Final Exam

- Open-book.
- Covers all of the course.
- Best four out of five questions.

Introduction to Time Series Analysis: Review

1. Time series modelling.
2. Time domain.
   (a) Concepts of stationarity, ACF.
   (b) Linear processes, causality, invertibility.
   (c) ARMA models, forecasting, estimation.
   (d) ARIMA, seasonal ARIMA models.
3. Frequency domain.
   (a) Spectral density.
   (b) Linear filters, frequency response.
   (c) Nonparametric spectral density estimation.
   (d) Parametric spectral density estimation.
   (e) Lagged regression models.

Objectives of Time Series Analysis

1. Compact description of data. Example:

$$X_t = T_t + S_t + f(Y_t) + W_t.$$

2. Interpretation. Example: Seasonal adjustment.
3. Forecasting. Example: Predict unemployment.
4. Control. Example: Impact of monetary policy on unemployment.
5. Hypothesis testing. Example: Global warming.
6. Simulation. Example: Estimate probability of catastrophic events.

Time Series Modelling

1. Plot the time series.

   Look for trends, seasonal components, step changes, outliers.

2. Transform data so that residuals are **stationary**.

   (a) Estimate and subtract T_t, S_t.

   (b) Differencing.

   (c) Nonlinear transformations (log, $\sqrt{\cdot}$).

3. Fit model to residuals.

1. Time series modelling.
2. **Time domain.**
   (a) Concepts of stationarity, ACF.
   (b) Linear processes, causality, invertibility.
   (c) ARMA models, forecasting, estimation.
   (d) ARIMA, seasonal ARIMA models.
3. Frequency domain.
   (a) Spectral density.
   (b) Linear filters, frequency response.
   (c) Nonparametric spectral density estimation.
   (d) Parametric spectral density estimation.
   (e) Lagged regression models.

Stationarity

$\{X_t\}$ is **strictly stationary** if, for all $k, t_1, \ldots, t_k, x_1, \ldots, x_k$, and h,

$$P(X_{t_1} \leq x_1, \ldots, X_{t_k} \leq x_k) = P(x_{t_1+h} \leq x_1, \ldots, X_{t_k+h} \leq x_k).$$

i.e., shifting the time axis does not affect the distribution.

We consider **second-order properties** only:
$\{X_t\}$ is stationary if its mean function and autocovariance function satisfy

$$\mu_x(t) = \mathrm{E}[X_t] = \mu,$$
$$\gamma_x(s,t) = \mathrm{Cov}(X_s, X_t) = \gamma_x(s-t).$$

NB: Constant variance: $\gamma_x(t,t) = \mathrm{Var}(X_t) = \gamma_x(0)$.

ACF and Sample ACF

The **autocorrelation function (ACF)** is

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \mathrm{Corr}(X_{t+h}, X_t).$$

For observations $x_1, \ldots, x_n$ of a time series,

the **sample mean** is $$\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t.$$

The **sample autocovariance function** is

$$\hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-|h|}(x_{t+|h|} - \bar{x})(x_t - \bar{x}), \qquad \text{for } -n < h < n.$$

The **sample autocorrelation function** is $\hat{\rho}(h) = \hat{\gamma}(h)/\hat{\gamma}(0)$.

Linear Processes

An important class of stationary time series:

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j W_{t-j}$$

where $\{W_t\} \sim WN(0, \sigma_w^2)$

and μ, ψ_j are parameters satisfying

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty.$$

e.g.: ARMA(p,q).

Causality

A linear process $\{X_t\}$ is **causal** (strictly, a **causal function of** $\{W_t\}$) if there is a

$$\psi(B) = \psi_0 + \psi_1 B + \psi_2 B^2 + \cdots$$

$$\text{with} \qquad \sum_{j=0}^{\infty} |\psi_j| < \infty$$

$$\text{and} \qquad X_t = \psi(B) W_t.$$

Invertibility

A linear process $\{X_t\}$ is **invertible** (strictly, an **invertible function of** $\{W_t\}$) if there is a

$$\pi(B) = \pi_0 + \pi_1 B + \pi_2 B^2 + \cdots$$

$$\text{with} \qquad \sum_{j=0}^{\infty} |\pi_j| < \infty$$

$$\text{and} \qquad W_t = \pi(B) X_t.$$

Polynomials of a complex variable

Every degree p polynomial $a(z)$ can be factorized as

$$a(z) = a_0 + a_1 z + \cdots + a_p z^p = a_p(z - z_1)(z - z_2)\cdots(z - z_p),$$

where $z_1, \ldots, z_p \in \mathbb{C}$ are called the roots of $a(z)$. If the coefficients $a_0, a_1, \ldots, a_p$ are all real, then c is real, and the roots are all either real or come in complex conjugate pairs, $z_i = \bar{z}_j$.

