Stationary Processes and Linear Systems

M. Deistler

Econometrics, Operations Research and Systems Theory

Institute of Econometrics, OR and Systems Theory
University of Technology, Vienna

manfred.deistler@tuwien.ac.at
http:\\www.eos.tuwien.ac.at
PART I: The general framework

1. Introduction

Time Series Analysis: Systematic approaches to extract information from time series, i.e., from observations ordered in time (no permutation invariance).

- Data driven modeling
- Signal and feature extraction

Observations may be “noisy”.

Questions: Trends,
Hidden periodicities,
Dependence on time (dynamics)

Models: Stationary processes
Linear systems
2. The History of Time Series Analysis

2.1. The Early History (1772 - 1920)

- Late 18\textsuperscript{th} century astronomy:
  - More accurate data from observation of the orbits of the planets
  - Kepler’s laws are based on the two body problem
  \[\rightarrow\] Are there deviations from the elliptic shape of the orbits (beside measurement noise)?
hidden periodicities or trends

Question of secular changes (Laplace 1787; Jupiter, Saturn) Harmonic analysis:

- J. L. Lagrange (1736 - 1813), Oeuvres, Vol 6, 1772
- L. Euler (1707 - 1783)
- J. B. J. Fourier (1768 - 1831), Théorie analytique de la chaleur
2. History of time series analysis

• Method of least squares for fitting a line into a scatter plot: A.M. Legendre and C.F. Gauss: Early 19th century.

• Periodogram: G.C. Stokes (1879), A. Schuster (1894) Detection of hidden periodicities:

\[
I_T(\lambda) = \frac{1}{T} \left| \sum_{t=1}^{T} x_t e^{-i\lambda t} \right|^2
\]

\( T \ldots \) sample size
Sunspot numbers, periodicity of earthquakes

• Empirical analysis of business cycles
W.S. Jevons: Periodic fluctuations in economic time series (\( \sim 1870, 1880 \))
H. Moore “Economic Cycles: Their Law and Cause” 1914
W. Beveridge 1922: Wheat price index
2.2. The formation of modern time series analysis (1920-1970)

- Business cycles: Not exactly periodic:
  Stochastic models: AR and MA process e.g.
  \[ y_t = a y_{t-1} + \epsilon_t, \quad y_t = \epsilon_t + b \epsilon_{t-1} \]
  
  \((\epsilon_t)\) white noise
  G.U. Yule (1921, 1927) E. Slutzky (1927)
  R. Frisch: Propagation and Impulse Problems in Dynamic Econometrics (1933)

- Theory of stationary processes:
  Concept: A. Ya Khinchin (1934)
  Spectral representation: A. N. Kolmogorov (1939, 1941)
  Wold representation: H. Wold (1938)
  Factorization of spectra; linear least squares forecasting and filtering A.N. Kolmogorov (1939, 1941)
Ergodic theory: G.D. Birkhoff (1931), A. Ya Khinchin (1932)

• Cowles Commission, Identifiability and ML estimation of (multivariate) ARX models.
  H.B. Mann and A. Wald (1943),
  T. Haavelmo (1944),
  T.C. Koopmann, H. Rubin and R.B. Leipnik (1950);
  Klein I model

• Spectral Estimation:
  Daniel (1946),
  R.B. Blackman and J. Tukey (1958),
  U. Grenander and M. Rosenblatt (1958),
  E. J. Hannan (1960)


2.3. The recent past (1970-1990)

- Box, G.E.P. and G.M. Jenkins (1970)
  Explicit instructions for SISO system identification:
  Differencing, Order determination, ML-estimation, validation

- Kalman: Structure theory for state space systems:
  Realization and parametrization. MIMO case

- Order estimation by information criteria such as AIC
  or BIC: Akakike, Hannan, Rissanen, Schwartz

- Asymptotic properties of ML-type estimation: E.J.
  Hannan (1973), W. Dunsmuir and E.J. Hannan
  (1976), P. Caines and L. Ljung (1979)

- Textbooks (late 80ies): Ljung, Caines, Hannan and
  Deistler, Söderström and Stoica
3. Areas of application

- Signal processing
- Control
- Econometrics: Macroeconometrics, finance, microeconometrics, marketing, logistics
- Medicine and biology
PART II: Stationary Processes

4. Stationary processes in time domain

For us a stochastic process is a model for random phenomena evolving in time 
\((\Omega, \mathcal{A}, \mathbb{P})\) probability space 
y\_t : \Omega \rightarrow \mathbb{C}^n \text{ random variable}
\((y_t | t \in T)\), random process, \(T \subset \mathbb{R}\)
in particular \(T = \mathbb{Z}\)

Def.: A stochastic process \((y_t)\) is called (weakly) stationary if:

(i) \(\mathbb{E} y_t^* y_t < \infty \quad t \in \mathbb{Z}\)

(ii) \(\mathbb{E} y_t = m = \text{const} \quad t \in \mathbb{Z}\)

(iii) \(\gamma(s) = \mathbb{E} y_{t+s} y_t^*\) does not depend on \(t\)
Covariance function

\[ \gamma : \mathbb{Z} \rightarrow \mathbb{C}^{n \times n} : \gamma(t) = E y_t y_0^* \]

describes all linear dependence relations between the one dimensional random variables \( y_t^{(i)}, y_s^{(j)} \).

