)\big)$ and

$$\int_{\mathbb{R}^2} |G_h(s,t) + \xi_h(s,t)|^2 \mu(ds,dt) \,/ T^Z(0) \succcurlyeq_{st} \int_{\mathbb{R}^2} |G_h(s,t)|^2 \mu(ds,dt) \,/ T^Z(0)$$

The ACDF of the residuals will be **stochastically bigger** than those of the noise when $k_{ra} \geq 2k_{ov}$.

ADCF and Sample Splitting: GARCH case

A similar result holds for **GARCH**, but in this case

$$\mathrm{Cov}\big(G_h(s_1,t_1) + \xi_h(s_1,t_1), G_h(s_2,t_2) + \xi_h(s_2,t_2)\big)$$

$$= Cov\big(G_h(s_1,t_1), G_h(s_2,t_2)\big) + [k_{ra} - k_{ov}(\lambda_Z(t_1) + \lambda_Z(t_2))]v_h(s_1,t_1)\mathrm{Var}(\mathrm{Q})\overline{v_h(s_2,t_2)},$$

where

$$\lambda_Z(t) = -\frac{2}{\mathbb{E}Z^4 - 1} \cdot \frac{\varphi_Z(t) + \varphi_Z''(t)}{t\varphi_Z'(t)}$$

Remarks:

If Z is Gaussian, then $\lambda_Z(t) = 1$ and term in **brackets** is zero whenever $k_{ra} = 2k_{ov}$.

If Z Laplace, $\lambda_Z(t) \leq 1 \implies$ term in **brackets** is positive whenever $k_{ra} \geq 2k_{ov}$. Thus,

$\mathrm{Cov}\big(G_h(s_1,t_1) + \xi_h(s_1,t_1), G_h(s_2,t_2) + \xi_h(s_2,t_2)\big) \succcurlyeq Cov\big(G_h(s_1,t_1), G_h(s_2,t_2)\big)$ and

$$\int_{\mathbb{R}^2} |G_h(s,t) + \xi_h(s,t)|^2 \mu(ds,dt) \,/T^Z(0) \succcurlyeq_{st} \int_{\mathbb{R}^2} |G_h(s,t)|^2 \mu(ds,dt) \,/T^Z(0)$$

The ACDF of the residuals will be **stochastically bigger** than those of the noise when $k_{ra} \geq 2k_{ov}$.

Two Lemmas

Previous results rely on a couple of lemmas.

Lemma 1: If X_1 and X_2 are two mean-zero Gaussian processes on $\mathbb{R}^2$ such that

$$\mathrm{Cov}\big(X_1(s_1,t_1),X_1(s_2,t_2)\big) \preccurlyeq \mathrm{Cov}\big(X_2(s_1,t_1),X_2(s_2,t_2)\big)$$

for all $(s_1,t_1),(s_2,t_2) \in \mathbb{R}^2$, then

$$\int_{\mathbb{R}^2} |X_1(s,t)|^2 \mu(ds,dt) \preccurlyeq_{st} \int_{\mathbb{R}^2} |X_1(s,t)|^2 \mu(ds,dt)$$

provided the integrals exist (expectations are finite).

Lemma 2: If $Z \sim t_\nu$ with $\nu > 2$, then

$$\frac{\varphi_Z'(t)}{\sigma^2 t \varphi_Z(t)} \geq -1.$$

Proof uses a Turan type inequality for modified Bessel functions which says

$$K_{\nu/2-2}(x)K_{\nu/2}(x) - K_{\nu/2-1}^2(x) > 0 \ \text{ for all } x \in \mathbb{R}.$$

A GARCH(1,1) Example

GARCH(1,1) process:

$$\begin{cases} X_t = \sigma_t Z_t, & \{Z_t\} \sim IID\ (0,1) \\ \sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2, & \end{cases}$$

with $\alpha_0 > 0, \alpha_1, \beta_1 \geq 0, \alpha_1 + \beta_1 < 1.$ Set $\boldsymbol{\beta} = (\alpha_0, \alpha_1, \beta_1)^{\mathsf{T}}$.