Autoregressive moving average models

An **ARMA(p,q) process** $\{X_t\}$ is a stationary process that satisfies

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = W_t + \theta_1 W_{t-1} + \cdots + \theta_q W_{t-q},$$

where $\{W_t\} \sim WN(0, \sigma^2)$.

Also, $\phi_p, \theta_q \neq 0$ and $\phi(z), \theta(z)$ have no common factors.

Properties of ARMA(p,q) models

Theorem: If ϕ and θ have no common factors, a (unique) *stationary* solution to $\phi(B)X_t = \theta(B)W_t$ exists iff

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p = 0 \;\Rightarrow\; |z| \neq 1.$$

This ARMA(p,q) process is *causal* iff

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p = 0 \;\Rightarrow\; |z| > 1.$$

It is *invertible* iff

$$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q = 0. \;\Rightarrow\; |z| > 1.$$

Properties of ARMA(p,q) models

$$\phi(B)X_t = \theta(B)W_t, \qquad \Leftrightarrow \qquad X_t = \psi(B)W_t$$

$$\text{so} \qquad \theta(B) = \psi(B)\phi(B)$$

$$\Leftrightarrow \; 1 + \theta_1 B + \cdots + \theta_q B^q = (\psi_0 + \psi_1 B + \cdots)(1 - \phi_1 B - \cdots - \phi_p B^p)$$

$$\begin{aligned} \Leftrightarrow \qquad 1 &= \psi_0, \\ \theta_1 &= \psi_1 - \phi_1\psi_0, \\ \theta_2 &= \psi_2 - \phi_1\psi_1 - \cdots - \phi_2\psi_0, \\ &\vdots \end{aligned}$$

This is equivalent to $\theta_j = \phi(B)\psi_j$, with $\theta_0 = 1$, $\theta_j = 0$ for $j < 0$, $j > q$.

Linear prediction

Given $X_1, X_2, \ldots, X_n$, the best linear predictor

$$X_{n+m}^n = \alpha_0 + \sum_{i=1}^n \alpha_i X_i$$

of X_{n+m} satisfies the **prediction equations**

$$\mathrm{E}\left(X_{n+m} - X_{n+m}^n\right) = 0$$

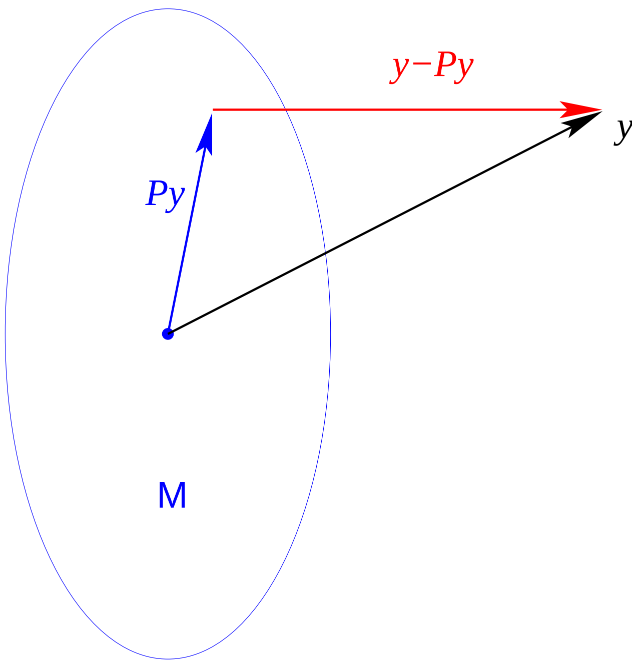
$$\mathrm{E}\left[\left(X_{n+m} - X_{n+m}^n\right) X_i\right] = 0 \qquad \text{for } i = 1, \ldots, n.$$

That is, the *prediction errors* ($X_{n+m}^n - X_{n+m}$) *are uncorrelated with the prediction variables* ($1, X_1, \ldots, X_n$).

