Financial Time Series

Topic 3: ARMA and Time Series Modeling

Hung Chen
Department of Mathematics
National Taiwan University
3/14/2002

OUTLINE

1. General AR Processes
2. General MA Processes
3. ARMA Processes
4. Preliminary Data Analysis
   - Time Series Plots
   - Data Transformation
   - Checking Normality
5. Time Series Modeling
   - Model Identification
   - Model Estimation
   - Diagnostic Checking

$AR(2)$ process:

$$X_t - \mu = \phi_1(X_{t-1} - \mu) + \phi_2(X_{t-2} - \mu) + a_t$$

or

$$(1 - \phi_1 B - \phi_2 B^2)(X_t - \mu) = a_t.$$

- 2nd moments and ACF:

$$\begin{aligned}
\gamma_k &= \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2}, \\
\rho_k &= \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}, \\
\rho_1 &= \phi_1 + \phi_2 \rho_1, \\
\rho_2 &= \phi_1 \rho_1 + \phi_2, \\
\rho_\ell &= \phi_1 \rho_{\ell-1} + \phi_2 \rho_{\ell-2}, \quad \ell > 0, \\
\rho_1 &= \frac{\phi_1}{1 - \phi_2}, \\
\rho_2 &= \frac{\phi_1^2}{1 - \phi_2} + \phi_2.
\end{aligned}$$

  Observe that

$$\begin{aligned}
V(X_t) &= \phi_1^2 V(X_{t-1}) + \phi_2^2 V(X_{t-2}) \\
&\quad + 2\phi_1 \phi_2 Cov(X_{t-1}, X_{t-2}) + \sigma^2, \\
V(X_t) &= \left[1 - \phi_1^2 - \phi_2^2 - \frac{2\phi_1^2 \phi_2}{1 - \phi_2} \right] \sigma^2.
\end{aligned}$$

- Recall that $(1 - \phi_1 B - \phi_2 B^2)\rho_\ell = 0$.

- Consider the second order polynomial $x^2 - \phi_1 x - \phi_2 = 0$
- If the two solutions are real-valued, then $1 - \phi_1 B - \phi_2 B^2$ can be factored as $(1 - w_1 B)(1 - w_2 B)$. The $AR(2)$ model can be regarded as an $AR(1)$ model operates on top of another $AR(1)$ model.
  The ACF of X_t is then a mixture of two exponential decays.
- If the solutions are complex numbers, the plot of ACF of X_t would show a picture of damping sine and cosine waves.
  In business and economic applications, they give rise to the behavior of business cycles.

- The model is stationary if the roots of $1 - \phi_1 B - \phi_2 B^2 = 0$ lie outside the unit circle. By restricting the roots outside the unit circle, we have

$$\phi_1 + \phi_2 < 1, \quad \phi_1 - \phi_2 < 1, \quad -1 < \phi_2 < 1.$$

- Mathematically speaking, we require $1 - \phi_1 B - \phi_2 B^2 \neq 0$ for all $|B| \leq 1$.

This condition implies that $(1-\phi_1 B-\phi_2 B^2)^{-1}$ has a power series expansion.

Example: Check Stationarity

- $X_t = 0.8X_{t-1} - 0.4X_{t-2} + a_t$
- $\phi(B) = 1 - 0.8B + 0.4B^2$
- Roots of $\phi(B)$ are $B = 1 \pm 1.22474i$ and $|B| = 1.58114$.
- Since the roots of $\phi(B)$ are all outside the unit circle, the process is stationary.

$AR(p)$ processes

- The model

$$X_t - \mu = \phi_1(X_{t-1} - \mu) + \cdots + \phi_p(X_{t-p} - \mu) + a_t$$

or

$$(1 - \phi_1 B - \cdots - \phi_p B^p)(X_t - \mu) = a_t.$$

- This model says that conditional on the past return $X_{t-1}, \ldots, X_{t-p}$, we have

$$\begin{aligned} E(X_t | X_{t-1}, \ldots, X_{t-p}) &= c + \sum_{i=1}^{p} \phi_i X_{t-i}, \\ Var(X_t | X_{t-1}, \ldots, X_{t-p}) &= Var(a_t) = \sigma_a^2. \end{aligned}$$

This is a Markov property.

- The $AR(p)$ process is always invertible.
- It is stationary if the roots of

$$1 - \phi_1 B - \cdots - \phi_p B^p = 0$$

lie outside the unit circle.

- ACF of the Model:

$$\gamma_k = \phi_1 \gamma_{k-1} + \cdots + \phi_p \gamma_{k-p}$$

for $k \geq 1$. Therefore,

$$\rho_k = \phi_1 \rho_{k-1} + \cdots + \phi_p \rho_{k-p}.$$

We have a system of Yule-Walker equations:

$$\begin{cases} \rho_1 &= \phi_1 + \phi_2\rho_1 + \phi_3\rho_2 + \cdots + \phi_p\rho_{p-1} \\ \rho_2 &= \phi_1\rho_1 + \phi_2 + \phi_3\rho_1 + \cdots + \phi_p\rho_{p-2} \\ \vdots & \vdots \quad \vdots \\ \rho_p &= \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \phi_3\rho_{p-3} + \cdots + \phi_p \end{cases}$$

- The plot of ACF of a statoionary $AR(p)$ model would show a mixture of damping sine and cosine patterns and damping exponential decays.

Example: ACF

- Consider the quarterly growth rate of U.S. real gross national product (GNP), seasonally adjusted, from 1947.II to 1991.I.
- Employ an $AR(3)$ model for the data.
- The fitted model is

  $r_t = 0.0047 + 0.35r_{t-1} + 0.18r_{t-2} - 0.14r_{t-3} + a_t$

  and $\hat{\sigma}_a = 0.0098$.
- The third-order difference equation is

$$1 - 0.35B - 0.18B^2 + 0.14B^3 = 0,$$

which can be factorized as

$$(1 + 0.52B)(1 - 0.87B + 0.27B^2) = 0.$$

- It suggests an exponentially decaying nature of the GNP growth rates.
  The second factor implies existence of stochastic business cycle in the quarterly growth rate of U.S. real GNP. The average length of the stochastic cycles is approximately 10.83 quarters.

