Chapter Two

Linear Time Series Analysis

Tingguo Zheng (zhengtg@gmail.com)
School of Economics & WISE, Xiamen University

(For SOE & WISE Msc and PhD Program, 2022-2023, Autumn Semester)

- **Reference:**
  - "Analysis of Financial Time Series (3ed)" by Ruey S. Tsay, 2010. (Chapter 2)

- **Outline:**
  - Data handling
  - ARIMA Model Building
  - ARFIMA model
  - Seasonal ARIMA model$*$
  - Generalized ARMA model

Note:

If you are not clear about the basic ARMA model, please read the textbooks such as Hamilton (1994) and Tsay (2010). I also provide you a supplement of basic ARMA model for further reading.

1. Data Handling

1.1 Data Sources

- Book Website.

  http://faculty.chicagobooth.edu/ruey.tsay/teaching/fts3/

- Yahoo Finance.

  http://finance.yahoo.com/

- Investing (外汇、股票、股市、石油、贵金属等).

  http://cn.investing.com/

- CEIC中国经济数据库.

  http://webcdm.ceicdata.com/

- 锐思（RESSET）金融研究数据库.
  http://www.resset.cn/

- 国泰安数据库.
  http://www.gtarsc.com/

- WIND咨讯金融终端.
  http://www.wind.com.cn/

- 东方财富Choice终端.
  http://data.eastmoney.com/

- Python + FRED (Federal Reserve Bank of St. Louis.)

  http://research.stlouisfed.org/fred2/

  `import pandas_datareader.data as web`

- Python + Tushare金融大数据.

  https://tushare.pro/

  `import tushare as ts`

- Python + web html

1.2 Data transformations

If the data is not normally distributed, it is often possible to normalize it by data transformation.

- Logarithmic transformation: $y_t = \log(x_t)$ or $y_t = \ln(x_t)$;
  For example, take a log of seasonal unadjusted GDP; take a log of realized volatility.

- Box-Cox transformation (Box and Cox, 1964)

$$y_t = h(x_t, \lambda) = \begin{cases} (x_t^{\lambda} - 1)/\lambda, & \text{if } \lambda \neq 0; \\ \log x_t, & \text{if } \lambda = 0. \end{cases} \tag{1}$$

  for $x_t > 0$, $t = 1, \ldots, n$.

- Variance stabilizing transformation.

By simple calculations for the exponential family distributions listed in Table 1.

Table 1 The variance stabilizing transformation of some exponential family distributions

	Normal	Lognormal	Gamma	Beta	Poisson	Binomial
Parameter	$(\mu, \sigma^2)'$	$(\mu, \sigma^2)'$	$(\alpha, \alpha/\mu)'$	$(\tau\mu, \tau(1-\mu))'$	μ	μ/n
$h(y)$	y	$\ln y$	$\ln y$	$\arcsin\sqrt{y}$	$\sqrt{y}$	$\arcsin\sqrt{y/n}$
$\mathrm{Var}[h(y)]$	σ^2	σ^2	$\psi_1(\alpha)$	$\approx 1/4(1+\tau)$	$\approx 1/4$	$\approx 1/4n$

Note: $\psi()$ and $\psi_1()$ denote the digamma and trigamma functions, respectively.

1.3 Detrending (Trend/Cycle decomposition)

Assume that a variable y_t can be decomposed as:

$$y_t = T_t + c_t,$$

where T_t and c_t denote trend component and cyclical component.

This is often used for decomposizing the GDP and unemployment rate.

Five univariate methods are commonly used to extract cyclical components in time series.

- Linear trend filter or quadratic trend filter.
- BN (Beveridge-Nelson, 1981) decomposition.
- HP (Hodrick-Prescott, 1981) filter.
- Frequency filtering techniques, for example BK (Baxter-King, 1999) filter, CF (Christiano-Fitzgerald, 2003) filter.
- Unobservable components models, for example Harvey (1985), Clark (1987), Harvey and Jaeger (1993).

See 郑挺国和王霞(《经济研究》2010年第10期)

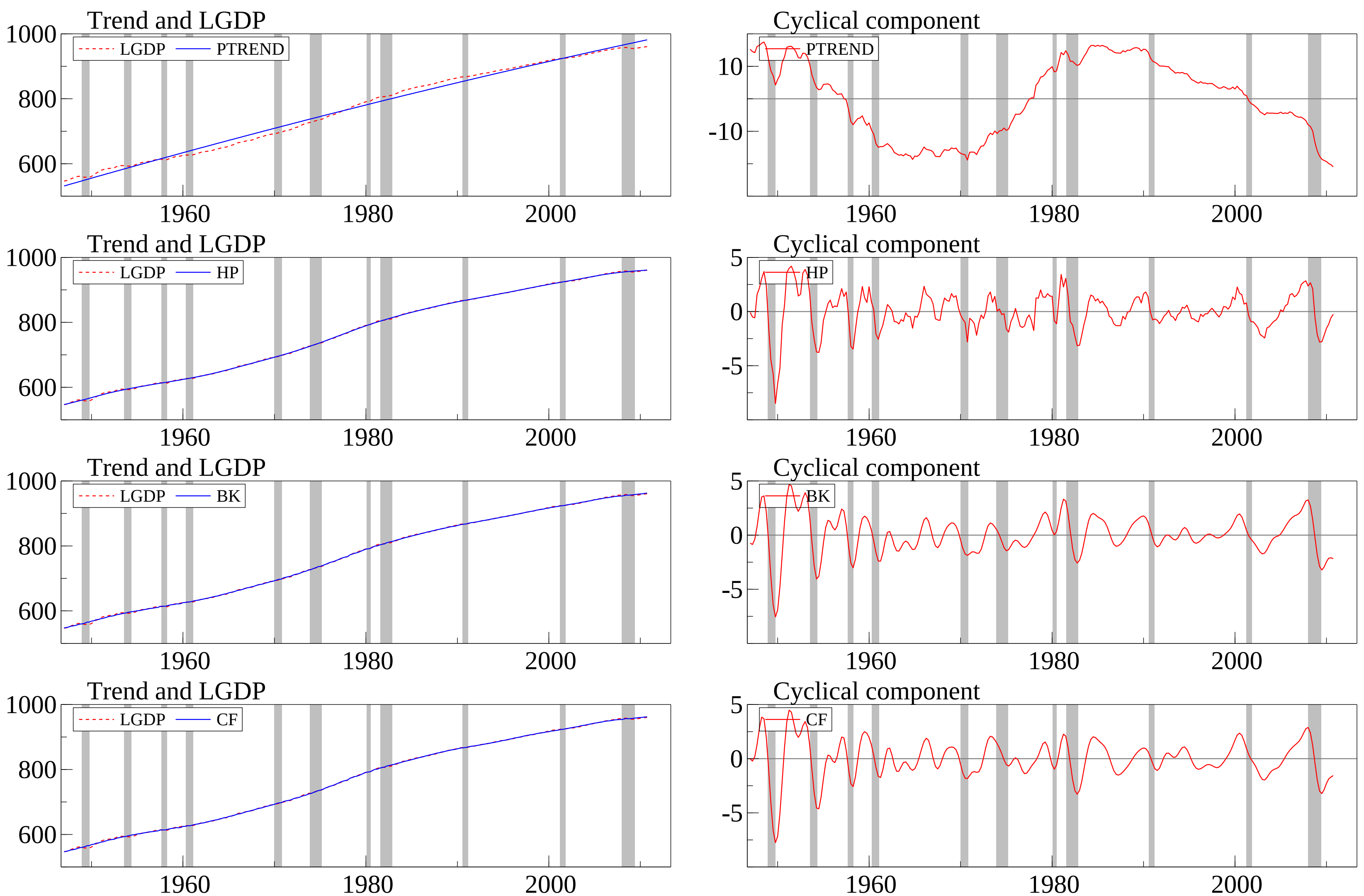


Figure 1 Trend/Cycle decomposition for the U.S. GDP (in logarithm $\times 100$) using quadratic filter, HP filter, BK filter, and CF filter.

