Time Series Analysis

Time Series Regression Models

Hossein Moradi Rekabdarkolaee

South Dakota State University

email: *hossein.moradirekabdarkolaee@sdstate.edu*.

Office: CAME Building, *Room: 240*.

Introduction

- So far, we have learned
  - Different components of time series
    - Trend,
    - Seasonality,
    - Error.
  - Different type of time series
    - Stationary (weak and strong)
    - Non-stationary.
  - Different models for time series
    - Autoregressive,
    - Moving Average,
    - ARMA,
    - ARIMA.

Introduction

- Different way to find the order of time series
  - ACF,
  - PACF,
  - EACF,
  - Information Criteria-based method (AIC, BIC, AICc, etc.)
- Different approaches to estimate the parameters of the model we choose to fit the time series.
  - Method of Moment,
  - Least Squares,
  - Maximum Likelihood.

Introduction

- After fitting the model, we did
  - Model diagnostics or model criticism,
  - What does it mean
  - How to do it.
- We studied the prediction at a future time for different time series models.
- We also studied how to incorporate seasonality in the model.
- In this section, we study how to incorporate external information into time series modeling.

Backshift operator

- The backshift operator, denoted **B**, operates on the time index of a series and shifts time back one time unit to form a new series.
- For this, define the backward shift operator **B** as follows:

$$\mathbf{B}y_t = y_{t-1}.$$

- Or generally,

$$\mathbf{B}^s y_t = y_{t-s}.$$

- The operator $(1 - \mathbf{B})$ is usually denoted by $\nabla = 1 - \mathbf{B}$ and is called the backward difference operator or simply difference operator.

Backshift operator

- The difference of the series $\{y_t\}$ is

$$(1 - \mathbf{B})y_t = \nabla y_t = y_t - y_{t-1}.$$

- Second-order differencing is defined as

$$\begin{aligned}(1 - \mathbf{B})^2 y_t &= \nabla^2 y_t \\ &= \nabla y_t - \nabla y_{t-1} \\ &= y_t - 2y_{t-1} + y_{t-2}\end{aligned}$$

- $k-$th order differencing can be defined in a similar manner.

- Using $\mathbf{B}$, the MA(q) model can be written as

$$Y_t = \Theta(\mathbf{B})e_t,$$

- where $\Theta(\mathbf{B})$ is the MA characteristic polynomial evaluated at $\mathbf{B}$.

Some properties of **B** Operators

- The AR(p) model can be expressed as

$$\phi(\mathbf{B})Y_t = e_t,$$

- where $\phi(\mathbf{B})$ is the AR characteristic polynomial evaluated at **B**.
- The ARMA(p,q) model can be expressed as

$$\phi(\mathbf{B})Y_t = \Theta(\mathbf{B})e_t.$$

- The ARIMA(p,d,q) can be expressed as

$$\phi(\mathbf{B})(1-\mathbf{B})^d Y_t = \Theta(\mathbf{B})e_t.$$

Some properties of **B** Operators

- Let c be a constant and $\{y_t\}$ a time series then

$$
\begin{aligned}
\mathbf{B}c &= c \\
\mathbf{B}cy_t &= c\mathbf{B}y_t \\
\mathbf{B}^i\mathbf{B}^j y_t &= \mathbf{B}^{i+j} y_t = y_{t-i-j} \\
(a_1\mathbf{B}^i + a_2\mathbf{B}^j)y_t &= a_1\mathbf{B}^i y_t + a_2\mathbf{B}^j y_t \\
\frac{1}{(1-a\mathbf{B})} y_t &= c\left[1 + a\mathbf{B} + a^2\mathbf{B}^2 + \cdots\right] y_t, \quad |a| < 1
\end{aligned}
$$

- An analogous set of properties hold for ∇, except that $\nabla c = 0$.

Intervention Analysis

- Sometimes exceptional external events, called interventions, affect the time series to be analyzed.
- It is assumed that the intervention affects the process by changing the mean function or trend of a time series.
- Examples of intervention
  - Airline travels.
  - Animal population.
  - Speed limit.

Intervention

- We refer to the events changing the behavior of a time series as intervention.
- An intervention can change the level of a series
- Changes the level after a short delay;
- It could make a series going downward causing it to drift upward or the other way around.
- A single intervention can also have different effects on different time series.

