Topics in time-series analysis: Models, seasonal adjustment, and imputation

Dr. Andreï V. Kostyrka
University of Luxembourg

1. Brief description

The goal of this course is to get Ph.D. students acquainted with the fundamentals of time-series modelling and several applied time-series methods that are actively used at major economic institutions (Eurostat, Statec, central banks etc.). The main focus of the course is on characterising the structural features of macroeconomic time series, on seasonal adjustment, and on the basics of imputation for multiple unbalanced time series. Computations are done in free and open-source software: R and JDemetra+.

Upon successful completion of this course, students will be able to:

- Specify many popular time-series models in the state-space framework to apply unified filtering algorithms and produce short-term signal and state forecasts of multiple macro-economic time series with R;
- Diagnose macroeconomic series and use data-driven methods to detect structural breaks and calendar effects;
- Filter out the seasonal component of time series by parametric and semi-parametric methods and evaluate the quality of seasonal adjustment in compliance with the latest Eurostat guidelines with JDemetra+;
- Extract the common dynamics of multiple time series by modelling a small number of unobserved factors;
- Apply state-of-the-art econometric methods to reconstruct partially incomplete macroeconomic data sets with R.

2. Intended audience

This course is intended for Ph.D. students in economics, management, and finance. However, all post-doctoral researchers, Ph.D. or master's students interested in learning these methods are also welcome.

3. Text and materials

The required books, manuals, and articles for this course are presented below by topic (ordered by relevance).

Time-series modelling.

- Hamilton, J. D. (1994). *Times series analysis*. Princeton University Press (chapters 3, 5, 6, 8, 13).
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control* (5th ed.). Wiley (chapters 1, 3, 4, 6, 7).
- Douc, R., Moulines, E., & Stoffer, D. (2014). *Nonlinear time series: Theory, methods and applications with R examples*. CRC press (chapters 1, 2).
- Pawitan, Y. (2001). *In all likelihood: Statistical modelling and inference using likelihood*. Oxford University Press (chapters 2, 4, 6, 9, 11, 13, 14).
- Brockwell, P. J., & Davis, R. A. (1991). *Time series: Theory and methods* (2nd ed.). Springer (chapters 3, 8–12).
- Shumway, R. H., & Stoffer, D. S. (2016). *Time series analysis and its applications with R examples* (4th ed.). Springer (chapters 3, 4, 6).
- Chatfield, C., & Xing, H. (2019). *The analysis of time series: An introduction with R* (7th ed.). Chapman; Hall/CRC (chapters 2, 3, 4).

- Kilian, L., & Lütkepohl, H. (2017). *Structural vector autoregressive analysis*. Cambridge University Press (chapters 16, 19).
- Cryer, J. D., & Chan, K.-S. (2008). *Time series analysis with applications in R* (2nd ed.). Springer (chapters 10, 11).
- Enders, W. (2015). *Applied econometric time series* (4th ed.). John Wiley & Sons (chapter 2).
- Brockwell, P. J., & Davis, R. A. (2016). *Introduction to time series and forecasting* (3rd ed.). Springer (chapters 6, 9).
- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd ed.). OTexts (chapters 3, 4, 9, 11, 12, 13). https://otexts.com/fpp3
- Hendry, D. F., Pagan, A. R., & Sargan, D. J. (1984). Dynamic specification. In *Handbook of econometrics* (pp. 1023–1100, Vol. 2). Elsevier.

Fundamentals of seasonal adjustment.

- Mazzi, G. L., Ladiray, D., & Rieser, D. A. (Eds.). (2018). *Handbook on seasonal adjustment: 2018 edition*. Publications Office of the European Union (chapters 2, 3, 5–8, 30).
- Dagum, E. B., & Bianconcini, S. (2016). *Seasonal adjustment methods and real time trend-cycle estimation*. Springer (chapters 1–7).
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control* (5th ed.). Wiley (chapter 9).
- European Commission & Eurostat. (2015). *ESS guidelines on seasonal adjustment: 2015 edition*. Luxembourg: Publications Office of the European Union.
- Young, A. (1965). *Estimating the trading-day variation in monthly time series*. U.S. Bureau of the Census.

Semi-parametric (X11-like) seasonal adjustment.

