Topics in time-series analysis

Models · Seasonal adjustment · Imputation

Compiled from session2.tex @ 2024-08-06 15:22:16+02:00.

Day 1: Introduction to time-series analysis

Andreï V. Kostyrka

3rd of April 2024

Université du Luxembourg

Presentation structure

1. Principles of forecasting

2. Non-stationary time series

3. Time-series models in open-source software

Principles of forecasting

Forecasting in ARMA models

Let Ω_t denote all the information available up to t: $(\{Y_t\}_{t=1}^T$, our smart guesses $\underline{Y}, \underline{U}$ and, thus, the conditional values of U_t denoted by $\hat{U}_t \coloneqq \hat{U}_t(\{Y_t\}_{t=1}^T, \underline{Y}, \underline{U})$.

Forecast: conditional expectation $\mathbb{E}(Y_{t+h} \mid \Omega_t)$.

- $\mathbb{E}(Y_{t+h} \mid \Omega_t)$ is the BLP in linear specifications
  - $\mathbb{E}(U_t \mid Y_{t-1}, \dots, Y_{t-p}) = 0 \Rightarrow$ assume $U_{t+h} = 0$ for $h \geq 1$
- Plug $\{(Y_t, \hat{U}_t)\}_{t=1}^T$ into $\hat{Y}_{t+1} \coloneqq \hat{\mu}_0 + \sum_{i=1}^p \hat{\varphi}_i Y_{t+1-i} + \sum_{j=1}^q \hat{\theta}_j \hat{U}_{t+1-j}$
  - Save $\hat{U}_{t+1} \coloneqq Y_{t+1} - \hat{Y}_{t+1}$ for further use
- 2-step forecast in AR(1) models: $\mathbb{E}(Y_t \mid \Omega_{t-2}) = \mathbb{E}(\mu_0 + \varphi_1 Y_{t-1} + U_t \mid \Omega_{t-2}) = \mu_0 + \varphi_1 \mathbb{E}(Y_{t-1} \mid \Omega_{t-2}) = \mu_0 + \varphi_1(\mu_0 + \varphi_1 Y_{t-2}) = \mu_0 + \varphi_1\mu_0 + \varphi_1^2 Y_{t-2}$
  - In AR models, $\mathbb{E}(Y_{t+h} \mid \Omega_t)$ converges to $\mu_0/(1 - \sum_{i=1}^p \varphi_i)$

Forecasting in SARMA models

In seasonal models, Y_{t-c} are in the equation ⇒ any shock to Y_{t-c} will be mechanistically repeated in the forecast.

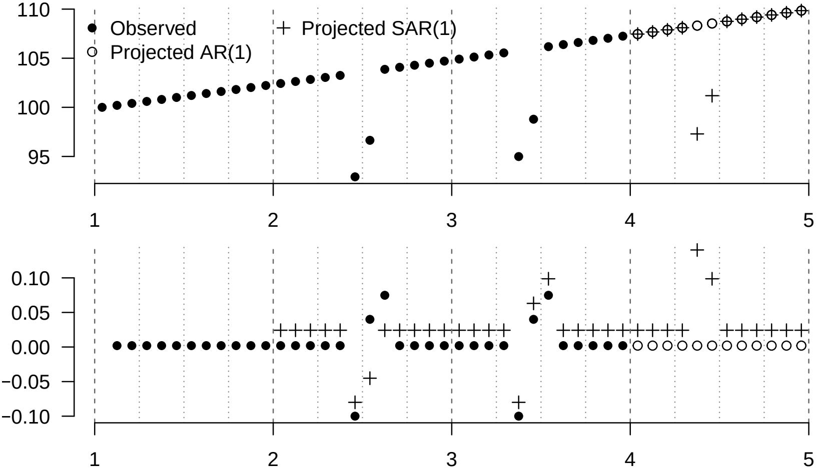
- SARMA models can induce *artificial seasonality* by reproducing the features of the previous year

Problem: year-on-year growth rates, $Y_t / Y_{t-c} - 1$, exhibit illusory 'miraculous recoveries' in economies. YoY figures worsen further if seasonality patterns change.

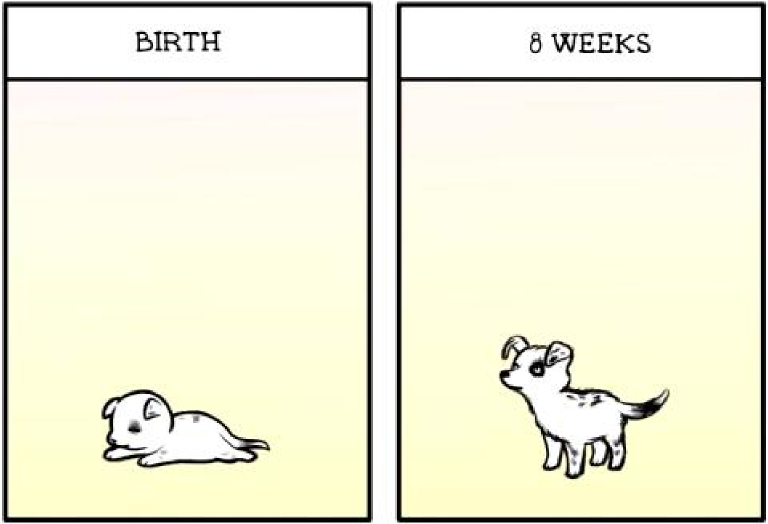
Example. Consumers organise fewer family gatherings ⇒ declining consumption of poultry on winter holidays ⇒ decreasing gap between December and the rest of the year.

