Lecture 1: Transfer Function Model

1 Introduction

Transfer function model is a statistical model describing the relationship between an output variable $Y$ and one or more input variables $X$. It has many applications in business and economics, especially in forecasting turning points. Examples of forecasting applications of the model include assessing the impact of monthly advertisement on the profit of a firm and the effect of monthly average daily temperature on gas bill of a household. In most applications, linear equation is used to describe the relationship, resulting in the distributed-lag model commonly known in the econometric literature.

For simplicity, we focus on discrete-time linear models. By discrete-time, we meant that the data are observed at discrete time points, even though the actual process may be continuous in time. The variables $Y$ and $X$ are typically continuous random variables. By linear model, we mean that the relationship between $Y$ and $X$ is linear and both $X$ and $Y$ are linear processes.

Let us start with the simple case that $X$ is a scalar variable and $Y$ has no additional innovation. In this case, the dynamic dependence of $Y_t$ on the current and past values of the $X$, namely $\{X_{t-j}\}_{j=0}^{\infty}$, can be written as

$$Y_t = v_0 X_t + v_1 X_{t-1} + v_2 X_{t-2} + v_3 X_{t-3} + \cdots,$$

where $v_0, v_1, \cdots$ are constant denoting the impact of $X_{t-j}$ on $Y_t$, and $v(B) = v_0 + v_1 B + v_2 B^2 + \cdots$ with $B$ denoting the backshift operator such that $BX_t = X_{t-1}$. In the economic literature, lag operator (L) is often used instead of the notation $B$. The coefficients $v_0, v_1, \cdots$ are referred to as the impulse response function of the system.

For the model in Eq. (1) to be meaningful, the impulse responses must satisfy certain condition. A simple condition is that $\sum_{j=0}^{\infty} |v_j| < \infty$, i.e., the impulse responses are absolutely summable. In this case, the system is said to be stable. The value

$$g = \sum_{j=0}^{\infty} v_j$$

is called the steady-state gain as it represents the impact on $Y$ when $X_{t-j}$ are held constant over time.

The function $v(B)$ determines the impact of input $X_t$ on output $Y_t$. It pays to study some simple examples of $v(B)$. See Table 10.6 of Box, Jenkins and Reinsel (1994, p. 389).

**Example 1.** Consider the model $Y_t = B^3 X_t$. What is the impulse response function? What is the cumulative response function?
Example 2. Consider the model \( Y_t = (0.5 + 0.5B)B^3X_t \). What is the impulse response function? What is the cumulative response function?

Example 3. Consider the model \((1 - 0.5B)Y_t = 0.5B^3X_t\). What is the impulse response function? What is the cumulative response function?

For the model in Eq. (1) to be practical, the number of impulse response coefficients \( v_j \) must satisfy certain constraints. For instance, one can use the same idea as the univariate autoregressive integrated moving-average (ARIMA) model to describe the function \( v(B) \). That is, one assumes that \( v(B) = \omega(B)B^b/\delta(B) \), where \( b \) is a non-negative integer, \( \omega(B) = \omega_0 + \omega_1B + \omega_2B^2 + \cdots + \omega_sB^s \) and \( \delta(B) = 1 - \delta_1B - \cdots - \delta_rB^r \) are finite-order polynomials in \( B \), and \( \omega_0 \neq 0 \). Obviously, \( \omega(B) \) and \( \delta(B) \) have no common factors.

The prior model says that \( v(B) = \omega(B)B^b/\delta(B) \) of Eq. (2) gives rise to a stable transfer function?

Question: Under what condition that \( v(B) = \omega(B)B^b/\delta(B) \) of Eq. (2) gives rise to a stable transfer function?

Question: Suppose that in Eq. (2) \( v(B) = \omega(B)B^b/\delta(B) \) is stable. What is the steady-state gain of the system?

