Introduction to time series analysis *

Jean-Marie Dufour †

First version: December 1998
Revised: January 2003
This version: January 14, 2003
Compiled: January 14, 2003, 10:48am

* This work was supported by the Canada Research Chair Program (Chair in Econometrics, Université de Montréal), the Alexander-von-Humboldt Foundation (Germany), the Canadian Network of Centres of Excellence [program on Mathematics of Information Technology and Complex Systems (MITACS)], the Canada Council for the Arts (Killam Fellowship), the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and the Fonds FCAR (Government of Québec).

† Canada Research Chair Holder (Econometrics). Centre interuniversitaire de recherche en analyse des organisations (CIRANO), Centre interuniversitaire de recherche en économie quantitative (CIREQ), and Département de sciences économiques, Université de Montréal. Mailing address: Département de sciences économiques, Université de Montréal, C.P. 6128 succursale Centre-ville, Montréal, Québec, Canada H3C 3J7. TEL: 1 514 343 2400; FAX: 1 514 343 5831; e-mail: jean.marie.dufour@umontreal.ca. Web page: http://www.fas.umontreal.ca/SCECO/Dufour.
1. Notion of time series

Most statistical methods are aimed to be applied to independent experiments or sample survey results: observation ordering has no special meaning (as occurs typically, e.g., in biology, agronomy, sociology, etc.).

In economics, data often take the form of sequences (series) of observations on one or several variables taken at different dates: these observations cannot typically be assumed to be independent.

1.1 Definition (Time series). We call a time series any (finite or finite) sequence observations \((X_t : t \in T)\) indexed by an ordered set ("time").

1.2 Important types of time series

1. Continuous series _ In certain fields (e.g., physics), some variables \(X_t\) can be observed continuously, i.e. the time index \(t\) can take all the values in an interval of real numbers.
   In such a case, we speak of a continuous series. Such series are however quite rare in economic data.

2. Discrete series _ A series is discrete when the set of the possible values \(t\) is a discrete set, i.e. \(T\) can be viewed as a subset of the integers.
   There are two important types of discrete series, depending on whether the observations represent

   (a) levels: series registered instantaneously (e.g., prices, stocks),
   or

   (b) flows : cumulated series over a time interval (e.g., income, income, consumption).
   When analyzing a flow series, it is important to be aware of the time interval involved.
2. **Examples of time series**

   Source: Barro (1987, Figure 1.1).
   Secular trend apparent.

2. Rate of growth of real gross national product in the United States, 1873-1985 (annual).
   Source: Barro (1987, Figure 1.2).
   No trend.

   Source: Barro (1987, Figure 1.3).
   No trend. Series less volatile since 1945.

   Source: Barro (1987, Figure 1.4).
   Strong trend. Possibility of a structural change.

   Source: Barro (1987, Figure 1.5).
   No trend. Possibility of a structural change around 1950.

   Source: Hillmer, Bell, and Tiao (1983, Figure 3.12, page 87).
   Trend plus seasonal fluctuations.
**Figure 1.1 The Behavior of Output in the United States, 1872–1985**

Sources for Figures 1.1–1.5:


For the unemployment rate—The figures are the number unemployed divided by the total labor force, which includes military personnel. Data since 1930 are from *Economic Report of the President*, 1985, Table B-29; 1983, Table B-29; 1970, Table C-22. The data from 1933–43 are adjusted to classify federal emergency workers as employed, as discussed in Michael Darby, "Three-and-a-Half Million U.S. Employees Have been Mislaid: Or, an Explanation of Unemployment, 1934–1941," *Journal of Political Economy*, February 1976. Values for 1890–1929 are based on Christina Romer, "Spurious Volatility in Historical Unemployment Data," *Journal of Political Economy*, February 1986, Table 9.

Figure 1. Real gross national product in the United States, 1872-1985 (annual)

Source: Barro (1987, Chapter 1)
Figure 2. Rate of growth of real gross national product in the United States, 1873-1985 (annual)
Source: Barro (1987, Chapter 1)
Figure 3. Unemployment rate in the United States, 1873-1985 (annual)
Source: Barro (1987, Chapter 1)
Figure 4. Price level in the United States, 1870-1985 (annual)
Source: Barro (1987, Chapter 1)
Figure 5. Inflation rate in the United States, 1870-1985 (annual)
Source: Barro (1987, Chapter 1)
Figure 6. Logarithm of retail sales of men’s and boys’ clothing in the United States, 1967-1979 (quarterly)
Source: Hillmer, Bell, and Tiao (1983)
3. Objectives and problems of time series analysis

3.1. General objectives

1. To develop models for describing the behavior of individual or multiple time series.

2. To propose a methodology for
   - specifying
   - estimating
   - validating (assessing)
   an appropriate model for specific data.

3.2. Important problems in time series analysis

3.2.1 Prediction

Given \( X_1, \ldots, X_T \), we wish to estimate an observed value \( X_{T+h} \).
Prediction can be pointwise, \( \hat{X}_T(h) \), or take the form of an interval (predictive interval):
\[
\left[ \hat{X}_{T-1}(h), \hat{X}_T^2(h) \right].
\]

3.2.2 Decomposition

The most frequent decomposition problems are:

1. to estimate a trend;

2. to eliminate a trend (detrending);

3. to estimate seasonal fluctuations (seasonal components);

4. to eliminate seasonal fluctuations (seasonal adjustment).

