Group Assignment 2

Advanced Econometrics

Etienne Wijler (Coordinator and Lecturer)

Noah Stegehuis (Coding instructor)

Gabriele Mingoli (Tutorial instructor)

Notes and instructions:

1. This assignment is mandatory.
2. The assignment is to be made in groups of 4 students. You can create your own group in Canvas > People > Assignment Groups, and then self-enroll or join an incomplete team. Please be professional and welcoming to new team members.
3. Only one of you needs to hand in all files.
4. The deadline for delivery of this assignment is on Tuesday, October 8, at 23:59h

   There will be no tolerance period for late deliveries. Deliveries after the assigned deadline imply that you have a final grade of zero for the assignment (GA2 = 0).
5. To get the full score for this assignment, the following three things must be done:

   (a) upload your final report as a PDF file in Canvas Assignments. Name the file **GA2report_2601842_2511351_2661510_2639486.pdf**, where the numbers are replaced by the VU student numbers of the 4 group members.To write you report according to academic standards follow the relevant tips that we have in the questions and also check the example report file under the name 'example_report.pdf' on Canvas > Assignments > Group Assignment 2: Instructions.

   (b) upload a zip file of your runnable R or Python code in Canvas Assignments. Name the file **GA2code_2601842_2511351_2661510_2639486_language.zip**, where the numbers are replaced by the VU student numbers of the 4 group members (or 3 if your group consists of 3 people) and language is either R or Python.

   The code file(s) should be clear, well commented, and directly runnable, so that it reads the datafile and obtains the results of all questions and prints them. Your initial comments in the file should hold your names and student numbers.

(c) upload a pdf of your entire code in Canvas Assignments. Name the file **GA2code_2601842_2511351_2661510_2639486_language.pdf**, where the numbers are replaced by the VU student numbers of the 4 group members (or 3 if your group consists of 3 people) and language is either R or Python. The file should be well readable, with proper indentations and should not contain pictures/photos/screenshots of code snippets.

6. As a standard anti-fraud measure, we will at random select a number of you to explain your code and answers. Any one of you must be able to explain any part of the code. Failure to explain your answers will result in a deduction of credits for this assignment.

7. For the support for the assignments, carefully read the announcement we put out at the start of the course and consult the discussion boards related to the assignments.

We wish you success!!

VU VRIJE UNIVERSITEIT AMSTERDAM

School of Business and Economics

Modelling unemployment rates

In this assignment you are tasked with uncovering the dynamic dependencies in US unemployment rates. The unemployment rate is a key macro-economic indicator that greatly affects an economy's resilience and therefore provides valuable guidance to monetary policy. You are required to download the data yourself for this assignment, but thanks to the hard work of McCracken and Ng (2016) this has now become incredibly easy.

From this website, you can download large collections of macro-economic variables observed at a monthly or quarterly frequency. The database is continuously maintained and updated with new data vintages by Dr. McCracken, senior economic policy advisor for the US Federal Reserve Bank. The database is of high quality and has found widespread usage in empirical work by practitioners and academics alike.

Data extraction and processing

Download the file "2024-7.csv" under the *monthly data* column. From this dataset, extract the columns containing the date ("sasdates"), unemployment rate ("UNRATE") and industrial production ("INDPRO"). Next, discard all observations *after* December 2019. When plotting the unemployment rates or the industrial production, it should be obvious to you that these series do not look stationary. Accordingly, you should take the first differences of both of these time series. For your reference, you may replicate the plots in Figure 1, to ensure that you are working with the correct data.

Estimation of dynamic models

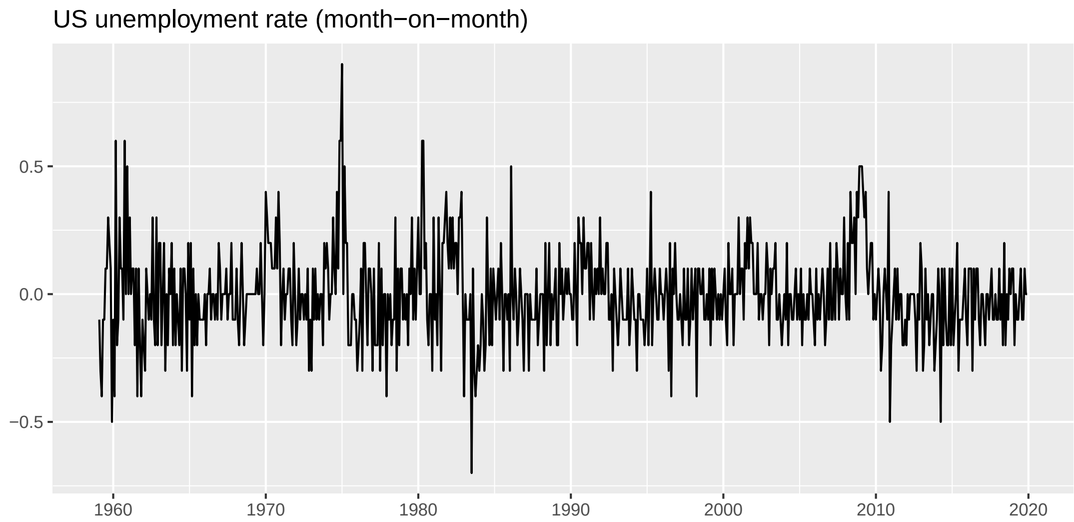
Let x_t and z_t denote the unemployment rate and industrial production at time t, respectively. Consider the following three models for the unemployment rate:

$$\text{AR:} \quad x_t = \mu + \delta x_{t-1} + \epsilon_t, \tag{1}$$

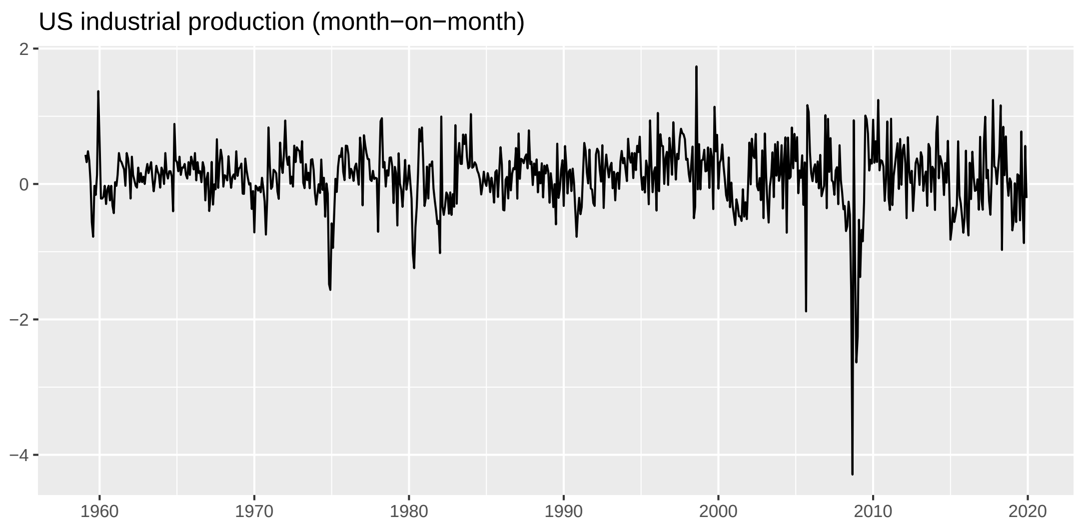
$$\text{SESTAR:} \quad x_t = \mu + \left(\delta + \frac{\gamma}{1 + e^{\alpha + \beta x_{t-1}}}\right) x_{t-1} + \epsilon_t, \tag{2}$$

$$\text{STAR:} \quad x_t = \mu + \left(\delta + \frac{\gamma}{1 + e^{\alpha + \beta z_{t-1}}}\right) x_{t-1} + \epsilon_t, \tag{3}$$

where ϵ_t is some independent and identically distributed sequence with $\mathbb{E}\left(\epsilon_t^2\right) < \infty$.



(a)



(b)

Figure 1: These plots display **(a)** the month-on-month changes in US unemployment rates and **(b)** the month-on-month changes in industrial production. The data ranges from February 1959 to December 2019.

Question 1: Are these models (semi)parametric or (semi)non-parametric? Is any of the models nested in one of the other?

Question 2: Code up the non-linear least-squares loss functions for all three models. Defining the parameter vector $\boldsymbol{\theta} = (\mu, \delta, \gamma, \alpha, \beta)'$, provide in the report the value of the total loss of all three models when evaluated at $\tilde{\boldsymbol{\theta}} = (0, 0.1, 2, 0, 3)'$ (clearly only the first two parameters are relevant for the AR(1) model).

- Tip: To help you on your way, we provide the total loss (L) you should obtain (exactly) with the parameter $\boldsymbol{\theta}^* = (0, 0.3, 1, 0, 2)'$:
  AR(1): $L = 23.7398$
  SESTAR: $L = 34.8042$
  STAR: $L = 33.2778$

Question 3: Estimate the three models with the use of non-linear least squares (NLS). For each model, make a separate line plot (3 panels in total) containing the unemployment rate with the *fitted values* of the unemployment rate. Make sure that your plot contains a legend. Discuss the differences between the model fits.