Projection Theorem

If $\mathcal{H}$ is a Hilbert space,
$\mathcal{M}$ is a closed linear subspace of $\mathcal{H}$,
and $y \in \mathcal{H}$,
then there is a point $Py \in \mathcal{M}$
(the **projection of** y **on** $\mathcal{M}$)
satisfying

1. $\|Py - y\| \leq \|w - y\|$ for $w \in \mathcal{M}$,
2. $\|Py - y\| < \|w - y\|$ for $w \in \mathcal{M}$, $w \neq y$
3. $\langle y - Py, w \rangle = 0$ for $w \in \mathcal{M}$.



One-step-ahead linear prediction

$$X_{n+1}^n = \phi_{n1}X_n + \phi_{n2}X_{n-1} + \cdots + \phi_{nn}X_1$$

$$\Gamma_n \phi_n = \gamma_n,$$

$$P_{n+1}^n = \mathrm{E}\left(X_{n+1} - X_{n+1}^n\right)^2 = \gamma(0) - \gamma_n' \Gamma_n^{-1} \gamma_n,$$

$$\text{with } \Gamma_n = \begin{bmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & & \gamma(n-2) \\ \vdots & & \ddots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \cdots & \gamma(0) \end{bmatrix},$$

$$\phi_n = (\phi_{n1}, \phi_{n2}, \ldots, \phi_{nn})', \quad \gamma_n = (\gamma(1), \gamma(2), \ldots, \gamma(n))'.$$

The innovations representation

Write the best linear predictor as

$$X_{n+1}^n = \theta_{n1} \underbrace{\left(X_n - X_n^{n-1}\right)}_{\text{innovation}} + \theta_{n2}\left(X_{n-1} - X_{n-1}^{n-2}\right) + \cdots + \theta_{nn}\left(X_1 - X_1^0\right).$$

The innovations are uncorrelated:
$\mathrm{Cov}(X_j - X_j^{j-1}, X_i - X_i^{i-1}) = 0$ for $i \neq j$.

Yule-Walker estimation

Method of moments: We choose parameters for which the moments are equal to the empirical moments.

In this case, we choose ϕ so that $\gamma = \hat{\gamma}$.

Yule-Walker equations for $\hat{\phi}$:
$$\begin{cases} \hat{\Gamma}_p \hat{\phi} = \hat{\gamma}_p, \\ \hat{\sigma}^2 = \hat{\gamma}(0) - \hat{\phi}' \hat{\gamma}_p. \end{cases}$$

These are the forecasting equations.
Recursive computation: Durbin-Levinson algorithm.

Maximum likelihood estimation

Suppose that $X_1, X_2, \ldots, X_n$ is drawn from a zero mean Gaussian ARMA(p,q) process. The likelihood of parameters $\phi \in \mathbb{R}^p$, $\theta \in \mathbb{R}^q$, $\sigma_w^2 \in \mathbb{R}_+$ is defined as the density of $X = (X_1, X_2, \ldots, X_n)'$ under the Gaussian model with those parameters:

$$L(\phi, \theta, \sigma_w^2) = \frac{1}{(2\pi)^{n/2} \left|\Gamma_n\right|^{1/2}} \exp\left(-\frac{1}{2} X' \Gamma_n^{-1} X\right),$$

where $|A|$ denotes the determinant of a matrix A, and Γ_n is the variance/covariance matrix of X with the given parameter values.

The maximum likelihood estimator (MLE) of ϕ, θ, σ_w^2 maximizes this quantity.

Maximum likelihood estimation

The MLE $(\hat{\phi}, \hat{\theta}, \hat{\sigma}_w^2)$ satisfies

$$\hat{\sigma}_w^2 = \frac{S(\hat{\phi}, \hat{\theta})}{n},$$

and $\hat{\phi}, \hat{\theta}$ minimize

$$\log\left(\frac{S(\hat{\phi}, \hat{\theta})}{n}\right) + \frac{1}{n}\sum_{i=1}^{n} \log r_i^{i-1},$$

where $r_i^{i-1} = P_i^{i-1}/\sigma_w^2$ and

$$S(\phi, \theta) = \sum_{i=1}^{n} \frac{\left(X_i - X_i^{i-1}\right)^2}{r_i^{i-1}}.$$

Integrated ARMA Models: ARIMA(p,d,q)