- Mazzi, G. L., Ladiray, D., & Rieser, D. A. (Eds.). (2018). *Handbook on seasonal adjustment: 2018 edition*. Publications Office of the European Union (chapters 10, 11, 12).
- Ladiray, D., & Quenneville, B. (2001). *Désaisonnaliser avec la méthode X-11*. Springer-Verlag (chapters 2–5).
- Findley, D. F., Monsell, B. C., Bell, W. R., Otto, M. C., & Chen, B.-C. (1998). New capabilities and methods of the X-12-ARIMA seasonal-adjustment program. *Journal of Business & Economic Statistics*, *16*(2), 127–152.
- Shiskin, J. (1967). *The X-11 variant of the census method II seasonal adjustment program*. U.S. Census.
- Dagum, E. B. (1980). *The X-II-ARIMA seasonal adjustment method*. Statistics Canada.
- Dagum, E. B. (1999). *X11ARIMA version 2000: Foundation and user's manual*. Statistics Canada.
- Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, *6*(1), 3–73.
- Plosser, C. I. (1979). Short-term forecasting and seasonal adjustment. *Journal of the American Statistical Association*, *74*(365), 15–24.

Parametric (TRAMO-SEATS-like) seasonal adjustment.

- Mazzi, G. L., Ladiray, D., & Rieser, D. A. (Eds.). (2018). *Handbook on seasonal adjustment: 2018 edition*. Publications Office of the European Union (chapters 10, 13).
- Gómez, V., & Maravall, A. (2000). Automatic modeling methods for univariate series. In *A course in time series analysis* (pp. 171–201). John Wiley & Sons.
- Findley, D. F., Lytras, D. P., & Maravall, A. (2016). Illuminating ARIMA model-based seasonal adjustment with three fundamental seasonal models. *SERIEs*, *7*(1), 11–52.
- Kaiser, R., & Maravall, A. (2000). *Notes on time series analysis ARIMA models and signal extraction. Documento de trabajo n. 0012*. Banco de España — Servicio de Estudios.

- Mélard, G. (2016). On some remarks about SEATS signal extraction. *SERIEs*, *7*(1), 53–98.
- Hillmer, S. C., & Tiao, G. C. (1982). An ARIMA-model-based approach to seasonal adjustment. *Journal of the American Statistical Association*, *77*(377), 63–70.
- Burman, P. J. (1980). Seasonal adjustment by signal extraction. *Journal of the Royal Statistical Society: Series A (General)*, *143*(3), 321–337.

Seasonal diagnostics.

- Mazzi, G. L., Ladiray, D., & Rieser, D. A. (Eds.). (2018). *Handbook on seasonal adjustment: 2018 edition*. Publications Office of the European Union (chapters 17, 20–23).
- McKenzie, R., & Gamba, M. (2008). *Interpreting the results of revision analyses: Recommended summary statistics*. OECD.
- Lothian, J. R., & Morry, M. (1978). *A set of quality control statistics for the X-11-ARIMA seasonal adjustment method*. Statistics Canada.
- Higginson, J. (1975). *An F test for the presence of moving seasonality*. Statistics Canada.
- Godfrey, M. D., & Karreman, H. F. (1964). *A spectrum analysis of seasonal adjustment. Research memorandum #64*. Princeton University.
- Cleveland, W. S., & Devlin, S. J. (1980). Calendar effects in monthly time series: Detection by spectrum analysis and graphical methods. *Journal of the American Statistical Association*, *75*(371), 487–496.
- Canova, F., & Hansen, B. E. (1995). Are seasonal patterns constant over time? A test for seasonal stability. *Journal of Business & Economic Statistics*, *13*(3), 237–252.
- Dagum, E. B., Quenneville, B., & Sutradhar, B. (1992). Trading-day variations multiple regression models with random parameters. *International Statistical Review*, 57–73.

Software implementations of SA techniques.

- Buono, D. (2017). Introduction to seasonal adjustment and JDemetra+. European statistical training programme slides.
- Grudkowska, S. (2017). *JDemetra+ user guide version 2.2*. Narodowy Bank Polski, Department of Statistics. Retrieved September 1, 2023, from https://cros-legacy.ec.europa.eu/system/files/jdemetra_user_guide_version_2.2.pdf https://cros-legacy.ec.europa.eu/system/files/jdemetra_user_guide_version_2.2.pdf
- Time Series Research Staff. (2021). *X-13ARIMA-SEATS reference manual, version 1.1*. U.S. Census Bureau.
- Maravall, A. (2003). *Notes on programs TRAMO and SEATS (parts I, II, III)*. Bank of Spain.
- Gómez, V., & Maravall, A. (1997). *Programs TRAMO and SEATS*. Bank of Spain.