The Hodrick – Prescott Filter

One popular example is the ad hoc trend extraction filter proposed by Hodrick and Prescott (1981$\rightarrow$1997).

HP define the estimator of the trend as the minimiser of the penalised least square criterion:

$$PLS = \sum_{t=1}^{T}(y_t - \mu_t)^2 + \lambda \sum_{t=3}^{T}(\Delta^2 \mu_t)^2,$$

where the first summand measures fidelity (精确度) and the second roughness (粗糙度); λ is the smoothness parameter governing the trade-off between them ($\lambda = 1600$ for quarterly series; $= 14400$ for monthly series).

Unobservable components models

A classic trend/cycle decomposition model is given as follows:

$$y_t = \tau_t + c_t \tag{2}$$

$$\tau_t = \tau_{t-1} + g_{t-1} + u_t \tag{3}$$

$$g_t = gc + \lambda g_{t-1} + w_t, \tag{4}$$

$$c_t = \phi_1 c_{t-1} + \phi_2 c_{t-2} + v_t \tag{5}$$

where y_t is log GDP; τ_t is its *trend component* follows a *random walk* with a *stochastic drift* or growth rate which is an autoregressive process; c_t is the *cycle component*.

When $\lambda = \sigma_w = 0$, this becomes the Watson (1986) model.

When $gc = 0$ and $\lambda = 1$, this becomes the Clark (1987) model.

1.4 Lag operator and Differencing

- Lag operator (L) or backshift operator (B). For example, given some time series $\{y_1, y_2, \ldots\}$, then

$$Ly_t = y_{t-1} \quad \text{for all } t > 1.$$

- Lag operator can be raised to arbitrary integer powers so that

$$L^{-1}y_t = y_{t+1}, \quad \text{and} \quad L^k y_t = y_{t-k}.$$

- First difference operator $\Delta = (1 - L)$ is a special case of lag polynomial. Then $\Delta y_t = y_t - y_{t-1} = (1 - L)y_t$.

For difference operator, we have the following results:

- The first difference operator: $\Delta y_t = (1 - L)y_t$;
- The second difference operator: $\Delta^2 y_t = (1-L)^2 y_t = (1-2L+L^2)y_t = y_t - 2y_{t-1} + y_{t-2}$;
- The above approach generalises to the i-th difference operator $\Delta^i y_t = (1 - L)^i y_t$;
- Log-First-Difference: $\Delta \ln y_t = \ln y_t - \ln y_{t-1}$.

Now if consider the change between y_t and y_{t-k}, we can use notation of difference operator by $\Delta_k y_t = (1 - L^k)y_t$. Similarly, we have the following results:

- The seasonal difference operator: $\Delta_{12}y_t = (1-L^{12})y_t$ if monthly data; $\Delta_4 y_t = (1 - L^4)y_t$ if quarterly data;

- Year over year change: $\Delta_{12} \ln y_t = (1 - L^{12}) \ln y_t = \ln(y_t/y_{t-12})$ if y_t is an aggregated monthly time series.

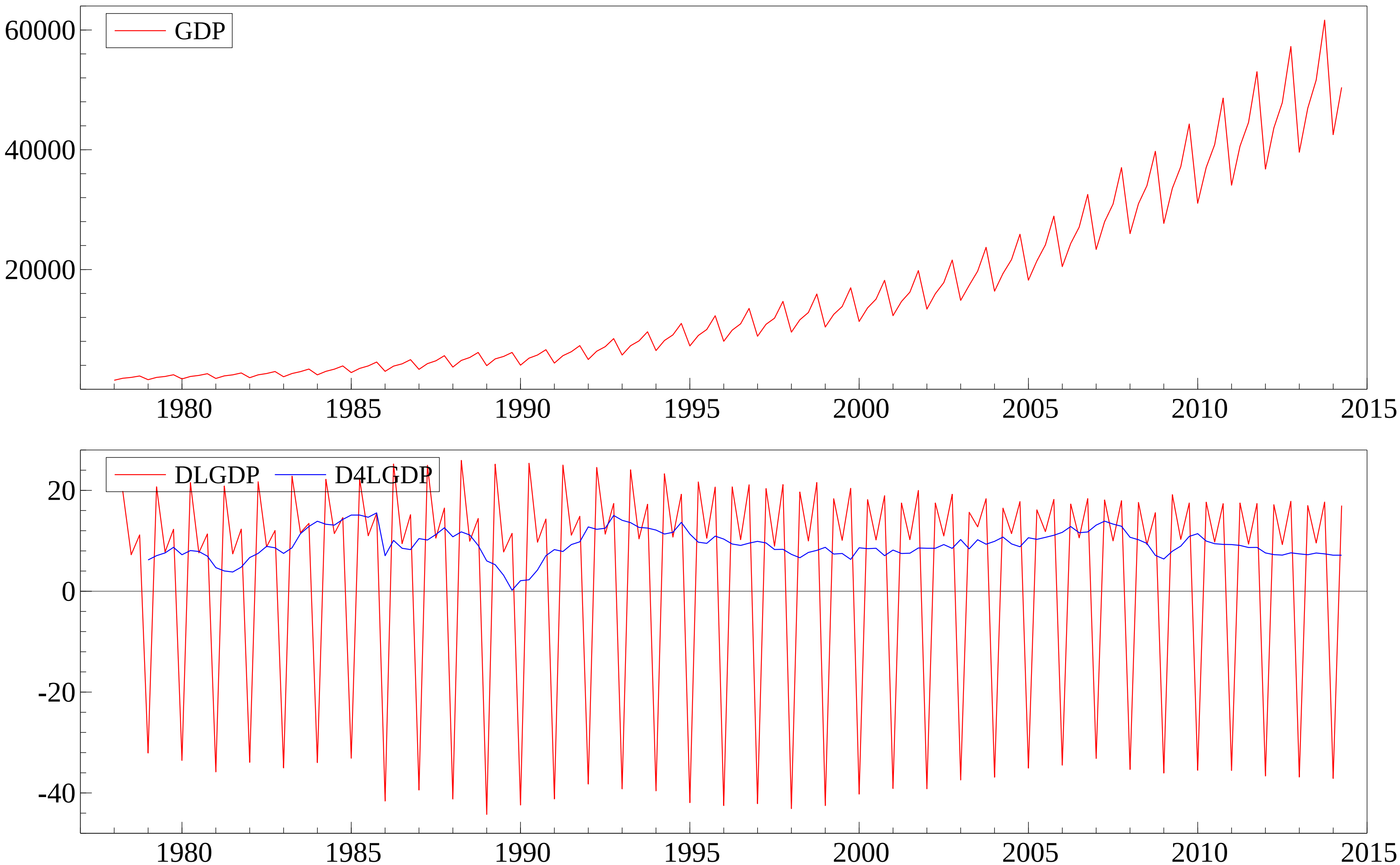


Figure 2 Comparison of $\Delta \log(GDP_t)$ and $\Delta_4 \log(GDP_t)$ for China's quarterly GDP.