Types of Intervention

- For simplicity, we assume we have just intervention on the mean function.
- We will distinguish several types of intervention.
  - Classification based on the time of intervention
    - Intervention time is known
    - The time of intervention is not known
  - Classification based on effect of intervention
    - Intervention with permanent effect
    - Intervention with temporary effect

Intervention models: known intervention time

- We consider two simple cases of modeling the effect of an intervention used in practice.

- The effect of a temporary intervention, that takes place at only one time can be entered as an input to the system by

$$P_t^T = \begin{cases} 1, & t = T \\ 0, & t \neq T \end{cases}$$

- We call this a pulse.

Intervention models

- The effect of a permanent intervention, that takes place at time T and remains in effect thereafter can be entered to a model by

$$S_t^T = \begin{cases} 0, & t < T \\ 1, & t \geq T \end{cases}$$

- We call this a step intervention.
- Note that

$$P_t^T = S_t^T - S_{t-1}^T = (1 - \mathbf{B})S_t^T \tag{1}$$

- The effects of interventions can be entered to any model of a time series.
- We will use *SARIMA* models.

SARIMA model with an intervention: Sudden fixed unknown impact

- Let $\{N_t\}$ be an $SARIMA(p,d,q) \times (P,D,Q)_s$ series.
- If the impact of an intervention is sudden and fixed, we can write
  - For the pulse intervention, P_t^T,

$$Y_t = \omega P_t^T + N_t$$

  - For an $ARMA(p,q)$ this can be written as:

$$Y_t = \omega P_t^T + \frac{\Theta_q(\mathbf{B})}{\Phi_p(\mathbf{B})} e_t,$$

SARIMA model with an intervention

- For the step intervention, S_t^T,

$$Y_t = \omega S_t^T + \frac{\Theta_q(\mathbf{B})}{\Phi_p(\mathbf{B})} e_t$$

- This model represents an immediate level change in the series.

- Let I_t denote P_t^T or S_t^T, as appropriate, in general we have

$$Y_t = \omega I_t + \frac{\Theta_q(\mathbf{B})}{\Phi_p(\mathbf{B})} e_t. \tag{2}$$

SARIMA model with an intervention: Fixed unknown impact with delay

- If the impact of an intervention is fixed with a time delay of b, a model of the form

$$Y_t = \omega \mathbf{B}^b I_t + \frac{\Theta_q(\mathbf{B})}{\Phi_p(\mathbf{B})} e_t \qquad (3)$$

- can be used.

- For the pulse intervention, P_t^T, the model would be

$$Y_t = \omega \mathbf{B}^b P_t^T + \frac{\Theta_q(\mathbf{B})}{\Phi_p(\mathbf{B})} e_t.$$

- For the step intervention, S_t^T, the model would be

$$Y_t = \omega \mathbf{B}^b S_t^T + \frac{\Theta_q(\mathbf{B})}{\Phi_p(\mathbf{B})} e_t.$$

General *SARIMA* model with an intervention

- We can generalize the previous models to a general intervention model as follows.
- Let N_t be an $SARIMA(p, d, q) \times (P, D, Q)_s$ time series and I_t be an intervention series, then a general *ARIMA* with intervention model has the form of

$$Y_t = \nu(\mathbf{B})I_t + N_t,$$

- where ν is called the impulse function of the model.

General *SARIMA* model with an intervention

- Various models can be produced by the form of ν.
- In practice, a rational function of the following form is considered as the impulse function:

$$\nu(\mathbf{B}) = \frac{\omega(\mathbf{B})\mathbf{B}^b}{\delta(\mathbf{B})},$$

- where

$$\omega(\mathbf{B}) = \omega_0 - \omega_1 \mathbf{B} - \cdots - \omega_k \mathbf{B}^k$$

- and

$$\delta(\mathbf{B}) = 1 - \delta_1 \mathbf{B} - \cdots - \delta_l \mathbf{B}^l$$

SARIMA model with an intervention

Gradual unknown impact

- If we take $\nu(\mathbf{B}) = \frac{\omega}{1-\delta\mathbf{B}}$, where $0 \le \delta \le 1$, we have a model for gradual impact of intervention.

- For *ARMA* model, with the intervention I_t the model reduces to

$$Y_t = \frac{\omega}{1-\delta\mathbf{B}} I_t + \frac{\Theta_q(\mathbf{B})}{\Phi_p(\mathbf{B})} e_t. \tag{4}$$

- Note that (4) reduces to (2) when $\delta = 0$.

