Assignment 2019–2020

Advanced Time Series Econometrics

Francisco Blasques

Notes and instructions:

1. The assignment is to be made in groups of two or three students. No more, no less.
2. The deadlines for delivery of this assignment are:
   - Part 1: Monday, 13 January, at 16:00
   - Part 2: Monday, 20 January, at 16:00
3. The assignment must be sent in pdf format to `f.blasquesvu.nl` before the deadlines mentioned above.
4. You must also submit all the MATLAB, R or Python files that you used to obtain the answers. The code should be clear and well commented. **Note:** Please inform me in advance if you will not be using MATLAB, Python or R.
5. The assignment must state the name and student number of each member of the group.

PART I: Filtering methods - Practice

Optimal Airbag Deployment

The use of airbags in modern cars has dramatically reduced the mortality of road accidents. Accelerometers which measure changes in the car's velocity, are used to determine if the airbag should be deployed. In particular, airbags are deployed when an accelerometer detects an abrupt negative variation in the vehicle's speed.

Unfortunately, when using the break pedal, in bumpy roads, or city traffic, the car's vibrations and changes in speed may lead the accelerometer to incorrectly deploy the airbags. This can be dangerous to both the occupants of the car and other nearby vehicles or pedestrians. Observation driven filters are thus used to create a *signal* s_t which is less sensitive to changes in the car's velocity, vibrations, and other noise components which influence the accelerometer's measurements.

The data set `airbag_data.csv` contains 10 seconds of accelerometer data from 10 different vehicles. The accelerometer provides data in milliseconds. Cars 1 to 5 did not have an accident. Cars 6 to 10 had road accidents occurring at the very end of the sample.

1. Plot the accelerometer data for the 10 different cars.

The 'deployment rule' for the airbag is that the airbag deploys if the signal s_t crosses below the threshold value of -1, otherwise, the airbag is not deployed,

$$\begin{cases} \text{Deploy airbag} & \text{if} \quad s_{i,t} \leq -1 \\ \text{Do not deploy airbag} & \text{if} \quad s_{i,t} > -1 \end{cases}$$

Ideally we would like the airbag to be deployed only in cars which actually experienced an accident.

2. Consider using the raw data as a signal. In other words, set

$$s_{i,t} = x_{i,t-1}$$

where $s_{i,t}$ denotes the signal for car i at time t, and $x_{i,t}$ denotes the raw accelerometer data for car i at time t. For which cars is the airbag deployed? What is the success rate of your deployment rule? i.e. for how many cars is the airbag deployment decision correct?

3. When answering the previous question, you may have noticed that using the raw signal is not a good idea simply because it is not well scaled for the selected threshold of value of -1. Consider the re-scaled signal given by

$$s_{i,t} = 10x_{i,t-1}.$$

For which cars is the airbag deployed? What is the success rate of your deployment rule?

4. You may have found that the (scaled) raw accelerometer data is too noisy to deliver an accurate signal for airbag deployment. With this in mind, we consider generating a signal using the following local-level observation driven filters:

$$s_{i,t+1} = -0.0135 + 0.4x_{i,t} + 0.975s_{i,t} \ ,$$

$$s_{i,t+1} = -0.0135 + 0.8x_{i,t} + 0.975s_{i,t} \ .$$

$$s_{i,t+1} = -0.0135 + 0.5x_{i,t} + 0.975s_{i,t} \ ,$$

   For each car, please plot the raw data and signals produced by each of these signals. Which of these signals provides the best airbag deployment performance? What is the success rate of the deployment rule for each signal?

Yaw Alignment of Wind Turbines

The output of a wind turbine depends, among other factors, on the proper alignment of the rotor. Ideally, the rotor should always be facing the wind directly as the turbine is less efficient when it 'sees' the wind arriving at an angle. In particular, the relative efficiency of a wind turbine is well approximated by the function

$$\mathrm{RE}_t = 100 \times \left(1 - \frac{c^2 y_t^2}{1 + c^2 y_t^2}\right)$$

where $c \approx 0.1$. The relative efficiency of the turbine at time t, denoted RE_t, is a number between 100 (maximum efficiency) and 0 (minimum efficiency). The variable $y_t = (w_t - r_t)$ is the *wind direction offset*, which is a measure of the discrepancy between the wind direction w_t and the position of the rotor r_t. A zero offset ($y_t = 0$) is optimal and occurs when the rotor's position r_t is perfectly aligned with the incoming wind's direction w_t, The extreme case of $y_t \to \pm\infty$, occurs as the position of the rotor's position r_t becomes progressively worse and approaches a 90 degree angle relative to the wind's direction w_t.