1.5 Temporal aggregation

We consider temporal aggregation of a variable with quarterly and monthly growth rates.

- If y_t^q is a quarterly stock variable, it is easily shown that

$$y_t^q = y_t^m + y_{t-1}^m + y_{t-2}^m.$$

- If y_t^q is a quarterly flow variable, we have

$$y_t^q = \frac{1}{3}\left(y_t^m + 2y_{t-1}^m + 3y_{t-2}^m + 2y_{t-3}^m + y_{t-4}^m\right).$$

- See Mariano and Murasawa (2003) and 郑挺国和王霞 (2013)

1.6 Seasonality and seasonal adjustment

The investigation of many economic time series becomes problematic due to seasonal fluctuations.

Time series are made up of four components:

- S_t: The seasonal component;
- T_t: The trend component;
- C_t: The cyclical component;
- I_t: The error, or irregular component.

Different statistical research groups have developed different methods of seasonal adjustment, for example

- *X-11-ARIMA, X-12-ARIMA* and *X-13-ARIMA-SEATS* in many softwares such as SAS, Eviews, OxMetrics, R and Python developed by the United States Census Bureau (美国人口调查局);

- *TRAMO/SEATS* in Eviews developed by the Bank of Spain;

- *STAMP* in OxMetrics developed by a group led by S. J. Koopman.

- *STL* in R and Python developed by Cleveland, et al., 1990, “STL: A Seasonal-Trend Decomposition Procedure Based on Loess”, Journal of Official Statistics, 6(1), 3-73.

BSM model in STAMP*

The basic structural model (BSM) is formulated in terms of trend, seasonal, and irregular components. It can be represented by

$$y_t = \tau_t + \gamma_t + \varepsilon_t, \quad \varepsilon_t \sim iidN(0, \sigma_\varepsilon^2), \tag{6}$$

where τ_t is a stochastic trend, γ_t is a stochastic seasonal component, and ε_t is an irregular component; see Harvey (1989).

The trend is specified in the following way:

$$\tau_t = \tau_{t-1} + \beta_{t-1} + \eta_t, \quad \eta_t \sim iidN(0, \sigma_\eta^2) \tag{7}$$

$$\beta_t = \beta_{t-1} + \zeta_t, \quad \zeta_t \sim iidN(0, \sigma_\zeta^2), \tag{8}$$

where β_t is the slope, η_t and ζ_t are assumed to be mutually independent.

The seasonal component γ_t is often specified in two ways such as dummy-variable seasonality and trigonometric seasonality, see Harvey et al. (1998) and Commandeur and Koopman(2007).

A time-varying dummy seasonal specification is given by

$$\sum_{i=0}^{s-1} \gamma_{t-i} = e_t, \tag{9}$$

where s is the seasonal length, e_t is an i.i.d. normal variable with mean zero and variance σ_e^2, i.e., $e_t \sim iidN(0, \sigma_e^2)$.

- In the limiting case $\sigma_e^2 = 0$, the seasonal effects are fixed over time.

- Commonly, $s = 4$ for quarterly data and $s = 12$ for monthly data.

2 ARIMA Model Building

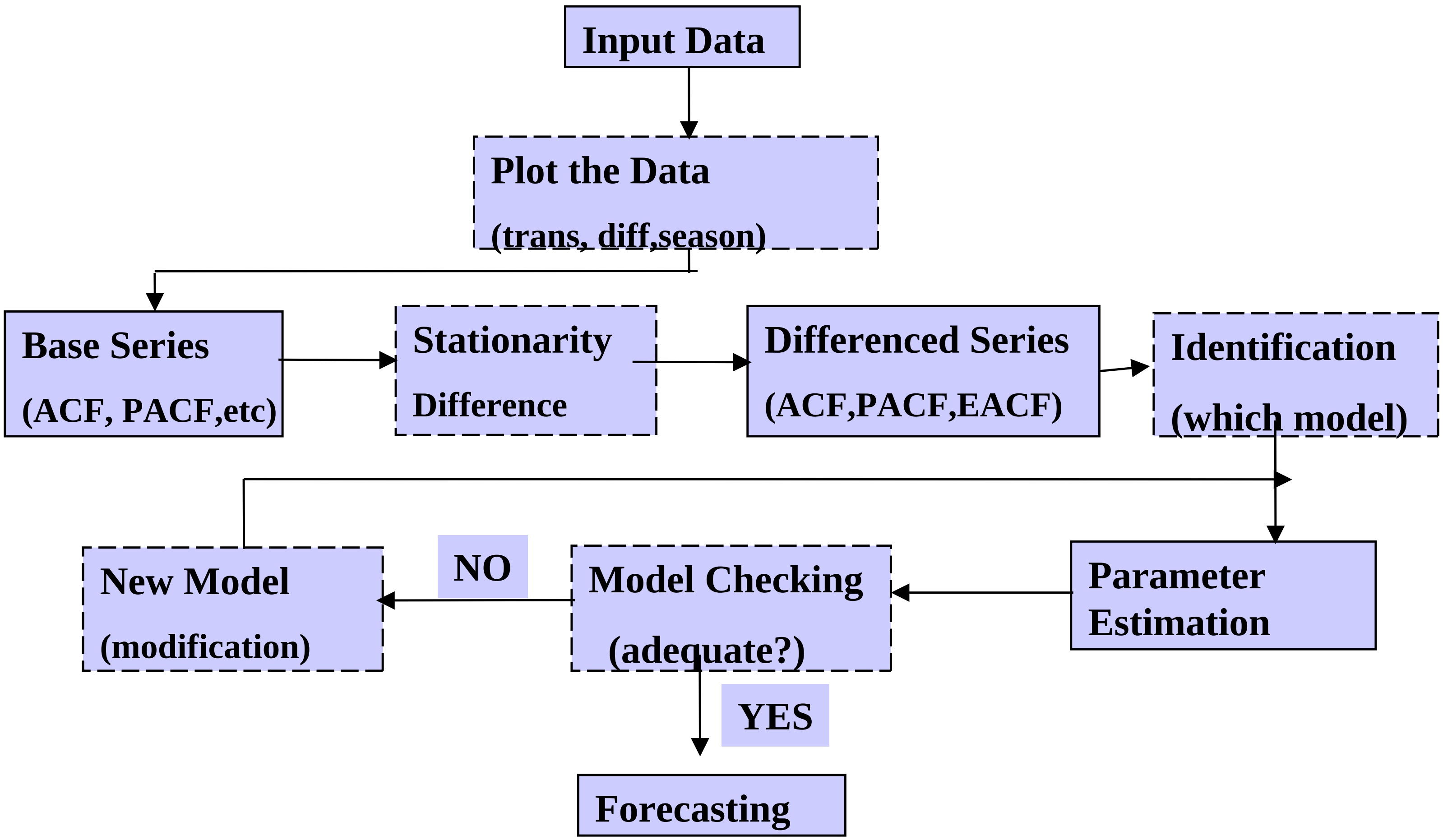
Detailed introduction of AR, MA and ARMA models are referred to the books in Hamilton (1994) and Tsay (2010). In this section, I will give you a short review for model building.