- If we take $\nu(\mathbf{B}) = \frac{\omega\mathbf{B}^b}{1-\delta\mathbf{B}}$, where $0 \le \delta \le 1$, we have a model for gradual impact of an intervention with b period delay.

- For *ARMA* model, with a pulse intervention, P_t^T, the model reduces to

$$Y_t = \frac{\omega\mathbf{B}^b}{1-\delta\mathbf{B}} P_t^T + \frac{\Theta_q(\mathbf{B})}{\Phi_p(\mathbf{B})} e_t.$$

SARIMA model with multiple interventions

- Models with one intervention can be extended to multiple interventions.
- The following general form is used in practice for multiple interventions.

$$Y_t = \sum_{j=1}^{m} \frac{\omega_j(\mathbf{B})\mathbf{B}^{b_j}}{\delta_j(\mathbf{B})} I_{jt} + N_t,$$

- where $I_{jt}, j = 1, 2, \cdots, m$ are intervention variables.

Fitting a model with intervention

- Assuming we know the form of the impulse function, the problem of fitting a *SARIMA* model with intervention is similar to a simple *SARIMA* model.
- Parameters can be estimated using least square or maximum likelihood methods.
- We can fit a *SARIMA* model with intervention in R using "arimax" function specifying the intervention variable and order of SARIMA model.

Outliers

- Crucial question: What is an outlier?
- As in other areas of statistics presence of outliers can disrupt a time series analysis.
- Most often outliers are clearly visible in the time plot of the data.

Causes of Outliers

- There are several ways outliers can be produced in time series data.
  - Data entry or measuring instrument errors can produce outliers. (Systematic errors)
  - Outliers could be produced by error distributions with ’thick’ (fat) tails, in which extreme observations occur with greater frequency than expected for a normal distribution.
  - Outliers could be caused because the underlying model is non-linear.

Solutions to the problem of Outliers

- If outliers are errors (mistakes) then they need to be adjusted.
- If it is not clear that the outlier is an error or a real extreme value, it would be misleading to remove the observation completely
- On the other hand leaving the outlier in the model could mess up the analysis.
- For this case the solution depends on the objective of time series analysis.
- For example, if the objective is forecasting, one approach is to use robust methods which downweight extreme observations.
- We will consider outlier analysis in the framework of *ARIMA* models.

Types of outliers

- It is customary to distinguish several types of outliers.
- Different types of outliers were defined as:
  - Additive outliers (AO)
  - Innovation outliers (IO)
  - Level shift (LS)
  - Temporary Change (TC)

Additive outliers (AO)

- Let X_t be an stationary invertible $ARMA(p,q)$ process with AR polynomial of $\phi_p(\mathbf{B})$ and MA polynomial of $\Theta_q(\mathbf{B})$
- An additive outlier (AO) model is defined as

$$Y_t = \begin{cases} X_t & t \neq T \\ X_t + \omega & t = T \end{cases}$$

$$= X_t + \omega I_t^T$$

- where

$$I_t^T = \begin{cases} 1 & t = T \\ 0 & t \neq T \end{cases}$$

Additive outliers (AO)

- Let X_t be an stationary invertible $ARMA(p,q)$ process.
- An additive outlier (AO) model is defined as

$$\begin{aligned} Y_t =& X_t + \frac{\Theta_q(\mathbf{B})}{\phi_p(\mathbf{B})}\omega I_t^T \\ =& \frac{\Theta_q(\mathbf{B})}{\phi_p(\mathbf{B})}(e_t + \omega I_t^T) \end{aligned}$$

- An additive outlier affects only the Tth observation.
- We call ω the outlier effect.

Innovational Outlier (IO)

- An innovation outlier occurs at time t if the error at time t is perturbed.
- That is
$$e_t = e_t + \omega P_t^T.$$
- An innovative outlier at T perturbs all observations on and after T.
- The effect of the IO diminishes as the observation is further away from the origin of the outlier.
- We call ω the outlier effect.

Level shift (LS) and Temporary Change (TC)

- The idea of gradual impact for intervention can be used for outlier detection.
- Level shift (LS) model is defined as

$$Y_t = X_t + \frac{\omega_L}{1 - \mathbf{B}} I_t^T.$$

- Temporary Change (TC) model is defined as

$$Y_t = X_t + \frac{\omega_C}{1 - \delta \mathbf{B}} I_t^T.$$

General Outlier model

- Similar to the intervention analysis we might have several outliers of different types.
- If we have k outliers of different types, we can use the following general model:

$$Y_t = \sum_{j=1}^{k} \omega_j \nu_j(\mathbf{B}) I_t^T + X_t.$$

- where $X_t = \frac{\Theta_q(\mathbf{B})}{\phi_p(\mathbf{B})} e_t$ and $\nu_j(\mathbf{B}) = 1$ for an AO and $\nu_j(\mathbf{B}) = \frac{\Theta_q(\mathbf{B})}{\phi_p(\mathbf{B})}$ for an *IO*.

