Time Series and Dynamic Econometrics
Assignment 2

Last updated: September 14, 2025

Instructions:

1. The assignment is made in groups of **exactly two students**. The same group is maintained for Assignments 1, 2, and 3.
2. The assignment is made using a programming language of choice, such as `python`/`R`/`matlab`/`Ox`. Software packages such as `Eviews` or `STATA` can **not** be used.
3. Within your chosen program language, you may use packages for computing statistical estimates (such as sample autocorrelations). However, forecasts and impulse response functions should be computed manually (using for loops).
4. The assignment must be **uploaded in PDF format** on Canvas before the **deadline**:

   Monday, September 29, 2025, at 23:59.

   The code must be uploaded as separate file(s) from the PDF.
5. The assignment is graded as either "complete" or "incomplete"; in the latter case the assignment must be revised. To pass the TSDE course, all three assignments must be graded as "complete."
6. On the first page of the assignment state the names and student numbers of the group members.
7. Do **not** copy the questions in your assignment (this invokes the plagiarism detection software). Simply write:
   1. <YOUR ANSWER TO SUBQUESTION 1>.
   2. <YOUR ANSWER TO SUBQUESTION 2>.
   3. ...

Remark (Plagiarism). *The VU has a* ***zero tolerance*** *policy for plagiarism, which can lead to exclusion from the Minor "Applied Econometrics: A Big Data Experience for All" without graduating. All assignments are checked for plagiarism using dedicated software, and cases that raise suspicion are automatically sent to the examination board for review.*

Part I: Forecasting, Parameter Estimation, and Model Selection

Let us go back in time to the first quarter of 2009. The world economy has just been hit by a major financial crisis. In just one year, the Dutch quarterly GDP growth rate has fallen from 1.4%, in the first quarter of 2008, to -2.7%, in the first quarter of 2009. In the first quarter of 2009, at the peak of the economic recession, suppose that government officials ask you to describe the dynamics of the Dutch GDP quarterly growth rate and deliver a forecast for the two years ahead. The available sample of observed GDP growth rates spans from the second quarter of 1987 to the first quarter of 2009. You can find this data set named `data_tsde_assignment_2_part_1.csv` on Canvas.

1. Plot the sample of *Dutch GDP quarterly growth rates* that you have at your disposal. Report and interpret the sample ACF up to 12 lags.

2. Estimate an AR(p) model with intercept for the given data. Use the Bayesian information criterion (BIC) to select p, with a maximum $p = 4$ lags. Report the final estimated AR(p) model. Comment on the estimated coefficients: what do they imply about the dynamic properties of the GDP quarterly growth rate? Based on the SACF, did you expect to obtain your selected model?

3. Use your estimated AR model to produce forecasts up to 2 years ahead (= 8 quarters, so $h = 1, 2, \ldots, 8$) for the Dutch GDP quarterly growth rate that ranges until the first quarter of 2011. Plot the forecasts.

4. Assume NID errors in the AR model.[1] Based on this assumption, produce 95% confidence intervals for your previous forecasts and add them to the plot. Next, investigate the residuals based on the estimated model. Do you find the assumption of NID errors reasonable? (Appendix A discusses tests for investigating this assumption.) How may this affect the confidence bounds for the forecasts?

[1] That is, $\{\varepsilon_t\} \sim NID(0, \sigma_\varepsilon^2)$, where the ε_t terms are typically referred to as the *errors* or *error terms* of the model. These should not be confused with the forecast errors, which are weighted sums of the ε_t.

Part II: Impulse Response Functions, Autoregressive Distributed Lag Models, and Granger Causality

By the beginning of 2014, in the face of fast rising unemployment, the Dutch Ministry of Social Affairs is concerned that the provisions for future government expenditure with social pensions may be severely underestimated. This is particularly true if the economy is hit again by a large negative shock. Suppose that you have been asked to analyze alternative unemployment scenarios. The available sample of observed unemployment rates and GDP growth rates ranges from the second quarter of 1987 to the first quarter of 2014. You can find this data set named `data_tsde_assignment_2_part_2.csv` on Canvas.

1. Plot the sample of *Dutch quarterly unemployment rates* and *Dutch GDP quarterly growth rates* that you have at your disposal. Estimate an AR model for the GDP growth rate and an ADL model for the unemployment rate using the GDP growth rate as exogenous explanatory variable. Use the Akaike information criterion (AIC) for model selection. For the AR model use a maximum of four lags of GDP; for the ADL model use a maximum of four lags for both variables. Report the final estimated AR and ADL models. Comment on the estimated coefficients: what do they imply regarding the dynamic properties of the unemployment rate and the GDP growth rate?