An effective procedure for building empirical time series models (ARIMA) is the Box-Jenkins approach, which consists of three stages:

- model specification,
- estimation,
- diagnostics checking.

Forecasts follow directly from the form of fitted model.

For ARIMA models, the modelling sequence can be usually described as follow (see the Figure):



2.1 Model-specification

Correlation approach

The basic tools used in this approach of model specification include (a) sample autocorrelation function (ACF), (b) sample partial autocorrelation function (PACF), and (c) extended autocorrelation function (EACF). The function of these tools can be summarized as

Function	Model	Feature
ACF	MA(q)	Cutting-off at lag q, i.e.
		$\rho_i \neq 0$ for $i <= q$ and $\rho_i = 0$ for $i > q$
PACF	AR(p)	Cutting-off at lag p
EACF	ARMA(p, q)	A triangle with vertex (p, q)

(1) Extended Autocorrelation Function (EACF)

See Tsay and Tiao (1984, JASA)

Consider the ARMA(p, q) model given by

$$(1 - \phi_1 L - \cdots - \phi_p L^p)(y_t - \mu) := w_t = (1 + \theta_1 L + \cdots + \theta_q L^q)\varepsilon_t.$$

The correlation coefficient ρ_j satisfies that

$$\rho_j - \phi_1 \rho_{j-1} - \cdots - \phi_p \rho_{j-p} = 0, \quad \text{for } j > q,$$

then the ACF satisfy the difference equation $\phi(L)\rho_j = 0$ for $j > q$ with $\rho_1, \ldots, \rho_q$ as initial conditions.

Consider the above Eqs. of ACF for $j = q+1, q+2, \ldots, q+p$, we have

$$\begin{bmatrix} \rho_{q+1} \\ \rho_{q+2} \\ \vdots \\ \rho_{q+p} \end{bmatrix} = \begin{bmatrix} \rho_q & \rho_{q-1} & \cdots & \rho_{q+2-p} & \rho_{q+1-p} \\ \rho_{q+1} & \rho_q & \cdots & \rho_{q+3-p} & \rho_{q+2-p} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{q+p-1} & \rho_{q+p-2} & \cdots & \rho_{q+1} & \rho_q \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix},$$

which is referred to as a p-order *generalized Yule-Walker equation* for the ARMA(p, q) process. It can be used to solve for ϕ_i's given the ρ_i's.

Given the sample ACF, the AR coefficients can be consistently estimated in the following cases!

- The lags p and q are correctly specified.
- The lag q is greater than the true value q^* under the correct p.

Again,

$$\phi(L)y_t = x_t = c + \theta(L)\varepsilon_t$$

Its basic idea is based on the "generalized" Yule-Walker equation. Conceptually, it involves two steps:

- In the first step, obtain consistent estimates of AR coefficients.
  - Given such estimates, we can transform the ARMA series into a pure MA process.

- The second step then uses the sample ACF of the transformed MA process to identify the MA order q.

Again,

$$\phi(L)y_t = x_t = c + \theta(L)\varepsilon_t$$

To make use of the EACF for model specification, we consider the two-way table:

AR	MA (or j)					
m	0	1	2	3	4	$\cdots$
0	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	$\cdots$
1	$\rho_{1,1}$	$\rho_{1,2}$	$\rho_{1,3}$	$\rho_{1,4}$	$\rho_{1,5}$	$\cdots$
2	$\rho_{2,1}$	$\rho_{2,2}$	$\rho_{2,3}$	$\rho_{2,4}$	$\rho_{2,5}$	$\cdots$
3	$\rho_{3,1}$	$\rho_{3,2}$	$\rho_{3,3}$	$\rho_{3,4}$	$\rho_{3,5}$	$\cdots$
$\vdots$	$\vdots$			$\vdots$		

where $\rho_{m,j+1}$ denotes the ACF($j+1$) for the MA(j) process of x_t when considering the ARMA(m, j) for y_t.

In practice, the EACF in the above table is replaced by its sample counterpart. Suppose that z_t is an ARMA(1,1) model, then the corresponding EACF table is

AR	MA (or j)						
m	0	1	2	3	4	5	$\cdots$
0	X	X	X	X	X	X	$\cdots$
1	X	O	O	O	O	O	$\cdots$
2	*	X	O	O	O	O	$\cdots$
3	*	*	X	O	O	O	$\cdots$
4	*	*	*	X	O	O	

where "X" and "O" denote non-zero and zero quantities, respectively, "*" represents a quantity which can assume any value between -1 and 1. Note that:

$$\text{ARMA(1,1):} \quad (1-\phi_1 L)y_t = (1+\theta_1 L)\varepsilon_t$$

$$\Longrightarrow \text{ARMA(2,2):} \quad (1-\phi_1 L)(1-\phi_2 L)y_t = (1+\theta_1 L)(1-\phi_2 L)\varepsilon_t.$$

```
[1] "EACF table"
           [,1]        [,2]        [,3]        [,4]         [,5]
[1,]  0.7851208  0.759087202  0.73012581  0.717539019  0.702734511
[2,] -0.4367669  0.009058918 -0.03741167  0.006104493  0.009847222
[3,] -0.4225225 -0.124097181 -0.04051274  0.013159662  0.014754117
[4,] -0.4927834 -0.150258522 -0.18880108  0.002506284 -0.006988006
[5,] -0.4729140  0.178555881 -0.17444239 -0.007166510 -0.011347496
[6,] -0.4307488 -0.284929820 -0.10113442 -0.048853062  0.219957080
[7,] -0.4654659  0.065633947 -0.16148225 -0.162236126  0.163174127
              [,6]        [,7]         [,8]
[1,]  0.6842451607  0.66754166  0.662508201
[2,] -0.0034095750 -0.02560240  0.006476825
[3,] -0.0018310784 -0.03179061  0.010449822
[4,]  0.0002235044 -0.02452108 -0.016740646
[5,] -0.0073935684 -0.01024074 -0.012380360
[6,]  0.0084169899 -0.01230994 -0.003858685
[7,]  0.1169535503 -0.01321116 -0.004599728
[1] " "
[1] "Simplified EACF: 2 denotes significance"
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,]    2    2    2    2    2    2    2    2
[2,]    2    0    2    0    0    0    0    0
[3,]    2    2    2    0    0    0    0    0
[4,]    2    2    2    0    0    0    0    0
[5,]    2    2    2    0    0    0    0    0
[6,]    2    2    2    2    2    0    0    0
[7,]    2    2    2    2    2    2    0    0
```

Figure 3 ESACF table for log realized volatility of S&P 500 index.