Estimation of ω

- Assume T is known, to estimate ω, let

$$\pi(\mathbf{B}) = \frac{\Theta_q(\mathbf{B})}{\phi_p(\mathbf{B})} = \sum_{j=1}^{\infty} \pi_j \mathbf{B}^j.$$

- For AO model we can estimate ω by

$$\hat{\omega}_{AO}^T = \frac{e_T - \sum_{j=1}^{n-T} \pi_j e_{T+j}}{\tau^2}, \quad \text{where } \tau = \sum_{j=1}^{n-T} \pi_j^2.$$

- It can be shown that

$$Var(\hat{\omega}_{AO}^T) = \frac{\sigma^2}{\tau^2}.$$

- For IO model we can estimate ω by

$$\hat{\omega}_{IO}^T = e_T$$

- And $Var(\hat{\omega}_{IO}^T) = \sigma^2$.

Testing for ω

- Having an estimate for ω various tests can be performed for the hypotheses

$$H_0 : Y_T \text{ is neither an AO nor an IO}$$

$$H_1 : Y_T \text{ is an AO}$$

$$H_2 : Y_T \text{ is an IO}$$

- The likelihood ratio test statistics for AO and IO are

$$H_0 \ vs H_1 \quad \lambda_{1,T} = \frac{\tau \hat{\omega}_{AO}^{T}}{\sigma}$$

$$H_0 \ vs H_1 \quad \lambda_{2,T} = \frac{\hat{\omega}_{IO}^{T}}{\sigma}$$

- Under the null hypothesis H_0, both $\lambda_{1,T}$ and $\lambda_{2,T}$ have standard normal distributions.

Testing for outliers when T is unknown

- If T is unknown but the time series parameters are known, we can calculate λ_{1t} and λ_{2t} for $t = 1, \cdots n$
- Then based on the distribution of the test statistics we can decide which observation is an outlier.
- In practice, the parameters are unknown and the estimates of the parameters in the presence of outliers is seriously biased.
- A Four-step iterative method proposed by Chang and Tiao (1983) to detect an unknown number of outliers.

Detection of outliers using an Iterative method

- Model the series Y_t assuming no outliers, compute the residuals
- Calculate $\hat{\lambda}_{1t}$ and $\hat{\lambda}_{2t}$ for $t = 1, \cdots n$.
- Define
$$\hat{\lambda}_T = \max_t \max_i \{|\hat{\lambda}_{i,t}|\}$$
- where T is the time that maximum occurs.
  - If $\hat{\lambda}_T = |\hat{\lambda}_{1,T}| > C$, where C is a predetermined positive constant, then there is an AO at time T.
  - Estimate ω_{AO} and modify the observation by $\tilde{Y}_t = Y_t - \hat{\omega}_{AO}^T$ and define the new residuals.
  - If $\hat{\lambda}_T = |\hat{\lambda}_{2,T}| > C$ then there is an IO at time T.
  - Estimate ω_{IO} and modify the observation by $\tilde{Y}_t = Y_t - \hat{\omega}_{AO}^T$ and define the new residuals.

Detection of outliers using an Iterative method

- Compute a new estimate for σ^2.
- Based on this new estimate, recompute $\hat{\lambda}_{1,t}$ and $\hat{\lambda}_{2,t}$.
- The initial model remains unchanged.
- Suppose in step 3 k outliers have been identified at times $T_1, \cdots, T_k$.
- Treat them as if they are known, and estimate the effect of the outlier $\omega_1, \cdot, \omega_k$ and the other parameters in

$$Y_t = \sum_{j=1}^{k} \omega_j \nu_j(\mathbf{B}) I_t^T + X_t.$$

- where $X_t = \frac{\Theta_q(\mathbf{B})}{\phi_p(\mathbf{B})} Z_t$ and $\nu_j(\mathbf{B}) = 1$ for an AO and $\nu_j(\mathbf{B}) = \frac{\Theta_q(\mathbf{B})}{\phi_p(\mathbf{B})}$ for an *IO*.

