Topics in time-series analysis

Models · Seasonal adjustment · Imputation

Compiled from session3.tex @ 2024-08-06 15:22:30+02:00.

Day 3: Seasonality in time series

Andreï V. Kostyrka

15th of April 2024

Université du Luxembourg

Presentation structure

1. Seasonality in time series

2. Seasonality adjustment

Seasonality in time series

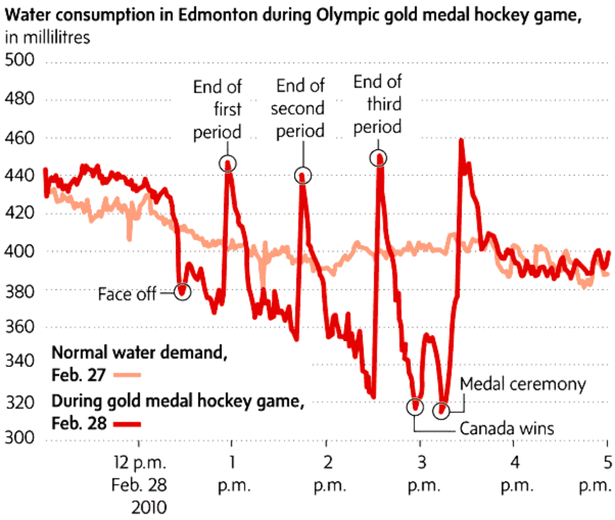
What are we adjusting for?

The movements of infra-annual real economic indicators that are of interest to macro-economists should not contain:

1. **Seasonal fluctuations** – the movements which recur with similar intensity in the same season each year and which can be expected to recur in the future
2. **Calendar effects** that arise from the differences in the number of working or trading days in a month / quarter or the existence of public holidays

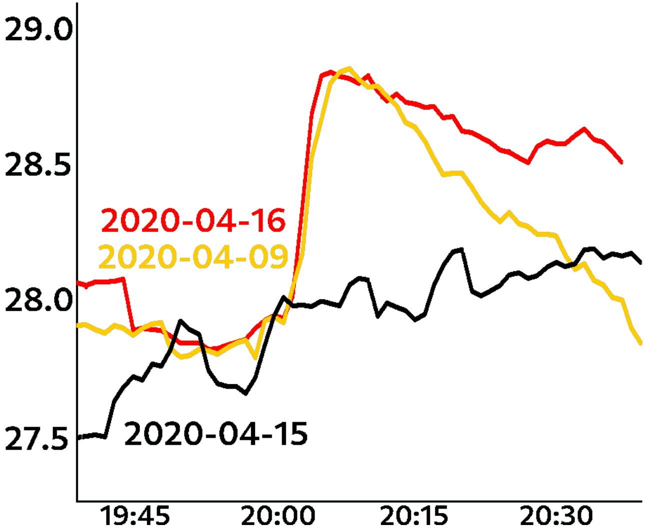
Example: ‘seasonally adj. figures’ vs. ‘only seasonally adj. figures’ (Deutsche Bundesbank). *Should read: ‘only-seasonally* adjusted’.

What is predictable should be predicted 1/2



Credit: EPCOR utilities, 2010.

What is predictable should be predicted 2/2



Credit: National Grid ESO, 2020.

Why is seasonality important?

- Regular sub-yearly observations are often affected by seasonal variation
- Seasonality usually accounts for most of the total variation within the year
- Seasonality affects consumption and production in most sectors: agriculture, trade, mining, tourism etc.
- Seasonal changes are mostly deterministic (can be extrapolated or extracted with little error in many real-world applications)

Why are we adjusting the series?

- Seasonal fluctuations can mask short- and long-term movements in time series and impede a clear understanding of underlying phenomena
- SA is a fundamental process in the interpretation of time series to inform policy-making
- Avoiding including calendar regressors in the models (Frisch–Waugh–Lovell theorem) without creating the omitted-variable bias or spurious relations
- Seasonal effects vary in time and intensity; economists are interested in trend-cycle + irregular

What is a good adjusted series?

SA is still an open issue. **A** good adjustment contains:

- Reasonable trend-cycle and irregular components
- Reasonable seasonal and calendar components
- Evidence that the seasonal + calendar components of the signal are strong (identifiable) and stable (reliably estimable with **linear methods**)

Multiple valid decompositions might be equally good for analysis.

Why is de-seasoning not easy? (1/2)

Seasonality is expected to **evolve slowly and gradually**, which is not always the case.