Order determination of AR models

- Use partial autocorrelation function.
- Recall PACF is determined by the regression coefficients determined by r_t on r_{t-1}, on r_{t-1}, r_{t-2}, and etc.
- $AR(1)$: $\psi_{11} = \rho_1 = \psi$,
  $\psi_{kk} = 0$ for $k > 1$.
- $AR(2)$: $\psi_{11} = \rho_1$,

$$\psi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2},$$

  $\psi_{kk} = 0$ for $k > 2$.
- $AR(p)$: $\psi_{11} \neq 0$, $\psi_{22} \neq 0$, $\ldots$, $\psi_{pp} \neq 0$, $\psi_{kk} = 0$ for $k > p$.

- ACF: tail-off at lag p
- PACF: cut-off at lag p

The linear filter representation $X_t - \mu = \phi(B)a_t$ can be obtained by equating coefficients in

$$(1 - \psi_1 B - \cdots - \psi_p B^p)\phi(B) = 1$$

or

$$(1 - \psi_1 B - \cdots - \psi_p B^p)(1 - \phi_1 B - \phi_2 B - \cdots) = 1.$$

Work out $AR(2)$ in the class.

MA(2) process:

- The model

$$X_t - \mu = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$$

or

$$X_t - \mu = (1 - \theta_1 B - \theta_2 B^2) a_t.$$

- The model is stationary if the roots of $1 - \theta_1 B - \theta_2 B^2 = 0$ lie outside the unit circle. Analogue to the stationary condition of an $AR(2)$ model, we require

$$\theta_1 + \theta_2 < 1, \quad \theta_2 - \theta_1 < 1, \quad -1 < \theta_2 < 1.$$

- 2nd moments and ACF:

$$V(X_t) = (1 + \theta_1^2 + \theta_2^2)\sigma^2.$$

$$\gamma_k = \begin{cases} (-\theta_1 + \theta_1\theta_2)\sigma^2 & k = 1 \\ -\theta_2\sigma^2 & k = 2 \\ 0 & k \geq 3 \end{cases}.$$

Therefore,

$$\rho_k = \begin{cases} \frac{(-\theta_1 + \theta_1\theta_2)}{1+\theta_1^2+\theta_2^2} & k = 1 \\ \frac{-\theta_2}{1+\theta_1^2+\theta_2^2} & k = 2 \\ 0 & k \geq 3 \end{cases}.$$

- ACF: cut-off at lag 2
- PACF: tail-off at lag 2

MA(q) process:

- The model

$$X_t - \mu = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$

  or

$$X_t - \mu = (1 - \theta_1 B - \cdots - \theta_q B^q) a_t = \theta(B) a_t.$$

- The model is stationary if the roots of

$$1 - \theta_1 B - \cdots - \theta_q B^q = 0$$

  lie outside the unit circle.

- The weights in the $AR(\infty)$ representation $\pi(B)(X_t - \mu) = a_t$ are given by $\pi(B) = \theta^{-1}(B)$ and can be obtained by equating coefficients of B^j in $\pi(B)\theta(B) = 1$.

- Variance is time-invariant.

$$Var(X_t) = (1 + \theta_1^2 + \cdots + \theta_q^2)\sigma_a^2.$$

- ACF:

$$\begin{aligned} \rho_k &= \frac{-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \cdots + \theta_q^2}, \\ \rho_k &= 0, \quad k > q. \end{aligned}$$

- ACF: cut-off at lag q
  - If the lag-q ACF is non-zero, but ACF beyond lag-q are all zero, then we have an $MA(q)$ model.
  - This property says that the $MA(q)$ model is linearly related to its first q lagged values only.
  - This model is said to have *short memory*.
- PACF: tail-off at lag q

$ARMA(p,q)$ process

- This is the most general class of mixed models, combining the $AR(p)$ and $MA(q)$ characteristics.
- The model

  $$(X_t - \mu) - \phi_1(X_{t-1} - \mu) - \cdots - \phi_p(X_{t-p} - \mu)$$
  $$= a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q},$$

  or

  $$(1 - \phi_1 B - \cdots - \phi_p B^p)(X_t - \mu)$$
  $$= (1 - \theta_1 B - \cdots - \theta_q B^q)a_t,$$

  i.e.

  $$\phi(B)(X_t - \mu) = \theta(B)a_t.$$
- The model is stationary if roots of $\phi(B)$ all lie outside the unit circle.
- The process is invertible if roots of $\theta(B)$ all lie outside the unit circle.
- ACF tail-off at lag p.
  It will eventually follow the same pattern as that of an $AR(p)$ process after $q - p$ initial values.

- PACF tail-off at lag q.
It will eventually behaves (for $k > p - q$) like that of an $MA(q)$ process.

Example:

Find the AR representation of the following $ARMA(1,1)$ model

$$(1 + 0.4B)X_t = (1 - 0.6B)a_t.$$

$$\begin{aligned} a_t &= \frac{(1 + 0.4B)}{(1 - 0.6B)} X_t \\ &= (1 + 0.4B)(1 + 0.6B + 0.6^2 B^2 + \cdots) X_t \\ &= X_t + \sum_{j=1}^{\infty} \pi^j X_{t-j}, \end{aligned}$$

where $\pi_j = (0.6)^{j-1}$.

- It should be noted that stationary and invertible ARMA models can be written as either $AR(\infty)$ or $MA(\infty)$.
- The MA representation shows the impact of the past shock a_{t-i}.
- The AR representation shows the dependence of the current X_t on the past X_{t-i}.

Unit-root Nonstationarity

- In some studies, price series are of interest which tend to be nonstationary.
  It is caused by the fact that there is no fixed level for price series.
- In the time series literature, such a nonstationary series is called unit-root time series. The best known example is the random walk model.
- Random walk

$$p_t = p_{t-1} + a_t,$$

  where p_0 is a real number denoting the starting value of the process and $\{a_t\}$ is a white noise series.

  - The MA representation of the random walk model is

$$p_t = a_t + a_{t-1} + a_{t-2} + \cdots.$$

    It tells us that we cannot make a long-term forecast.
  - The impact of any past shock p_{t-i} on p_t does not decay over time.

This means that the series has strong memory as it remembers all of the past shocks. (Such shocks have a permanent effect on the system.)

– The stock price is not predictable.