Solution: accept that SARMA models are ultra-short-term, and *filter seasonality out* instead of adding distant lags.

SARMA forecast problem illustration

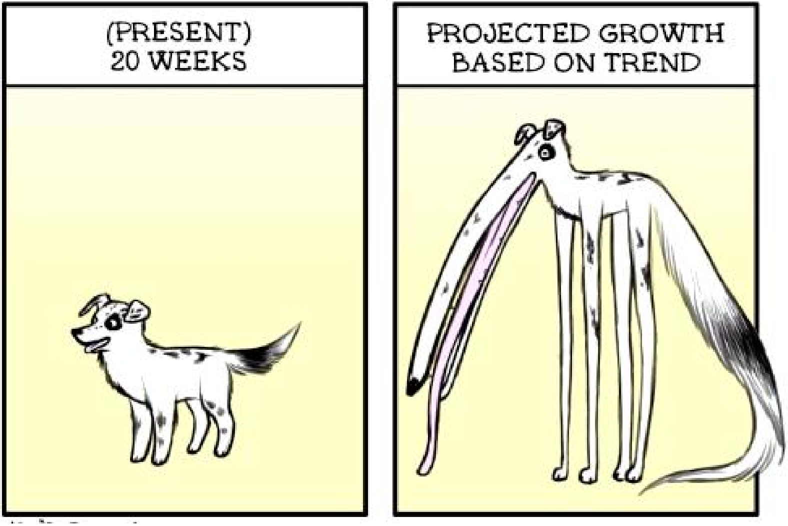


Perils of extrapolation



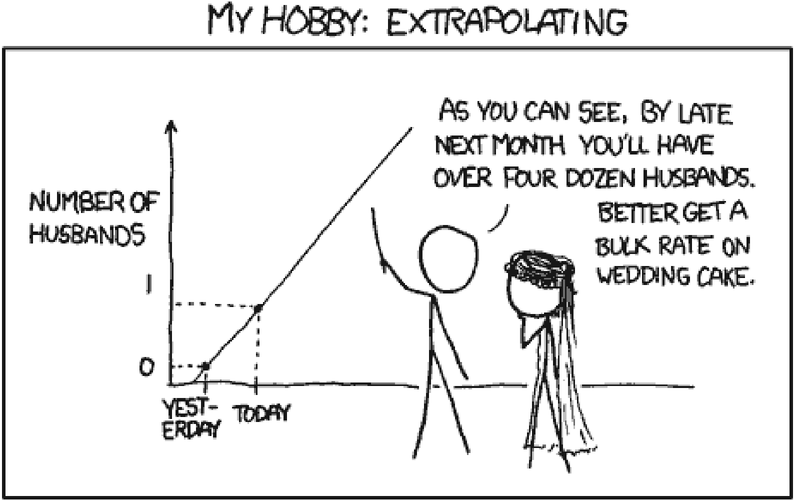
Credit: Jessie Robinson.

Perils of extrapolation



Credit: Jessie Robinson.

Apply extrapolation correctly



Credit: xkcd. Technical explanation.

Specification and distribution tests

- Specification: Ramsey's RESET (1969), Härdle–Mammen (1993), Zheng (1996), Ellison–Ellison (2000), Racine–Hart–Li (2006)…
- Distribution: Kolmogorov–Smirnov (1948), Shapiro–Wilk (1965), Jarque–Bera (1987), Diebold–Günther–Tay (1998), Berkowitz (2001), Kheifets (2014)…
- **Normality tests are silly:** all non-rejections in small samples, all rejections in large samples, normality is optional and yields nothing, and there are worse problems in small samples than non-normality
  - Never make decisions based on normality tests
- Visual tests are the best: look at Q-Q plots, density plots, histograms, scatter plots…

Adequacy test: Ljung–Box

In ARMA models, U_t should be white noise:

$$Y_t = \mu_0 + U_t + \sum_{i=1}^{p} \varphi_i Y_{t-i} + \sum_{j=1}^{q} \theta_j U_{t-j}$$

$\mathcal{H}_0 : \{U_t\}_{t=1}^T \sim \text{WN} \ \Rightarrow \ \text{Cor}(U_t, U_{t-h}) = 0 \quad \forall h \geq 1$
$\Rightarrow \ \forall i \in \overline{1, h}: \ \text{Cor}(\hat{U}_t, \hat{U}_{t-i}) = 0 \Rightarrow \sum_{i=1}^h \text{Cor}^2(\hat{U}_t, \hat{U}_{t-i}) = 0.$

- Obtain the residuals $\hat{U}_t$ and their autocorrelation estimates $\hat{\rho}$ up to order h
- Compute $\hat{Q} \coloneqq (T+2)\sum_{i=1}^h \frac{T}{T-i}\hat{\rho}_i^2$
- Reject $\mathcal{H}_0$ if $\hat{Q} > Q_{\chi^2_h}(1-\alpha)$ (choose $\alpha = 5\%$ or any other)

t-test for individual correlations: $\frac{\hat{\gamma}(h) - \gamma(h)}{\sqrt{\frac{1-\hat{\gamma}(h)^2}{n-h-1}}} \xrightarrow[T\to\infty]{d} \mathcal{N}(0, 1).$

Lag selection

Choosing p and q in ARMA models is up to the researcher.