Spurious Correlation

- We consider two series Y_t and X_t.
- An extension to more than two series is usually straightforward.
- Our model will be a pair of stochastic processes, which we write as $\{X_t, Y_t\}$.
- To analyze the inter-dependence between $\{X_t\}$ and $\{Y_t\}$ we assume that bivariate process is stationary.

Cross covariance function

- The cross-covariance function of a bivariate random process $\{X_t, Y_t\}$ is defines as

$$\gamma_{t,s}(XY) = cov(X_t, Y_s).$$

- $\{X_t, Y_t\}$ are jointly (weakly) stationary if their means are constant and $\gamma_{t,s}(XY)$ is a function of the times difference $t - s$.
- The cross-correlation function (CCF) of $\{Y_t\}$ and $\{X_t\}$ is

$$\rho_k(XY) = \frac{\gamma_k(XY)}{\sqrt{\gamma_X(0)\gamma_Y(0)}},$$

- where γ_Y and γ_X be the acf of $\{Y_t\}$ and $\{X_t\}$, respectively.

- We have

$$\gamma_{-k}(XY) = \gamma_k(XY).$$

Estimating CCF

- For a given bivariate time series data (x_t, y_t), $t = 1, \cdots n$ the CCF is estimated by

$$\hat{\rho}_{XY}(k) = r_{XY} = \frac{g_{XY}(k)}{\sqrt{g_{XX}(0)g_{YY}(0)}},$$

- where

$$g_{XY}(k) = \begin{cases} \frac{1}{n}\sum_{t=k+1}^{n}(x_t - \bar{x})(y_{t-k} - \bar{y}), & k \geq 0 \\ \frac{1}{n}\sum_{t=1}^{n+k}(x_t - \bar{x})(y_{t-k} - \bar{y}), & k < 0 \end{cases}$$

- and

$$g_{XX}(0) = \hat{\gamma}_{XX}(0) \text{ and } g_{YY}(0) = \hat{\gamma}_{YY}(0).$$

Confidence interval for CCF

- To test for the nonzero values of $\rho_{XY}(k)$, we compare the sample CCF, $r_{XY}(k)$, with their standard errors.
- Under the normal assumption, and the hypothesis that two series $\{X_t\}$ and $\{Y_t\}$ are uncorrelated, if the series $\{X_t\}$ is white noise then

$$Var(r_{XY}(k)) \approx \frac{1}{n-k}.$$

- In practice, the series $\{X_t\}$ usually is not white noise, and we have to **prewhiten** it.

Transfer Function models

- The first model in the analysis of bivariate time series is the transfer function model.
- The transfer function model is used where $\{X_t\}$ is an explanatory times series and $\{Y_t\}$ a dependent time series.
- This model is a generalization of regression to time series data.
- The transfer function model is an extension of intervention model to any kind of function rather than pulse or step functions.
- Some authors refer to transfer function model as **dynamic regression**.

Transfer Function models

- In economy and engineering the independent or the explanatory time series is called an input and the dependent time series is called the output.
- Assume that X_t and Y_t are properly transformed series so that they are both stationary.
- In a single input, single output linear system, the output series Y_t and the input series X_t are related in a linear form as

$$Y_t = \nu(\mathbf{B})X_t + N_t \tag{5}$$

- where $\nu(\mathbf{B}) = \sum_{j=0}^{\infty} \nu_j \mathbf{B}^j$ is called the transfer function, and N_t is the noise series of the system.
- We assume $\{X_t\}$ and $\{N_t\}$ are independent.
- Note that $\{N_t\}$ is not necessarily a white noise.

Transfer Function models

- The coefficients ν_j are called the impulse function weights.
- If $\{X_t\}$ and $\{N_t\}$ follow *ARMA* models, then the model(5) is also known as *ARMAX* model.
- As an example consider the simple regression model

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \cdots + \beta_k x_{t-k} + N_t,$$

- In this case, $\nu(\mathbf{B}) = \beta_0 + \beta_1 \mathbf{B} + \cdots + \beta_k \mathbf{B}^k$.

Transfer Function models

- In practice, a rational transfer function of the following form is used

$$\nu(\mathbf{B}) = \frac{\omega(\mathbf{B})\mathbf{B}^b}{\delta(\mathbf{B})},$$

- where

$$\omega(\mathbf{B}) = \omega_0 - \omega_1 \mathbf{B} - \cdots - \omega_l \mathbf{B}^l,$$

and

$$\delta(\mathbf{B}) = 1 - \delta_1 \mathbf{B} - \cdots - \delta_m \mathbf{B}^m.$$

- Similar to the intervention model, b is a delay parameter.
- For a stable system we assume the roots of $\delta(\mathbf{B})$ are outside the unit circle.
- Now, our aim is to fit a transfer function model to a bivariate time series.