Time Series Plots

- The first step in time series analysis is data plotting.
- Classical Decomposition Model:

$$X_t = m_t + s_t + Y_t,$$

  where

  - m_t: a slowly changing function known as a **trend** component
  - s_t: a function with known period d referred to as a **seasonal** component
  - Y_t: a **random noise** component which is stationary

- It gives us important information about the time series data concerning:
  - outliers/typos
  - stationarity (in mean and/or variance)
  - seasonality/cycles/patterns

- Aim
  - Estimate and extract the deterministic components m_t and s_t.

- Hope that the residual or noise component Y_t will turn out to be a stationary random process.
- Find a satisfactory probabilistic model for the process $\{Y_t\}$.
- Use the properties of probability model in conjunction with m_t and s_t for purposes of prediction and control of $\{X_t\}$.

- Box and Jenkins (1970) developed an alternative approach to remove m_t and s_t.
  It is to apply difference operators repeatedly to the data $\{x_t\}$ until the differenced observations resemble a realization of some stationary process.

Outliers

- Time series observations are sometimes influenced by interuptive events, such as wars, crises, natural disaster, or even errors of typing or recording.
  The consequences of these events could create spurious observations that are inconsistent with the rest of the series.
  They are called outliers.

- Time series outliers need special treatment.
- Example:

Stationarity

- Nonstationarity in mean (i.e., the mean follows a time trend) can be easily detected from the graph.
- For an example, the Australian Retail Prices Index exhibits a clear upward time trend.
  Its mean is not a constant and depends on time.
  Conclude from the graph that the series is nonstationary in mean.
- Nonstationarity in variance (i.e., variation of the time series depends on time) can be also easily detected from the graph.
- For an example, consider the U.S. Tobacco Production from 1871 to 1984.
  The series is nonstationary in variance as well as nonstationary in mean.

Seasonality

- Many business and economic time series contain a seasonal phenomenon that repeats itself after a regular period of time.
- There are some special time series models that can be used to capture the seasonal cycles.

Data Transformations

- If the raw data are not stationary, proper transformation(s) might be used to alleviate the problems.
- Commonly used transformations in time series analysis include:
  - Differencing: ARIMA models Remove mean trend.
    Remove polynomial mean trend effectively since the data are observed in equally spaced intervals.
    Determine the order of differencing (d) by the trial-and-error method.
    Try $d = 1, \ldots$ and plot the series to see whether the transformed series is reduced to stationarity.
    Refer to Chapter 2.6.2.
  - Power transformation: Stabilize the variance.
    The power transformations are defined only on positive series. However, a constant can always be added to the series (make the series all positive) without af-

fecting the correlation structure of the data.
Refer to Chapter 2.6.1.

Checking Normality

- If the time series is normally distributed, all distribution results usually work better.
- Frequently, the power transformation not only stabilizes the variance, but also improves the approximation to normality.
- Use probability plot (QQ plot) to check the normality.

Model Identification

- The goal is to match patterns of the sample ACF and sample PACF from the data with the known patterns of ACF and PACF for the ARMA models.
- The result from the Model Identification stage is a selected $ARMA(p,q)$ model for the data (i.e., we identify the values of p and q for the data).
- Under the stationarity and ergodicity assumptions, μ and σ^2 can be estimated by the sample mean and variance as following:

$$\begin{aligned} \bar{x} &= T^{-1} \sum_{t=1}^{T} x_t, \\ s^2 &= T^{-1} \sum_{t=1}^{T} (x_t - \bar{x})^2. \end{aligned}$$

- lag k sample autocorrelation (SACF)

$$r_k = (Ts^2)^{-1} \sum_{t=k+1}^{T} (x_t - \bar{x})(x_{t-k} - \bar{x}).$$

  – If the data are generated from a white noise, then

$$V(\hat{\rho}_k) = T^{-1/2}.$$

When T is large, $\sqrt{T} r_k \xrightarrow{L} N(0,1)$.
An absolute value of r_k in excess of $2T^{-1}$ may be regarded as significantly different from zero.
Plot the sample autocorrelation function r_k as a function of k, approximately 0.95 of the sample autocorrelations should lie between the bounds $\pm 1.96 T^{-1/2}$.
This can be used as a check that the observations truly are from a white noise process.

– Consider the general case that $\rho_k = 0$ for $k > q$.
The approximate variance of r_k, for $k > q$, is

$$V(r_k) = T^{-1}(1 + 2\rho_1^2 + \cdots + 2\rho_q^2).$$

A conservative scheme for handling the case that q is unknown.
(i) Use SACF plot to check whether it is a realization from a white noise process.
(ii) Increase successively the value of q and replace the ρ_js by r_js.
(iii) The variances of $r_1, r_2, \ldots, r_k$ can

be estimated as $T^{-1}, T^{-1}(1+2r_1^2), \ldots,$
$T^{-1}(1+2r_1^2+\cdots+2r_{k-1}^2)$.

- Calculation of Sample partial autocorrelation (SPACF): (i) Fit AR models of increasing order.
  (ii) The estimate of the last coefficient in each model is $\hat{\psi}_{kk}$.

- If the data follow an $AR(p)$ process, then for $k > p$, $\sqrt{T}\hat{\psi}_{kk} \xrightarrow{L} N(0,1)$.

- SACF and SPACF plot of the data are basis for model identification.
  We identify the model by match the SACF and SPACF plots with the theoretical patterns of ACF and PACF of known ARMA models.

- Theoretical patterns of ACF and PACF:

ACF
Whi
Clean
$AR(p$
Tail-off
$MA(q$
Cut-off lag q
$ARMA($
Tail-off at lag $q -$

Procedure for Model Identification

1. Plot the time series data and choose proper transformations.
2. Compute and examine the SACF and SPACF of the original series (raw or variance-stabilized) to further confirm a necessary degree of differencing.
   If the SACF decays very slow and there is a spike of SPACE at lag 1 only, it indicates that differencing is needed.
   Try $d = 1$ first and re-examine the SACF and SPACF of the differenced series.
   If the SACF are still not decaying, try higher values of d.

It should be noted that d is seldom greater than 2 in practice.

3. Compute the SACF and SPACF of the properly transformed and differenced series to tentatively identify the orders of p and q.

Model Estimation

After specifying an $ARMA(p, q)$ model for the data, the next step is to estimate the parameters in the model.
There are $(p+q+1)$ parameters (i.e., $\phi_1, \cdots, \phi_p$, $\theta_1, \cdots, \theta_q, \sigma^2$).