- Too many lags ⇒ poor identification + estimation noise ⇒ the fitted ARMA process might become ill-behaved or close to non-stationary
  - ARMA$(2, 3)$ models often do not even converge
- Too few lags ⇒ unaccounted-for inertia ⇒ biased estimates of φ and θ
- Should one increase p in the AR part or q in the MA part for a better fit? Which one creates persistence?

Methods: (1) pulling out of thin air, (2) theory-backed, (3) based on visualisations of ACF/PACF, or (4) **data-driven**.

Information criteria for ARMA models

Higher p, q = more irrelevant old variables = same problem as R^2 in linear models: rubbish variables increase R^2 whilst contributing nothing but estimation noise and over-fitting.

General idea: encourage both goodness of fit and parsimony (i. e. penalise the number of parameters). Common penalties (the lower, the better):

- Akaike (1974): $\mathrm{AIC}(\theta) \coloneqq T^{-1}[-2\mathcal{L}_T(\theta) + 2\dim\theta]$
- Schwarz (1978): $\mathrm{BIC}(\theta) \coloneqq T^{-1}[-2\mathcal{L}_T(\theta) + \dim\theta \cdot \log T]$

In the Gaussian case, $-\frac{2}{T}\mathcal{L}_T(\theta) = \ln \widehat{\mathrm{Var}}\, U_t(\theta) = \ln \hat{\sigma}_U^2$.

NB. BIC is asymptotically consistent; AIC suggests over-parametrised models but may work better for small T.

Model selection

Iterative approach: choose a class of models ⇒ identify a preliminary model ⇒ estimate ⇒ check adequacy ⇒ revise if necessary ⇒ use in the application. The **most parsimonious** adequate model wins.

IC approach: estimate multiple models *on the same sample* (remove observations if necessary), calculate information criteria (AIC, BIC), select the one that minimises either.

Cross-validation approach: estimate multiple models on the train data ⇒ predict values on the test data ⇒ compare forecast accuracy (e. g. Diebold–Mariano) ⇒ choose the simplest model indistinguishable from the 'best' one.

Backtesting

Backtesting / TS cross-validation: assessing hypothetical historical performance as if one had estimated the model up to time $\tau < T$ and used forecasts for $\tau + 1, \dots, T$.

Example. How to choose a model out of these: AR(3), MA(2), ARMA(1, 1), and SARMA(0, 1)(1, 1), given $\{Y_t\}_{t=1}^{200}$?

1. Estimate all models on $t = 1, \dots, 150$
2. Using these estimates, produce rolling forecasts $\{\hat{Y}_t\}_{t=151}^{200}$
3. Make a decision based on these forecasts ('invest' in portfolios, re-balance inventory etc.), calculate the outcomes using real $\{Y_t\}_{t=151}^{200}$
4. Choose the model that led to the 'best' decision

General forecasting aspects

- Forecast combination: average the predictions of the best k models from a set with or without weights (Bates & Granger, 1969)
  - Forecast combination puzzle (Stock & Watson, 2004): in comparisons of point forecasts, a simple average forecast from multiple models with equal weights often out-performs more complicated weighting schemes where the optimal forecast weights are estimated
- In TSCV, re-estimating every k points is optional
  - Updating the specification / parameters may result in forecasts with greater variability (Spiliotis & Petropoulos, 2024)

Non-stationary time series

Stationary processes are lame

- It is easy to handle stationary processes: the theory is solid, and virtually all statistical techniques work well (WLLN, CLT, CMT etc.)
- Forecasting is somewhat meaningless: exponential(-ish) decay of the influence of the past
  - Forecasts converge to the unconditional mean pretty quickly $\Rightarrow$ uncommon to use horizons longer than 2

The struggle is about transforming real observed processes (with trends, seasonal fluctuations, jumps, volatility clustering etc.) into stationary ones, applying simple analysis / forecasting methods, and then ‘un-transforming’ the processed data.

Transforming inappropriate series

Real deal: finding a transformation that would

1. Be invertible
2. Convert the inputs into stationary processes
3. Not lose too much information in the process
4. Not suffer too much from Jensen's inequality: $f(\mathbb{E}X) \le \mathbb{E}f(X)$ for convex f, i. e. be not too non-linear
   - If $\ln Y_t = \tilde{X}_t'\theta + U_t$, the fitted values are $\widehat{\ln Y_t}$, but $\mathbb{E} \exp \widehat{\ln Y_t} \neq \mathbb{E}Y_t$! (The medians and quantiles are invariant to monotone transformations, though.)

If the model is estimated via Gaussian ML but the residuals do not look normal, consider the Box–Cox transform:

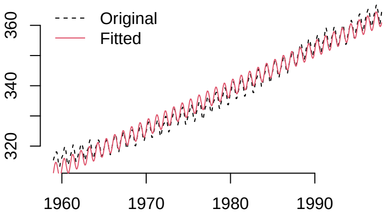
$$Y_t^{\text{BC}} = (Y_t^\lambda - 1)/\lambda \quad \text{or} \quad Y_t^{\text{BC}} = \ln Y_t$$

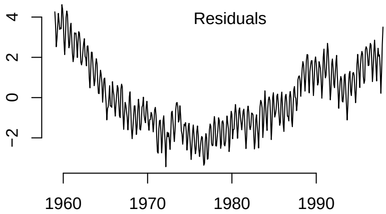
Deterministic processes

Processes with linear trends or harmonic components can be modelled by including deterministic regressors.