Dynamic factor models.

- Stock, J. H., & Watson, M. W. (2016). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In *Handbook of macroeconomics* (pp. 415–525, Vol. 2). Elsevier.
- Petris, G., Petrone, S., & Campagnoli, P. (2009). *Dynamic linear models with R*. Springer (chapters 1, 2, 3).
- Zivot, E., & Wang, J. (2006). *Modeling financial time series with S-PLUS* (2nd ed.). Springer (chapters 15, 16).
- Hyndman, R. J., Koehler, A. B., Keith Ord, J., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: The state space approach*. Springer (chapters 1–9, 13, 14).
- Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer (chapter 18).
- Stock, J. H., & Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, *97*(460), 1167–1179.

- Mariano, R. S., & Murasawa, Y. (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics*, *18*(4), 427–443
- Solberger, M., & Spånberg, E. (2020). Estimating a dynamic factor model in EViews using the Kalman filter and smoother. *Computational Economics*, *55*(3), 875–900.
- Doz, C., Giannone, D., & Reichlin, L. (2011). A two-step estimator for large approximate dynamic factor models based on Kalman filtering [Annals Issue on Forecasting]. *Journal of Econometrics*, *164*(1), 188–205.
- Auger-Méthé, M., Field, C., Albertsen, C. M., Derocher, A. E., Lewis, M. A., Jonsen, I. D., & Mills Flemming, J. (2016). State-space models' dirty little secrets: Even simple linear gaussian models can have estimation problems. *Scientific Reports*, *6*(1), 26677.

Imputation.

- Little, R. J. A., & Rubin, D. B. (2019). *Statistical analysis with missing data* (3rd ed.). John Wiley & Sons (chapters 1, 2, 5, 10, 11).
- Enders, C. K. (2022). *Applied missing data analysis* (2nd ed.). The Guilford Press (chapters 4, 5, 7).
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (3rd ed.). CRC press (chapters 1, 2, 6, 7, 10–13, 18).
- Hastie, T., Tibshirani, R., & Friedman, J. (2017). *The elements of statistical learning* (2nd ed.). Springer (chapters 8, 9).
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control* (5th ed.). Wiley (chapter 13).
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, *39*(1), 1–38.
- King, G., Honaker, J., Joseph, A., & Scheve, K. (2001). Analyzing incomplete political science data: An alternative algorithm for multiple imputation. *American political science review*, *95*(1), 49–69.
- Honaker, J., & King, G. (2010). What to do about missing values in time-series cross-section data. *American journal of political science*, *54*(2), 561–581.
- Blackwell, M., Honaker, J., & King, G. (2015). A unified approach to measurement error and missing data: Overview and applications. *Sociological Methods & Research*, *46*(3), 303–341.
- Honaker, J., King, G., & Blackwell, M. (2011). Amelia II: A program for missing data. *Journal of statistical software*, *45*, 1–47.
- Gómez, V., Maravall, A., & Peña, D. (1999). Missing observations in ARIMA models: Skipping approach versus additive outlier approach. *Journal of Econometrics*, *88*(2), 341–363.
- van Ginkel, J. R. (2023). Handling missing data in principal component analysis using multiple imputation. In L. A. van der Ark, W. H. M. Emons & R. R. Meijer (Eds.), *Essays on contemporary psychometrics* (pp. 141–161). Springer International Publishing.
- Josse, J., Pagès, J., & Husson, F. (2011). Multiple imputation in principal component analysis. *Advances in data analysis and classification*, *5*(3), 231–246.
- Josse, J., & Husson, F. (2016). missMDA: A package for handling missing values in multivariate data analysis. *Journal of Statistical Software*, *70*(1), 1–31.
- Beckers, J. M., & Rixen, M. (2003). EOF calculations and data filling from incomplete oceanographic datasets. *Journal of Atmospheric and Oceanic Technology*, *20*(12), 1839–1856.
- Taylor, M. H., Losch, M., Wenzel, M., & Schröter, J. (2013). On the sensitivity of field reconstruction and prediction using empirical orthogonal functions derived from gappy data. *Journal of Climate*, *26*(22), 9194–9205.
- Nguyen, C. D., Carlin, J. B., & Lee, K. J. (2017). Model checking in multiple imputation: An overview and case study. *Emerging Themes in Epidemiology*, *14*(1), 1–8.
- Josse, J., & Husson, F. (2012). Handling missing values in exploratory multivariate data analysis