The Method of Moments

- This method consists of substituting sample moments such as mean, sample variance and SACF for their corresponding model counterparts and solving the resultant equations.
- Consider an $AR(p)$ process,

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + a_t.$$

Recall the Yule-Walker equations:

$$\begin{cases} \rho_1 &= \phi_1 + \phi_2\rho_1 + \phi_3\rho_2 + \cdots + \phi_p\rho_{p-1} \\ \rho_2 &= \phi_1\rho_1 + \phi_2 + \phi_3\rho_1 + \cdots + \phi_p\rho_{p-2} \\ \vdots & \vdots \quad \vdots \\ \rho_p &= \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \phi_3\rho_{p-3} + \cdots + \phi_p \end{cases}$$

Replacing the unknown ρ_k by sample $\hat{\rho}_k$ and solving the equations and we get $\hat{\phi}_1, \cdots, \hat{\phi}_p$.

$$\hat{\sigma}^2 = s^2(1 - \hat{\phi}_1\hat{\rho}_1 - \cdots - \hat{\phi}_p\hat{\rho}_p.$$

- Unlike the $AR(p)$ case, the moment equations for $MA(q)$ and the mixed $ARMA$ models are nonlinear.
  It is very difficult to solve.
- The moment estimators are also sensitive to rounding errors.
  They are used as initial estimates for other more advanced estimation methods.

Estimation in MA Models

- Use maximum likelihood estimation.
- The first approach assumes that the initial shocks, i.e. X_t for $t \leq 0$, are zero.
  - The shocks are computed recursively from the model, starting with $a_1 = X_1 - c_0$.
  - Parameter estimates obtained by this approach are called the conditional maximum likelihood estimates.
- The second approach treats the initial shocks, i.e. X_t for $t \leq 0$, are additional parameters of the model and estimate them jointly with other parameters.

  This approach is referred to as the exact likelihood method.

Conditional Maximum Likelihood Estimation

- For a general $ARMA(p,q)$ model,

$$a_t = (X_t - \mu) - \phi_1(X_{t-1} - \mu) - \cdots$$
$$-\phi_p(X_{t-p} - \mu) + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q}.$$

- Suppose that $\{a_t\}$ are IID $N(0, \sigma_a^2)$.
- $(a_1, \cdots, a_T)^T$ can be computed if $(a_{1-q}, \cdots, a_0)$, $(X_1, \cdots, X_T)^T$, and $(X_{1-q}, \cdots, X_0)$ are given.
- The joint pdf of $(a_1, \cdots, a_T)^T$ is

$$(2\pi\sigma_a^2)^{-T/2} \exp\left\{-\frac{1}{2\sigma_a^2}\sum_{t=1}^{T} a_t^2\right\}.$$

- The word **conditional** means that the estimator is conditional on the assumed initial values.
- Requires an optimization algorithm.

Unconditional Maximum Likelihood Estimation

- We should consider all the past squared a_t^2's.
- Since the past a_t^2's are unknown, we shall replace them by their expected values.
- Use the backcasting method to get those expected values.
- Both conditional and unconditional likelihood functions are only approximations. The exact likelihood function of a general ARMA model is complicated.

Diagnostic Checking

Objectives:

- Assess model adequacy by checking whether the model assumptions are satisfied.
- If the model is inadequate, this stage will provide some information for us to re-identify the model.