Iterating the GARCH recursion, we obtain

$$\sigma_t^2(\boldsymbol{\beta}) = c_0(\boldsymbol{\beta}) + \sum_{k=1}^{\infty} c_k(\boldsymbol{\beta}) X_{t-k}^2$$

Now this volatility can be approximated by

$$\hat{\sigma}_t^2(\boldsymbol{\beta}) = c_0(\boldsymbol{\beta}) + \sum_{k=1}^{t-1} c_k(\boldsymbol{\beta}) X_{t-k}^2$$

from which the estimated residuals are given by:

$$\hat{Z}_t(\boldsymbol{\beta}) = \frac{X_t}{\hat{\sigma}_t(\boldsymbol{\beta})}.$$

A GARCH(1,1) Example

GARCH(1,1) process:

$$X_t = \sigma_t Z_t, \qquad \{Z_t\} \sim IID\ (0,1)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

with $\alpha_0 > 0, \alpha_1, \beta_1 \geq 0, \alpha_1 + \beta_1 < 1.$ Set $\boldsymbol{\beta} = (\alpha_0, \alpha_1, \beta_1)^{\mathsf{T}}$.

- Estimate $\boldsymbol{\beta}$ using the first f_n observations via ***quasi-likelihood***:

$$\widehat{\boldsymbol{\beta}}_{f_n} = \arg\max_{\boldsymbol{\beta}} \sum_{t=1}^{f_n} l_t(\boldsymbol{\beta}); \qquad l_t(\boldsymbol{\beta}) = -\frac{1}{2}\log \hat{\sigma}_t^2(\boldsymbol{\beta}) - \frac{X_t^2}{\hat{\sigma}_t^2(\boldsymbol{\beta})}$$