2 Transfer Function Model

In practice, the output \( Y_t \) is not a deterministic function of \( X_t \). It is often disturbed by some noise or has its own dynamic structure. We denote the noise component as \( N_t \). The noise may be serially correlated, and we assume that \( N_t \) follows an ARMA(\( p, q \)) model, i.e.

\[
\phi(B)N_t = \theta(B)a_t,
\]

where \( \theta(B) = 1 - \theta_1B - \cdots - \theta_qB^q \) and \( \phi(B) = 1 - \phi_1B - \cdots - \phi_pB^p \) are polynomials in \( B \) of degree \( q \) and \( p \), respectively, and \( \{a_t\} \) is a sequence of independent and identically distributed random
variables with mean zero and variance $\sigma^2_a$. Often we also assume that $a_t$ is Gaussian. Note that for the ARMA model in Eq. (3), $E(N_t) = 0$ and the usual conditions of stationarity and invertibility apply.

Putting together, we obtain a simple transfer function model as

$$Y_t = c + v(B)X_t + N_t = \frac{\omega(B)}{\delta(B)}B^bX_t + \frac{\theta(B)}{\phi(B)}a_t,$$

where $c$ is a constant, $\theta(B), \phi(B), \omega(B)$ and $\delta(B)$ are defined as before with degree $q, p, s$ and $r$, respectively, and $\{a_t\}$ are Gaussian white noise series. The noise component $N_t$ should be independent of $X_t$; otherwise, the model is not identifiable.

Note that when $b > 0$ the transfer function model is useful in predicting the turning points of $Y_t$ given those of $X_t$.

When there are multiple input variables, say two, the transfer function model becomes

$$Y_t = c + \frac{\omega_1(B)}{\delta_1(B)}B^{b_1}X_{1t} + \frac{\omega_2(B)}{\delta_2(B)}B^{b_2}X_{2t} + \frac{\theta(B)}{\phi(B)}a_t,$$

where $\omega_i(B)$ and $\delta_i(B)$ are similarly defined as in Eq. (4).

3 An Example

Consider the Gas-Furnace example of Box, Jenkins and Reinsel (1994, Chapter 11). The data consist of 296 observations of (a) input gas rate in cubic feet per minute and (b) the percentage of $CO_2$ in outlet gas. The time interval used is 9 seconds and the actual feed rate is $Z_t = 0.6 - 0.04X_t$, where $X_t$ is the input series. What is the dynamic relationship between the input gas rate $X_t$ and the output $CO_2$ measurement $Y_t$? Figure 1 gives the time plot of the data.

Given the data set, our goal is to specify an adequate model for making inference. One approach to achieve this objective is to adopt the iterated modeling procedure of Box and Jenkins (1976) which consists of the following steps:

1. Model specification,
2. Estimation,
3. Model checking (residual analysis).

If a fitted model is judged to be inadequate via model checking statistics, the procedure is iterated to refine the model. A model that passed rigorous model checking can then be used to make inference, e.g. forecasting or policy simulation.

4 Model Building

The task of model specification in the case of a single input variable involves

- estimation of the impulse response function $v_i$’s,

- specification of the noise model $N_t$,
identification of the rational polynomials $\omega(B)$ and $\delta(B)$ and the delay $b$ to best approximate $v(B)$.

We shall briefly discuss methods and statistics that are useful in specifying a transfer function model.

4.1 Preliminary estimation of $v(B)$

Consider the TFM in Eq. (4). Since $X_t$ and $N_t$ might be serially dependent, the regression

$$Y_t = c + v_0 X_t + v_1 X_{t-1} + \cdots + v_h X_{t-h} + e_t,$$

where $h$ is a large positive integer, would, in general, not provide consistent estimates of the $v_i$’s. In the literature, pre-whitening has been proposed as a tool to obtain consistent estimates of $v_i$. The idea of pre-whitening is to remove the serial dependence in $X_t$. Suppose that $X_t$ follows the univariate ARMA model

$$\phi_x(B)X_t = \theta_x(B)\eta_t,$$

where $\{\eta_t\}$ is a sequence of white noises (i.e. iid random variables). Applying the operator $\frac{\phi_x(B)}{\theta_x(B)}$ to Eq. (4), we obtain