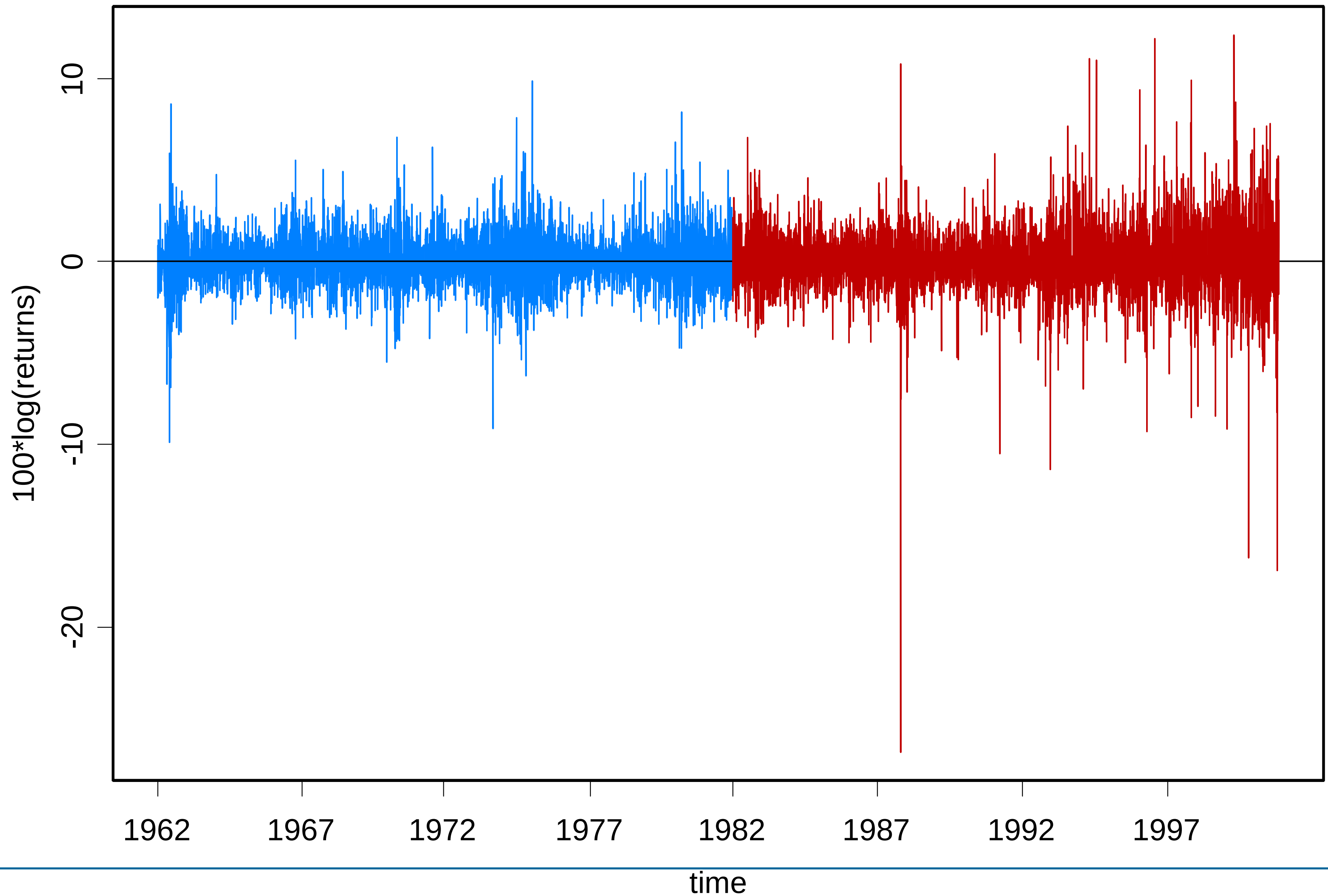
Checking Model Assumptions

- The basic assumption of an ARIMA model is that $a_t \sim IID\ N(0, \sigma_a^2)$.
- If the fitted model is adequate to represent the data, the fitted residuals $\{\hat{a}_t\}$ should have the same assumed behavior as the theoretical innovations $\{a_t\}$.
- Checking normality of the residuals
  (a) plotting histogram of $\hat{a}_t$
  (b) normal probability plot of $\hat{a}_t$
- Checking constant variance assumption
  (a) plotting the standardized residuals $\hat{a}_t/\hat{\sigma}_t$ with error bounds
  (b) Evaluate the effect of different λ values

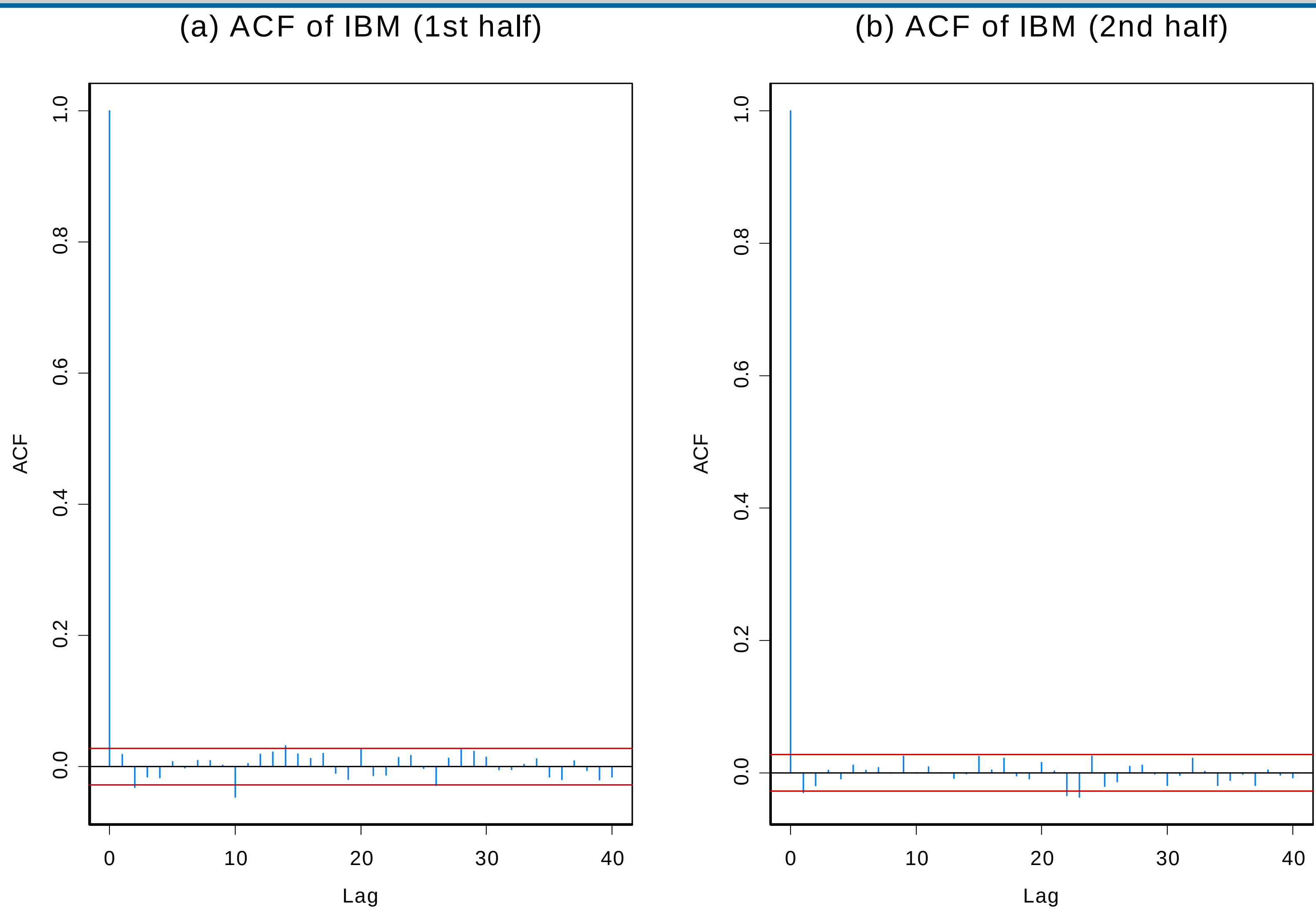
of the power transformation via Box-Cox method

- Checking independence assumption among residuals
  - SACF and SPACF of the residuals should be clean (i.e., all insignificant)
  - Portmanteau Test: Refer to Example 2.2.