Changes in the root causes of seasonality:

1. Weather cycles
   - New construction materials ⇒ building in the winter
   - New crops that better resist cold and dry weather
   - Trade of fresh groceries can be increased in the winter
2. Composition of the calendar
   - But national holidays might change
   - Different holidays in different countries

Why is de-seasoning not easy? (2/2)

Changes in the root causes of seasonality:

3. Institutional changes and deadlines
   - Changes in legislation / reporting can break patterns
   - More workers can do remote freelance jobs
4. Expectations
   - But the central bank can partially offset them
   - But changes in methodology can subvert them

Sometimes, SA data do not yield consistently improved predictions, and forecasts are less accurate than those from unadjusted data (Plosser, 1979)!

Economic diagnosis of seasonality

- Some months or quarters are more important in terms of activity or level
- Variability of trading days per period
  - January 2022 ≠ February 2022 (# days)
  - January 2022 ≠ January 2021 (# Fridays)
- Variability of holidays per period
- Moving holidays
  - Easter falls on March with ≈4/17 chance
  - International trade depends on the Lunar NY, Ramadan, festivals etc.

Remaining: minor regular events (e. g. consumption during the Euro cup) and stochastic seasonality

Principles stated in the ESS guidelines

- To avoid misleading results, seasonal adjustment should be applied **only when seasonal effects can be properly explained, identified, and estimated**
  - Where none of these effects can be identified and estimated, unadjusted and SA series are identical
- RegARIMA models should be used to estimate and remove outliers before estimating seasonal effects
  - Calendar adjustment factors in RegARIMA should rely on appropriate national calendars
- SA series should have neither residual seasonality nor residual calendar effects
- The quality of SA data must be regularly checked; the results should be made available to the public

Software implementations

- X13(-ARIMA-SEATS) – actively developed by the U.S. Census, written in Fortran
- TRAMO-SEATS – almost dead (unavailable online, web page deleted and will not be restored – confirmed by BdE, the last binaries redistributed with packages)
- **JDemetra+** – has X13**-like** and TRAMO**-like** algorithms with the same general workflow and slightly different tweaking parameters
- Various interfaces (e. g. X13 in R, JDemetra+ in R, X13 in Python or EViews etc.) to call the 3 methods above

Which method to use

Eurostat recommends that Luxembourg use JDemetra+ with **X13 and RSA 5c specification**:

1. Outliers: automatic data-driven (+ user-provided)
2. Additive / multiplicative: test (automatic data-driven)
3. National calendar
4. Trading days: test (automatic data-driven)
5. Leap-year adjustment: test (automatic data-driven)
6. Easter effect: test (automatic data-driven)
7. ARIMA specification: automatic model identification

Source: ‘Notes on Seasonal and Calendar Adjustment’ (2023-07)

Should everything be automatic?

- Luxembourgish indicators can be quite volatile ⇒ ‘auto-detect all’ is a reasonable recommendation
- However, for some key variables, the methodology should not be changed
  - Eurostat allows domain-specific SA policies (some sectors of the economy require special treatment)
- When the data are updated and adjusted, model changes should be logged

Example: the total employment ARIMA should be stable; for aeroplane ticket sales, a model change is not critical

Why not always use JDemetra+?

- Sometimes, a workflow may be inflexible / have a lot of legacy code base
- It can be cumbersome to connect all data sources to R / JDemetra+ and to teach everyone how to use it
- JDemetra+ is not a development environment, and some operations are hard to implement
- Java installations on Windows can be unstable ⇒ prepare to be flexible to do similar things with the Census X13
  - As of 2024-04, this link works for Windows

Cautionary tale: the horrors of EViews

Do not rely on technically obsolete, closed-source, black-box products with no extensibility.

- EViews does not store manageable X13 output
  - Merely calls the X13 binaries with parameters, imports the values, and shows the unformatted log-file bedsheet
- EViews does not have text-handling capabilities to conveniently parse the X13 output
- This makes it impossible to meaningfully diagnose time series within EViews
  - Automating extra diagnostics requires a lot of programming with very limited capabilities

Convince your advisor / boss that shoehorning the workflow into a dead software ecosystem is bound to backfire.