Consider the CO_2 concentration model for monthly data (cycle length $c = 12$): $Y_t = \beta_0 + \beta_1 t + \beta_2 \sin(\pi t/6 + \beta_3)$.

Estimates: $\hat\beta_1 = 0.11$, $\hat\beta_2 = 2.8$ (wave amplitude), $\hat\beta_3 = -0.66$.





Next step: tweak the trend specification (maybe spline?).

Lag operator

Introducing the lag operator L (in some books, B for **b**ackwards) that shifts the series by one period:

- $LY_t := Y_{t-1}$, $L^2 Y_t = Y_{t-2}$, …
- Distributive law: $(L^m + L^n)Y_t = Y_{t-m} + Y_{t-n}$
- Associative law: $L^m(L^n Y_t) = Y_{t-n-m} = L^n(L^m y_t)$
- $L^{-m} Y_t = Y_{t+m}$ (forward operator, F)

Powers of L simplify notation in ARMA models:

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + U_t \iff (1 - \varphi_1 L - \varphi_2 L^2) Y_t = U_t$$

NB. For brevity, we drop μ because Y_t can be de-meaned.

Lag polynomial

Fundamental theorem of algebra: the polynomial $1 - \varphi_1 x - \varphi_2 x^2$ has 2 real or complex roots $\tilde{\varphi}_1, \tilde{\varphi}_2$, therefore,

$$(1 - \varphi_1 L - \varphi_2 L^2)Y_t = (1 - \tilde{\varphi}_1 L)(1 - \tilde{\varphi}_2)Y_t,$$

Another leap of faith (do not worry if this is confusing): one can treat these lag polynomials as simple algebraic polynomials, e. g. divide by them.

Example: an AR(2) process can be written as

$$(1 - \tilde{\varphi}_1 L)(1 - \tilde{\varphi}_2 L)Y_t = U_t \iff Y_t = \frac{U_t}{(1 - \tilde{\varphi}_1 L)(1 - \tilde{\varphi}_2 L)}$$

Lag-operator form

Factoring the polynomial of the AR(2) model:

$$Y_t = 0.8Y_{t-1} - 0.12Y_{t-2} + U_t$$

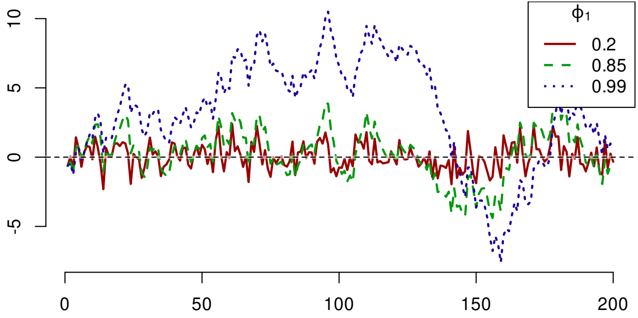
$$(1 - 0.8L + 0.12L^2)Y_t = U_t$$

$$(1 - 0.6L)(1 - 0.2L)Y_t = U_t$$

Stationarity of an AR(1) process

$$Y_t = \varphi_1 Y_{t-1} + U_t, \quad U_t \sim \mathcal{N}(0, 1)$$

Consider various degrees of autocorrelation:



The higher φ_1, the less frequently the process crosses the zero line (the higher the persistence).

Example: a dog on a leash

Example: a dog on a leash

Stationarity of ARMA models

ARMA(p, q) models in lag-operator form:

$$\underbrace{(1 - \varphi_1 L - \ldots - \varphi_p L^p)}_{\Phi(L)} Y_t = \underbrace{(1 - \theta_1 L - \ldots - \theta_q L^p)}_{\Theta(L)} U_t$$

Since $U_t \sim \text{WN}$, any finite sum of its lags is stationary (by the properties of white noise). Therefore, an ARMA process is stationarity if the AR part is stationary.

Example: $Y_t = 1.1 Y_{t-1} + U_t$ is non-stationary (explosive).

- For $p = 1$, Y_t is stationary if $|\varphi_1| < 1$
- For $p > 1$, Y_t is stationary if the roots of the polynomial $1 - \varphi_1 x - \ldots - \varphi_p x^p$ lie *outside* the unit circle, i. e. in the form $(1 - \tilde\varphi_1 x)(1 - \tilde\varphi_2 x) \cdots (1 - \tilde\varphi_p x)$, all $|\tilde\varphi_i| < 1$

Unit roots

Consider the process

$$Y_t = Y_{t-1} + U_t \iff (1 - \varphi_1 L) Y_t = U_t \quad \text{where } \varphi_1 = 1$$

It is called a **random walk** (drunkard's walk).

If in the lag-operator form of the AR part of an ARMA process, $(1 - \tilde{\varphi}_1 L) \cdots (1 - \tilde{\varphi}_p L) Y_t = \ldots$ contains at least one $\tilde{\varphi}_i = 1$, then this process is said to contain a **unit root**.

