MULTIVARIATE TIME SERIES ANALYSIS AND FORECASTING

Manfred Deistler

Econometrics and Systems Theory

Institute for Mathematical Methods in Economics
University of Technology Vienna
Singapore, May 2004
Introduction

**Time Series**: Observations ordered in time, information contained in ordering: Results are not permutation-invariant in general.

Discrete time, equally spaced data: $x_t, t = 1, 2, \ldots, T; x_t \in \mathbb{R}^n$

**Main questions in TSA:**

- Data driven modelling (system identification)
- Signal and feature extraction (e.g. seasonal adjustment)
Theory and methods concern:
- Model classes
- Estimation and inference
- Model selection
- Model evaluation

Areas of application:
- Signal processing e.g. speech, sonar and radar signals
- Data driven modelling for simulation and control of technical systems and processes; monitoring
- Time series econometrics: Macroeconometrics, finance econometrics, applications to marketing and firm data
- Medicine and biology: Genetics, EEG data, monitoring ...
Stationary processes are an important model class for time series

Def: A stochastic process \((x_t)(t \in \mathbb{R}), x_t : \Omega \to \mathbb{R}^n\), is called (wide sense stationary) if

(i) \(E x'_t x_t < \infty\) for all \(t\)
(ii) \(E x_t = m = \text{const}\) for all \(t\)
(iii) \(\gamma(s) = E(x_{t+s} - m)(x_t - m)'\) does not depend on \(t\)

“Shift invariance” of first and second moments

\(\gamma : \mathbb{Z} \to \mathbb{R}^{n \times n}\) covariance function of \((x_t)\)
contains all linear dependence relations between process variables
A function $\gamma : \mathbb{Z} \rightarrow \mathbb{R}^{n \times n}$ is called **nonnegative definite** if $\Gamma_T \geq 0 \quad \forall T$

Mathematical characterization of covariance functions of stationary processes:

$\gamma$ is a covariance function if and only if $\gamma$ is nonnegative definite
Examples for stationary processes

(1) *White noise*

\[ E\varepsilon_t = 0 \quad E\varepsilon_s\varepsilon_t = \delta_{st} \Sigma, \quad \Sigma \geq 0 \]

no (linear) memory

(2) *Moving average (MA) process*

\[ x_t = \sum_{j=0}^{q} b_j \varepsilon_{t-j}; \quad b_j \in \mathbb{R}^{n \times m} \]

finite memory

(3) *Infinite moving average process*

\[ x_t = \sum_{j=-\infty}^{\infty} b_j \varepsilon_{t-j} \]

large class of stationary processes
(4) **(Stationary) Autoregressive (AR) process**

Steady state solution of stable VDE of the form

$$\sum_{j=0}^{p} a_j x_{t-j} = \varepsilon_t; \quad a_j \in \mathbb{C}^{n \times n} \quad \text{det} a(z) \neq 0 \quad |z| \leq 1, a(z) = \sum_{j=0}^{p} a_j z^j$$

(5) **(Stationary) ARMA process**

Steady state solution of stable VDE of the form

$$a(z)x_t = b(z)\varepsilon_t \quad \text{...backwardshift;} \quad b(z) = \sum_{j=0}^{q} b_j z^j$$
(6) **Harmonic process**

\[ x_t = \sum_{j=0}^{h} e^{i \lambda_j t} z_j = \sum_{j=1}^{\lceil (h+1)/2 \rceil} a_j \cos \gamma_j t + b_j \sin \gamma_j t \]

where \( \lambda_j \in [-\pi, \pi] \) (angular) frequencies,

\[ \gamma_j = \lambda_j \left( h+1 \right) / 2 \]

\[ z_j : \Omega \to \mathbb{C}^n, a_j, b_j : \Omega \to \mathbb{C}^n \]

\( \pi \) : Nyquist frequency; \((x_t)\) is a weighted sum of harmonic oscillations with random weights (amplitudes and phases)