$$\frac{\phi_x(B)}{\theta_x(B)}Y_t = c^* + v(B)\frac{\phi_x(B)}{\theta_x(B)}X_t + \frac{\phi_x(B)}{\theta_x(B)}N_t$$

$$= c^* + v(B)\eta_t + \frac{\phi_x(B)}{\theta_x(B)}N_t,$$
where $c^*$ is a constant given by $c^* = \frac{\phi_x(1)}{\theta_x(1)}c$. Define

$$y_t = \phi_x(B) Y_t, \quad n_t = \phi_x(B) N_t.$$  

The prior equation reduces to

$$y_t = c^* + v(B)\eta_t + n_t.$$  \hspace{1cm} (5)

Notice that $\{n_t\}$ is independent of $\{\eta_t\}$ and $\eta_t$ is a white noise series. Multiplying Eq. (5) by $\eta_{t-j}$ for $j \geq 0$, we have

$$y_t \eta_{t-j} = c^* \eta_{t-j} + [v(B)\eta_t] \eta_{t-j} + n_t \eta_{t-j}.$$  

Taking expectation, we obtain

$$\text{Cov}(y_t, \eta_{t-j}) = v_j \text{Var}(\eta_{t-j}).$$  

Consequently, we have

$$v_j = \frac{\text{Cov}(y_t, \eta_{t-j})}{\text{Var}(\eta_t)}.$$  

In term of cross-correlation, we have

$$v_j = \text{Corr}(y_t, \eta_{t-j}) \frac{\text{std}(y_t)}{\text{std}(\eta_t)}.$$  

In practice, the model for $X_t$ can be specified via the univariate time series analysis (e.g., Bus 41910 or Bus 41202). One can then apply the model to obtain $y_t$. This process is called pre-whitening or filtering in the time series literature.

**Discussion:** Some comments on pre-whitening are in order.

1. In finite samples, the accuracy of $v_j$ estimates might be affected by the noise term $n_t$.
2. Pre-whitening becomes complicated when there are multiple input variables.

### 4.2 A rough approximation

Experience shows that the effect of $N_t$ on the estimation of $v(B)$ can often be reduced when a simple model is assumed for $N_t$. In theory, the resulting estimates of $v_j$ are biased. However, such estimates can often serve the purpose of model specification. The approximate models for $N_t$ include the following:

- An AR(1) model if $Y_t$ is not a seasonal time series. Here we use the approximate model

$$N_t = \frac{1}{1 - \phi_1 B} a_t.$$  

- A seasonal ARIMA(1,0,0)(1,0,0) model if $Y_t$ is seasonal. Here the approximate model is

$$N_t = \frac{1}{(1 - \phi_1 B)(1 - \phi_k B^k)} a_t,$$

where $k$ is the number of periods in a year, e.g. $k=4$ for quarterly data.
4.3 Specification of model for \(N_t\)

Construct the estimate of \(N_t\) by
\[
\hat{N}_t = Y_t - \hat{c} - \hat{v}(B)X_t.
\]

Apply the usual univariate time series methods to identify a model for \(N_t\) using \(\hat{N}_t\) as the observed series.

4.4 Specification of transfer function

The goal is to find a rational form for \(v(B)\). To this end, we can use the Corner method, which is based on the Padé approximation of a polynomial. From
\[
v(B) = \frac{\omega(B)B^b}{\delta(B)},
\]
we obtain
\[
v_0 + v_1 B + v_2 B^2 + \cdots = \frac{\omega_0 B^b + \omega_1 B^{b+1} + \cdots + \omega_s B^{b+s}}{1 - \delta_1 B - \cdots - \delta_r B^r}.
\]

By equating the coefficients of \(B^j\), it is easy to see that
- \(v_j = 0\) for \(j < b\) if \(b\) is positive.
- \(v_b, v_{b+1}, \cdots, v_{b+s-r}\) follow no fixed pattern (no such values occur if \(s < r\)),
- \(v_j\) with \(j \geq b + s - r + 1\) follows a \(r\)th order difference equation
  \[
  v_j = \delta_1 v_{j-1} + \cdots + \delta_r v_{j-r}, \quad \text{or} \quad \delta(B)v_j = 0,
  \]
  with starting values \(v_{b+s}, \cdots, v_{b+s-r+1}\).