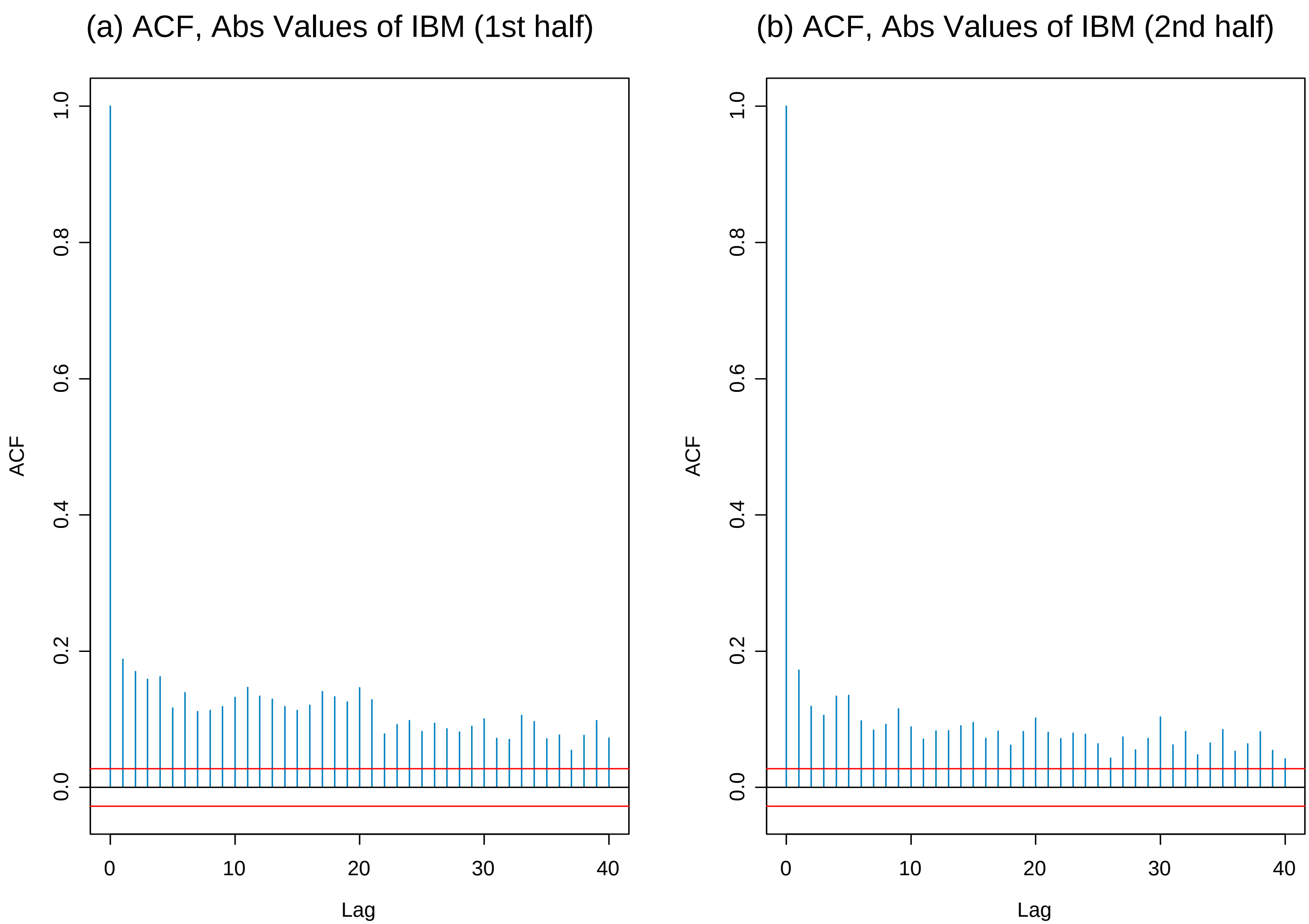
Log returns for IBM 1/3/62-11/3/00 (blue=1961-1981)



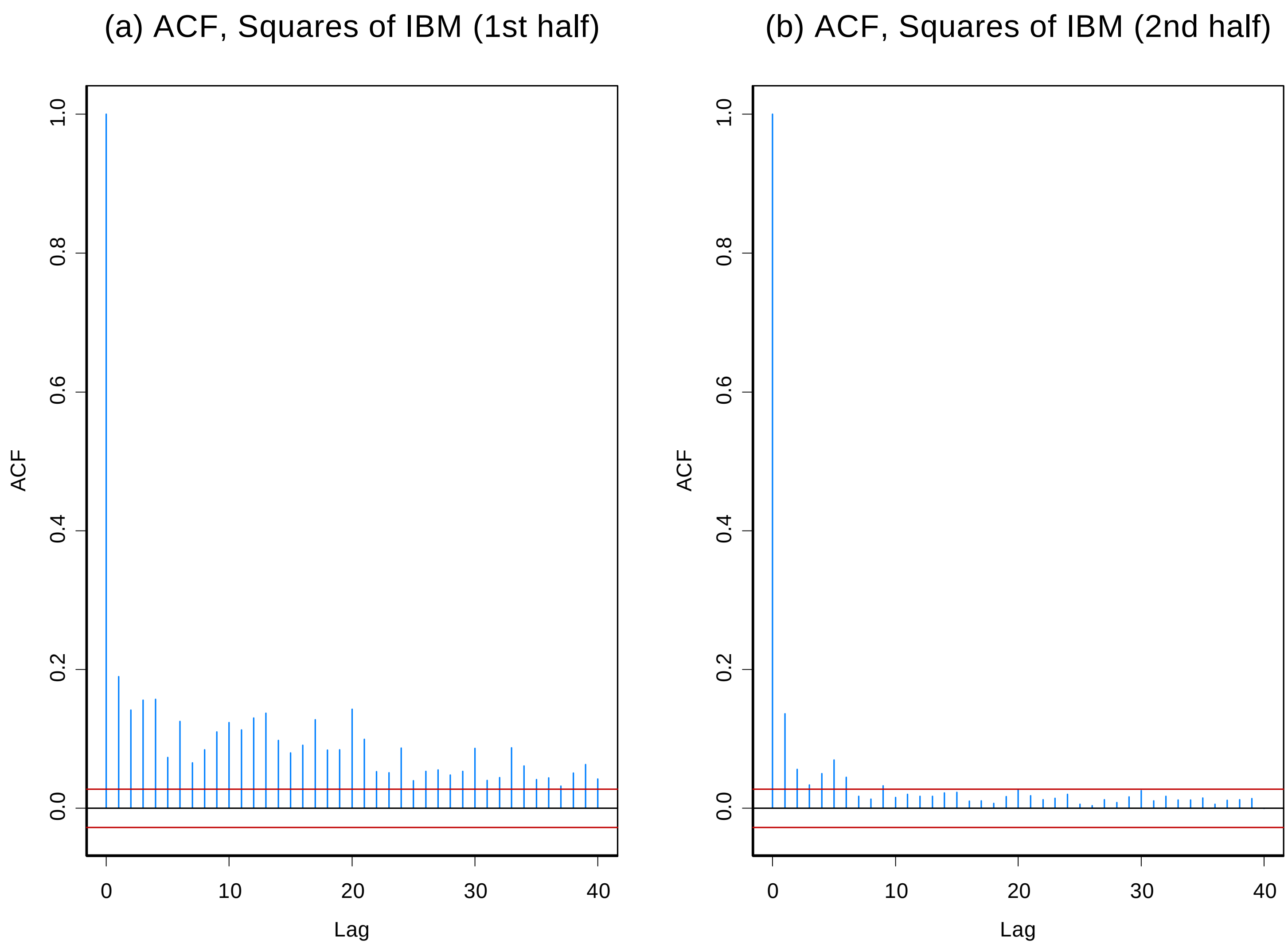
Sample ACF IBM (a) 1962-1981, (b) 1982-2000