- Tip: Pay particular attention to three periods: Nov 1973 - March 1975, January 1980 - July 1980 and December 2007 - June 2009. What happened in these periods? How does this help explain your results?
- Tip: For this question, simply the plot is enough. The reporting of the estimated parameters and optimal loss will happen later in Question 6 in which you have to provide a table.
- Tip: As an initial value for optimisation, you can use $\boldsymbol{\theta} = (\mu, \delta, \gamma, \alpha, \beta) = (0, 0.3, 0, 0, 6)$. The models might be sensitive to the initial value and the optimisation algorithm you choose. You can try different settings, but always check whether the optimiser converged and if the final function value is indeed an improvement compared to standard configurations.
- Tip: for requirements on the layout of the figures, please consult our extensive tips in previous assignments.

Question 4: For the STAR model, plot the values of the (differenced) industrial production index and the value of the time varying autoregressive parameter, i.e.

$$G(z_{t-1}, \boldsymbol{\theta}) = \delta + \frac{\gamma}{1 + e^{\alpha + \beta z_{t-1}}},$$

in a single line plot. Comment on the observed relationship between the persistence in the unemployment rates and industrial production. Motivate this relationship based on your own intuition and/or economic theory.

Statistical inference

Question 5: Define $\hat{\epsilon}_t$ as the residual of the SESTAR model. Estimate the auto-correlations $\gamma_1, \ldots, \gamma_{12}$ with $\gamma_p = \mathbb{E}(\hat{\epsilon}_t \hat{\epsilon}_{t-p})$ and provide an ACF plot. Based on this plot, do you believe this model to be correctly specified? Motivate your answer.

- Tip: both R and Python have build-in functions that directly compute and plot the auto-correlations. It is recommended to use these. In R for example the `Acf` function in the `forecast` package can be used, and in Python `plot_acf` from the `statsmodels.graphics.tsaplots` package. The default settings in these functions do not always provide the most informative plots, so play around with the function parameters.

Question 6: Taking into account the conclusions you made in Question 5, now estimate the *robust* standard errors of all estimated parameters. Following the slides on Chapter 8, you may estimate

$$\boldsymbol{\Sigma} := \mathbb{V}\mathrm{ar}\left(\frac{1}{\sqrt{T}} \sum_{t=2}^{T} \nabla q(x_t, x_{t-1}, \boldsymbol{\theta}_0)\right)$$

with the famous Newey-West estimator

$$\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\Sigma}}_0 + \sum_{j=1}^{p} \left[1 - \frac{j}{p+1}\right] \left(\hat{\boldsymbol{\Sigma}}_j + \hat{\boldsymbol{\Sigma}}_j'\right),$$

where

$$\hat{\boldsymbol{\Sigma}}_j = \frac{1}{T} \sum_{t=j+2}^{T} \hat{\nabla} q\left((x_t, x_{t-1}, \hat{\boldsymbol{\theta}}\right) \hat{\nabla} q\left((x_{t-j}, x_{t-1-j}, \hat{\boldsymbol{\theta}}\right)'.$$

In the report, provide a mathematical expression of your variance estimator. Furthermore, put your results in a nice and informative table, as in Table 1, in which you also report the estimates themselves (with your standard errors in parentheses), the values of the optimal loss, the R^2 and adjusted R^2.

- Tip: The value for p is related to the degree of serial correlation in the score. This would normally have to be chosen in a data-driven way, but for the sake of simplification you may set $p = 12$ in this exercise.

Question 7: Formally test whether there are non-linear time-dependence effects in the SESTAR and STAR model. Write down your hypothesis tests in symbols and report the associated p-values of your hypothesis test.

Table 1: Example

Here I explain everything that is in the table without interpreting the table of course. This text can actually be pretty long to ensure that the table is really stand alone.

	AR	SESTAR	STAR
μ	0.11	0.11	0.11
	(0.20)	(0.20)	(0.20)
δ	0.11	0.11	0.11
	(0.20)	(0.20)	(0.20)
γ		0.11	0.11
		(0.20)	(0.20)
α		0.11	0.11
		(0.20)	(0.20)
β		0.11	0.11
		(0.20)	(0.20)
$Loss$	10	10	10
R^2	0.10	0.10	0.10
R^2_{adj}	0.09	0.09	0.09

Question 8: Finally, based on your output from Question 6 and answer to Question 7, comment on which model you believe to fit the data best.

References

McCracken, M. W. and Ng, S. (2016). Fred-md: A monthly database for macroeconomic research. Journal of Business & Economic Statistics, 34(4):574–589.