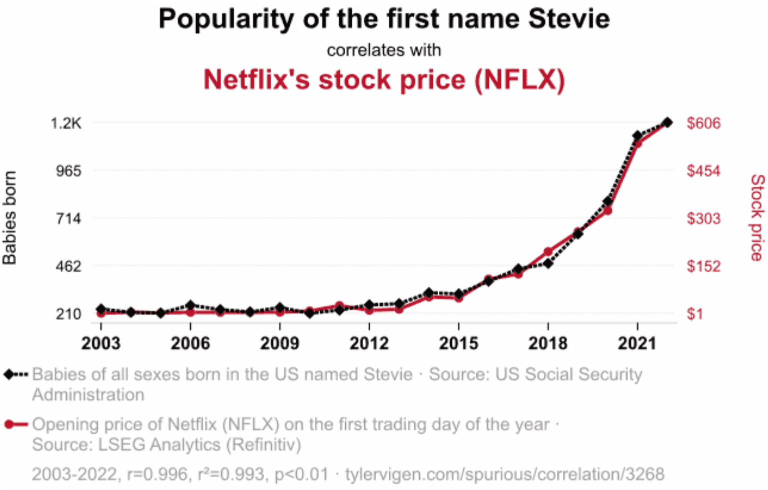
Any process with a unit root is **non-stationary**.

Problems with unit roots

- For AR(1), $\hat{\varphi}_1 = \hat{\gamma}_1 / \hat{\gamma}_0$ is consistent and asymptotically normal: $\sqrt{T}(\hat{\varphi}_1 - \varphi_1) \xrightarrow[T\to\infty]{d} \mathcal{N}(0, 1 - \varphi_1^2)$ for $|\varphi_1| < 1$
  - Inference breaks if $\varphi_1 = 1$
- When $Y_t = Y_{t-1} + U_t = \sum_{t=1}^{T} U_t$ (if $Y_0 = 0$), $U_t \sim \mathrm{WN}(0, \sigma^2)$, the variance of Y_t explodes: $\operatorname{Var} Y_t = t\sigma^2 \xrightarrow{t\to\infty} \infty$
- Regression estimators have non-standard distributions involving Wiener processes, Brownian bridges, and non-$\sqrt{T}$ rate of convergence
- Shocks persist infinitely, the influence of the past does not decay, the process is sensitive to initial conditions

Goal: convert non-stationary processes to stationary ones for the purposes of analysis.

Implications of non-stationarity



Credit: https://www.tylervigen.com/spurious-correlations.

Spurious regression

Seminal article: Granger & Newbold (1974).

> *If one's variables are random or near-random walks, and one includes variables which should not be included, it will be the rule rather than the exception to find spurious relationships.*
> *Also: a high value for R^2, combined with a high value of $\hat{\rho}$, is no indication of a true relationship.*

Nelson & Plosser (1982) analysed 14 series commonly used in business-cycle analysis and concluded:

> *If we are observing stationary deviations from linear trends in these series, then, the tendency to return to the trend must be so weak as to avoid detection even in samples as long as 60–100+ years.*

Witnessing spurious relationships

Simulation time!

- Trends create spurious relationships (bias)
- Seasonality creates spurious relationships (bias)
- Non-stationarity creates spurious relationships (non-standard distribution, inflated significance)

Dickey–Fuller test idea

Consider the potentially non-stationary process

$$Y_t = \rho_1 Y_{t-1} + U_t, \quad U_t \sim \mathrm{WN}(0, \sigma^2)$$

Subtract Y_{t-1} from both sides:

$$\Delta Y_t = (\rho_1 - 1) Y_{t-1} + U_t \coloneqq \alpha_1 Y_{t-1} + U_t$$

$\rho_1 = 1$ (unit root) $\iff \alpha_1 = 0$

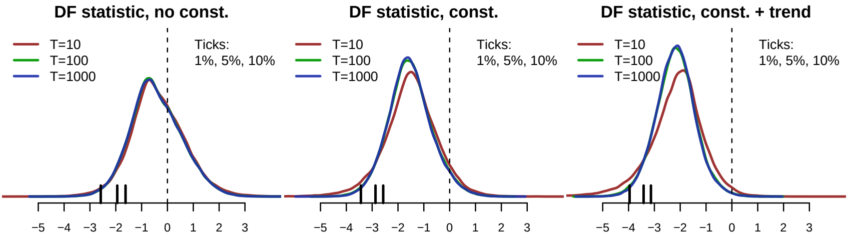
Null hypothesis: $\alpha_1 = 0$ (deviation from the standard paradigm '$\mathcal{H}_0$ = there is no effect / no complication / the world is beautiful').

DF test statistics

We can estimate this AR(1) process with an OLS regression of ΔY_t onto Y_{t-1}. Then, '$\mathcal{H}_0 \colon \alpha_1 = 0$' can be verified via a t-test.