For $p, d, q \geq 0$, we say that a time series $\{X_t\}$ is an **ARIMA (p,d,q) process** if $Y_t = \nabla^d X_t = (1 - B)^d X_t$ is ARMA(p,q). We can write

$$\phi(B)(1 - B)^d X_t = \theta(B) W_t.$$

Multiplicative seasonal ARMA Models

For $p, q, P, Q \geq 0$, $s > 0$, $d, D > 0$, we say that a time series $\{X_t\}$ is a **multiplicative seasonal ARIMA model** (ARIMA(p,d,q)×(P,D,Q)$_s$)

$$\Phi(B^s)\phi(B)\nabla_s^D \nabla^d X_t = \Theta(B^s)\theta(B)W_t,$$

where the *seasonal difference operator of order* D is defined by

$$\nabla_s^D X_t = (1 - B^s)^D X_t.$$

1. Time series modelling.
2. Time domain.
   (a) Concepts of stationarity, ACF.
   (b) Linear processes, causality, invertibility.
   (c) ARMA models, forecasting, estimation.
   (d) ARIMA, seasonal ARIMA models.
3. **Frequency domain.**
   (a) Spectral density.
   (b) Linear filters, frequency response.
   (c) Nonparametric spectral density estimation.
   (d) Parametric spectral density estimation.
   (e) Lagged regression models.

Spectral density and spectral distribution function

If $\{X_t\}$ has $\sum_{h=-\infty}^{\infty} |\gamma_x(h)| < \infty$, then we define its **spectral density** as

$$f(\nu) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i \nu h}$$

for $-\infty < \nu < \infty$. We have

$$\gamma(h) = \int_{-1/2}^{1/2} e^{2\pi i \nu h} f(\nu)\, d\nu = \int_{-1/2}^{1/2} e^{2\pi i \nu h}\, dF(\nu),$$

where $dF(\nu) = f(\nu) d\nu$.

f measures how the variance of X_t is distributed across the spectrum.

Frequency response of a linear filter

If $\{X_t\}$ has spectral density $f_x(\nu)$ and the coefficients of the time-invariant linear filter ψ are absolutely summable, then $Y_t = \psi(B)X_t$ has spectral density

$$f_y(\nu) = \left|\psi\left(e^{2\pi i\nu}\right)\right|^2 f_x(\nu).$$

If ψ is a rational function, the transfer function is determined by the locations of its poles and zeros.

Sample autocovariance

The sample autocovariance $\hat{\gamma}(\cdot)$ can be used to give an estimate of the spectral density,

$$\hat{f}(\nu) = \sum_{h=-n+1}^{n-1} \hat{\gamma}(h) e^{-2\pi i \nu h}$$

for $-1/2 \leq \nu \leq 1/2$.

This is equivalent to the periodogram.

Periodogram

The periodogram is defined as

$$I(\nu) = |X(\nu)|^2 = \frac{1}{n}\left|\sum_{t=1}^{n} e^{-2\pi i t\nu} x_t\right|^2 = X_c^2(\nu) + X_s^2(\nu).$$

$$X_c(\nu) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \cos(2\pi t\nu) x_t,$$

$$X_s(\nu) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sin(2\pi t\nu_j) x_t.$$

Asymptotic properties of the periodogram

Under general conditions (e.g., gaussian, or linear process with rapidly decaying ACF), the $X_c(\nu_j)$, $X_s(\nu_j)$ are all asymptotically independent and $N(0, f(\nu_j)/2)$, and $f(\hat{\nu}^{(n)}) \to f(\nu)$, where $\hat{\nu}^{(n)}$ is the closest Fourier frequency (k/n) to the frequency ν.

In that case, we have

$$\frac{2}{f(\nu)} I(\hat{\nu}^{(n)}) = \frac{2}{f(\nu)} \left(X_c^2(\hat{\nu}^{(n)}) + X_s^2(\hat{\nu}^{(n)}) \right) \xrightarrow{\text{d}} \chi_2^2.$$

Thus, $\mathbf{E} I(\hat{\nu}^{(n)}) \to f(\nu)$, and $\text{Var}(I(\hat{\nu}^{(n)})) \to f(\nu)^2$.

Smoothed periodogram

If $f(\nu)$ is approximately constant in the band $[\nu_k - L/(2n), \nu_k + L/(2n)]$, the average of the periodogram over the band will be unbiased.

$$\hat{f}(\nu_k) = \frac{1}{L} \sum_{l=-(L-1)/2}^{(L-1)/2} I(\nu_k - l/n)$$

$$= \frac{1}{L} \sum_{l=-(L-1)/2}^{(L-1)/2} \left(X_c^2(\nu_k - l/n) + X_s^2(\nu_k - l/n) \right).$$

Then $\mathbf{E}\hat{f}(\hat{\nu}^{(n)}) \to f(\nu)$ and $\mathrm{Var}\hat{f}(\hat{\nu}^{(n)}) \to f^2(\nu)/L$.