Relation between the CCF and The transfer Function

- Using (5) for time $t-k$ we can write

$$Y_{t-k} = \nu_0 X_{t-k} + \nu_1 X_{t-k-1} + \cdots + N_{t-k} \tag{6}$$

- W.L.G., assume mean of Y and mean of X are zero.

- Multiplying both sides of (6) by X_t and taking expectation, then dividing by $\sigma_X \sigma_Y$ we get

$$\rho_{XY}(k) = \frac{\sigma_X}{\sigma_Y}\left[\nu_0 \rho_X(k) + \nu_1 \rho_X(k-1) + \cdots\right] \tag{7}$$

- If the input series $\{X_t\}$ is white noise, i.e. $\rho_X = 0$ for $k \neq 0$, then we have

$$\nu_k = \frac{\sigma_Y}{\sigma_X}\rho_{XY}(k). \tag{8}$$

Relation between the CCF and The transfer Function

- In general, if we know ρ_{XY} and ρ_X we can use (7) to solve for ν_j.
- Solving (7) for ν_j is usually complicated.
- But, if the input series $\{X_t\}$ is white noise, we can simply use (8) to solve for ν_j.
- Usually $\{X_t\}$ is not white noise, and we use prewhitening procedure to get a white noise.

Prewhitening

- In the transfer function model (5), we assume that
  - $\{X_t\}$ and $\{Y_t\}$ are jointly stationary.
  - the input series $\{X_t\}$ follows an *ARMA* model

$$\phi_X(\mathbf{B})X_t = \Theta_X(\mathbf{B})U_t,$$

  - where $\{U_t\}$ is a white noise.
- Since $\{X_t\}$ is stationary, we can write

$$U_t = \frac{\phi_X(\mathbf{B})}{\Theta_X(\mathbf{B})}X_t$$

- U_t is called the prewhitened series and $\frac{\phi_X(\mathbf{B})}{\Theta_X(\mathbf{B})}$ is called the prewhitening transformation.

Prewhitening

- If we apply the prewhitening transformation on the output series $\{Y_t\}$, we obtain

$$V_t = \frac{\phi_X(\mathbf{B})}{\Theta_X(\mathbf{B})} Y_t,$$

- Let $Z_t = \frac{\phi_X(\mathbf{B})}{\Theta_X(\mathbf{B})} N_t$, then the transfer function model reduces to

$$V_t = \nu(\mathbf{B}) U_t + Z_t \tag{9}$$

- The weights ν_j can be found as

$$\nu_j = \frac{\sigma_V}{\sigma_U} \rho_{UV}(j).$$

Relationships between ν_j and ω and δ

- The orders l, m, and b and their relationships to the ν_j can be found by equating the coefficients of $\mathbf{B}^j$ in both sides of the equation

$$\delta(\mathbf{B})\nu(\mathbf{B}) = \omega(\mathbf{B})\mathbf{B}^b.$$

- Expanding this equation, we get:

$$\begin{aligned}
\nu_j &= 0 && j < b \\
\nu_j &= \delta_1\nu_{j-1} + \delta_2\nu_{j-2} + \cdots + \delta_m\nu_{j-m} + \omega_0 && j = b \\
\nu_j &= \delta_1\nu_{j-1} + \delta_2\nu_{j-2} + \cdots + \delta_m\nu_{j-m} - \omega_{j-b} && \\
& \qquad j = b+1, ..., b+l && \\
\nu_j &= \delta_1\nu_{j-1} + \delta_2\nu_{j-2} + \cdots + \delta_m\nu_{j-m} && j > b+l
\end{aligned} \tag{10}$$

Relationships between ν_j and ω and δ

- The equations imply that ν_j consist of the following :

  1. b zero weights $\nu_0, \cdots, \nu_{b-1}$.
  2. $l - m + 1$ weights $\nu_b, \nu_{b+1}, \cdots, \nu_{b+l-m}$ that do not follow a fixed pattern
  3. m starting weights, $\nu_{b+l-m+1}, \nu_{b+l-m+2}, \cdots, \nu_{b+l}$
  4. ν_j, for $j > b + l$, that follows (10).