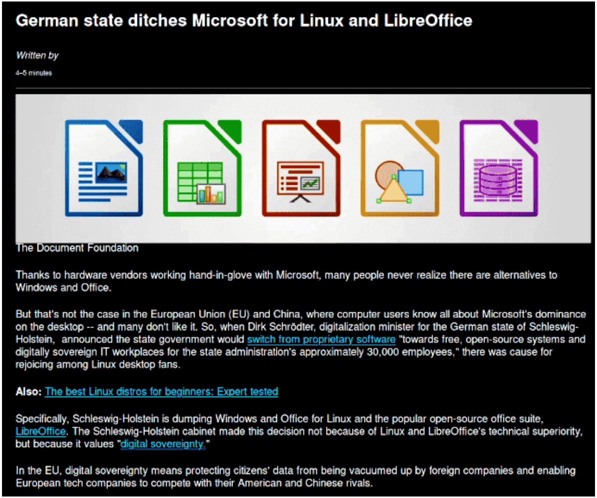
EViews and ESS guideline compliance

- Detailed pre-treatment based on RegARIMA models
- Detailed graphical analysis: correlograms, spectra, zeros and outliers, breaks
- Pre-tests: regressors, plausibility, outliers, stability

This cannot be automated in EViews (partially because the outsourced SA is disjoint from other TS diagnostics).

Recommended: ‘Freely available up-to-date software officially released by statistical institutions, preferably open-source, which fully contains the various recommended methods, follows a clear release strategy and has been thoroughly tested’.

Europe is getting more pro-FOSS



German state ditches Microsoft for Linux and LibreOffice

Written by

4–5 minutes

The Document Foundation

Thanks to hardware vendors working hand-in-glove with Microsoft, many people never realize there are alternatives to Windows and Office.

But that's not the case in the European Union (EU) and China, where computer users know all about Microsoft's dominance on the desktop -- and many don't like it. So, when Dirk Schrödter, digitalization minister for the German state of Schleswig-Holstein, announced the state government would switch from proprietary software "towards free, open-source systems and digitally sovereign IT workplaces for the state administration's approximately 30,000 employees," there was cause for rejoicing among Linux desktop fans.

Also: The best Linux distros for beginners: Expert tested

Specifically, Schleswig-Holstein is dumping Windows and Office for Linux and the popular open-source office suite, LibreOffice. The Schleswig-Holstein cabinet made this decision not because of Linux and LibreOffice's technical superiority, but because it values "digital sovereignty."

In the EU, digital sovereignty means protecting citizens' data from being vacuumed up by foreign companies and enabling European tech companies to compete with their American and Chinese rivals.

Credit: Schleswig-Holstein official web site, ZDnet news 2014-04-04.

Focus of the harmonisation

- It is not about using JDemetra+
  - ESS guidelines: ‘The recommended SA methods are **based on** <...> TRAMO-SEATS and <...> X13.’
- *A* uses SEATS, *B* uses X13 in JD+ ⇒ incomparable
- *C* uses X13 in JD+ but forces SA when there is no seasonality ⇒ detrimental
- *D* uses JD+ but with the default calendar (instead of LU, FR, ...– without 9 holidays) ⇒ wrong
- *E* adjusts 12-year chunks and *F* adjusts 15-year chunks ⇒ incompatible
- *G* uses old Demetra with binary `.dem` files ⇒ non-transparent and irreplicable
- **Focus: visual clarity & comparability over time**

Eurostat compliance

In the narrowest sense, Eurostat compliance means

1. JDemetra+ compatibility (same tweaking parameters regardless of the software used)
2. Accessible history of SA results / diagnostics
3. Well-documented deviations from the fully automated procedure based on the specificity of the time series in question

Therefore, for most applications, a wrapper for X13 binaries may be a near-perfect substitute to the JD+ X13 call with identical parameters. The plots and diagnostic statistics can be computed in any software and require only the decomposed series.

Proposed workflow

1. Use any software to make X13-like SA routine calls
   - It should compare multiple models and auto-choose the best one according to the JD+ rules *(i. e. based on AICc)*, or respect the user's forced parameters
2. At the end, save the parameters, diagnostics and SA series to the archive / folder / spreadsheet
   - If necessary, compare the new SA series with the archived ones
   - Examine the revision discrepancy

Seasonality adjustment

Expected output

Decompose the observed series Y_t into 3–5 components:

- Estimated **trend** $\{T_t\}$
  - Old sources: 'trend-cycle', or TC
- Estimated **seasonal** component $\{S_t\}$
  - Hopefully, relatively stable
- Estimated **irregular** component $\{I_t\}$
- Optional deterministic components:
  1. Calendar effects $\{C_t\}$ (if found: week-day effects, Easter, holidays)
  2. Pre-treatment effects $\{P_t\}$ for shifts or strong anomalies

Economists are interested in I_t and T_t (sometimes, P_t).