For example, suppose a series \( X_t \) can be represented in the form:
\[
X_t = Z_t + S_t + u_t
\]

where:

\( Z_t \) is a trend (smooth function of time),
\( S_t \) is a seasonal component,
\( u_t \) is an irregular component (random perturbation).

The four decomposition problems mentioned above may then be interpreted as follows:
1. estimate $Z_t$;
2. estimate $X_t - Z_t$;
3. estimate $S_t$;
4. estimate $X_t - S_t$.

3.2.3 Detection and modeling of breakpoints (structural change analysis).

3.2.4 Analysis of the dynamic links between:

1. causality;
2. lead-lag relationships.

3.2.5 Separation between short-run and long-run relations (e.g., through the concept of cointegration).

3.2.6 Analysis of expectations.

3.2.7 Control.

4. Types of models

4.1. Deterministic models

A deterministic model is a model not bases on probability theory. There are two main types of deterministic models:

1. deterministic functions of time:

$$ X_t = f(t) ; \quad (4.1) $$

2. recurrence equations:

$$ X_t = f(t, X_{t-1}, X_{t-2}, \ldots) . \quad (4.2) $$

Provided $f(\cdot)$ and (if required) past values of $X_t$ are known, a deterministic model for $X_t$ allows one to predict perfectly the future of $X_t$.
4.2. Stochastic models

A stochastic model is a model where the variables $X_t$ in a series are viewed as *random variables*.

When we consider a series $(X_t : t \in T)$ of random variables, we say we have a *stochastic process* (or a *random function*). The theory of *stochastic processes* is the theoretical mathematical foundation for studying stochastic time series models.

4.3. Important types of deterministic trends

Different types of deterministic trends can be obtained by varying the functional form of $f(t)$. Especially important ones the following.

1. **Trigonometric trend**:
   \[
   f(t) = A_0 + \sum_{j=1}^{q} [A_j \cos(\omega_j t) + B_j \sin(\omega_j t)] . \tag{4.3}
   \]
   This function is periodic (or quasi-periodic). From the very start of time series analysis, such models were considered in order to represent series whose behavior appeared to exhibit periodicities. An important issue in such analyses consists in determining the important frequencies $\omega_j$ (*harmonic analysis* or *spectral analysis*).

2. **Linear trend**:
   \[
   f(t) = \beta_0 + \beta_1 t . \tag{4.4}
   \]

3. **Polynomial trend**:
   \[
   f(t) = \beta_0 + \beta_1 t + \cdots + \beta_k t^k . \tag{4.5}
   \]

4. **Exponential curve**:
   \[
   f(t) = \beta_0 + \beta_1 r^t . \tag{4.6}
   \]

5. **Logistic curve**:
   \[
   f(t) = \frac{1}{\beta_0 + \beta_1 r^t} , \text{ where } r > 0 . \tag{4.7}
   \]

6. **Gompertz curve**:
   \[
   f(t) = \exp \{\beta_0 + \beta_1 r^t\} , \text{ where } r > 0 . \tag{4.8}
   \]
4.4. Important categories of stochastic models

4.4.1. Adjustment models

\[ X_t = f(t, u_t) \]  

(4.9)

where:

\[ t \] represents time,

\[ u_t \] is a random disturbance.

Usually, it is assumed that the \( u_t \)'s are mutually independent or uncorrelated.

4.4.1.1. Important types of adjustment models.

1. **Additive trend**: 

\[ X_t = f(t) + u_t \]  

(4.10)

2. **Multiplicative trend**: 

\[ X_t = f(t) u_t \]  

(4.11)

where \( f(t) \) is independent of (or uncorrelated with) \( u_t \). Usually, it is assumed that \( f(t) \) is a deterministic (non random) function of time as considered above. In certain cases, \( f(t) \) can be viewed as random (unobserved components models).

4.4.1.2. Trend estimation and elimination. Methods for estimating or eliminating trends belong to two basic types:

1. **global adjustment methods**, where all the observations play equivalent roles;

2. **local adjustment methods**, where nearby observations (in time) play more important roles:

   (a) moving averages;

   (b) exponential smoothing.

4.4.1.3. **Persons decomposition.** In economics, the following standard decomposition (called the Persons decomposition) has often been used:

\[ X_t = Z_t + C_t + S_t + u_t \]  

(4.12)
where

- $Z_t$ is a secular (long-run) trend,
- $C_t$ is a relatively smooth deviation from the secular trend (business cycle),
- $S_t$ is a seasonal component,
- $u_t$ is a random perturbation (unpredictable).

4.4.2. Filtering models (generalized moving averages)

\[ X_t = f (\ldots, u_{t-1}, u_t, u_{t+1}, \ldots) \]  \hspace{1cm} (4.13)

where the $u_t$’s are random disturbances (independent or mutually uncorrelated random variables).

Important case – Moving average of order $q$ :

\[ X_t = \bar{\mu} + u_t - \sum_{j=1}^{q} \theta_j u_{t-j} . \]  \hspace{1cm} (4.14)

4.4.3. Autopredictive models

\[ X_t = f (X_{t-1}, X_{t-2}, \ldots, u_t) \]  \hspace{1cm} (4.15)

where the $u_t$’s are random disturbances.

Important case – Autoregressive process of order $p$ :

\[ X_t = \bar{\mu} + \sum_{j=1}^{p} \phi_j X_{t-j} + u_t . \]

4.4.4. Explanatory models

\[ X_t = f (Z_t^*, u_t) \]  \hspace{1cm} (4.16)

where $Z_t^*$ contains various explanatory variables (exogenous variables) and (possibly) lagged values of $X_t$. 
References