methods. *Journal de la Société Française de Statistique*, *153*(2), 79–99.
- Audigier, V., Husson, F., & Josse, J. (2016). Multiple imputation for continuous variables using a bayesian principal component analysis. *Journal of statistical computation and simulation*, *86*(11), 2140–2156.
- Kamperis, S. (2021). *The expectation-maximization algorithm*. https://ekamperi.github.io/mathematics/2021/07/03/expectation-maximization-part1.html
- Anderton, T. (2017). *Imputing missing values with PCA*. https://asymptoticlabs.com/blog/posts/other_use_for_PCA_part2.html

Other topics.

- Eurostat. (2023). *Notes on seasonal and calendar adjustment*. Retrieved March 25, 2024, from https://ec.europa.eu/eurostat/statistics-explained/images/3/3d/Metadata_on_seasonal_adjustment_2023Q1.pdf
- Dagum, E. B., & Cholette, P. A. (2006). *Benchmarking, temporal distribution, and reconciliation methods for time series*. Springer.
- Liu, X., Huang, J., Li, C., Zhao, Y., Wang, D., Huang, Z., & Yang, K. (2021). The role of seasonality in the spread of COVID-19 pandemic. *Environmental research*, *195*, 110874.
- Stats NZ. (2023). *Seasonal adjustment and automatic outliers in time series after COVID-19*.
- Bógalo, J., Llada, M., Poncela, P., & Senra, E. (2022). Seasonality in COVID-19 times. *Economics Letters*, *211*.
- Mazzi, G. L., & Savio, G. (2005). The seasonal adjustment of short time series.
- Barska, M. (2014). Seasonality testing for macroeconomic time series: Comparison of X-12-ARIMA and TRAMO/SEATS procedures. *Śląski Przegląd Statystyczny*, *18*(12), 121–140.

The following open-source software are used in this course:

- The R project for statistical computing (https://www.r-project.org);
- RStudio (https://posit.co);
- JDemetra+ (https://github.com/jdemetra).

Other course material (lecture notes, additional data sets, computer exercises) will be made available on the Moodle site of this course.

4. Lecture schedule

The course will be given in a hybrid format, which means that the participants who cannot be present physically due to the limited room capacity / health issues can join the session via this Webex link.

#	Date	Room	Time	Webex link
1	03/04/2024	BLG 001	14:00 – 17:00	Session 1
2	10/04/2024	BLG 001	14:00 – 17:00	Session 2
3	15/04/2024	BLG 001	14:00 – 17:00	Session 3
4	19/04/2024	BLG 001	14:00 – 17:00	Session 4
5	22/04/2024	BLG 001	14:00 – 17:00	Session 5

1. Introduction to time-series analysis (TSA). Stationary and non-stationary processes. Wold decomposition. Linear TS models: non-seasonal and seasonal ARIMA. Estimation of TS models in R.
2. Seasonality and seasonal adjustment (SA). Comparison of SA methods. Determining whether the series need any adjustment. Evaluating SA quality. Basics of model-driven outlier detection.
3. Practical implementation of SA. Producing diagnostic plots and tables in R and JDemetra+. Spectral diagnostics of seasonality.

4. Principal component analysis (PCA). Dynamic factor models (DFM). State-space models and Kalman filtering. Mixed-frequency estimation of coincident indices. Basics of time-series imputation, assumptions, and limitations.
5. Univariate model-based time-series imputation. Multivariate PCA-based imputation. Multiple imputation (MI). Expectation-maximisation (EM) algorithm. Imputation with trends. Bayesian imputation, bootstrapping, and the Amelia II algorithm. Imputation diagnostics.

5. Grading

Regular class participation: 10% (every class meeting yields 2%).

Students need to come to the classroom; active participation and questions from the audience are encouraged.

Assignment: 30%. Due: 30$^{\text{th}}$ of April 2024. The first assignment tests the participants' ability to work with highly irregular indicators on the Luxembourg economy. Each student will be assigned 2 time series; the task is to give economic rationale behind the observed variability, to apply a seasonal adjustment procedure and evaluate the output quality.