(2) Model Selection Criteria

Alternatively, statistical model selection criteria may be used. The idea is to fit all ARMA(p, q) models with orders $p \leq p_{\max}$ and $q \leq q_{\max}$ and choose the values of p and q which minimizes some *model selection criteria*.

Model selection criteria for ARMA(p, q) models have the form

$$\text{MSC}(p, q) = \ln(\tilde{\sigma}^2(p, q)) + c_n \cdot \psi(p, q)$$

where $\tilde{\sigma}^2(p, q)$ is the MLE of $\mathrm{Var}(\varepsilon_t) = \sigma^2$ without a degrees of freedom correction from the ARMA(p, q) model, c_n is a sequence indexed by the sample size n, and $\psi(p, q)$ is a *penalty function* which penalizes large ARMA(p, q) models.

The two most common *information criteria* are the Akaike (AIC) and Schwarz- Bayesian (BIC): (Akaike, 1974; Schwarz, 1978)

$$\begin{aligned} \text{AIC}(p,q) &= \ln(\tilde{\sigma}^2(p,q)) + \frac{2}{n}(p+q), \\ \text{BIC}(p,q) &= \ln(\tilde{\sigma}^2(p,q)) + \frac{\ln n}{n}(p+q). \end{aligned}$$

- The AIC criterion asymptotically overestimates the order, since the penalty term is smaller for AIC than BIC ($2 < \ln n$).

- The BIC estimates the order consistently under fairly general conditions if the true orders p and q are less than or equal to $p_{\max}$ and $q_{\max}$.

- However, in finite samples the BIC generally shares no particular advantage over the AIC.

Example 1. AR model of U.S. $\triangle$inflation, lags 0 - 6, see Table 2

Table 2 Information criteria

#Lags	BIC	AIC	R^2
0	1.095	1.076	0.000
1	1.067	1.030	0.056
2	**0.955**	0.900	0.181
3	0.957	**0.884**	0.203
4	0.986	0.895	0.204
5	1.016	0.906	0.204
6	1.046	0.918	**0.204**

- BIC chooses 2 lags, AIC chooses 3 lags.
- For R^2, you would (always) select the largest possible number of lags.

2.2 Estimation methods

Several estimation method are available for AR models and ARMA models. We do not expand this.

- AR models: Original least square; Unconditional least square; Yule-Walker estimator; Conditional maximum likelihood estimator; Maximum likelihood estimator.

- ARMA models: Conditional least square; Unconditional least square; Conditional maximum likelihood estimator; Maximum likelihood estimator.

- MA models: the same as ARMA models

3.3 Diagnostic testing on residuals

(1) Autocorrelation

1. Sample ACF of the residuals, given by

$$r_k(\hat{\varepsilon}) = \frac{\sum_{t=k+1}^{n} \hat{\varepsilon}_t \hat{\varepsilon}_{t-k}}{\sum_{t=k+1}^{n} \hat{\varepsilon}_t^2} \sim N(0, 1/n), \tag{10}$$

2. Joint significance of the first m residual autocorrelations. The test-statistic developed by Ljung and Box (1978), given by (for an ARMA(p, q) model)

$$\mathrm{LB}(m) = n(n+2) \sum_{k=1}^{m} (n-k)^{-1} r_k^2(\hat{\varepsilon}) \sim \chi^2(m). \tag{11}$$

(2) Homoscedasticity

Neglecting heteroscedasticity of the residuals leads to

- ordinary t-statistics cannot be used.
- confidence intervals for forecasts can no longer be computed in the usual manner.

Davidson and MacKinnon (1985) and Wooldridge (1990, 1991) discuss general principles for constructing heteroscedasticity-consistent test statistics.

In standard statistical or econometric software, it provides heteroscedasticity-consistent t-statistic for parameter estimates.

Testing for heteroscedasticity

Portmanteau test statistic on (standardized) squared residuals developed by McLeod and Li (1983):

$$\mathrm{McL}(m) = n(n+2)\sum_{k=1}^{m}(n-k)^{-1}r_k^2(\hat{\varepsilon}^2). \tag{12}$$

When applied to the residuals from an ARMA(p, q) model, the McL test has an asymptotic $\chi^2(m)$ distribution, again provided that m/n is small and m is moderately large.

(3) Normality

Defining the jth moment of the estimated (standardized) residuals as

$$\hat{m}_j = \frac{1}{n}\sum_{t=1}^{n} \hat{\varepsilon}_t^j,$$

the skewness and kurtosis of $\hat{\varepsilon}_t$ can be calculated as

$$\widehat{SK}_{\hat{\varepsilon}} = \frac{\hat{m}_3}{\sqrt{\hat{m}_2^3}}, \quad \text{and} \quad \widehat{K}_{\hat{\varepsilon}} = \frac{\hat{m}_4}{\hat{m}_2^2}.$$

Under the null hypothesis of normality, $\sqrt{n/6} \cdot \widehat{SK}_{\hat{\varepsilon}} \sim N(0,1)$ and $\sqrt{n/24} \cdot (\widehat{K}_{\hat{\varepsilon}} - 3) \sim N(0,1)$.

A joint test for normality (Jarque and Bera, 1987) is then given by

$$\text{JB} = \frac{n}{6}\widehat{SK}_{\hat{\varepsilon}}^2 + \frac{n}{24}(\widehat{K}_{\hat{\varepsilon}} - 3)^2 \sim \chi^2(2). \qquad (13)$$

3. Long Memory Time Series

- $\{y_t\}$ is $I(0)$: its ACF declines at a *geometric rate*. As a result, $I(0)$ process have *short memory* since observations far apart in time are essentially independent.

- $\{y_t\}$ is $I(1)$: its ACF declines at a *linear rate* and observations far apart in time are not independent. See the supplement of basic ARMA model.

- $\{y_t\}$ is *fractionally integrated* $I(d)$, where $0 < d < 1$: The ACF declines at a *polynomial (hyperbolic) rate*, which implies that observations far apart in time may exhibit *weak but non-zero correlation*. This weak correlation between observations far apart is often referred to as *long memory*.

(1) Fractionally integrated process

A fractionally integrated white noise process y_t has the form

$$(1-L)^d y_t = \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma^2), \tag{14}$$

where the *differencing operator* $(1-L)^d$ has the binomial series expansion representation (valid for any $d > -1$)

$$(1-L)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-L)^k, \tag{15}$$

where the binomial coefficients $\binom{d}{k}$ are defined by

$$\binom{d}{k} = \frac{d(d-1)(d-2)\cdots(d-k+1)}{k!},$$

and we have

$$(-1)^k \binom{d}{k} = \frac{-d(1-d)(2-d)\cdots(k-1-d)}{k!} = \frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)}$$

since $\Gamma(t+1) = t\Gamma(t)$ and $k!$ (read as k factorial) is

$$k! = \Gamma(k+1) = 1 \cdot 2 \cdot 3 \cdots (k-1) \cdot k, \quad \text{with } 0! = 1.$$

Using these definitions, we can rewrite (11) as

$$\begin{aligned}(1-L)^d &= 1 - dL + \frac{d(1-d)}{2!}L^2 - \frac{d(1-d)(2-d)}{3!}L^3 + \cdots \\ &= \sum_{k=0}^{\infty} \frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)} L^k.\end{aligned}$$

- If $d = 1$ then y_t is a random walk and if $d = 0$ then y_t is white noise.