Relationships between ν_j and ω and δ

- In simple words:
  - b is determined by ν_0 for $j < b$ and $\nu_b \neq 0$.
  - m by the pattern of ν_j, in a similar manner to identification of p for a univariate *ARIMA*.
  - for a given value of b,
    - if $m = 0$, then the value of l can be found using that $\nu_j = 0$ for $j > b + l$.
    - if $m \neq 0$, then the value of l is found by checking where the pattern of decay for ν_j starts.

Identification of Transfer Function Models

- In transfer function models we should identify the transfer function and a model for noise.
- Identifying a transfer function model has three main steps
  1. Identify a model to describe the input series x
  2. Identify a preliminary transfer function
  3. Use the residuals of the preliminary model to identify a model describing the noise structure of the preliminary model and to form a final transfer function model.

Identification of Transfer Function Models

- After fitting a model to x_t, step 2 to identify a transfer function model $\nu(\mathbf{B})$ is done in the following simple steps

1. Prewhiten x_t to get $u_t = \frac{\phi_x(\mathbf{B})}{\Theta_x(\mathbf{B})} x_t$
2. apply the prewhitening transformation on y_t to have v_t
$$v_t = \frac{\phi_x(\mathbf{B})}{\Theta_x(\mathbf{B})} y_t,$$
3. calculate the sample CCF, $\hat{\rho}_{uv}$ between u_t and v_t, to estimate ν_k,
$$\nu_k = \frac{\hat{\sigma}_v}{\hat{\sigma}_u} \hat{\rho}_{uv}(k),$$
4. We can test for the significance of ν_k by comparing $\hat{\nu}_k$ with $(n-k)^{-1/2}$.
5. Identify $b, l, m,, \omega_i, i = 1 \cdots l$ and $\delta_j, j = 1 \cdots m$ by matching the pattern of $\hat{\nu}_k$ with the known patterns.
6. Then we have $\hat{\nu}(\mathbf{B})$.

Identification of the Noise model

- After identifying the transfer function, we calculate the estimated noise

$$\hat{n}_t = y_t - \hat{\nu}(\mathbf{B})x_t,$$

- Then the model for the noise can be identified by univariate time series identification tools, ending with

$$\phi(\mathbf{B})\hat{n}_t = \Theta(\mathbf{B})e_t.$$

- We finish up by identifying a complete model of the form

$$y_t = \frac{\omega(\mathbf{B})}{\delta(\mathbf{B})}x_{t-b} + \frac{\Theta(\mathbf{B})}{\phi(\mathbf{B})}e_t.$$

Estimation of Transfer function models

- After identifying the transfer function model

$$y_t = \frac{\omega(\mathbf{B})}{\delta(\mathbf{B})} x_{t-b} + \frac{\Theta(\mathbf{B})}{\phi(\mathbf{B})} e_t.$$

- we need to estimate the parameters of the model, $(\omega_1, \cdots, \omega_l)$, $(\delta_1, \cdots, \delta_m)$, $(\Theta_1, \cdots, \Theta_q)$, $(\phi_1, \cdots, \phi_p)$ and σ^2.

- In general, this can be done by using ML or CLS method.

Diagnostic Checking of Transfer function models

- After a transfer function model fitted to the bivariate time series, we need to check for the model adequacy before we can use it.
- The assumptions to be checked are:
  - noise model is adequate i.e. the e_t are white noise.
  - Independence of the noise series, n_t and input series,x_t.
  - This can be done equivalently, by checking the independence of e_t and u_t.

Diagnostic Checking of Transfer function models

- To check for the Independence of u_t and e_t we can use
  - sample CCF, $\hat{\rho}_{u\hat{z}}(k)$ between u_t and $\hat{e}_t$. They should be within their two standard errors and show no pattern.
  - the following portmanteau test statistics with $K + 1 - M$ degrees of freedom,
$$Q_0 = n(n+2)\sum_{j=0}^{K} \frac{1}{n-j}\hat{\rho}^2_{u\hat{e}}(j),$$
  - where n is the number of residuals calculated, and M is the number of parameters estimated for the transfer function.