Desirable decomposition properties

- $\hat{T}_t$, $\hat{S}_t$, and $\hat{I}_t$ should be uncorrelated
  - No seasonal effects must remain in the trend and irregular
  - The trend must be sufficiently smooth, the seasonal and irregular must be trend-free
- For additive decompositions: $\mathbb{E}S_t = \mathbb{E}I_t = 0$
  - The mean / level effect is captured entirely by the trend
- I_t should be a weakly stationary time series
  - Should be well approximated by ARMA processes

Practical questions: is there any seasonality at all (magnitude of $\hat{S}_t$), it is stable (stationarity of $\hat{S}_t$), and are linear filters applicable (linear TS model diagnostics for $\hat{I}_t$)?

Weighted moving average

Suppose that Y_t = signal$_t$ + WN$_t$. If the signal is a linear function of t, it can be estimated with moving averages: since $\mathbb{E}\mathrm{WN}_t = 0$, under WLLN,

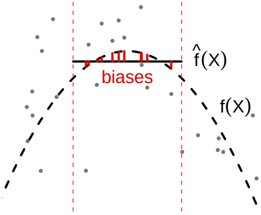
$$\mathrm{WMA}_k(Y_t) \coloneqq \frac{1}{2k+1} \sum_{i=-k}^{k} w_i Y_{t+i} \xrightarrow{k\to\infty} \mathrm{signal}_t.$$

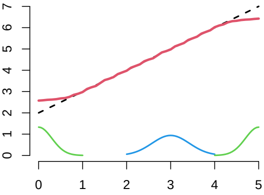
Conceptual problem: it is forward-looking. However, one-sided (backwards-looking) filters perform much worse.

- Analogy: two-sided numerical derivatives have a lower order or error than one-sided derivatives: $\frac{f(x+h)-f(x-h)}{2h}$ is better than $\frac{f(x+h)-f(x)}{h}$: the error order is $O(h^3)$ vs. $O(h^2)$

Properties of moving averages

- Local smoother of degree 0
  - For higher degrees, see local regression (LOESS)
- Creates biases if there are non-linear functions
- In finite samples, creates biases at the ends due to one-sided averaging
  - Can be mitigated by clever extrapolation or by negative weights





What is a good input series

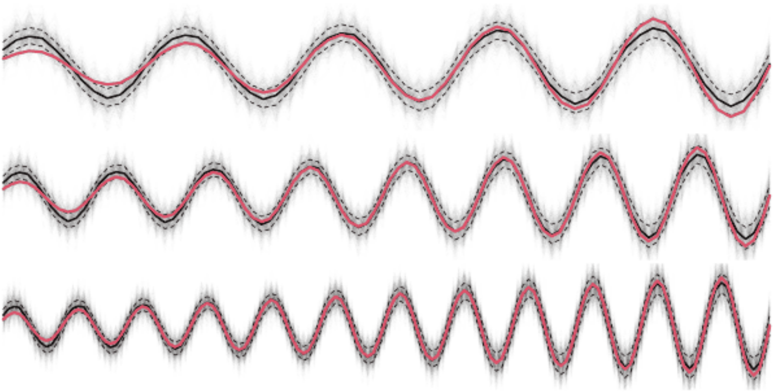
- Length 5–15 years *(more on that later)*
- A process that is well-approximated by a RegARIMA
  - RegARIMA = ARIMA after **pre-treating** for outliers (shifts, breaks etc.), calendar effects, Easter, leap year
- Either no seasonality or stable seasonality
  - The seasonal fluctuation in linear units or % should not change much from year to year
  - If the seasonal component is evolving, then the adjustment procedure *may actually introduce seasonality* due to its slow adaptation to changes

Statistical seasonality diagnosis

- Significance of seasonal dummies (regression *F* test)
  - $\mathcal{H}_0$: seasonal dummies are jointly insignificant
- Rank test (Kruskal—Wallis, Friedman)
  - $\mathcal{H}_0$: observation ranks grouped by period are similar
- Spectral tests
  - $\mathcal{H}_0$: no sharp peaks different from their neighbours
- Seasonal autocorrelation (Ljung—Box-like, e. g. QS)
  - $\mathcal{H}_0$: seasonal lags do not correlate
- Seasonal unit root test
  - $\mathcal{H}_0$: there is a seasonal UR (DF-type distribution)
- Seasonal stability over time (Canova—Hansen)
  - $\mathcal{H}_0$: deterministic seasonality, no stochasticity
- Evolutive seasonality test
  - $\mathcal{H}_0$: the magnitude of fluctuations is constant