- If $d < 0.5$, then y_t is a weakly stationary process and has the infinite MA representation

$$y_t = \varepsilon_t + \sum_{i=1}^{\infty} \psi_i \varepsilon_{t-i}$$

with

$$\psi_k = \frac{d(1+d)\cdots(k-1+d)}{k!} = \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)}.$$

- If $d > -0.5$, then y_t is invertible and has the infinite AR representation

$$y_t = \sum_{i=1}^{\infty} \pi_i y_{t-i} + \varepsilon_t$$

with

$$\pi_k = \frac{-d(1-d)\cdots(k-1-d)}{k!} = \frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)}.$$

- For $-0.5 < d < 0.5$ it can be shown that

$$\rho_k = \frac{d(1+d)\cdots(k-1+d)}{(1-d)(2-d)\cdots(k-d)} = \frac{\Gamma(k+d)}{\Gamma(1+k-d)}\frac{\Gamma(1-d)}{\Gamma(d)},$$

  for $k = 1, 2, \ldots$. In particular, $\rho_1 = d/(1-d)$ and

$$\rho_k \simeq \frac{\Gamma(1-d)}{\Gamma(d)} k^{2d-1} \propto k^{2d-1}$$

  as $k \to \infty$ so that the ACF for y_t declines *hyperbolically* to zero at a speed that depends on d.

- For $-0.5 < d < 0.5$, the PACF of y_t is $\phi_{k,k} = d/(k-d)$ for $k = 1, 2, \ldots$.

- Further, it can be shown y_t is stationary and ergodic for $0 < d < 0.5$ and that the variance of y_t is infinite for $0.5 \leq d < 1$.

- The process is said to exhibit intermediate memory (anti-persistence), or long-range negative dependence, for $d \in (-0.5, 0)$.

(2) Testing long memory

Geweke and Porter-Hudak (1983) proposed a semiparametric procedure to obtain an estimate of the memory parameter d of a fractionally integrated process X_t in a model of the form

$$(1-L)^d X_t = \varepsilon_t,$$

where ε_t is stationary with zero mean and continuous spectral density $f_\varepsilon(\lambda) > 0$.

The estimate $\hat{d}$ is obtained from the application of OLS to the regression

$$\log(I_X(\lambda_s)) = c - d\log\left|1 - e^{i\lambda_s}\right|^2 + \epsilon_s, \tag{16}$$

where the frequencies $\lambda_s = \frac{2\pi s}{n}$, $s = 1, \ldots, m$, $m \ll n$.

We define $\omega_X(\lambda_s) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} X_t e^{it\lambda_s}$ as the discrete Fourier transform (dft) of the time series X_t, $I_X(\lambda_s) = \omega_X(\lambda_s)\omega_X(\lambda_s)$ as the periodogram, and $x_s = \log \left|1 - e^{i\lambda_s}\right|$. Ordinary least squares on (12) yields

$$\hat{d} = 0.5 \frac{\sum_{s=1}^{g(T)} x_s \log I_X(\lambda_s)}{\sum_{s=1}^{g(T)} x_s^2}.$$

- A choice of $m = g(T) = \sqrt{T}$ is often employed.

- Two estimates of the d coefficient's standard error are commonly employed:

  - the regression standard error, giving rise to a standard t-test;
  - an asymptotic standard error, based upon the theoretical variance of the log periodogram of $\pi^2/6$.

(3) ARFIMA(p, d, q) model

A fractionally integrated process with stationary and ergodic ARMA(p, q) errors

$$(1 - L)^d y_t = u_t, \quad u_t \sim ARMA(p, q)$$

is called an *autoregressive fractionally integrated moving average* (ARFIMA) process, i.e.

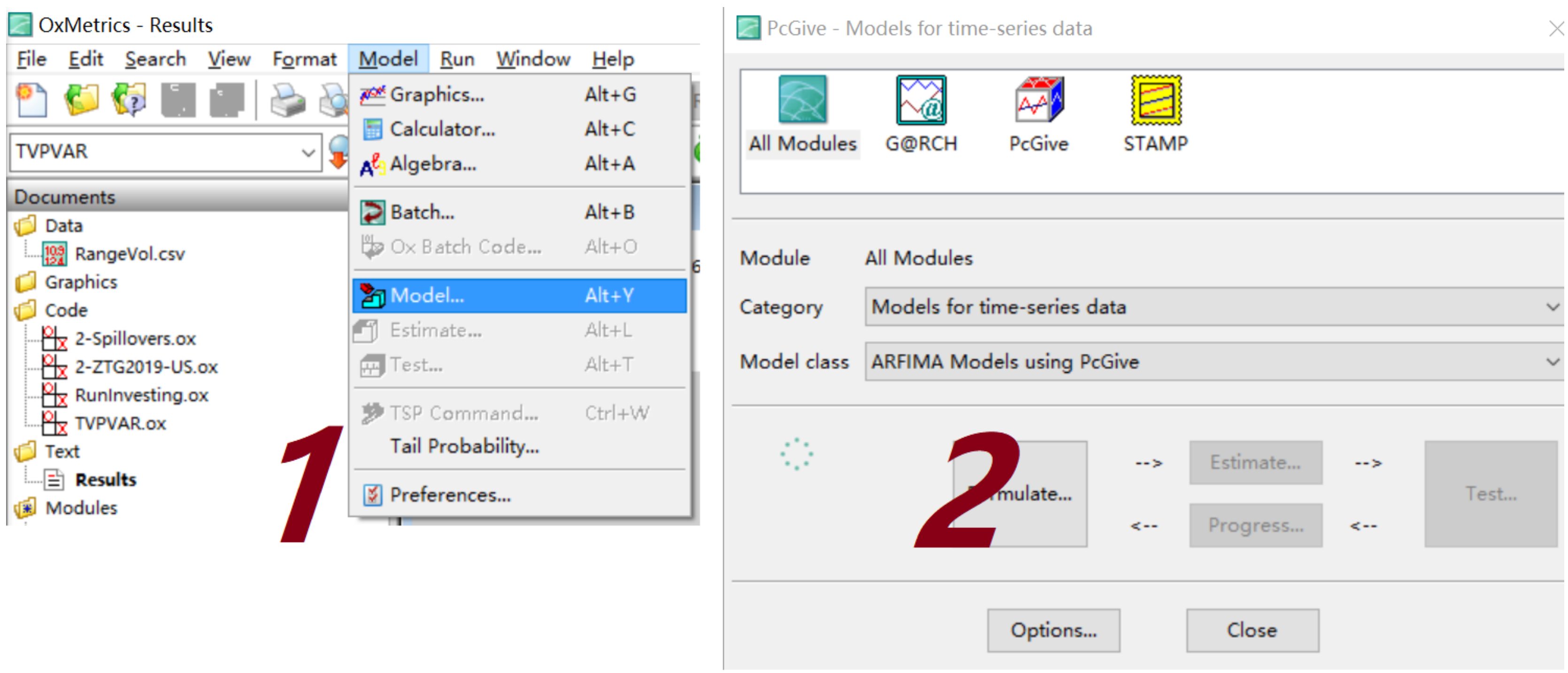
$$\phi(L)(1 - L)^d y_t = \theta(L)\varepsilon_t.$$

Estimation routines for general ARFIMA models, which include additional AR and MA parts in (10) are proposed in Sowell (1992) and Beran (1995).