Example. Consider the case \(\omega(B) = \omega_0 + \omega_1 B + \omega_2 B^2\), \(b = 1\), and \(\delta(B) = 1 - \delta_1 B\). Here \((r, s, b) = (1, 2, 1)\). We have
\[
v_0 + v_1 B + v_2 B^2 + \cdots = \frac{\omega_0 B + \omega_1 B^2 + \omega_2 B^3}{1 - \delta_1 B}.
\]

Therefore,
\[
v_0 + v_1 B + v_2 B^2 + \cdots = (\omega_0 B + \omega_1 B^2 + \omega_2 B^3)(1 + \delta_1 B + \delta_1^2 B^2 + \cdots).
\]

By equating coefficients, we have
- \(v_0 = 0\),
- \(v_1 = \omega_0\) and \(v_2 = \delta_1 \omega_0 + \omega_1 = \delta_1 v_1 + \omega_1\),
- \(v_3 = \delta_1^2 \omega_0 + \delta_1 \omega_1 + \omega_2 = \delta_1 v_2 + \omega_2\), (starting value)
- \(v_4 = \delta_1^3 \omega_0 + \delta_1^2 \omega_1 + \delta_1 \omega_2 = \delta_1 v_3\), and \(v_5 = \delta_1 v_4\), etc.
The last result can be written as \((1 - \delta_1 B)v_j = 0\) for \(j \geq 4\) with starting value \(v_3\).

In general, any polynomial \(v(B)\) can be approximated as accurately as possible by some ratio of two finite-order polynomials by increasing the orders of the two finite-order polynomials. In practice, we seek to find suitable \((r, s, b)\) so that the approximation is adequate. The property that the coefficients \(v_j\) satisfy a \(r\)th order difference equation is used in the Corner Method to specify \((r, s, b)\).

**Corner Method.** Corner method is a two-way table designed to show the pattern of \(v_j\). The rows are numbered 0, 1, 2, ... and the columns 1, 2, 3, ... Also, for numerical purpose, one uses \(u(B) = v(B)/v_{\text{max}}\), where \(v_{\text{max}} = \max_j\{|v_j|\}\). The \((i, j)\)-th element of the two-way table is the determinant of the \(j \times j\) matrix

\[
M(i, j) = \begin{bmatrix}
u_i & u_{i-1} & \cdots & u_{i-j+1} \\
u_{i+1} & u_i & \cdots & u_{i+j+2} \\
\vdots & \vdots & \ddots & \vdots \\
u_{i+j-1} & u_{i+j-2} & \cdots & u_i
\end{bmatrix},
\]

where \(u_h = 0\) if \(h < 0\). From the pattern of \(v_j\) discussed earlier, the table should exhibit the following pattern to show \((r, s, b)\):

<table>
<thead>
<tr>
<th>((i, j))</th>
<th>1</th>
<th>2</th>
<th>...</th>
<th>(r - 1)</th>
<th>(r)</th>
<th>(r + 1)</th>
<th>(r + 2)</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>...</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>...</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>...</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>...</td>
</tr>
<tr>
<td>(b - 1)</td>
<td>0</td>
<td>0</td>
<td>...</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>(b)</td>
<td>(X)</td>
<td>(X)</td>
<td>...</td>
<td>(X)</td>
<td>(X)</td>
<td>(X)</td>
<td>(X)</td>
<td>...</td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>...</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>...</td>
</tr>
<tr>
<td>(s + b)</td>
<td>(X)</td>
<td>(X)</td>
<td>...</td>
<td>(X)</td>
<td>(X)</td>
<td>(X)</td>
<td>(X)</td>
<td>...</td>
</tr>
<tr>
<td>(s + b + 1)</td>
<td>*</td>
<td>*</td>
<td>...</td>
<td>*</td>
<td>(X)</td>
<td>0</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>(s + b + 2)</td>
<td>*</td>
<td>*</td>
<td>...</td>
<td>*</td>
<td>(X)</td>
<td>0</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>...</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>...</td>
</tr>
</tbody>
</table>

**Discussion:** The variance of the determinant of a random matrix is not available. As such, no statistics are available to judge the significance of the elements in the two-way table. This is a drawback of the Corner method. The reading of the two-way table is rather subjective.