Importance of non-evolutive seasonality

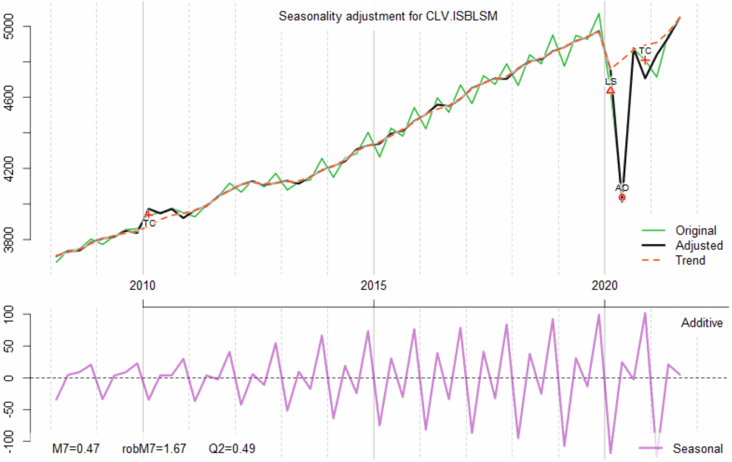
True S_t (magnitude 1 → 4) vs. **best routine output**.



If the log transformation does not equalise the swing amplitude, some systematic distortions are expected.

Real case of evolutive seasonality: ISBLSM

Non-profit institutions serving households in LU:



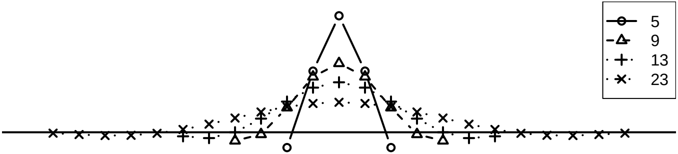
Credit: Statec.

Baseline X11 method (Shiskin et al., 1967)

- Non-parametric, no explicit model
  - Weighted moving averages only
- Decomposition:
  1. Trend + business cycles (T_t)
  2. Seasonal component (S_t)
  3. Trading-day effect and Easter effect (C_t)
  4. Irregular component (I_t)
- Significance tests of seasonality and trading-days variation
- Bad forecasts at the end (due to phase shifts)
- Can treat only quarterly or monthly data

Useful features of X11

- Additive or multiplicative models
- Trading days via zero-mean dummy variables or *a priori* weights (AIC selection)
- Re-weighting of outliers within a 5-year span ($w = 1$ if $|Y_t| \le 1.5\sigma$, $w = 0$ if $|Y_t| > 2.5\sigma$, linear w in between)
- Adaptive Henderson MA (9, 13, 23 for monthly, 5 for quarterly) preserving quadratic polynomials



‘I would rather have Bob Solow than an econometric model,

but I’d rather have Bob Solow with an econometric model than without one.’

Paul Samuelson.

X11-ARIMA (Dagum, 1980, 1988)

- The series are extended with SARIMA fore- and backcasts and treated with symmetric filters
- **The SARIMA model is used to test the data quality**
- Achieves minimum-MSE seasonal adjustment
- Extra tests:
  - Varying seasonality magnitude (reliability)
  - Combined identifiable seasonality (stable, variable, Kruskal—Wallis)
  - Periodicity in residuals (duration of runs)
  - Combined quality control statistics (Q)

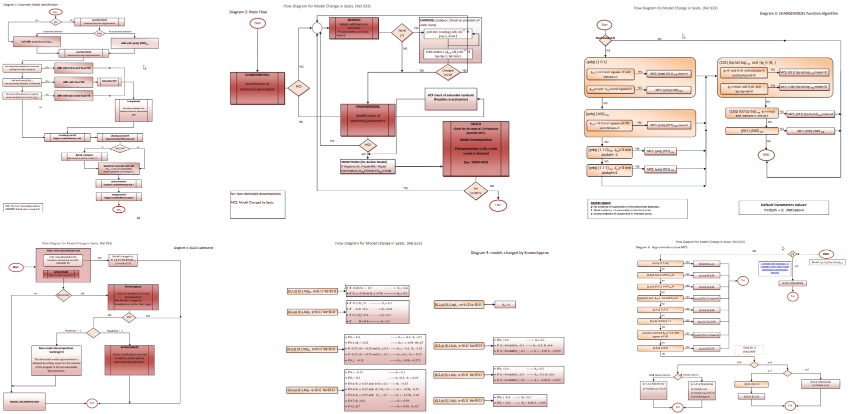
TRAMO-SEATS (Gómez & Maravall, 1996)