Reading

The readers are recommend to refer to ”A Package for Estimating, Forecasting and Simulating Arfima Models: Arfima package 1.04 for Ox” by JURGEN A. DOORNIK AND MARIUS OOMS.

We can use this model in OxMetrics to forecast empirical time series.



4 Seasonal ARIMA*

4.1 Pure seasonal ARMA

Let us assume that there is seasonality in the data, but no trend. Then we could model the data as

$$x_t = s_t + y_t, \tag{17}$$

where y_t is a stationary process. The seasonal component is such that

$$s_t = s_{t-h},$$

where h denotes the length of the period and $\sum_{k=1}^{h} s_k = 0$.

By differencing at lag h, we can remove the seasonality from the data

$$\Delta_h x_t = x_t - x_{t-h} = x_t - L^h x_t = (1 - L^h) x_t$$
$$= s_t + y_t - s_{t-h} - y_{t-h} = \Delta_h y_t.$$

Model: This fact leads to introducing the seasonal ARMA model, denoted by ARMA$(P, Q)_h$, which is of the form

$$\Phi(L^h) x_t = \Theta(L^h) \varepsilon_t, \tag{18}$$

where

$$\Phi(L^h) = 1 - \Phi_1 L^h - \Phi_2 L^{2h} - \cdots - \Phi_P L^{Ph},$$

and

$$\Theta(L^h) = 1 - \Theta_1 L^h - \Theta_2 L^{2h} - \cdots - \Theta_Q L^{Qh}.$$

Remark 1. Analogously to ARMA(p, q), the ARMA$(P, Q)_h$ model is causal only when the roots of $\Phi(z^h)$ lie outside the unit circle, and it is invertible only when the roots of $\Theta(z^h)$ lie outside the unit circle.

4.2 Mixed Seasonal ARMA

When we combine seasonal and non-seasonal operators we obtain a model

$$\Phi(L^h)\phi(L)x_t = \Theta(L^h)\theta(L)\varepsilon_t,$$

which is called mixed seasonal ARMA and it is denoted by

$$ARMA(p, q) \times ARMA(P, Q)_h$$

4.3 Seasonal ARIMA

Mixed seasonal ARMA is a stationary process. In practice however we often have nonstationary processes. Seasonal nonstationarity can occur when the process is nearly periodic in the season and the seasonal component varies slowly from period to period (say from year to year) according to a random walk, that is

$$s_t = s_{t-h} + v_t$$

where v_t is a white noise.

We can subtract the effect of the season (say month) using the lag operator L^h to obtain seasonal stationarity

$$x_t - x_{t-h} = (1 - L^h)x_t$$

This is a seasonal difference of order 1. In general we define a seasonal difference of order D as

$$\Delta_h^D x_t = (1 - L^h)^D x_t,$$

where $D = 1, 2, \ldots$. Usually $D = 1$ is sufficient to obtain seasonal stationarity.

This leads to a very general seasonal autoregressive integrated moving average (SARIMA) model written as follows

$$\Phi(L^h)\phi(L)\Delta_h^D \Delta^d x_t = \alpha + \Theta(L^h)\theta(L)\varepsilon_t, \tag{19}$$

and denoted by ARIMA$(p, d, q) \times (P, D, Q)_h$.

Example 2. The model ARIMA$(0, 1, 1) \times (0, 1, 1)_{12}$ with $\alpha = 0$ is often applied for various economic data. Using formula (15) we obtain

$$(1 - L^{12})(1 - L)x_t = (1 + \Theta L^{12})(1 + \theta L)\varepsilon_t$$

or, when expanded, we get the following form

$$(1 - L - L^{12} + L^{13})x_t = (1 + \theta L + \Theta L^{12} + \Theta\theta L^{13})\varepsilon_t,$$

or

$$x_t = x_{t-1} + x_{t-12} - x_{t-13} + \varepsilon_t + \theta\varepsilon_{t-1} + \Theta\varepsilon_{t+12} + \Theta\theta\varepsilon_{t-13}.$$

5 Generalized ARMA Models*

Referenced paper:

- Benjamin, M. A., Rigby, R. A., Stasinopoulos, D. M., (2003), Generalized autoregressive moving average models. Journal of the American Statistical Association.

- Zheng, TG, Xiao, H., and Chen, R., (2015), Generalized ARMA models with martingale difference errors. Journal of Econometrics.

5.1 **GARMA Models** - Benjamin et al. (2003)

Let $\{y_t\}$ be a *(non-Gaussian) time series* and $\mathcal{F}_t = \{y_t, y_{t-1}, \ldots\}$ be a past information set up to time t. The GARMA model is given by

$$f(y_t|\mathcal{F}_{t-1}) = \exp\left\{\frac{y_t\vartheta_t - b(\vartheta_t)}{\varphi} + a(y_t, \varphi)\right\} \longrightarrow \text{Exponetial Family}, \quad (20)$$

$$\eta_t \equiv g(\mu_t) = \nu + \sum_{j=1}^{p} \phi_j g(y_{t-j}) + \sum_{j=1}^{q} \delta_j [g(y_{t-j}) - \eta_{t-j}], \quad (21)$$

where ϑ_t and φ are the canonical and scale parameters, and $\mu_t = b'(\vartheta_t) = E(y_t|\mathcal{F}_{t-1})$ and $\mathrm{Var}(y_t|\mathcal{F}_{t-1}) = \varphi b''(\vartheta_t)$, respectively. The function $g(\cdot)$ is called a *link function*. It is assumed that the transformed mean follows a seemingly ARMA process. The quantity η_t is called the *linear predictor*. The link function $g(\cdot)$ is restricted to a one-to-one function hence it can be inverted to obtain $\mu_t = g^{-1}(\eta_t)$.

By adding $g(y_t) - \eta_t$ to both sides of (17), we have

$$g(y_t) = \nu + \sum_{j=1}^{p} \phi_j g(y_{t-j}) + \varepsilon_t + \sum_{j=1}^{q} \delta_j \varepsilon_{t-j}, \tag{22}$$

where $\varepsilon_t = g(y_t) - \eta_t = g(y_t) - g(\mu_t)$.

- Obviously (18) shows that under GARMA model, $g(y_t)$ assumes exactly a standard ARMA model formulation Benjamin et al. (2003).

- The only difference is that the error sequence ε_t is not a *MDS* in most of the cases. Note that

$$E(\varepsilon_t \mid \mathcal{F}_{t-1}) = E[g(y_t) \mid \mathcal{F}_{t-1}] - g(\mu_t) \neq 0,$$

  unless $g(\cdot)$ is an identity function.