The results indicate that \((r, s, b) = (1,2,3)\) and \(N_t\) follows an AR(2) model. A transfer function model is then specified.

**5 Estimation**

Conditional or exact maximum likelihood method is used to perform a joint estimation of all the parameters of a specified transfer function model. Typically, the innovations \(a_t\) is assumed to be Gaussian.
The difference between conditional and exact likelihood methods will be discussed later in vector ARMA models.

6 Model Checking

Check for possible outliers and serial correlations in the residuals of a fitted model. The Box-Ljung statistics of the residuals can be used to check the serial correlations.

**R demonstration**: Gas-Furnace example, including the two R scripts “ccm.R” and “tfm.R”.

```r
> setwd("C:/Users/rst/teaching/mts/sp2013")  # Set working directory on my pc.
> da=read.table("gasfur.dat")  # Load data into R.
> dim(da)  # Find the size of the data.
[1] 296 2
> x=da[,1]
> y=da[,2]
> acf(x)  # identify a simple model for x
> pacf(x)

> m1=arima(x,order=c(3,0,0))
> m1

Call:
arima(x = x, order = c(3, 0, 0))

Coefficients:
ar1 ar2 ar3 intercept
1.9691 -1.3651 0.3394 -0.0606
s.e. 0.0544 0.0985 0.0543 0.1898

sigma^2 estimated as 0.03530: log likelihood = 72.57, aic = -135.14
```

```r
> tsdiag(m1)  # Model checking
>
> source("ccm.R")  # Load the command ‘ccm’.
> ccm(da,lags=20)  # Plot is omitted from the output. You should read the plot.

[1] "Covariance matrix:"
   V1    V2
V1 1.15 -1.66
V2 -1.66  9.25

[1] "CCM at lag: " 0"
   [,1]   [,2]
[1,] 1.000 -0.484
[2,] -0.484  1.000

> f1=c(1,-m1$coef[1:3])  # Create a filter to transform ‘y’.
```

8
```r
> f1
  ar1    ar2    ar3
 1.000000 -1.9690658  1.3651431 -0.3394045

> yf=filter(y,f1,method=c("convo"),sides=1)  # Obtain filtered y-series
> xf=m1$residuals  # Residuals of 'x' is the filtered x-series.
> z=cbind(xf[4:296],yf[4:296])  # The first 3 data in 'yf' are missing due to filtering
> ccm(z, lags=20)

[1] "Covariance matrix:"
     [,1]      [,2]
[1,] 0.035737 -0.000229
[2,] -0.000229 0.132980
[1] "CCM at lag: " "0"
     [,1]      [,2]
[1,] 1.00000 -0.00332
[2,] -0.00332 1.00000

> source("tfm.R")  # Load the command 'tfm' to estimate transfer function models in R.
> mm=tfm(y,x,3,4,1)

[1] "ARMA coefficients & s.e."
   ar1
  coef.arma 0.9730
  se.arma  0.0175
[1] "Transfer function coefficients & s.e."
   intercept     X
  v      53.73 -0.4845 -0.637 -0.839 -0.428 -0.378
  se.v   0.62   0.0929  0.130  0.132  0.130  0.093
> acf(mm$residuals)  # Residuals ACF indicates the model is not adequate.

> names(mm)
[1] "coef" "se.coef" "coef.arma" "se.arma" "nt" "residuals"

> pacf(mm$nt)  # Identifies Nt as an AR(2) process.

> mm=tfm(y,x,3,4,2)

[1] "ARMA coefficients & s.e."
   ar1    ar2
  coef.arma 1.5379 -0.6291
  se.arma  0.0470  0.0509
[1] "Transfer function coefficients & s.e."
   intercept     X
  v      53.376 -0.5558 -0.6445 -0.860 -0.484 -0.3633
  se.v   0.155   0.0778  0.0812  0.081  0.081  0.0773
```
7 Forecasting