- The asymptotics of $\hat{t} \coloneqq \frac{\hat{\alpha}_1}{\sqrt{\widehat{\mathrm{Var}}\,\hat{\alpha}_1}}$ are non-standard (tables should be used)
- If the RW has a non-zero mean, i. e. $\Delta Y_t = \mu + \alpha_1 Y_{t-1} + U_t$, then, the observed $\hat{t}$, often denoted by $\hat{t}_\mu$, has a different non-standard distribution
- If the RW has a trend, i. e. $\Delta Y_t = \mu_0 + \tau_0 t + \alpha_1 Y_{t-1} + U_t$, then, the observed $\hat{t}$, often denoted by $\hat{t}_\tau$, has another non-standard distribution

DF statistic distribution



How does one select between $\hat{t}$, $\hat{t}_\mu$, and $\hat{t}_\tau$?

- Visually: no trend in the original series $\Rightarrow \hat{t}$, linear trend $\Rightarrow \hat{t}_\mu$, quadratic trend $\Rightarrow \hat{t}_\tau$
- Formally: if the differences are stationary ($\alpha_1 = 0$), $(\hat{\alpha}, \hat{\mu}, \hat{\tau})$ are asymptotically normal $\Rightarrow$ F-test for $(\alpha_1, \mu_0, \tau_0) = 0$ or $(\alpha_1, \tau_0) = 0$

Augmented Dickey–Fuller (ADF) test (1984)

- The DGP in the standard DF test is too restrictive: very few real phenomena can be accurately approximated by an AR(1) process
- Generalisations: consider an AR(p) model for differenced series:

$$\Delta Y_t = [\mu +]\ [\tau t +] \alpha_1 Y_{t-1} + \varphi_1 \Delta Y_{t-1} + \dots + \varphi_p \Delta Y_{t-p} + U_t$$

- Same critical values of $\hat{\alpha}_1 / (\widehat{\mathrm{SE}}\, \hat{\alpha}_1)$ as the DF statistic
- To choose p, start with a large p^*, test the significance of φ_{p^*}, reduce p^* until a significant lag appears (Hall, 1994)
  - Low size distortion but low power
- Alternative: determine the number of lags via AIC
  - Higher power but larger distortions

Phillips–Perron test (1988)

In DF and ADF tests, it is assumed that $U_t \sim \mathrm{WN}(0, \sigma^2)$, but what if U_t is heteroskedastic and autocorrelated?

- Assume only $Y_t = \varphi_1 Y_{t-1} + U_t$
- For inference, use a heteroskedasticity- and autocorrelation-robust (HAC) estimator of $\widehat{\mathrm{Var}}\, U_t$
  - Same as Newey–West (1987) HAC estimator
- Same $\mathcal{H}_0$ and asymptotic critical values as in the DF test
- Slightly worse performance than ADF test in finite samples, but a modification (Perron & Ng, 1996) exists

KPSS stationarity test (1992)

- $\mathcal{H}_0$: $Y_t = \mu_0 + \tau_0 t + U_t$, U_t is stationary
- $\mathcal{H}_1$: $Y_t = \mu_0 + \tau_0 t + U_t + \sum_{i=1}^t V_i$, V_t is stationary, $\operatorname{Var} V_i > 0$

Test statistic for $\operatorname{Var} V = 0$ using the HAC-robust (Newey–West-like) $\widehat{\operatorname{Var}}\, U$:

$$\widehat{\mathrm{KPSS}} := \sum_{t=1}^{T} \frac{\left(T^{-1} \sum_{i=1}^t \hat{U}_i\right)^2}{T^{-1} \sum_{i=1}^t U_t + \text{covariances}} = \sum_{t=1}^{T} \frac{\left(T^{-1} \sum_{t=1}^T \hat{U}_t\right)^2}{\widehat{\operatorname{Var}}_{\mathrm{HAC}}\, U}$$

For stationary processes, the denominator converges to $\operatorname{Var} U$, the numerator is bounded $\Rightarrow$ do not reject $\mathcal{H}_0$ for $\widehat{\mathrm{KPSS}}$ (another tabulated distribution).

Frisch–Waugh–Lovell theorem

$$Y = \alpha + \beta_1 X_1 + \ldots + \beta_k X_k + \gamma Z + U, \quad \mathbb{E}(U \mid X_1, \ldots, X_k, Z) = 0$$

Consider two estimation approaches:

- One linear regression: project Y onto $X_1, \ldots, X_k, Z$, get $\hat\beta_1, \ldots, \hat\beta_k, \hat\gamma$
- Many linear regressions
  1. Regress each $Y, X_1, \ldots, X_k$ onto Z individually, get projection residuals $\tilde{A} \coloneqq A - \mathrm{BLP}(A \mid Z)$
  2. Regress $\tilde{Y}$ onto $\tilde{X}_1, \ldots, \tilde{X}_k$ (without Z), get $\tilde\beta_1, \ldots, \tilde\beta_k$

Theorem (Yule, 1907). $\hat\beta_i \equiv \tilde\beta_i$ for all i.