**Stationarity conditions:**

\[ Ez_j^* z_j < \infty, Ez_j = \begin{cases} 0 & \text{for } j : \lambda_j \neq 0 \, \text{Ex}_t \, \text{for } j : \lambda_j = 0, Ez_j z_j^* = 0 \, j \neq 1 \end{cases} \]
Spectral distribution function for a harmonic process

\[ F : \ [-\pi, \pi] \rightarrow \mathbb{R}^{nxn} : F(\lambda) = \sum_{j: \lambda_j \leq \lambda} F_j ; F_j = Ez_jz_j^* \]

\[ F \leftrightarrow \gamma \]

has the same information as \( \gamma \) about the process, however displayed in a different way.
Spectral representation of stationary process

Every stationary process is the (pointwise in $t$) limit of a sequence of harmonic processes:

**Theorem:** Every stationary process $(x_t)$ can be represented as

$$x_t = \int_{[-\pi, \pi]} e^{i\lambda t} dz(\lambda)$$

where $(z(\lambda) \mid \lambda \in [-\pi, \pi]), \ z(\lambda) : \Omega \rightarrow \mathbb{C}^n$ satisfies

$z(-\pi) = 0, \ z(\pi) = x_0, \ Ez(\lambda)z^*(\lambda) < \infty, \ \lim_{\varepsilon \downarrow 0} z(\lambda + \varepsilon) = z(\lambda),$

$E \left\{ (z(\lambda_4) - z(\lambda_3))(z(\lambda_2) - z(\lambda_1))^* \right\} = 0 \quad \text{for} \quad \lambda_1 < \lambda_2 \leq \lambda_3 < \lambda_4$

and is unique for given $(x_t)$.
Spectral distribution of a general stationary process

\[ F : [-\pi, \pi] \rightarrow \mathbb{C}^{n \times n} : F(\lambda) = E z(\lambda) z^*(\lambda) \]

Spectral density

If \( \sum || \gamma(t) || < \infty \)

then there exists a function \( f : [-\pi, \pi] \rightarrow \mathbb{C}^{n \times n} \)

s.t. \( F(\lambda) = \int_{-\pi}^{\lambda} f(\omega) \, dw \), called the spectral density

We have

\[ \gamma(t) = \int e^{i\lambda t} f(\lambda) d\lambda \]

\[ f(\lambda) = (2\pi)^{-1} \sum_{t=-\infty}^{\infty} e^{-i\lambda t} \gamma(t) \]
$f$ is characterized by $f \geq 0 \quad \lambda \text{ a.e.}, \| \int f(\lambda)d\lambda \| < \infty$ and $f(\lambda) = f(-\lambda)$.

In particular we have $\gamma(0) = \int_{-\pi}^{\pi} f(\lambda)d\lambda$: Variance decomposition

The diagonal elements of $f$ show the contributions of the frequency bands to the variance of the respective component process and the off-diagonal elements show the frequency band specific covariances and expected phase shifts between component processes.
<table>
<thead>
<tr>
<th>Parametric Estimation</th>
<th>Seminonparametric Estimation</th>
<th>Nonparametric Estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>e.g. AR estimation for given $p$</td>
<td>e.g. AR estimator where in addition $p$ is estimated</td>
<td>e.g. Windowed spectral estimation</td>
</tr>
</tbody>
</table>

“The curse of dimensionality”: E.g. for AR estimation (with given $p$) the dimension of the parameter space is $n^2 p$ (for the $a_j$) plus $\frac{n(n+1)}{2}$ for $\Sigma$. 
Linear transformations of stationary processes

\[ y_t = \sum_{j=-\infty}^{\infty} a_j x_{t-j} ; a_j \in \mathbb{R}^{nxm}, \sum_{j} |a_j|^2 < \infty, \quad (x_t) \text{ stationary} \]

stationary, linear, dynamic, time invariant, stable system

Weighting function \( (a_j | j \in \mathbb{Z}) \)

Causality: \( a_j = 0 \quad j < 0 \)
\[ y_t = \int e^{i2\pi z_y} \, dz_y = \int e^{i2\pi z_x} \, dz_x = \int e^{i2\pi x} \, dx \]

transfer function

\[ \lambda \]

\[ \sum_{j=-\infty}^{\infty} a_j e^{-i\lambda j} \, dz_x(\lambda) \]

frequency-specific gain and phase shift

\[ k(z) = \sum_{j \in \mathbb{Z}} a_j z^j \]