ARIMA regression with missing observations and outliers + signal extraction (removing deterministic components)

- Pre-testing (log/level, TD, Easter, leap year)
- Automatic model identification: ARMA order ≤ 3, seasonal ≤ 2, $I \le 2$, $I_s \le 1$, prefers models with $p = q$
- Handles missingness via dummy variables (with interpolated values and AO)
- Outlier testing: additive, shift, or transitory (forward-backward iterative)

TRAMO flowchart

A highly empirical procedure from Maravall & Caporello (2014) tailored to work well with real-world data:



X12-ARIMA (Findley et al., 1998)

Workflow slightly different from X11-ARIMA:

1. ARIMA (more models) + regressors
   - Emphasis: model + diagnostics + re-adjustment
   - TD, Easter, length of month, pre-set holidays
   - Outlier testing: additive or shift (forward-backward)
   - AIC-based model selection (incl. adjustments)
   - Handles missingness via dummy variables (AO)
2. SA with adaptive smoothing strength
3. Post-adjustment diagnostics
   - Spectral methods (seasonality and TD significance)
   - Stability in sliding spans (more accurate than Q)
   - Revision stability (end of sample)
   - AIC selection history stability

X13-ARIMA-SEATS (Monsell, 2007)

Merger of X11-ARIMA, TRAMO, SEATS (based on source codes from all).

- Two modules under one hood
  - Same baseline ARIMA estimation, non-parametric (X11) or parametric (SEATS) SA
- TRAMO flowchart for model selection
- New outlier types: seasonal, ramps, temporary shifts (not detected automatically)

(J)Demetra+ (NBB, DB, Eurostat, 2012, 2015)

- Same 2 modules for SA: X11**-like** ans SEATS**-like**
- TRAMO**-like** model selection
- Regression and spectral seasonality tests, stability analysis
- Open-source, official support for integration with **R**, JDBC, ODBC, XML export
  - Any step of any procedure examinable in source codes
  - Easy to see changes between versions
- Results very close yet not identical to X13AS and TRAMO/SEATS!

Main steps of X13-ARIMA-SEATS

1. Pre-treatment to eliminate deterministic regressor effects (all of the below simultaneously)
   1.1 Estimation of ARIMA, selection of $(p, d, q)(p_s, d_s, q_s)$
   1.2 Estimation of the calendar regressor effects (trading days, Easter, leap year, user inputs), removing if found significant
   1.3 Auto-detection of outliers via 'outlier regressors' ($3T$ regressions at each iteration)
2. Series of moving averages on the pre-treated series to get the preliminary I_t and its robust SD $\hat{\sigma}(I_t)$
3. Computation of weights to reduce the effect of noisy observations by comparing to a multiple of $\hat{\sigma}(I_t)$
4. Final decomposition via robustly weighted averages
   - Removing only the seasonal + calendar components and adding back the regressor effects

Comparability problem: sample length

AICc: compare models on the **same response data**.

Example: $T = 100$. ARIMA(1,0,1) has $T_{\text{eff}} = 99$, ARIMA(2,0,1) has $T_{\text{eff}} = 98$. The models are incomparable: different lag length $\Rightarrow$ **different effective sample size**.

Correct: compare the averages, AICc / T_{eff} (X13 does this)

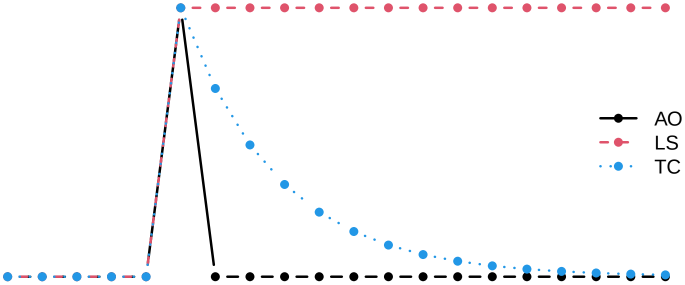
Ideal: trim the sample start to match the smallest T_{eff}, re-estimate the models, compare by AICc *(cumbersome)*

Wrong: compare AICc's without checking T_{eff}

Parameter 1: outliers

Eurostat recommends auto-detection of

- Additive outliers (AO)
- Level shifts (LS)
- Temporary changes (TC, exponential decay at rate 0.7)