The fitted model, if adequate, can be used to produce forecast of $Y_t$ provided that the needed $X$ values are given. In practice, if some $X$ values are not available, then they can be predicted using the univariate time series model for $X_t$. For the Gas-Furnace data set, $X_t$ follows a zero-mean AR(3) model.

Similarly to other time series analysis, the minimum mean squared error criterion is commonly used to produce point forecasts in transfer function modeling.

In time-series forecasting, the fitted model is often treated as the “true” model. As such, the variability in parameter estimation is not considered in producing forecasts. For large samples, this simplification is not a major issue. However, it can underestimate the interval forecasts. This comment also applies to transfer function forecasts.

8 Granger Causality

In using transfer function models, one assumes that $X_t$ is the input that does not depend on the output variable $Y_t$. This means $X_t$ is an exogenous variable and $Y_t$ is an endogenous variable. Care must be exercised in practice because the exogenous assumption might not be valid. Thus, certain tests are often used to verify the unidirectional relationship from $X_t$ to $Y_t$ before using a transfer function model. This is related to the well-known Granger Causality test.

From the transfer function model, $Y_t$ depends on the current and/or past values of $X_t$, but $X_t$ does not depend on any past value of $Y_t$. The issue then is how to conduct such a test.

A straightforward approach is to test $v(B)$ being zero in the model

$$X_t = c + v(B)Y_t + \frac{\theta(B)}{\phi(B)}a_t,$$

where the noise term denotes a model for $X_t$.

Another approach is to analyze the bivariate process $(X_t,Y_t)'$ jointly and perform the unidirectional test based on the fitted bivariate model. This latter approach also applies to the case of multiple input variables.

Remark: The CCF of filtered series can also be used to check the unidirectional relation.
Review of matrix operations useful in multivariate time series analysis

Vectorization: Let \( A_{p \times q} = [a_1, \ldots, a_q] \) be a \( p \times q \) matrix with columns \( a_i \). Then, \( \text{vec}(A) = [a'_1, a'_2, \ldots, a'_q]' \) is a \( pq \)-dimensional column vector.

Kronecker product: Let \( A = [a_{ij}] \) and \( C \) are \( p \times q \) and \( m \times n \) matrices. Then, \( A \otimes C \) is an \((pm) \times (qn)\) matrix given by

\[
A \otimes C = \begin{bmatrix}
a_{11}C & a_{12}C & \cdots & a_{1q}C \\
a_{21}C & a_{22}C & \cdots & a_{2q}C \\
& \vdots & \ddots & \vdots \\
a_{p1}C & a_{p2}C & \cdots & a_{pq}C \\
\end{bmatrix}.
\]

Some properties: (Assume dimensions are proper.)

1. \((A \otimes C)' = A' \otimes C'\).
2. \(A \otimes (C + D) = A \otimes C + A \otimes D\).
3. \((A \otimes C)(F \otimes G) = (AF) \otimes (CG)\).
4. If \( A \) and \( C \) are invertible, then \((A \otimes C)^{-1} = A^{-1} \otimes C^{-1}\).
5. For square matrices \( A \) and \( C \), \( tr(A \otimes C) = tr(A)tr(C)\).
6. \( \text{vec}(A + C) = \text{vec}(A) + \text{vec}(C)\).
7. \( \text{vec}(ABC) = (C' \otimes A)\text{vec}(B)\).
8. \( tr(AC) = \text{vec}(C')'\text{vec}(A) = \text{vec}(A')'\text{vec}(C)\).
9. \( tr(ABC) = \text{vec}(A')'(C' \otimes I)\text{vec}(B)\).

See the appendix of Lüttpohl (2005).