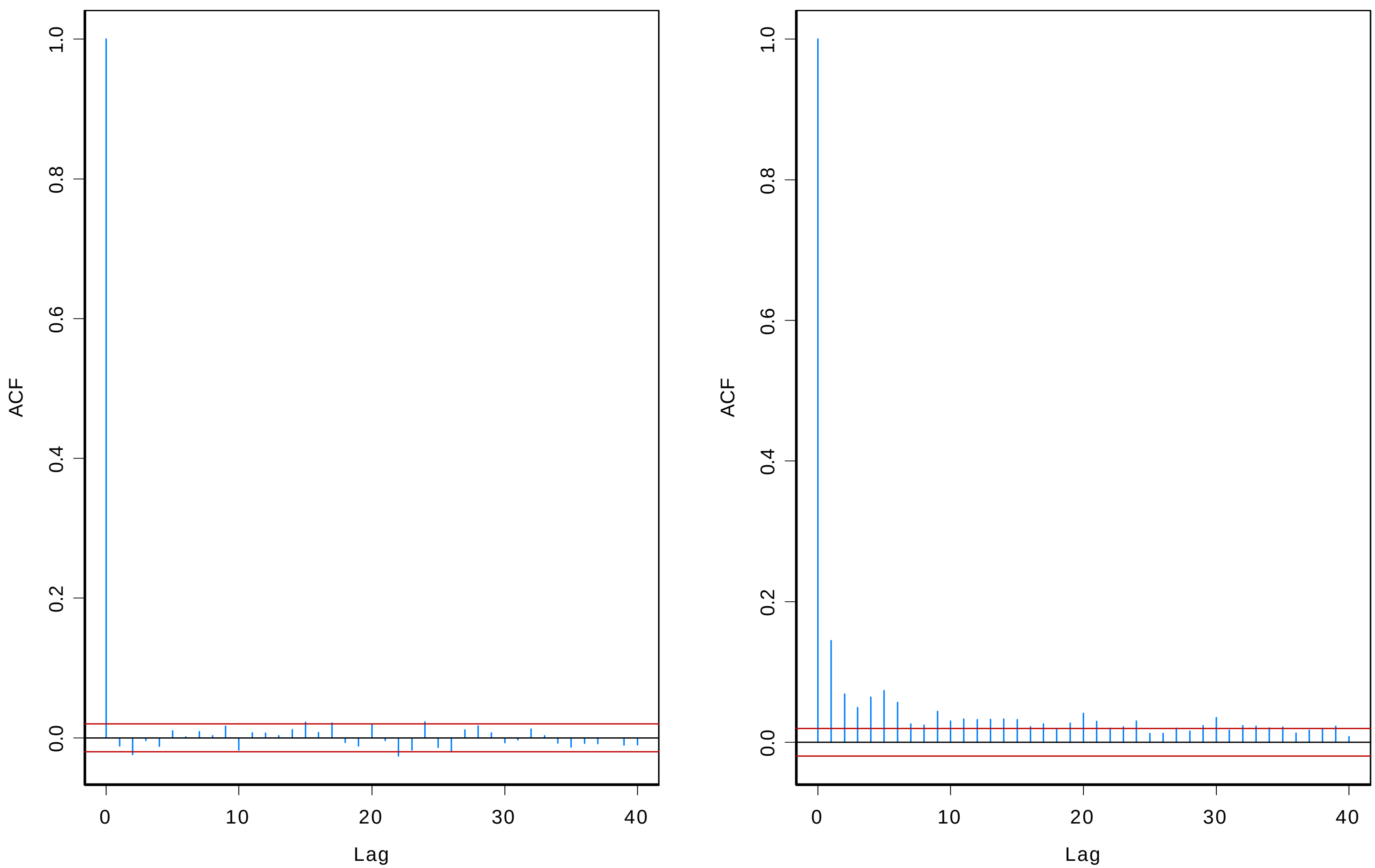
Sample ACF of abs values for IBM (a) 1961-1981, (b) 1982-2000