Auto-detecting outliers

For any time series $\{Y_t\}_{t=1}^T$:

- Run a candidate regression with a variable for AO, LS, and TC variables for all time points ($3T$ variables)
- Compute t-statistics for the significance of each outlier $\Rightarrow$ compare to the critical values
  - **NB.** The critical values are higher than the 5% $Q_\mathrm{t}(0.975) = 1.96$ to control for false positives / multiple comparisons, and increase as more outliers are found (e. g. CV = 2.74 for 5 regressors and 3.09 for 10 regressors)
  - Instead of $\sqrt{\hat\sigma^2}$, a robust measure of dispersion ($1.48 \cdot$ the median absolute deviation of the residuals) is used
- If the outlier regressors is significant, add the variable to the model and re-run $3T$ regressions again
  - Stop if no new significant outliers found

Comparability problem 2: dummies

(1): $Y_t = \alpha + X_t'\beta + U_t, \quad t = 1, \dots, 120.$

(2): $Y_t = \alpha + X_t'\beta + \sum_{i=117}^{120} \gamma_i \cdot \mathbb{I}(t = i) + U_t$, $t = 1, \dots, 120$ (dummies for $t = 117, \dots, 120$).

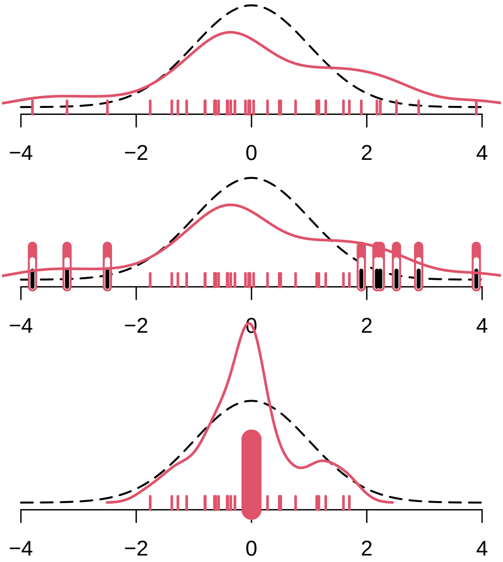
(3): $Y_t = \alpha + X_t'\beta + U_t$, $t = 1, \dots, 116.$

$(\hat\beta, \hat U_t)$ of (2) and (3) are identical.

- (2) yields $\hat U_{117}, \dots, \hat U_{120} \equiv 0 \Rightarrow$ lower avg. reg. SE
- The AICc contrib. of each $t = 117, \dots 120$ in (2) is $-2\phi(0) \approx -0.8$ (deterministic $\Rightarrow$ downwards bias)

Additive period-wise outliers imply non-Gaussian error distribution (atom at 0) $\Rightarrow$ AIC and BIC **not applicable**.

Comparability problem 2: illustration



Solution: Identical outlier regressor sets

The set of outlier variables should be fixed in all models.

Unless the models being compared have the same outliers (to cancel out their similar effects on the AICc), the outliers can largely determine the model selection rather than more relevant data properties.

Danger: the model with the most outliers will often be the one with the smallest criterion value (even with the Akaike or Schwarz penalty).

A preliminary run of multiple models should give an idea about which outliers to enforce for comparison.

How to handle the outlier problem

To ensure the comparability of k models by AIC,

1. Estimate all k variants with auto-detected outliers (call the estimates *Set 1*)
2. Extract the detected outliers from all models and combine them into a single list
3. Re-estimate all models with this enforced list of all outliers (call the estimates *Set 2*)
   - If estimation fails, reduce the list of outliers to those that are present in $k - 1$ or fewer models
4. Choose the best model type by AICc from *Set 2*
5. Use the model of this type from *Set 1*

Outliers for known one-time aberrations

- **Example:** in 2021, the winter sales were moved to 20.01–17.02.2021, which affected *core inflation*
  - Year before: 02.01–25.01.2020
  - Year after: 03.01.2022–29.01.2022
- The regular pattern was broken for 1 year

Solution: forcibly include AO dummies for 2021.01 and 2021.02 **into the SA routine** (not the subsequent analysis).

More complex breaks (change of law regarding the dates henceforth) cannot be tackled as easily; at least 3 years must pass before this change becomes identifiable.