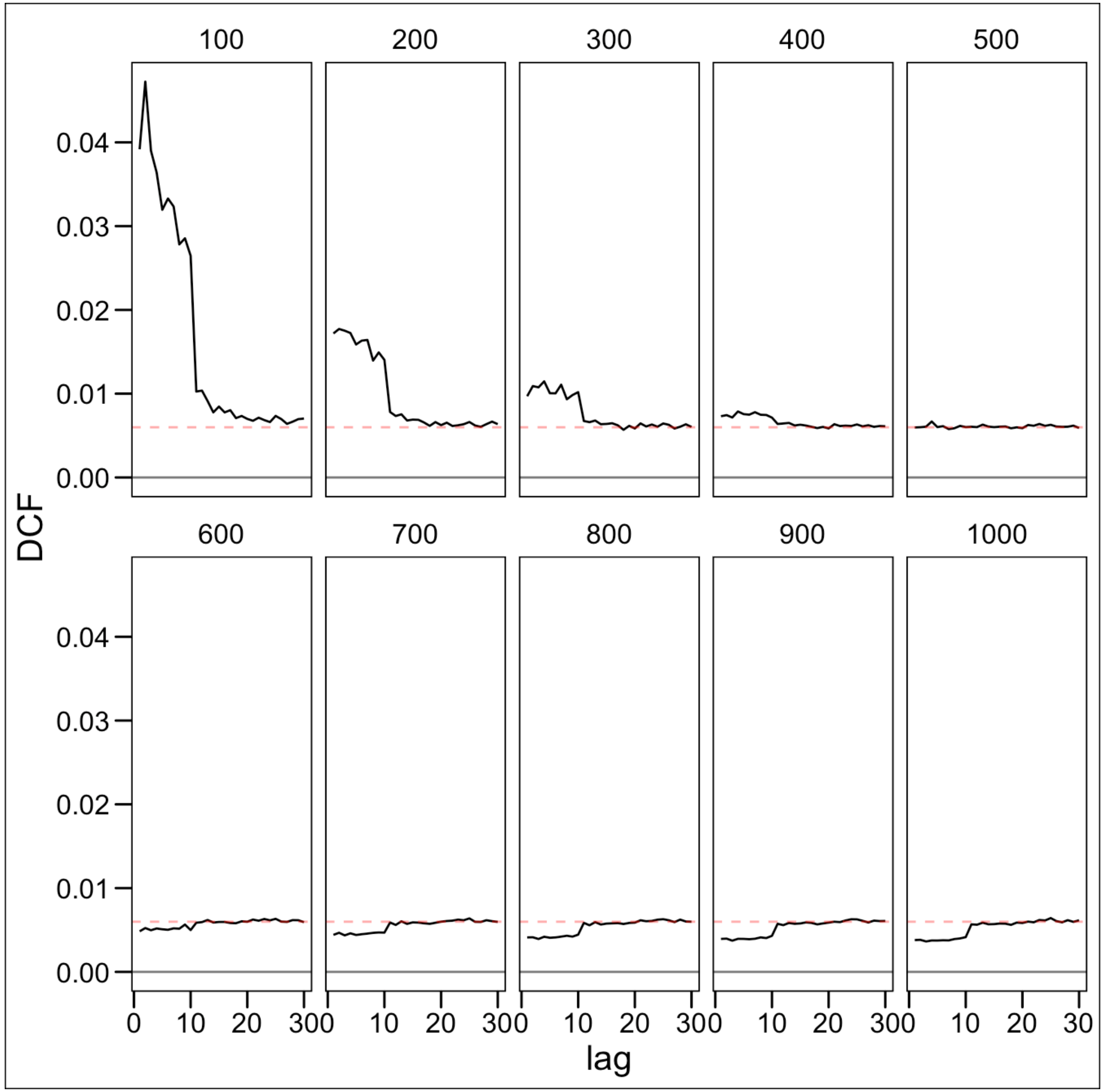
- Calculate residuals

$$\hat{Z}_t\left(\widehat{\boldsymbol{\beta}}_{f_n}\right) = \frac{X_t}{\hat{\sigma}_t\left(\widehat{\boldsymbol{\beta}}_{f_n}\right)}, \qquad for\ t = n - l_n + 1, \dots, n$$

- Previous asymptotic results now hold for the ACF and ADCF applied to $\hat{Z}_t\left(\widehat{\boldsymbol{\beta}}_{f_n}\right)$.

DACF of $\hat{Z}_t(\widehat{\boldsymbol{\phi}}_{f_n})$ $AR(10)$

Sample splitting: First $f_n = k * 100, k = 1, \dots, 10$ of the observations to estimate $\boldsymbol{\phi}$; $l_n = n$. Gaussian noise.



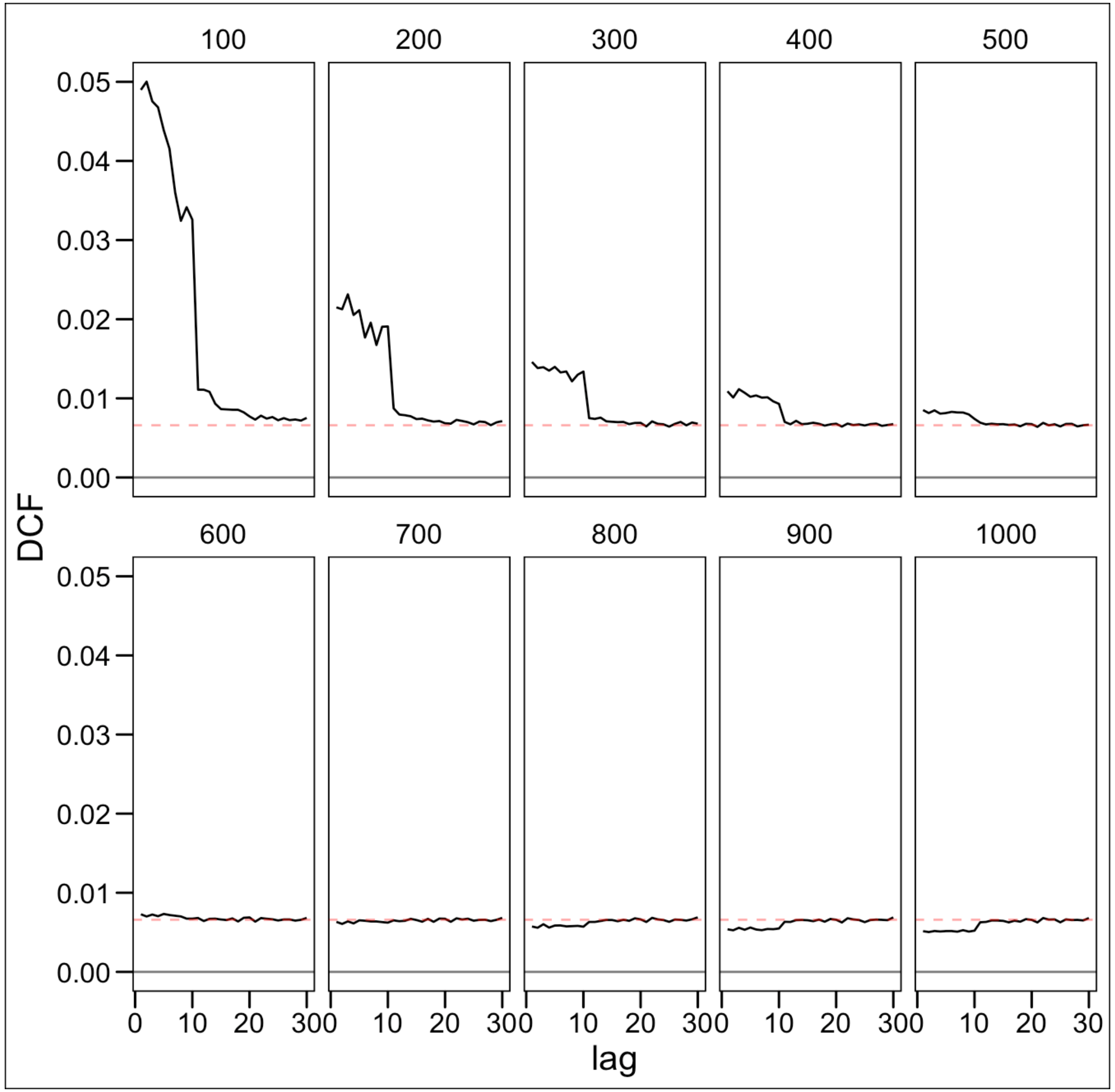
Note:

$$k_{ra} = \lim_n l_n/f_n = \frac{10}{k}; \quad k_{ov} = 1$$

$$k_{ra} - 2k_{ov} \begin{cases} > 0, & k < 5 \\ = 0, & k = 5 \\ < 0, & k > 5 \end{cases}$$

DACF of $\hat{Z}_t(\widehat{\boldsymbol{\phi}}_{f_n})$: $AR(10)$

Sample splitting: First $f_n = k * 100, k = 1, \dots, 10$ of the observations to estimate $\boldsymbol{\phi}$; $l_n = n$. Laplace noise.



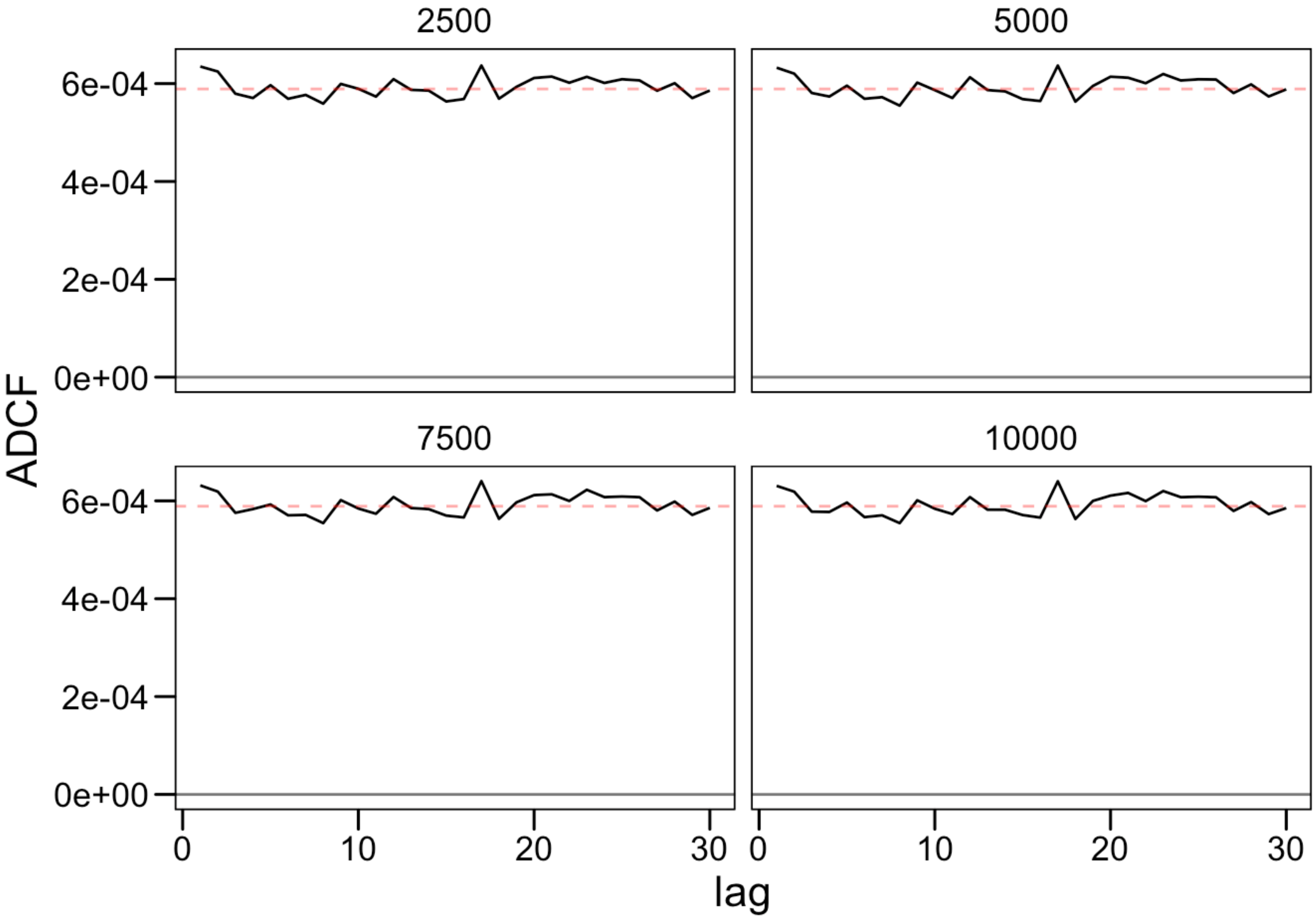
Note:

$$k_{ra} = \lim_n l_n/f_n = \frac{10}{k}; \quad k_{ov} = 1$$

$$k_{ra} - 2k_{ov} > 0, if\ k < 5$$

DACF of $\hat{Z}_t(\widehat{\boldsymbol{\beta}}_{f_n})$ $GARCH(1,1)$

Sample splitting: First $f_n = k * 2500, k = 1, \dots, 4$ of the observations for estimation; $l_n = n$. Gaussian noise.

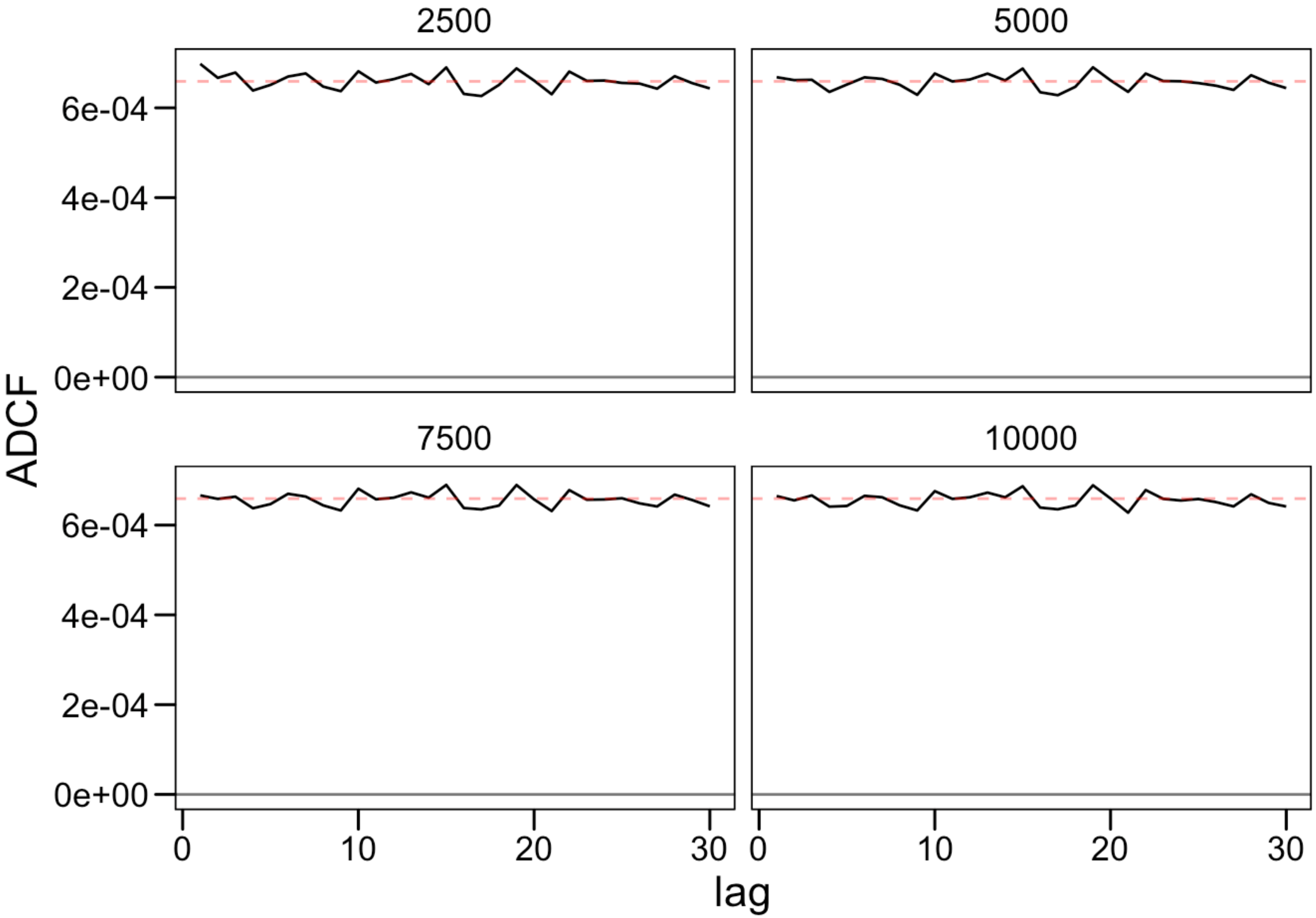


Note:

$$k_{ra} = \frac{4}{k} > 2, when\ k < 2$$

DACF of $\hat{Z}_t(\widehat{\boldsymbol{\beta}}_{f_n})$: $GARCH$ (1,1)

Sample splitting: First $f_n = k * 2500, k = 1, \ldots, 4$ of the observations for estimation; $l_n = n$. Laplace noise.