\( \gamma \) is a covariance function if and only if \( \gamma \) is nonnegative definite

\[ y_t^{(i)} \in L_2 \text{ (over } (\Omega, \mathcal{A}, P) \text{)} \]

Note \( L_2 \) with inner product

\[ < x, y > = E x \bar{y} \]

is a Hilbert space

**Def.**: The *time domain* \( H \subset L_2 \) of a stationary process \( y_t \) is the Hilbert space spanned by

\[ \{ y_t^{(i)} | t \in \mathbb{Z}, i = 1, \ldots, n \} \]
5. Stationary processes in frequency domain

As a consequence of the “translation invariance” of the covariances we have:

**Theorem:** For every stationary process \((y_t)\) there is a unique unitary operator \(U : H \rightarrow H\) such that
\[y_t^{(i)} = U^t y_0^{(i)}, \quad i = 1, \ldots, n, \quad t \in \mathbb{Z}\]
holds.

From this we obtain

**Theorem:** (Spectral representation of stationary processes). For every stationary process \((y_t)\) there exists a process \((z(\lambda)|\lambda \in [-\pi, \pi])\) (called process with orthogonal increments)
(i) \( z(-\pi) = 0, z(\pi) = x_0 \)

(ii) \( \lim_{\epsilon \downarrow 0} z(\lambda + \epsilon) = z(\lambda) \)

(iii) \( \mathbb{E} z^*(\lambda) z(\lambda) < \infty \)

(iv) \( \mathbb{E}\{(z(\lambda_4) - z(\lambda_3))(z(\lambda_2) - z(\lambda_1))^*\} = 0 \)

for \( \lambda_1 < \lambda_2 \leq \lambda_3 < \lambda_4 \)

such that

\[
y_t = \int e^{i\lambda t} \, d\lambda
\]

holds.

Thus, every stationary process is obtained as a limit of harmonic processes

\[
y_t = \sum_{j=1}^{h} e^{i\lambda_j t} z(\lambda_j)
\]
Second moments in frequency domain:

Spectral distribution function

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

Spectral representation of covariance function

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

\[ \gamma \longleftrightarrow F \]

Spectral density (w. r. t. L-measure)

\[ F(\lambda) = \int_{-\pi}^{\lambda} f(\omega)d\omega \]

exists e.g. if \( \sum \|\gamma(t)\|^2 < \infty \)

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

\[ f(\lambda) = (2\pi)^{-1} \sum \gamma(t)e^{-i\lambda t} \]

\( f \) (if it exists) contains the same information as \( \gamma \) but, often displayed in a more convenient form.

Peaks of \( f \) indicate dominating frequency bands.
6. Linear transformations of stationary processes

Let \((x_t)\) be stationary; a linear transformation of \((x_t)\) is given by
\[
y_t = \sum_{j=-\infty}^{\infty} k_j x_{t-j}; \quad k_j \in \mathbb{R}^{n \times m};
\]
\[
\sum \|k_j\| < \infty
\]
then \((x'_t, y'_t)'\) is jointly stationary.

\((k_j | t \in \mathbb{Z})\) weighting function
\[
y_t = \int e^{i\lambda t} dz_y(\lambda) = \sum k_j \int e^{i\lambda(t-j)} dz_x(\lambda) =
\]
\[
\int e^{i\lambda t} \left( \sum_{j=-\infty}^{\infty} k_j e^{-i\lambda j} \right) dz_x(\lambda)
\]
\[
k \longleftrightarrow (k_j)
\]
The transfer function describes the linear transformation in frequency domain.
Linear system

\[
(x_t) \rightarrow k \rightarrow (y_t)
\]

stable, time invariant

\((x_t)\) input
\((y_t)\) output

Linear system with noise

\[
(x_t) \rightarrow k \rightarrow (u_t)\]

\[
(x_t) \rightarrow 1 \rightarrow \hat{y}_t + \rightarrow (y_t)
\]