5.2 M-GARMA Model

Zheng et al. (2015) assume that the conditional distribution $p(y_t \mid \mathcal{F}_{t-1})$ can be parametrized as

$$p(y_t \mid \mathcal{F}_{t-1}) = f(y_t \mid \mu_t, \varphi), \tag{23}$$

where φ is a collection of time invariant parameters hence all past information is summarized in μ_t. In addition, let

$$g_\varphi(\mu_t) = \nu + \sum_{j=1}^{p} \phi_j h(y_{t-j}) + \sum_{j=1}^{q} \delta_j [h(y_{t-j}) - g_\varphi(\mu_{t-j})], \tag{24}$$

where $g_\varphi(\mu_t) = E[h(y_t) \mid \mathcal{F}_{t-1}]$ serves as the *link function* in the terminology of GLM (generalized linear model).

By adding $h(y_t) - g_\varphi(\mu_t)$ to both sides of (20), we have

$$h(y_t) = \nu + \sum_{j=1}^{p} \phi_j h(y_{t-j}) + \varepsilon_t + \sum_{j=1}^{q} \delta_j \varepsilon_{t-j}, \tag{25}$$

where $\varepsilon_t = h(y_t) - g_\varphi(\mu_t)$.

- It is clear by this construction of the pair of link functions $(h(\cdot), g_\varphi(\cdot))$ that ε_t is now a *MDS*.

- In the following we refer to $g_\varphi(\cdot)$ as the *link function* and $h(\cdot)$ the *y-link function*.

Table 3 Some commonly used conditional distributions, their recommended y-link functions, the corresponding link functions, and the conditional variances of the resulting MDS.

	density	$E[y_t\|\mathcal{F}_{t-1}]$	$\mathrm{Var}[y_t\|\mathcal{F}_{t-1}]$
Lognormal	$\log N(\log(\mu_t)-\sigma^2/2,\sigma^2)$	μ_t	$(e^{\sigma^2}-1)\mu_t^2$
Gamma	Gam$(c\mu_t^d, c\mu_t^{d-1})$	μ_t	μ_t^{2-d}/c
Inverse-Gamma	Inv-Gam$(c\mu_t^{d-1}+1, c\mu_t^d)$	μ_t	$c\mu_t^{1+d}/(c\mu_t^{d-1}-1)$
Weibull	Weibull$(k,\mu_t/\Gamma(1+k^{-1}))$	μ_t	$\mu_t^2\left[\frac{\Gamma(1+2k^{-1})}{\Gamma^2(1+k^{-1})}-1\right]$
Beta	Beta$(\tau\mu_t,\tau(1-\mu_t))'$	μ_t	$\frac{\mu_t(1-\mu_t)}{1+\tau}$
Poisson	Poisson(μ_t)	μ_t	μ_t
	$h(y_t)$	$g_\varphi(\mu_t)$	$\mathrm{Var}[h(y_t)\|\mathcal{F}_{t-1}]$
Lognormal	$\log(y_t)$	$\log\mu_t-\frac{1}{2}\sigma^2$	σ^2
Gamma	$\log(y_t)$	$\psi(c\mu_t^d)-(d-1)\log(\mu_t)-\log(c)$	$\psi_1(c\mu_t^d)$
Inverse-Gamma	$\log(y_t)$	$d\log(\mu_t)+\log(c)-\psi(c\mu_t^{d-1}+1)$	$\psi_1(c\mu_t^{d-1}+1)$
Weibull	$\log(y_t)$	$\approx\log\mu_t-\frac{1}{2}\left[\frac{\Gamma(1+2k^{-1})}{\Gamma^2(1+k^{-1})}-1\right]$	$\approx\frac{\Gamma(1+2k^{-1})}{\Gamma^2(1+k^{-1})}-1$
Beta	$\log(y_t/1-y_t)$	$\psi(\tau\mu_t)-\psi(\tau(1-\mu_t))$	$\psi_1(\tau\mu_t)+\psi_1(\tau(1-\mu_t))$
Poisson	$\sqrt{y_t}$	$\approx\sqrt{\mu_t}$	$\approx\frac{1}{4}$

Note: The functions $\psi(\cdot)$ and $\psi_1(\cdot)$ are the digamma and trigamma functions, respectively.

5.3 Two specific M-GARMA models

(a) Log Gamma-M-GARMA model: Consider the following Gamma-M-GARMA(p, q) model with the y-link function $h(y_t) = \log y_t$,

$$y_t|\mathcal{F}_{t-1} \sim \mathrm{Gam}(c\mu_t^d, c\mu_t^{d-1}), \quad \log y_t = \nu + \sum_{j=1}^{p} \phi_j \log y_{t-j} + \varepsilon_t + \sum_{j=1}^{q} \delta_j \varepsilon_{t-j},$$

with $\varepsilon_t = \log y_t - g_{c,d}(\mu_t)$, where

- $c\mu_t^d$ and $c\mu_t^{d-1}$ are the shape and rate parameters of Gamma distribution;
- $\mu_t = E[y_t|\mathcal{F}_{t-1}]$;
- The link function $g_{c,d}(\mu_t) = \psi(c\mu_t^d) - (d-1)\log \mu_t - \log c$;

- The conditional variance $\mathrm{Var}[\varepsilon_t|\mathcal{F}_{t-1}] = \psi_1(c\mu_t^d)$, where $\psi(\cdot)$ and $\psi_1(\cdot)$ are digamma and trigamma functions, respectively.

- When $d = 0$, the link function $g_{c,d}(\mu_t) = \log \mu_t + \psi(c) - \log c$ differs from the y-link function by a constant, and the conditional variance is also a constant, i.e. $\mathrm{Var}[\varepsilon_t|\mathcal{F}_{t-1}] = \psi_1(c)$.

- When using an identity transformation, i.e., $h(y_t) = y_t$, the M-GARMA becomes the *multiplicative error models (MEM)* given by Engle (2002), Engle and Gallo (2006) and Brownlees et al. (2012).

(2) Logit-Beta-M-GARMA model:

The model can be used for *proportion time series* where the observations take value in $(0,1)$. We consider the *logit y-link* $h(y_t) = \text{logit}(y_t) = \log[y_t/(1-y_t)]$, and the model is given

$$y_t|\mathcal{F}_{t-1} \sim \text{Beta}(\tau\mu_t, \tau(1-\mu_t)), \tag{26}$$

$$\text{logit}(y_t) = \nu + \sum_{j=1}^{p} \phi_j \text{logit}(y_{t-j}) + \varepsilon_t + \sum_{j=1}^{q} \delta_j \varepsilon_{t-j}, \tag{27}$$

with $\varepsilon_t = \text{logit}(y_t) - g_\tau(\mu_t)$, where $\tau\mu_t$ and $\tau(1-\mu_t)$ are two positive shape parameters of Beta distribution.

The link function and conditional variance are given by $g_\tau(\mu_t) = \psi(\tau\mu_t) - \psi(\tau(1-\mu_t))$ and $\text{Var}[h(y_t)|\mathcal{F}_{t-1}] = \text{Var}(\varepsilon_t|\mathcal{F}_{t-1}) = \psi_1(\tau\mu_t) + \psi_1(\tau(1-\mu_t))$ respectively.