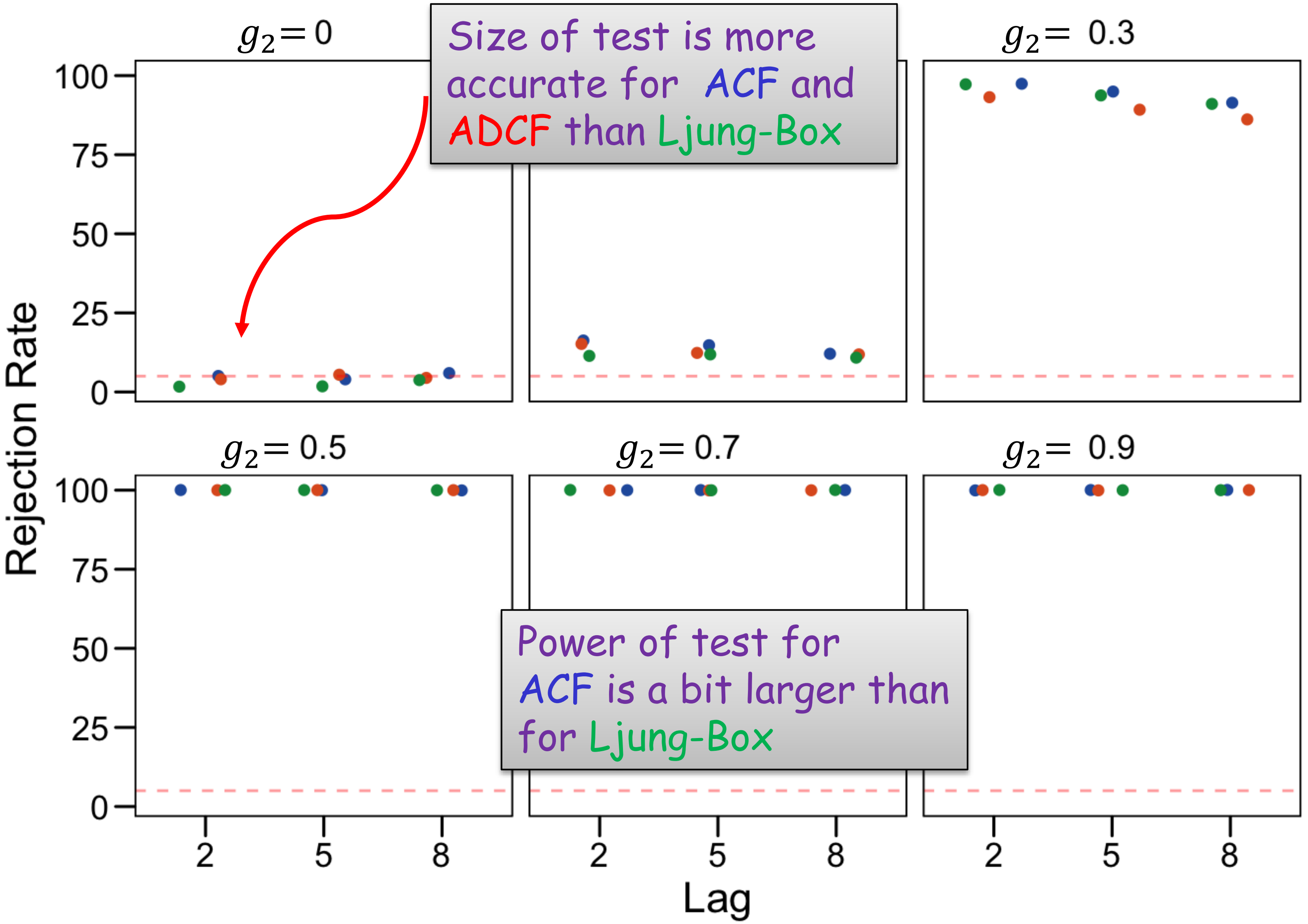


Note:

$$k_{ra} = \frac{4}{k} > 2, when\ k < 2$$

What about power?

Rejecting rates (in percentages): True model AR(1); alternative model AR(2): $(1-.7B)(1-g_2B)X_t = Z_t$



blue = ACF (sample split)
red = ADCF (sample split)
green= Ljung-Box test

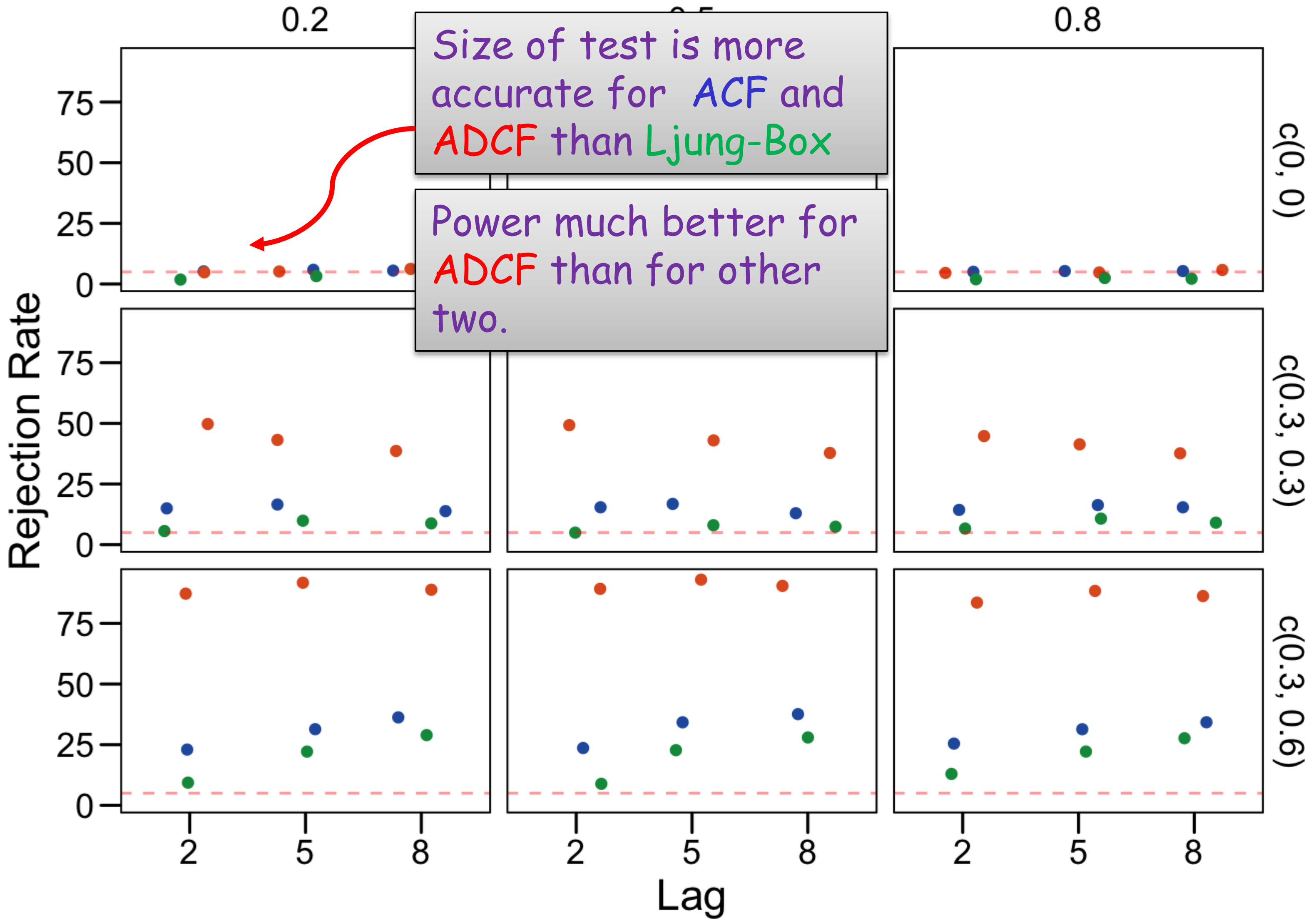
ACF test: $n\sum_{k=1}^{h}\left(\hat{\rho}_n^{\hat{Z}}(k)\right)^2 \sim \chi_k^2$