\((\epsilon_t)\) white noise, i.e., \(\mathbb{E} \epsilon_t = 0, \mathbb{E} \epsilon_s \epsilon'_t = \delta_{st} \cdot \Sigma\)

\[
f_{\epsilon} = (2\pi)^{-1} \cdot \Sigma
\]
"d\gamma_y(\lambda) = k(e^{-i\lambda})d\gamma_x(\lambda)"

Transformation of second moments in frequency domain

\[ f_y = kf_xk^* \]
\[ f_{yx} = kf_x \]

7. The Wold decomposition

Let \((x_t)\) be stationary

\[ H_x(t) = sp\{x_s^{(i)} | i = 1, \ldots, n, s \leq t\} \subset H_x \]

is called the past of \((x_t)\)

**Def.**: A stochastic process \((x_t)\) is called (linear) **regular** if \(\cap_t H_x(t) = \{0\}\)

and (linear) **singular** if \(\cap_t H_x(t) = H_x\)
Theorem: (Wold decomposition)

(i) Every stationary process \((x_t)\) can be uniquely decomposed as

\[ x_t = y_t + z_t \]

where \((y_t)\) is regular, \((z_t)\) is singular, \(\mathbb{E}y_t z_s^* = 0\),
\(y_t^{(i)} \in H_x(t), z_t^{(i)} \in H_x(t)\)

(ii) Every regular process \((y_t)\) can be represented as

\[ y_t = \sum_{j=0}^{\infty} k_j \epsilon_{t-j}, \quad \sum \|k_j\|^2 < \infty \quad (1) \]

where \((\epsilon_t)\) is white noise satisfying \(\epsilon_t^{(i)} \in H_y(t)\),
\(i = 1, \ldots, n\)

Thus, “practically every” stationary process is obtained as an output of a linear system whose input is the “simplest” random process, namely white noise.
\((\epsilon_s, s \leq t)\) constitutes an “orthonormal” basis for \(H_y(t)\), (1) is an abstract Fourier series.

**Linear least squares forecasting:**

of \(y_{t+\tau}\) based on \(y_s, s \leq t\)

project \(y_{t+\tau}^{(i)}\) on \(H_y(t)\):

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

\(\underline{\text{forecast } \hat{y}_{t,\tau}}\)

\(\underline{\text{forecasting error}}\)

**Spectral factorization:**

The spectral density of \((y_t)\) is given by

\[
f_y = (2\pi)^{-1} \cdot \left( \sum_j k_j e^{-i\lambda j} \right) \cdot \Sigma \cdot k(e^{-i\lambda})^* \quad (2)
\]

\(\underline{k(e^{-i\lambda})}\)

**Question:** Obtain \(k\) (and \(\Sigma\)) from \(f_y\) (in 2).
8. Estimation I

Estimation of mean
\[ \bar{y}_T = \frac{1}{T} \sum_{t=1}^{T} y_t \]

Estimation of covariances (ass. \( \mathbb{E} y_t = 0 \))
\[ \hat{\gamma}(t) = \frac{1}{T} \sum_s y_{t+s} y_s^* \]

Estimation of spectra
Periodogram
\[ I_T(\lambda) = \sum \hat{\gamma}(t) e^{-i\lambda t} \]
is not consistent; smoothed periodogram
9. Rational spectra, ARMA and state-space systems

ARMA system

\[ \sum_{j=0}^{p} a_j y_{t-j} = \sum_{j=0}^{p} b_j \epsilon_{t-j} \]

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

\( z \): backward shift as well as complex variable

Stability condition
\[ \det a(z) \neq 0 \quad |z| \leq 1 \]

Miniphase condition
\[ \det b(z) \neq 0 \quad |z| < 1 \]

Normalization
\[ a_0 = b_0 \]

Steady state solution
\[ y_t = a^{-1}(z)b(z) \epsilon_t \]

Transfer function \( k(z) \)
This transfer function corresponds to the Wold decomposition.

State space system

\[
\begin{align*}
    x_{t+1} & = Ax_t + B \epsilon_t \\
    y_t & = Cx_t + \epsilon_t
\end{align*}
\]

\(x_t\): state (n-dimensional)

Stability condition

\[|\lambda_{max}(A)| < 1\]

Miniphase condition

\[|\lambda_{max}(A - BC)| \leq 1\]

Steady state solution

\[
y_t = (C(Iz^{-1} - A)^{-1} B + I) \epsilon_t
\]

again corresponds to the Wold decomposition
Theorem:

(i) Every rational and a.e. nonsingular spectral density matrix may be uniquely factorized (as in (2)), where $k(z)$ is rational (in $z \in \mathbb{C}$), analytic within a circle containing the closed unit disk, $\det k(z) \neq 0$, $|z| < 1$ and $\Sigma > 0$.