Parameter 2: additive / multiplicative

The **t**rend, **s**easonal, **c**alendar, and **i**rregular components can be combined in two ways:

$$Y_t = T_t + S_t + C_t + I_t \quad \text{vs.} \quad \underset{(\ln Y_t = \ln T_t + \ln S_t + \ln C_t + \ln I_t)}{Y_t = T_t \cdot S_t \cdot C_t \cdot I_t}$$

Decision rule: estimate both models with auto-ARIMA (6 TD, include the Easter and leap-year regressors) and compare the AICc's. Then, log-transform if

$$\text{AICc(add. model)} - \text{AICc(mult. model)} > -2$$

Technical remark. Since AICc is only for comparing models with the **same** dependent variable, a correction must be applied to the log-likelihood of the transformed model. X13 does it automatically.

The outlier regressor caveat applies.

Parameter 3: calendar

The series for each country should be adjusted with a proper calendar for that country.

Additionally, if specific industries have specific calendars (with unique holidays, e. g. school holidays or days between a public holiday and a weekend used to extend the weekend via compensation), they can be used.

Obligatory rule: use the Luxembourgish calendar for Luxembourgish series.

Suggestion: if possible, use national calendars wherever possible. If not, use the default one but supply the holidays manually.

Parameter 4: trading / working days

The calendar component may depend on the number of distinct trading days:

1. 7 distinct days = 6 TD dummy regressors
2. The total number of working days (week-day or week-end = 1 WD dummy regressor)
3. No trading day regressors

Decision rule: estimate the 3 models and compare the AICc's. Choose the one with the lowest AICc.

The outlier regressor caveat applies.

Calendar: technical remarks

Use JDemetra+ to generate national calendars with long-term mean correction.

- By default, the 6 TD dummies are written as #Mon – #Sun, …, #Sat – #Sun
  - A separate regressor is thus needed for the leap years
- With holidays, such differences create problems; only long-run-mean-corrected calendars should be used

What should a calendar have?

1. Date
2. 6 week-day regressors `Monday`, ..., `Saturday`
3. Leap-year regressor `LeapYear`
4. Working-day regressor `WorkingDays`

Date	Monday	Tuesday	Wednes	Thursda	Friday	Saturday	LeapYea	Working y
01/01/2022	0	-1	-1	-1	-1	-1	0	-1.5
01/02/2022	0	0	0	0	0	0	-0.25	0
01/03/2022	0.406358	1.203179	1.203179	1.203179	0.203179	0.203179	0	3.711127
01/04/2022	-0.40152	-0.19834	-0.19834	-0.1935	0.801658	0.801658	0	-2.1942
01/05/2022	1.177511	1.567974	0.567974	0.52641	0.567974	0.567974	0	2.987909
01/06/2022	-1.18235	-0.57281	0.427189	-0.53609	-0.57281	-0.57281	0	-1.00484
01/07/2022	-1	-1	-1	-1	0	0	0	-4
01/08/2022	0	1	1	0	0	0	0	2
01/09/2022	0	0	0	1	1	0	0	2
01/10/2022	0	-1	-1	-1	-1	0	0	-4
01/11/2022	0	0	1	0	0	0	0	1
01/12/2022	0	1	1	2	2	2	0	1
01/01/2023	1	1	0	0	0	0	0	2
01/02/2023	0	0	0	0	0	0	0.25	0

Parameters 5, 6: Leap year and Easter

Leap year: 29 days in February.

The Easter is a floating holiday; it may be affecting economic indicators throughout multiple preceding days (1–14, default: 8).

Decision rule: let JD+ choose the best model based on the lowest AICc (estimate a full model, and use the built-in AICc test for LY and Easter[8]).

Parameter 7: RegARIMA specification

JDemetra+ and Census X13 have a complex routine to determine the best RegARIMA model that iterates back and forth within the set limits.

Nothing to change or implement here.

Under the hood: max. model order, unit root tolerance, lag polynomial simplification tolerance, significance testing, Ljung–Box acceptance level etc.

Suggested X13-like workflow

1. Choose the best model (under fixed outlier regressors):
   1.1 With 6 TD + Easter + LY, test log-transform vs. none
      - If estimation fails (short series), use 1 or 0 TD

   1.2 With this transformation, test 6 vs. 1 vs. 0 TD

   1.3 With this TD set, test Easter and leap year
2. Create SA series if: calendar effects exist, or the diagnostic statistics are convincing, or if the plots show something, or if it is required for specific series
   - Similar series should have similar treatment
3. Check if the adjusted values are reasonable (e. g. not negative), check SA plot and outlier types and dates
   - Add outlier regressors if the trend looks inadequate
4. Save the results + plot to the archive

Thank you for your attention!