2. For the ADL model, report the p-values for the estimated coefficients. At a 10% level of significance, does GDP growth *Granger-cause* the unemployment rate?

3. Based on the estimated ADL model, derive the long-run equilibrium relation between the unemployment rate and the GDP growth rate.

4. Use *impulse response functions* (IRFs) to analyze two different scenarios for the Dutch unemployment rate:

   (a) In the 'good scenario' the GDP quarterly growth rate is hit by a positive shock of 2%.[2]

   (b) In the 'bad scenario' the GDP quarterly growth rate suffers a negative shock of 2%.

   Use the final observed value of the unemployment rate and the GDP growth rate as origins for your IRFs. Plot the IRFs for the unemployment *and* GDP growth rates in both scenarios (four IRFs in total). Use a sufficient number of periods to show that the IRFs converge to the origin.

[2] Since the data are already in percentages, the shock to be applied to the GDP quarterly growth rate is simply 2. (So **not** 0.02 or 0.02 times the final GDP growth value.)

A Testing Model Assumptions

A.1 Testing for normality

To test for normality of a time series, we can use the *Jarque-Bera test*, which has hypotheses

$$H_0\colon \text{The data are normally distributed} \qquad \text{vs} \qquad H_1\colon \text{The data are } \textit{not} \text{ normally distributed}$$

For multiple regression analysis, the test statistic based on the least squares residuals as data is:

$$JB = \frac{T-k}{6}\left(\hat{S}^2 + \frac{1}{4}(\hat{K}-3)^2\right),$$

with

k: the number of covariates (regressors),

T: the sample size,

$\hat{S}$: the sample skewness,

$\hat{K}$: the sample kurtosis.

The test is asymptotically chi-squared with two degrees of freedom under the null hypothesis,

$$JB \sim \chi^2_2 \quad \text{under} \quad H_0 \quad \text{as} \quad T \to \infty.$$

By applying the test to the residuals of the AR(p) model, with $k = p$, it can be used to test the normality assumption for ε_t.

A.2 Testing for autocorrelation

Consider the following time series regression model

$$X_t = \beta_0 + \beta_1 Z_{1,t} + \cdots + \beta_p Z_{p,t} + \varepsilon_t, \qquad \mathbb{E}[\varepsilon_t] = 0,$$

for $t = 1, \ldots, T$, with dependent variable X_t and covariates $Z_{i,t},\ i = 1, \ldots, p$. The AR($p$) model is an important special case with

$$Z_{i,t} = X_{t-i}, \qquad i = 1, \ldots, p \qquad \text{and} \qquad \{\varepsilon_t\} \sim WN(0, \sigma^2_\varepsilon).$$

We can investigate the white noise assumption for the ε_t by testing for autocorrelation using the *Breusch-Godfrey test*. For $k \in \mathbb{N}$, the test has hypotheses

$$H_0\colon \{\varepsilon_t\} \text{ has no autocorrelation up to lag } k \qquad \text{vs} \qquad H_1\colon \{\varepsilon_t\} \text{ has autocorrelation up to lag } k$$

Denote the least squares residuals from a regression based on the above time series model by e_t. The test consists of two parts:

1. *Auxiliary regression*: Perform a regression of e_t on the covariates and k of its own lags:

$$e_t = \alpha_0 + \alpha_1 Z_{1,t} + \cdots + \alpha_p Z_{p,t} + \delta_1 e_{t-1} + \cdots + \delta_k e_{t-k} + u_t.$$

2. Let R^2 denote the R-squared from the auxiliary regression and T the original sample size. Then the test statistic $BG \equiv T \times R^2$ is asymptotically chi-squared distributed with k degrees of freedom:

$$BG \equiv T \times R^2 \sim \chi^2_k \quad \text{under} \quad H_0 \quad \text{as} \quad T \to \infty.$$

There are no general guidelines on how to choose k, but a simple rule of thumb is to set $k = \sqrt{T}$. Alternatively, k may be chosen by minimizing an information criterion (e.g., AIC/BIC).

Note: In the case of an AR(p) model, the auxiliary regression

$$e_t = \alpha_0 + \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + \delta_1 e_{t-1} + \cdots + \delta_k e_{t-k} + u_t$$

can be started at time $t = \max\{p, k\} + 1$ to ensure that X_{t-p} and e_{t-k} are available.