Notice the *bias-variance trade off.*

Smoothed spectral estimators

$$\hat{f}(\nu) = \sum_{|j| \leq L_n} W_n(j) I(\hat{\nu}^{(n)} - j/n),$$

where the *spectral window function* satisfies $L_n \to \infty$, $L_n/n \to 0$, $W_n(j) \geq 0$, $W_n(j) = W_n(-j)$, $\sum W_n(j) = 1$, and $\sum W_n^2(j) \to 0$.

Then $\hat{f}(\nu) \to f(\nu)$ (in the mean square sense), and asymptotically

$$\hat{f}(\nu_k) \sim f(\nu_k) \frac{\chi_d^2}{d},$$

where $d = 2/\sum W_n^2(j)$.

Parametric spectral density estimation

Given data $x_1, x_2, \ldots, x_n$,

1. Estimate the AR parameters $\phi_1, \ldots, \phi_p, \sigma_w^2$.
2. Use the estimates $\hat{\phi}_1, \ldots, \hat{\phi}_p, \hat{\sigma}_w^2$ to compute the estimated spectral density:

$$\hat{f}_y(\nu) = \frac{\hat{\sigma}_w^2}{\left|\hat{\phi}\left(e^{-2\pi i\nu}\right)\right|^2}.$$

Parametric spectral density estimation

For large n,

$$\mathrm{Var}(\hat{f}(\nu)) \approx \frac{2p}{n} f^2(\nu).$$

Notice the *bias-variance trade off*.

Advantage over nonparametric: better *frequency resolution* of a small number of peaks. This is especially important if there is more than one peak at nearby frequencies.
Disadvantage: inflexibility (bias).

Lagged regression models

Consider a lagged regression model of the form

$$Y_t = \sum_{h=-\infty}^{\infty} \beta_h X_{t-h} + V_t,$$

where X_t is an observed input time series, Y_t is the observed output time series, and V_t is a stationary noise process.

This is useful for

- Identifying the (best linear) relationship between two time series.
- Forecasting one time series from the other.

Lagged regression in the time domain

$$Y_t = \alpha(B)X_t + \eta_t = \sum_{j=0}^{\infty} \alpha_j X_{t-j} + \eta_t,$$

1. Fit an ARMA model (with $\theta_x(B), \phi_x(B)$) to the input series $\{X_t\}$.
2. *Prewhiten* the input series by applying the inverse operator $\phi_x(B)/\theta_x(B)$.
3. Calculate the cross-correlation of $\tilde{Y}_t$ with W_t, $\gamma_{\tilde{y},w}(h)$, to give an indication of the behavior of $\alpha(B)$ (for instance, the delay).
4. Estimate the coefficients of $\alpha(B)$ and hence fit an ARMA model for the noise series η_t.

Coherence

Define the *cross-spectrum* and the *squared coherence* function:

$$f_{xy}(\nu) = \sum_{h=-\infty}^{\infty} \gamma_{xy}(h) e^{-2\pi i \nu h},$$

$$\gamma_{xy}(h) = \int_{-1/2}^{1/2} f_{xy}(\nu) e^{2\pi i \nu h} d\nu,$$

$$\rho^2_{y,x}(\nu) = \frac{|f_{yx}(\nu)|^2}{f_x(\nu) f_y(\nu)}.$$

Lagged regression models in the frequency domain

$$Y_t = \sum_{j=-\infty}^{\infty} \beta_j X_{t-j} + V_t,$$

We compute the Fourier transform of the series $\{\beta_j\}$ in terms of the cross-spectral density and the spectral density:

$$B(\nu) f_x(\nu) = f_{yx}(\nu).$$

$$MSE = \int_{-1/2}^{1/2} f_y(\nu) \left(1 - \rho_{yx}^2(\nu)\right) d\nu.$$

Thus, $\rho_{yx}(\nu)^2$ indicates how the component of the variance of $\{Y_t\}$ at a frequency ν is accounted for by $\{X_t\}$.

Introduction to Time Series Analysis: Review

1. Time series modelling.
2. Time domain.
   (a) Concepts of stationarity, ACF.
   (b) Linear processes, causality, invertibility.
   (c) ARMA models, forecasting, estimation.
   (d) ARIMA, seasonal ARIMA models.
3. Frequency domain.
   (a) Spectral density.
   (b) Linear filters, frequency response.
   (c) Nonparametric spectral density estimation.
   (d) Parametric spectral density estimation.
   (e) Lagged regression models.