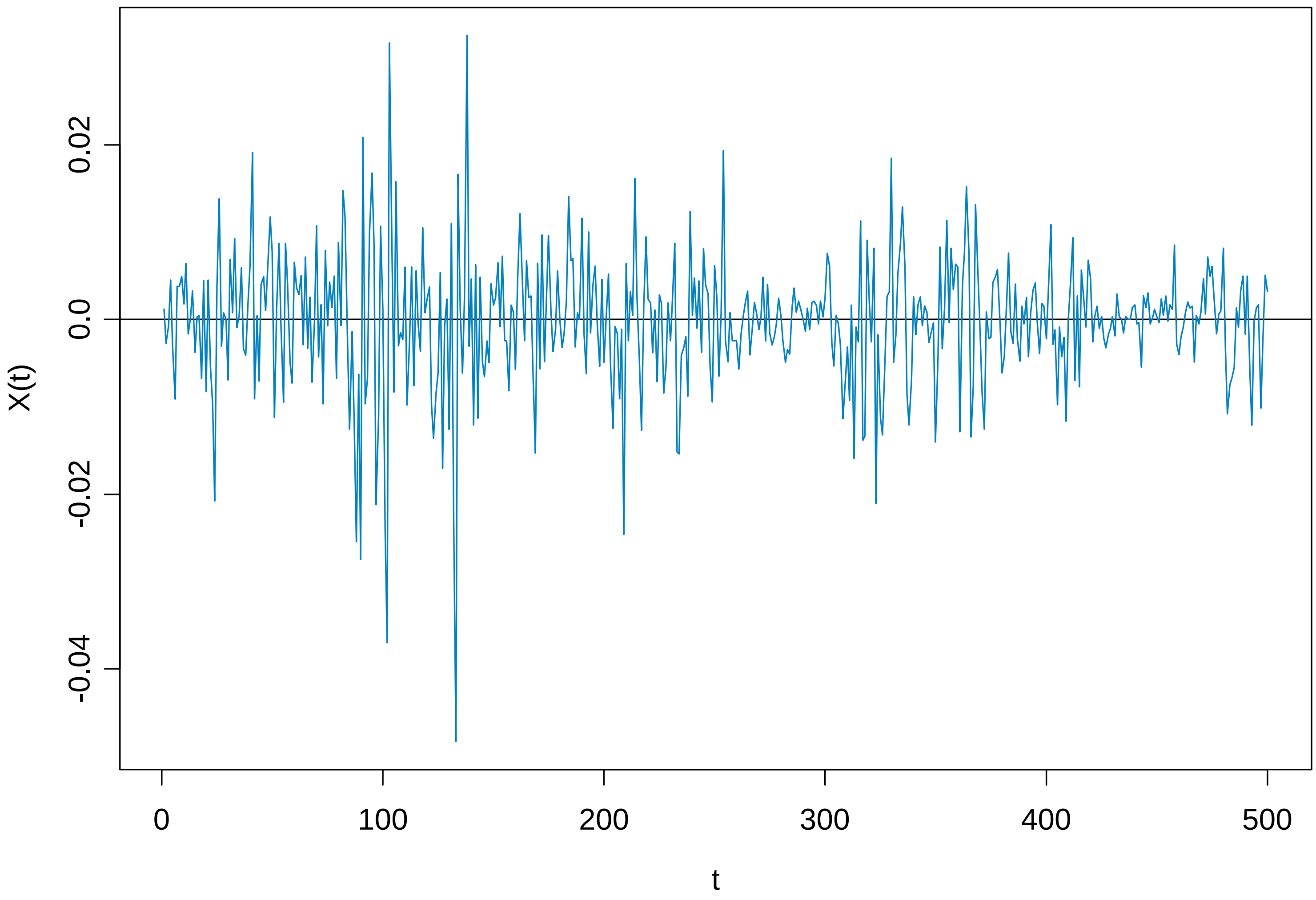
Sample ACF of squares for IBM (a) 1961-1981, (b) 1982-2000



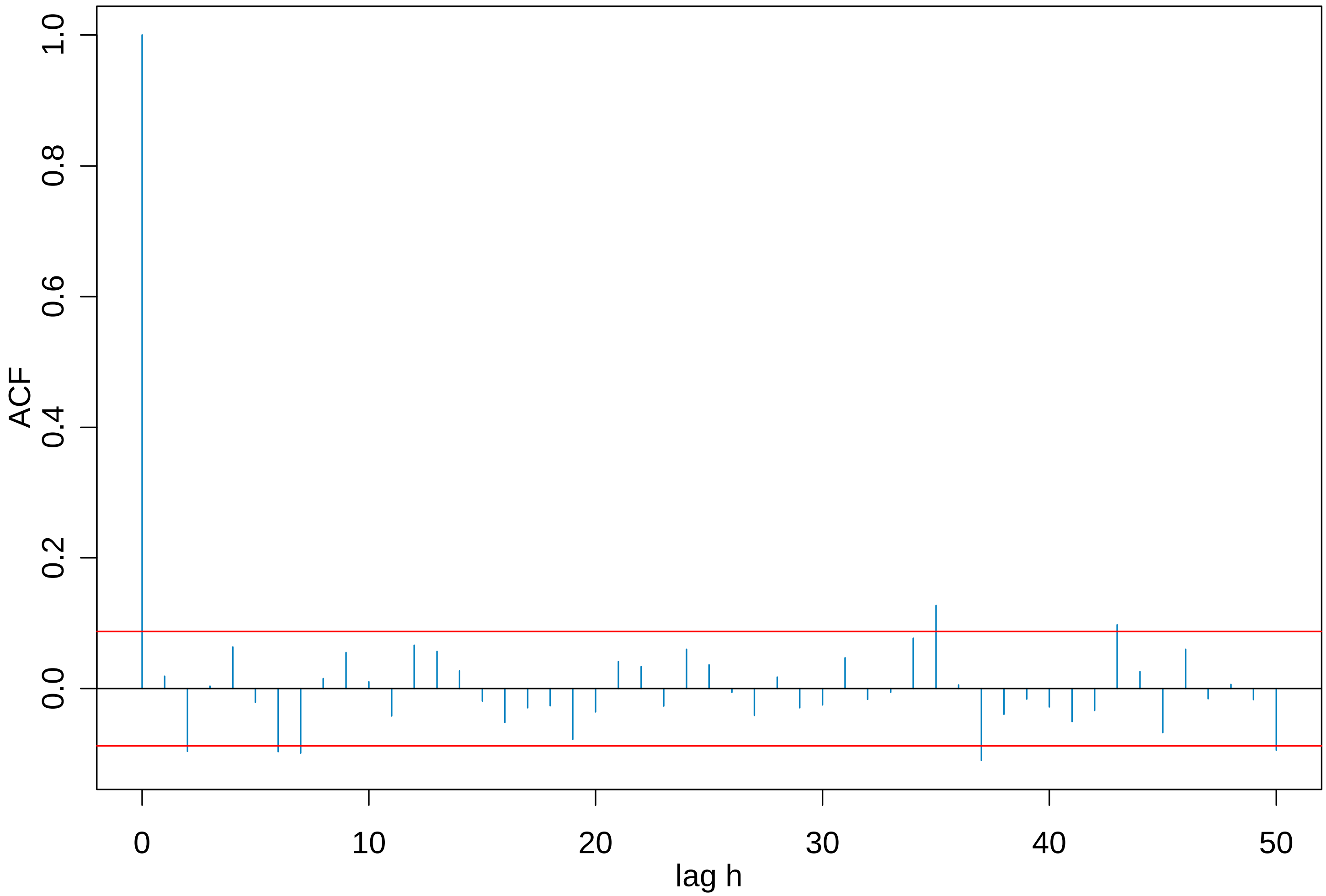
Sample ACF of original data and squares for IBM 1962-2000



500-daily log-returns of NZ/US exchange rate



ACF of X(t)=log-returns of NZ/US exchange rate



ACF of $X^2(t)$