ADCF test: $n\sum_{k=1}^{h}\hat{R}_n^{\hat{Z}}(k)$

LB test: $n(n+2)\sum_{k=1}^{h}\frac{\left(\hat{\rho}_n^{\hat{Z}}(k)\right)^2}{n-k} \sim \chi_k^2$

What about power?

Rejecting rates (in percentages): True model AR(1); noise generated rom a GARCH(1,1): $\sigma_t^2 = 1 + \alpha Z_{t-1}^2 + \beta \sigma_{t-1}^2$



blue = ACF (sample split)
red = ADCF (sample split)
green= Ljung-Box test

ACF test: $n \sum_{k=1}^{h} \left(\hat{\rho}_n^{\hat{Z}}(k)\right)^2 \sim \chi_k^2$

ADCF test: $n \sum_{k=1}^{h} \hat{R}_n^{\hat{Z}}(k)$

LB test: $n(n+2) \sum_{k=1}^{h} \frac{\left(\hat{\rho}_n^{\hat{Z}}(k)\right)^2}{n-k} \sim \chi_k^2$

Postscript

Durbin, J. "Kolmogorov-Smirnov tests when parameters are estimated." Empirical Distributions and Processes: Selected Papers from a Meeting at Oberwolfach, March 28–April 3, 1976. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006.

Method 4 (half-sample device). Similar to a random substitution method described earlier in paper. Let $x_1, \dots, x_n$ be observations in random order where $n = 2m$.

- $F(x;\theta)$ is the cdf.
- $\hat{F}_{\boldsymbol{n}}(x;\theta)$ = empirical cdf., i.e., $\hat{F}_{\boldsymbol{n}}(x;\theta) = \#\{i: F(x_i;\theta) \le x\}/n$

Estimation of θ: Let $\hat{\theta}_{\boldsymbol{m}}$ be can *efficient estimator* of θ based on $x_1, \dots, x_m$

Kolmogorov-Smirnov test statistic:

$$\sqrt{n}\left(\hat{F}_n(t;\theta_0) - t\right) \xrightarrow{\text{d}} B(t), \qquad B(\cdot) \sim \text{Brownian bridge on } [0,1]$$

Then

$$\sqrt{n}\left(\hat{F}_n\left(t;\hat{\theta}_m\right) - t\right) \xrightarrow{\text{d}} B(t).$$

Take Home Message

- Estimate the parameters of model using ***first half,*** $X_1, \dots, X_{\frac{n}{2}}$, of the data.

- Calculate the residuals based on the ***entire stream of data,***

$$\hat{Z}_j(\widehat{\boldsymbol{\beta}}) = h\big(X_{-\infty:j}; \widehat{\boldsymbol{\beta}}\big), \qquad j = 1, \dots, n.$$

- If model is correct, then

$$\hat{\rho}_{\hat{Z}}(h) \sim AN(0, \frac{1}{n})$$

Should this become the ***standard method*** for inspecting residuals in time series modeling? **Easy** to compute.

- In some cases,

$$\hat{R}^{\hat{Z}}(h) \text{ is asymptotically the same as } \hat{R}^{Z}(h)$$

Thank you for your attention!