Generalisation: works with any regressor partition – regress each $Y, X_1, \ldots, X_k$ onto $(Z_1, \ldots, Z_l)$

Partially linear model

Consider the partially linear model:

$$Y = X'\theta_0 + f(Z) + U, \quad \mathbb{E}(U \mid X, Z) = 0$$

$f(\cdot)$ is a completely unknown function, yet we can consistently estimate θ_0 (Robinson, 1988, Ecta):

$$Y - \mathrm{BP}(Y \mid Z) = [X - \mathrm{BP}(X \mid Z)]'\theta_0 + U$$

1. Regress Y non-parametrically onto Z, get the residuals $\tilde{Y}$
   - Best predictor = conditional expectation; estimate via local methods, e. g. kernel or spline regression
2. Regress X non-parametrically onto Z, get the residuals $\tilde{X}$
3. Regress $\tilde{Y}$ onto $\tilde{X}$ to get $\hat{\theta}$
   - Regressing Y onto $\tilde{X}$ is asymptotically equivalent because $\tilde{X}$ is orthogonal to any $f(Z)$; finite-sample results might vary

FWL theorem and seasonality

Consider a linear relation where each variable has its own seasonal component orthogonal to the non-seasonal one:

$$\underbrace{(Y_t^{[NS]} + Y_t^{[S]})}_{Y_t} = \underbrace{(X_t^{[NS]} + X_t^{[S]})'}_{X_t} \theta_0 + U_t, \quad \mathbb{E}(U \mid X) = 0$$

Any non-zero correlation between $S_t^{[Y]}$ and $S_t^{[X]}$ results in non-zero $\hat{\theta}$ even if the co-movement between Y and X is solely due to $\mathrm{Cor}(S_t^{[Y]}, S_t^{[X]}) \neq 0$. Avoid spurious conclusions:

L'hiver partit, l'été arriva — merci au parti communste pour cela.
The winter's gone, the summer's ablaze – it is the Party that we should praise. *(Old joke.)*

Causal analysis **requires** seasonal adjustment.

Partial auto-correlation function

Partial correlation: correlation between the parts of X and Y not explained by linear functions of Z, i. e. between projection residuals after *partialling out* the effects of other variables.

$$\mathrm{Cor}(X, Y \mid Z) \coloneqq \mathrm{Cor}[X - \mathrm{BLP}(X \mid Z), Y - \mathrm{BLP}(Y \mid Z)]$$

In the TS context, it is the correlation between the parts of Y_t and Y_{t-h} that do not correlate with $Y_{t-h+1}, \dots, Y_{t-1}$.
E. g. $\mathrm{PCor}(Y_t, Y_{t-1}) = \mathrm{Cor}(Y_t, Y_{t-1})$, but

$$\begin{aligned}\mathrm{PCor}(Y_t, Y_{t-2}) &\coloneqq \mathrm{Cor}(Y_t, Y_{t-2} \mid Y_{t-1}) = \\ &= \mathrm{Cor}[Y_t - \mathrm{BLP}(Y_t \mid Y_{t-1}), Y_{t-2} - \mathrm{BLP}(Y_{t-2} \mid Y_{t-1})]\end{aligned}$$

Seasonal unit roots

- We have discussed classical UR problems, but there can be seasonal unit roots, too
- Classical UR: $Y_t = \varphi_1 Y_{t-1} + U_t$, $\varphi_1 = 1$
- Seasonal UR (quarterly frequency): $Y_t = \varphi_4 Y_{t-4} + U_t$, $\varphi_4 = 1$
  - In the polynomial form, $(1 - \varphi_4 L^4)Y_t = U_t$, but this polynomial has 4 complex roots $\Rightarrow$ harder theory
- Define $\Delta^{[s]} Y_t \coloneqq Y_t - Y_{t-s}$
- Seasonal UR cause the same problems as classical URs (impossible to proceed with any kind of analysis using standard theory)

ARIMA and SARIMA models

- ARMA models for stationary difference of non-stationary processes are called ARIMA(p, d, q), where d is the number of times Y_t needs to be differenced before $\Delta^{[]d} Y$ it is stationary (usually 0 or 1)
  - Estimate ARMA(p, q) for the differenced $\Delta^d Y$
- Similarly, taking seasonal differences or including seasonal AR or MA parts, one may obtain a SARIMA(p, d, q)(p_s, d_s, q_s) model
  - Popular initial choice: 'airline model' = SARIMA$(0, 1, 1)(0, 1, 1)_s$, adequate for 50% of Eurostat series (Fischer and Plana, 2000)

Seasonal-unit-root testing

- In practice, it is often simpler to estimate a seasonal model like SAR(4); if $\hat{\varphi}_s \approx 1$, then there may be problems $\Rightarrow$ take seasonal differences before proceeding
- There are tests used to detect seasonal URs, but they are more complex
- Dickey, Hasza & Fuller (1984) propose a DF-like statistic from the regression $\Delta^s Y_t = \alpha_1 Y_{t-s} + U_t$ for $\mathcal{H}_0\colon \alpha_1 = 0$ (DHF test)
- HEGY (1990): OLS regression with multiple clever regressors to detect any of the 4 or 12 seasonal roots

Borderline stationarity

What to do if one is not sure about (seasonal) UR?

- Taking differences of already-stationary processes yields the **over-differencing** problem (increased variance) of estimators
- Not taking differences of non-stationary processes breaks the statistical inference
  - Taking a difference is the lesser of two evils
- Seasonal URs can create false positive results in classical UR tests
  - Jointly test for seasonal and non-seasonal URs (Li 1991)
  - If one suspects both types of URs, check for the seasonal UR first, then the non-seasonal one

Estimation of (S)ARIMA models

Consider a SARIMA$(p, d, q)(p_s, d_s, q_s)$ model.

Take d differences and estimate ARMA for $\Delta^d Y_t$.