The wind direction itself is typically measured using a simple *wind vane* installed at the top of (or near) the wind turbine. Unfortunately, measured wind direction w_t is often noisy and an attempt to align the rotor in reaction to any gush of wind coming sideways would render the wind turbine less efficient. Instead, observation-driven filters such as the one below are used to filter the wind direction carefully and determine the optimal position of the rotor

$$r_{t+1} = \alpha(w_t - r_t) + \beta r_t.$$

Obviously, a good filter is capable of separating the noise component from the fundamental slow shift in wind direction. This allows wind turbines to be more efficient in producing electricity.

A wind turbine manufacturer is currently setting the parameters α and β to the values $\alpha = 0.5$ and $\beta = 1$. They argue that, with this parameter configuration, their wind turbines can smoothly adjust to changes in the wind direction, achieve a small offset y_t, and an average relative efficiency above 90. They are however interested in improving further the efficiency of their turbines by fine-tuning these parameters.

The dataset `wind_data.csv` contains a one hour time series of wind direction w_t, measured in seconds, at a specific location.

5. Plot the wind direction data.

6. Calculate the efficiency obtained when the lagged raw wind direction data is used directly to set the rotor's position; i.e. set $r_t = w_{t-1}$ and calculate

$$\frac{1}{T}\sum_{t=1}^{T} 100 \times \left(1 - \frac{c^2 y_t^2}{1 + c^2 y_t^2}\right).$$

7. Calculate the in-sample relative efficiency of the turbine using the company's default filter parameters $\alpha = 0.5$ and $\beta = 1$. Is it true that the relative efficiency of the wind turbine is above 90?

8. Use the available data to fine-tune the filtering parameters $\boldsymbol{\theta} = (\alpha, \beta)$ that determine the position of the rotor and optimize the expected relative efficiency of the wind turbine.

$$\hat{\boldsymbol{\theta}}_T = \arg\max_{\boldsymbol{\theta}} \frac{1}{T}\sum_{t=1}^{T} 100 \times \left(1 - \frac{c^2 y_t^2}{1 + c^2 y_t^2}\right)$$

   Which parameter values did you obtain for α and β? Did you manage to improve the relative performance of the wind turbine?

Impressed by your excellent performance, the wind turbine manufacturer has asked you to fine-tune the parameters using as much data as you wish. They claim that they can set the parameters α and β with an accuracy level of 0.0001.

9. Suppose that the estimator $\hat{\boldsymbol{\theta}}_T$ of the unknown optimal vector of parameters $\boldsymbol{\theta}_0 = (\alpha_0, \beta_0)$ is consistent and asymptotically Gaussian,

$$\sqrt{T}(\hat{\alpha}_T - \alpha_0) \xrightarrow{d} N(0, 0.039^2) \ ,$$

$$\sqrt{T}(\hat{\beta}_T - \beta_0) \xrightarrow{d} N(0, 0.021^2) \ ,$$

where 0.039^2 and 0.021^2 are estimates of the asymptotic variance of the estimator. How much data do you need in order to deliver estimates of α_0 and β_0 with an accuracy level of 0.0001, at a 99.9% confidence level? i.e. what is the sample size T that allows you to state with 99.9% confidence that your estimate has an accuracy of 0.0001,

$$P(|\hat{\alpha}_T - \alpha_0| < 0.0001) = 0.999 \ ,$$

$$\text{and} \quad P(|\hat{\beta}_T - \beta_0| < 0.0001) = 0.999 \ .$$

The data contained in the file `wind_data.mat` shows the behavior of a wind vane under 'normal weather'. In stormy conditions however, it is typically useful to consider robust filters, or to combine multiple measurements of wind direction.