Transformation of second moments

\[ f_y(x) = k(e^{-i\lambda}) f_x(x) \]

\[ f_y(x) = k(e^{-i\lambda}) \]
Solution of linear vector difference equations (VDE’s)

\[ a_0 y_t + a_1 y_{t-1} + \ldots + a_p y_{t-p} = b_0 x_t + \ldots + b_q x_{t-q} ; a_j \in \mathbb{R}^{n \times n} \]

\[ b_j \in \mathbb{R}^{n \times m} , t \in \mathbb{Z} \]

or:

\[ a(z)y_t = b(z)x_t \]

where \( a(z) = \sum_{j=0}^{p} a_j z^j \), \( b(z) = \sum_{j=0}^{q} b_j z^j \)

\( z : z \in \mathbb{R} \) as well as backward-shift: \( z(y_t | t \in \mathbb{Z}) = (y_{t-1} | t \in \mathbb{Z}) \)
Steady state solution: $z$–transform

If $\det a(z) \neq 0$ and $|z| \leq 1$ then there exists a causal stable solution

$$y_t = \sum_{j=0}^{\infty} k_j x_{t-j}$$

where

$$\sum_{j=0}^{\infty} k_j z^j = k(z) = a^{-1}(z)b(z) = (\det a(z))^{-1} \text{adj}(a(z))b(z); \ z \in \mathbb{C}$$
Forecasting for stationary processes

Problem: Approximation of a future value $x_{t+h}, h > 0$ from the past $x_s, s \leq t$

Linear least squares forecasting:

$$\min_{\mathbf{a} \in \mathbb{R}^{nxn}} E(x_{t+h} - \sum_{j \geq 0} a_j x_{t-j})^* (x_{t+h} - \sum a_j x_{t-j})$$

Projection interpretation

Prediction from a finite past; $x_t, x_{t-1}, \ldots, x_{t-r}$

$$E(x_{t+h} - \sum_{j=0}^r a_j x_{t-j}) x_{t-s} = 0, \ s = 0, \ldots, r$$
leads to

\[
(a_0, \ldots, a_r) \begin{pmatrix}
\gamma(0) & \cdots & \gamma(r) \\
\vdots & & \vdots \\
\gamma(-r) & \gamma(0)
\end{pmatrix} = (\gamma(h), \ldots, \gamma(h + r))
\]

\[\hat{x}_{t,h} = \sum a_j x_{t-j}\] Predictor

Prediction from an infinite past; \(x_t, x_{t-1}, \ldots\)

A stationary process is called (linearly) singular if

\[\hat{x}_{t,h} = x_{t+h}\] for some and hence for all \(t, h > 0\)

Here \(\hat{x}_{t,h}\) denotes the best linear least squares predictor from the infinite past.
A stationary process is called (linearly) regular if
\[ \lim_{h \to \infty} \hat{x}_{t,h} = 0 \]
for one and hence for all \( t \).

**Theorem (Wold)**
(i) Every stationary process \( (x_t) \) can be represented in a unique way as \( x_t = y_t + z_t \) where \( (y_t) \) is regular, \( (z_t) \) is singular, \( Ey_t z'_s = 0 \) and \( y_t \) and \( z_t \) are causal linear transformations of \( (x_t) \).

(ii) Every regular process \( (y_t) \) can be represented as:
\[
y_t = \sum_{j=0}^{\infty} k_j \varepsilon_{t-j} ; \sum ||k_j||^2 < \infty, (\varepsilon_t) \text{ white noise and where } \varepsilon_t \text{ is a causal linear transformation of } (y_t)\]
Consequences for forecasting:
(i) \((y_t)\) and \((z_t)\) can be forecasted separately
(ii) for the regular process \((y_t)\) we have:

\[
y_{t+h} = \sum_{j=0}^{\infty} k_j \varepsilon_{t+h-j} = \sum_{j=h}^{\infty} k_j \varepsilon_{t+h-j} + \sum_{j=0}^{h-1} k_j \varepsilon_{t+h-j}
\]

Note: Every regular process can be forecasted with arbitrary accuracy by an \((AR)MA\) process

How do we obtain the Wold representation: Spectral factorization

\[
f_y = (2\pi)^{-1} k(e^{-i\lambda}) \Sigma k(e^{-i\lambda})^*, \quad \Sigma = E\varepsilon_t\varepsilon_t^*
\]