Final project: 60%. Due: 30$^{\text{th}}$ of June 2024. The participants may choose one of the following tasks:

- *Batch processing.* Fetch a collection of 20 (in total) monthly and quarterly indicators of any EU economy, carry out seasonal adjustment, and describe the results. Pick such indicators that there be both short and long observed series. If there are closely related series, enforce similar treatment (e. g. petrol and diesel sales) based on some preliminary diagnostics. Produce rich and information-dense plots and write down SA recommendations for these series.
- *Stress-testing Eurostat recommendations.* Read the 2023 notes on seasonal and calendar adjustment for 10 EU countries and check if an improvement can be made. For the selected countries, pick one major indicator that is mentioned in the document, adjust it via the recommended method and then, try a different one (e. g. complement X13 with SEATS and vice versa). Compare the adjustment quality statistics. Describe the findings, highlight the discrepancies, and give a plausible economic explanation.
- *Effect of SA on predictive performance.* Following Plosser (1979), take 5 time series of high economic importance and compare the accuracy of 1-, 3-, 6-, and 12-step-ahead forecasts with and without adjustment using univariate and multivariate prediction methods. Exclude the last 3 years from estimation and compare multiple rolling forecasts against these values. Compare any two seasonality-based and two seasonality-agnostic forecasting methods. Describe your findings in a small report and give practical recommendations.
- *Studying false-positive rates.* Write a small Monte-Carlo study: simulate various series where a certain feature is absent and check how often various black-box procedures falsely 'capture' it in the data, i. e. make the Type-I error. Possible features (choose one or two): SARIMA specification, trading-day, week-day and/or Easter effects, stable seasonality, moving seasonality, outliers of various types (additive, shifts, temporary). Present the findings about false positive rates in multiple tables for various sample sizes for some economically reasonable data-generating processes.
- *Pushing the envelope.* Since monthly and quarterly series are not the only type of data affected by seasonality, there is ongoing research in adjustment of high-frequency or mixed data: daily, hourly, intra-weekly etc. Do a literature review, showcase the most recent advances (in the last 5 years) in seasonal adjustment literature, and implement any of the newest methods in a free and open-source statistical programming language (a proof of concept on simulated data and one application to any real series will suffice).
- *Reconstructing EU economic indicator panels.* Choose 3–5 economic indicators available from Eurostat that constitute imbalanced panels (i. e. time-series for different countries having different lengths or even frequencies). Alternatively, choose a balanced panel of 3–5 series and introduce overall 20% missingness by trimming some time series at the beginning and/or the end. Use DFM, PCA, and MI to impute the panel ends and compute EU aggregate indicators. Compare them with

the Eurostat-provided aggregates if they are available. Perform imputation diagnostics. Describe your findings in a small report and give practical recommendations.

Hint: **choose the task that is the most relevant for your research** or the one that can be later reused in other projects. You may reuse your existing material in the assignment.

6. Technical requirements for the final project

1. The submission must consist of a report and script / source code for the invocation of the methods used for estimation. Optionally, data files, or graphics files (PDF or PNG), or videos (in MP4 format with H264, H265, or AV1 codec) can be included. Compress multiple files into a single ZIP archive.
   Example: `smith.zip` containing `smith.Rmd`, `SP500.RData`, and `forecast.png`.
2. The practical part must be done in an **open-source** statistical package.
   Suggestions: R, JDemetra+, Python, Julia. Should the participants have relevant experience, they may also choose gretl, JASP, PSPP, Scilab, Octave, C/C++, Fortran, Java, or any language with an open-source compiler.[1]
3. The report itself can be written in any of these formats: plain text / Markdown (in any editor of your preference), LaTeX, Jupyter, knitr, or Sweave (a readable PDF is optional, but the compilation / rendering / knitting / weaving must terminate successfully without errors).
   `.doc[x]` and `.odt` are **not accepted**; `.jpg` graphics are strongly discouraged.
4. The analysis must be reproducible from the source code. **If code execution stops due to an error, the assignment will not be graded.**

7. Contact information

- Office: Campus Kirchberg, room G214
- Email: `andrei.kostyrka@uni.lu`
- Course homepage: moodle.uni.lu
- Office hours: By appointment

[1]The university is unlikely to purchase licence keys for technologically inferior and methodologically obsolete commercial software for the sake of assignment grading given the availability of excellent free alternatives officially endorsed by the European Commission, Eurostat, and ECB.