- If $q = 0$, estimate the model by OLS
- If $q > 0$, estimate the model by ML (using popular software implementations)

Model selection for forecasting

Compare models in terms of forecasting power.

Choose an appropriate criterion: RMSE, MAE, MAPE, tick loss…

- RMSE: $\sqrt{\frac{1}{T_{\text{test}}} \sum_{t=\tau+1}^{\tau+T_{\text{test}}} (Y_t - \hat{Y}_t)^2}$
- MAE: $\frac{1}{T_{\text{test}}} \sum_{t=\tau+1}^{\tau+T_{\text{test}}} |Y_t - \hat{Y}_t|$
- MAPE: $\frac{1}{T_{\text{test}}} \sum_{t=\tau+1}^{\tau+T_{\text{test}}} \left|\frac{Y_t - \hat{Y}_t}{Y_t}\right|$

Automatic model selection

In practice, one is often forced to choose only one model. As perilous as it sounds, algorithms for **'automated'** ARIMA(p, d, q) are often applied:

- TRAMO (Gomez & Maravall)
- Hyndman–Khandakar algorithm

Ideas:

- d is determined according to UR tests (ADF, KPSS)
- p and q are determined using AIC
  - Try initial models: $(0, d, 0)$, $(2, d, 2)$, $(1, d, 0)$, $(0, d, 1)$
- Choose the 'best' model by AIC, vary p and q by ±1, and repeat until convergence

Breaks

- Assumption: homogeneous and stationary time series; however, it is not true for all cases
- There are three main types of breaks: instant jump, temporary jump, permanent jump (shift)
- Instant jump: a short-term change in time series (with reversal approximately to the previous level), modelled as a one-period dummy variable
- Shift: a permanent change in the level, modelled using a dummy variable for the period length

We consider break testing in Session 3. Other types of breaks (e. g. parameter change) exist.

Time-series models in open-source software

R as a programming language

1. Install base R
   - `https://cran.r-project.org`
2. Install RStudio (integrated development environment)
   - `https://posit.co`
3. To compile non-pre-compiled packages (just in case):
   - Windows: `http://cran.r-project.org/bin/windows/Rtools`
   - Mac: install Xcode and a Fortran compiler
   - Linux: for Debian-based (Ubuntu, Mint etc.), install `r-base`, `r-base-dev`, `build-essential`; otherwise, just the C/C++/Fortran compilers (e. g. GCC)

JDemetra+ as a SA tool

JDemetra+: cross-platform graphical software for seasonal adjustment officially endorsed by Eurostat.

- Both point-and-click graphical interface and support for model specifications written in code
- Reusable and extensible Java components
  - Possible to check all the tweaking parameters and algorithms in case of under-performance
- Free and Open Source Software (FOSS), EUPL licence

Developers: National Bank of Belgium, Deutsche Bundesbank, INSEE.

Download JDemetra from GitHub: `https://github.com/jdemetra/jdplus-main`

Time series in R

- Many ways to represent time series as objects; different types = different methods
- Can be index- or time-based (the user decides how to treat them)
  - Multi-firm stock return data with gaps = which type?
  - Weeks of the year = which type? How many?
- Default: `ts` (index-based with frequency), extensions: `zoo`, `xts`
- Some analyses can be carried out without any TS attributes (e. g. `sandwich::vcovHAC` assumes ordered residuals)

Functions

The real power of R!

- Direct productivity gain if you have a piece of code that you use at least twice
- Very easy to create flexible functions, zero costs
- Functions can be generalised, wrapped, and nested to create super-convenient, user-friendly wrappers
- Declare once, use everywhere; if you like it, upload it to a repository or package it

Creating functions

- A set of instructions applied to objects given in input
- Mandatory and optional (with default values) arguments
- May have multi-object return via a list
- Do not rely on global objects; create functions that rely only on inputs
- Many tutorials on good practices

Numerical optimisation in R

Solving a model = estimating the unknown coefficients = numerically optimising some function (distance, discrepancy, deviation, divergence).

- Write a function that accepts the parameter as input, generates the series according to the specification, and computes the desired measure (goodness / badness of fit).
- Pass this function to a numerical optimiser (`optim`, `constrOptim`, `DEoptim::DEoptim`, `hydroPSO::hydroPSO`)

I can share slides on numerical optimisation in R.

Simulation of TS processes in R

Show time with `arima.sim()`.

Statistics in R

- `var()`, `cov()`, `cor()` for variance, covariance, correlation
- `acf()` and `pacf()` for the auto-correlation and partial auto-correlation

Built-in methods

- `arima()` + `tsdiag()` for its output
- `hist()`, `qqplot()`, `plot(density(x))`
- `monthplot(x)` if `x` is a `ts` object with a set frequency

Useful packages

Apart from the built-in `arima()`:

- `forecast` for TS modelling
- `urca` for unit-root tests
- `rugarch` for conditional volatility modelling
- `RJDemetra` and `seasonal` for SA

AR(1) estimation

- Via OLS
- Via ML

Surprisingly, the results might be different due to the implementation specifics (MLE is done via Kalman filtering with some assumptions about the state variable values).

Use OLS for AR models:

- Adding extra regressors is trivial
- `sandwich::vcovHAC()` allows one to try various robust variance estimators

Thank you for your attention!