10. Please explain why the following robust filter given by,

$$r_{t+1} = \alpha \tanh(\gamma(w_t - r_t)) + \beta r_t,$$

might perform better than a linear filter in stormy conditions,

$$r_{t+1} = \alpha(w_t - r_t) + \beta r_t.$$

11. Please explain why a robust filter that combines the wind direction information from several wind vanes might further improve your results. You can take the following dynamic factor filter,

$$r_{t+1} = \alpha \tanh\left(\frac{1}{9}\sum_{i=1}^{9} w_{i,t} - r_t\right) + \beta r_t$$

where $w_{i,t}$ denotes the wind direction as measured by the ith wind vane.

PART 2: Filtering methods - Theory

Industrial Production Index

The Industrial Production Index (IPI) is an economic indicator that measures the real production output of industrial establishments and covers sectors such as mining, manufacturing, electricity, oil, and gas. This index, along with other industrial indices and construction, accounts for the bulk of the variation in national output over the duration of the business cycle.

IPI data is useful for managers and investors within specific lines of business, while the composite index is an important macroeconomic indicator for governments, central banks, economists and investors.

The IPI plays also a crucial role in the forecasting of economic activity, as well as the development of recession warning systems since it is available at a monthly frequency, which contrasts with the quarterly frequency at which GDP figures are published. Fluctuations within the industrial sector account for most of the variation in overall economic growth, so a monthly metric helps keep investors apprised of shifts in output.

Unlike GDP growth rates, however, the IPI growth rates are more volatile and contain outliers. The presence of fat-tailed innovations can render linear Gaussian filters quite inaccurate, and hence, robust nonlinear filtering methods play an important role in modeling IPI. Below, we compare two filters of the conditional expectation of the series of 1st differences in IPI.

Linear Gaussian observation-driven model:

$$x_t = \mu_t + \varepsilon_t \ , \quad \{\varepsilon_t\}_{t\in\mathbb{Z}} \sim \mathrm{NID}(0,\sigma^2) \ ,$$
$$\mu_{t+1} = \omega + \alpha(x_t - \mu_t) + \beta\mu_t \ .$$

Robust non-Gaussian observation-driven model:[1]

$$x_t = \mu_t + \varepsilon_t \ , \quad \{\varepsilon_t\}_{t\in\mathbb{Z}} \sim \mathrm{TID}(0,\sigma^2,\lambda) \ ,$$
$$\mu_{t+1} = \omega + \alpha\frac{x_t - \mu_t}{1+\delta(x_t-\mu_t)^2} + \beta\mu_t \ .$$

[1]Below $\mathrm{TID}(0,\sigma^2,\lambda)$ denotes data that is independently distributed with generalized student-t distribution, with mean zero, scale parameter σ^2 and λ degrees of freedom. The probabily density function given by

$$f(\varepsilon_t) = \frac{1}{\sigma}K(\lambda)\Big(1+\frac{1}{\lambda}\Big|\frac{\varepsilon_t}{\sigma}\Big|^2\Big)^{-(\lambda+1)/2} \quad \text{where} \quad K(\lambda) = \frac{1}{\lambda^{1/2}}\frac{1}{B(\eta/2,1/2)}$$

where $\mathrm{B}(\cdot,\cdot)$ is the beta function.

The dataset `IPI_data.csv` contains the monthly growth rate of the IPI index for the US spanning 100 years from 1919 to 2019.

1. Plot the available IPI data.
2. Analyze the linear Gaussian observation-driven filtering model as a data generating process. In particular, give conditions for $\{x_t\}_{t\in\mathbb{Z}}$ to be SE and have four bounded moments $\mathbb{E}|x_t|^4 < \infty$.
3. Give conditions for the invertibility of the linear Gaussian filter observation-driven assuming the correct specification of the model.
4. Give conditions for the consistency and asymptotic normality of the MLE for the parameters of the linear Gaussian observation-driven model assuming the correct specification of the model.
5. Estimate the parameters of the linear Gaussian observation-driven model by MLE.
6. Analyze the robust nonlinear non-Gaussian observation-driven filtering model as a data generating process. In particular, give conditions for $\{x_t\}_{t\in\mathbb{Z}}$ to be SE and have four bounded moments $\mathbb{E}|x_t|^4 < \infty$.
7. Give sufficient conditions for the invertibility of the robust non-linear non-Gaussian observation-driven filter assuming the correct specification of the model.
8. Give conditions for the consistency of the MLE for the parameters of the robust nonlinear non-Gaussian model assuming the correct specification of the model.
9. Estimate the parameters of the robust nonlinear non-Gaussian model by MLE.