Diagnostic Checking of Transfer function models

- To check for the adequacy of noise model,
    - both sample ACF and PACF of $\hat{e}_t$ should not show any pattern, and lie within two standard errors.
    - and we also can use portmanteau test statistics with $K - p - q$ degrees of freedom,

$$Q_1 = n(n+2)\sum_{j=0}^{K}\frac{1}{n-j}\hat{\rho}^2_{\hat{e}}(j),$$

Diagnostic Checking of Transfer function models

- In diagnostic checking we may have the following two cases
  - **Case 1:** the transfer function $\nu(\mathbf{B})$ is not adequate.
    - In this case, for some k $\hat{\rho}_{u\hat{e}}(k) \neq 0$ and $\hat{\rho}_{\hat{e}}(k) \neq 0$.
    - The remedy is first to re-identify $\nu(\mathbf{B})$ and then the noise model.
  - **Case 2:** transfer function $\nu(\mathbf{B})$ is adequate, and only the noise model is inadequate.
    - In this case, $\hat{\rho}_{u\hat{e}}(k) = 0$ for all k, but for some k, $\hat{\rho}_{\hat{e}}(k) \neq 0$.
    - We can use the pattern of $\hat{\rho}_{\hat{e}}(k)$ to modify the noise model.

Some remarks on using transfer function models

- For constructing the transfer function model, we assumed that x_t and y_t are stationary. If this is not the case we should use some transformation or differencing to achieve stationarity.

- Prewhitening is applied to x_t to whiten it.

- To fit a causal transfer function model to a bivariate time series we should have $\rho_{xy}(k) = 0$ for all $k > 0$ or $\rho_{xy}(k) = 0$ for all $k < 0$.

- If $\rho_{xy} \neq 0$ for both negative and positive values of k, this means relationship between x and y is not causal.

- In engineering and economy, this is called “feedback relationship”.

- In this case, we should use multivariate modeling.

Time Series Regression

- Suppose we have the following time series model for $\{Y_t\}$:

$$Y_t = m_t + \eta_t,$$

- where
  - m_t captures the mean of $\{Y_t\}$, i.e., $E(Y_t) = m_t$.
  - $\{\eta_t\}$ is a zero mean stationary process with ACVF $\gamma_\eta(\cdot)$.
- The component $\{m_t\}$ may depend on time t, or possibly on other explanatory series.

Time Series Regression

- Constant trend model: For each t let $m_t = \beta_0$ for some unknown parameter β_0.
- Simple linear regression: For unknown parameters β_0 and β_1,

$$m_t = \beta_0 + \beta_1 x_t,$$

- where $\{x_t\}$ is some explanatory variable indexed in time (may just be a function of time or could be other series)
- Harmonic regression: For each t let

$$m_t = Acos(2\pi ft + \phi),$$

- where $A > 0$ is the amplitude (an unknown parameter), $f > 0$ is the frequency of the sinusoid (usually known), and $\phi \in (-\pi, \pi]$ is the phase (usually unknown). We can rewrite this model as

$$m_t = \beta_0 x_{1,t} + \beta_1 x_{2,t},$$

- where $x_{1,t} = cos(2\pi ft)$ and $x_{2,t} = sin(2\pi ft)$.

Time Series Regression

- Suppose there are p explanatory series$\{x_{j,t}\}_{j=1}^{p}$, the time series model for $\{Y_t\}$ is
$$Y_t = m_t + \eta_t,$$
- where
$$m_t = \beta_0 + \sum_{j=1}^{p} \beta_j x_{j,t},$$
- and $\{\eta_t\}$ is a mean zero stationary process with ACVF $\gamma_\eta(\cdot)$.
- We can write the linear model in matrix notation:
$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta},$$
- where $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$ is the observation vector, the coefficient vector is $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$, $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^T$ is the error vector.

Time Series Regression

- When dealing with time series the errors $\{\eta_t\}$ are typically correlated in time

  Assuming the errors $\{\eta_t\}$ are a stationary Gaussian process, consider the model
  $$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta},$$
  where $\boldsymbol{\eta}$ has a multivariate normal distribution, i.e.,
  $$\boldsymbol{\eta} \sim N(0, \boldsymbol{\Sigma})$$
  The generalized least squares (GLS) estimate of $\boldsymbol{\beta}$ is
  $$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X}.$$

Applying GLS In Practice

- The main problem in applying GLS in practice is that $\mathbf{\Sigma}$ depends on ϕ and θ and we have to estimate these
- A two-step procedure
  1. Estimate β by OLS, calculating the residuals $\hat{\eta} = Y - X\beta_{OLS}$,and fit an ARMA to $\hat{\eta}$ to get $\mathbf{\Sigma}$.
  2. Re-estimate β using GLS
- Alternatively, we can consider one-shot maximum likelihood methods