(ii) For every rational transfer function $k$ satisfying the above mentioned properties there is a stable and miniphase ARMA system with $a_0 = b_0$ and conversely every such ARMA system has a transfer function with the properties mentioned in (i).

(iii) A completely analogous statement holds for state space systems
PART III: Identification of linear systems

10. Problem statement

Data driven modeling: Find a good model from (noisy) data

One has to specify:

• The model class, i.e. the class of all a priori feasible candidate systems to be fitted to the data. Here the model class is the set of all stable and miniphase ARMA or state space systems (for given $s$)

• The class of feasible data

• An identification procedure, which is a set of rules - in the fully automatized case a function - attaching to every feasible data string $y_t, t = 1, \ldots, T$ a system from the model class.

The theory of identification is mainly concerned with the development and evaluation of identification algorithms.
Steps in actual identification

- Data generation and preprocessing of data (e.g. removing outliers)
- Description of the model class using the prior knowledge available
- Identifying the model
- Model validation

In identification in general the following parameters have to be determined from data

- Integer-valued parameters such as the state dimension of a minimal state space system; this defines a subclass, namely the class of all systems of order \( n \)
- Real-valued parameters, such as the entries in \((A,B,C)\)

Semi-nonparametric estimation problem
3 modules of the problem:

- Structure theory: Idealized identification, we commence from the population second moments of the observations or from the (“true”) transfer functions rather than from data

- Estimation of real-valued parameters for given integer-valued parameters

- Model selection: Estimation of integer-valued parameters
11. Structure Theory

Note: (2) defines a one-to-one relation between $f$ and $k$, $\Sigma$ under our assumptions.

We restrict ourselves to state space systems:
Let $U_A$ denote the set of all rational $s \times s$ transfer functions $k(z)$ satisfying our assumptions and let $T_A$ denote the set of all state space systems $(A, B, C)$ (for fixed $s$ but variable $n$) satisfying our assumptions; finally let the mapping $\pi : T_A \rightarrow U_A$ be defined by

$$\pi(A, B, C) = C(Iz^{-1} - A)^{-1}B + I$$
• \( \pi \) is not injective (no identifiability)
  \( \pi^{-1}(k') \) the class of observationally equivalent systems

• There exists no continuous selection of representatives from the equivalence classes

“illposedness”
Identifiability and continuity of the parametrization are desirable: $U_A$ and $T_A$ are broken into bits $U_\alpha$ and $T_\alpha$ respectively s.t. $\pi/T_\alpha$ is injective and surjective and its inverse, the parametrization

$$\psi_\alpha : U_\alpha \rightarrow T_\alpha \subset \mathbb{R}^{d_\alpha}$$

is continuous.

Free parameters $\mathbb{R}^{d_\alpha} \ni \tau_\alpha \leftrightarrow (A, B, C)$ for given $T_\alpha$. 
12. Estimation of real-valued parameters

(Gaussian) Likelihood function \((-2T^{-1} \times \log)\)

\[
L_T(\tau_\alpha, \Sigma) = T^{-1} \log \det \Gamma_T(\tau_\alpha, \Sigma) \\
+ T^{-1} y'(T) \Gamma_T^{-1}(\tau_\alpha, \Sigma) y(T)
\]

where

\[
y(T) = (y'_1, \ldots, y'_T)' \text{ (stacked sample)}
\]

\[
\Gamma_T(\tau_\alpha, \Sigma) = \left( \int e^{-i\lambda(r-t)} f_y(\lambda; \tau_\alpha, \Sigma) d\lambda \right)_{r,t=1,\ldots,T}
\]

ML estimators:

\[
(\hat{\tau}_\alpha, T, \hat{\Sigma}_T) = \arg \min_{\tau_\alpha \in T_\alpha, \Sigma \in \Sigma} L_T(\tau_\alpha, \Sigma)
\]
Coordinate free MLE: $\hat{k}_T$

Asymptotic properties:

- **Consistency:**
  \[ \hat{k}_T \to k_0 \]
  \[ \hat{\Sigma}_T \to \Sigma_0 \]

- **Asymptotic normality**
  \[ \sqrt{T}(\hat{\tau}_{\alpha,T} - \tau_{\alpha,0}) \to N(0, V) \]
13. Model Selection

Example: Estimation of $n$, analogous for $\alpha$

Information criteria: Tradeoff between fit and complexity

$$I_T(n) = \log \det \hat{\Sigma}_T(n) + (2n_s) \frac{c(T)}{T}$$

Measure of fit  dimension, measures complexity

$$\hat{n}_T = \arg \min I_n$$

AIC criterion $c(T) = 2$

BIC criterion $c(T) = \log(T)$

BIC is consistent, AIC not

Post model selection properties