Applied Macroeconomic Modelling

Building a Sample for Time Series Analysis

Gauthier Vermandel

Objectives

- Downloading data in real time in MATLAB;
- Getting a stationary sample for macroeconomic analysis.

Additional reading list

- Nelson, Charles R., and Charles R. Plosser. "Trends and random walks in macroeconmic time series: some evidence and implications." Journal of monetary economics 10.2 (1982): 139-162.
- Canova, Fabio. Methods for applied macroeconomic research. Vol. 13. Princeton university press, 2007.

Back in time

- In the 70s, revolution in forecasting analysis: simple models perform much better than large macroeconometrics models [Nelson, 1972].
- Influencial revolution, including in macroeconomic theory: pushed toward much pacimonious models (Lucas' Revolution).
- Development of a rich set of time series models: AR, MA, VAR, ARCH, ... up to DSGE models (that we will estimate later on).

Back in time

- In the 70s, revolution in forecasting analysis: simple models perform much better than large macroeconometrics models [Nelson, 1972].
- Influencial revolution, including in macroeconomic theory: pushed toward much pacimonious models (Lucas' Revolution).
- Development of a rich set of time series models: AR, MA, VAR, ARCH, ... up to DSGE models (that we will estimate later on).

Back in time

- In the 70s, revolution in forecasting analysis: simple models perform much better than large macroeconometrics models [Nelson, 1972].
- Influencial revolution, including in macroeconomic theory: pushed toward much pacimonious models (Lucas' Revolution).
- Development of a rich set of time series models: AR, MA, VAR, ARCH, ... up to DSGE models (that we will estimate later on).

Data in forecasting divisions

- Policy making institutions (Banque de france, French Treasury) operates in real time analysis.
- As new data are released, applied economists must re-run their forecasting models taking into account the latest observation released.
- In recent times, this policy work operates with real time data to facilitate the repetitive work of the econometrician.
- We will use DB·nomics to get data in real time.

Data in forecasting divisions

- Policy making institutions (Banque de france, French Treasury) operates in real time analysis.
- As new data are released, applied economists must re-run their forecasting models taking into account the latest observation released.
- In recent times, this policy work operates with real time data to facilitate the repetitive work of the econometrician.
- We will use DB·nomics to get data in real time.

Data in forecasting divisions

- Policy making institutions (Banque de france, French Treasury) operates in real time analysis.
- As new data are released, applied economists must re-run their forecasting models taking into account the latest observation released.
- In recent times, this policy work operates with real time data to facilitate the repetitive work of the econometrician.
- We will use DB·nomics to get data in real time.

Data in forecasting divisions

- Policy making institutions (Banque de france, French Treasury) operates in real time analysis.
- As new data are released, applied economists must re-run their forecasting models taking into account the latest observation released.
- In recent times, this policy work operates with real time data to facilitate the repetitive work of the econometrician.
- We will use DB·nomics to get data in real time.

Reading list:

I assume you have a good knowledge in standard time series models, the reading list can be shortened to:

1. Basic assessment of the stationarity in MATLAB: ADF Test
2. A guide for DBnomics.

Outline

Plan

Core assumption in conventional models

- A time series Y_t is a process in sequence over time $t \in \{1, 2, ..T\}$:

$$\{y_1, y_2, ..., y_T\}$$

  where $y_1, y_2, ..., y_T$ are interpreted as random variables.
- How was this sample generated?
- Understanding how this sample was generated is a fundamental task for an applied economist.
- By identifying the **underlying data-generating process**, we can develop meaningful quantitative analyses (e.g. causality analysis, forecasting, counterfactuals, etc).

Core assumption in conventional models

- A time series Y_t is a process in sequence over time $t \in \{1, 2, ..T\}$:

$$\{y_1, y_2, ..., y_T\}$$

  where $y_1, y_2, ..., y_T$ are interpreted as random variables.
- How was this sample generated?
- Understanding how this sample was generated is a fundamental task for an applied economist.
- By identifying the **underlying data-generating process**, we can develop meaningful quantitative analyses (e.g. causality analysis, forecasting, counterfactuals, etc).

Core assumption in conventional models

- A time series Y_t is a process in sequence over time $t \in \{1, 2, ..T\}$:

$$\{y_1, y_2, ..., y_T\}$$

  where $y_1, y_2, ..., y_T$ are interpreted as random variables.
- How was this sample generated?
- Understanding how this sample was generated is a fundamental task for an applied economist.
- By identifying the **underlying data-generating process**, we can develop meaningful quantitative analyses (e.g. causality analysis, forecasting, counterfactuals, etc).

Core assumption in conventional models

- A time series Y_t is a process in sequence over time $t \in \{1, 2, ..T\}$:

$$\{y_1, y_2, ..., y_T\}$$

  where $y_1, y_2, ..., y_T$ are interpreted as random variables.

- How was this sample generated?
- Understanding how this sample was generated is a fundamental task for an applied economist.
- By identifying the **underlying data-generating process**, we can develop meaningful quantitative analyses (e.g. causality analysis, forecasting, counterfactuals, etc).

Core assumption in conventional models

- One can assume a specific **Data Generating Process** (i.e. a time series model) with specific distribution for prediction errors. These models typically relate the present value of a series to past values (AR model) and past prediction errors (MA):

$$y_t = f\left(y_{t-1}, y_{t-2}, .., y_1\right)$$

- One can use the model for forecasting purpose: mean-wise h horizon forecast $E\left(y_{t+h}|I_t\right)$ with $I_t = \{Y_1, Y_2, ..., Y_t\}$. In practice, computation of $E\left(y_{t+h}|I_t\right)$ numerically costly (e.g. monte-carlo)...
- ... but under certains conditions (stationarity, gaussian errors, linearity) the computation of $E\left(y_{t+h}|I_t\right)$ is **straightforward**.
- In this class, our focus will be only on **linear gaussian time series models**, which imply to be careful on the features of the sample we utilize.

Core assumption in conventional models

- One can assume a specific **Data Generating Process** (i.e. a time series model) with specific distribution for prediction errors. These models typically relate the present value of a series to past values (AR model) and past prediction errors (MA):

$$y_t = f\left(y_{t-1}, y_{t-2}, .., y_1\right)$$

- One can use the model for forecasting purpose: mean-wise h horizon forecast $E\left(y_{t+h}|I_t\right)$ with $I_t = \{Y_1, Y_2, ..., Y_t\}$. In practice, computation of $E\left(y_{t+h}|I_t\right)$ numerically costly (e.g. monte-carlo)...
- ... but under certains conditions (stationarity, gaussian errors, linearity) the computation of $E\left(y_{t+h}|I_t\right)$ is **straightforward**.
- In this class, our focus will be only on **linear gaussian time series models**, which imply to be careful on the features of the sample we utilize.

Core assumption in conventional models

- One can assume a specific **Data Generating Process** (i.e. a time series model) with specific distribution for prediction errors. These models typically relate the present value of a series to past values (AR model) and past prediction errors (MA):

$$y_t = f\left(y_{t-1}, y_{t-2}, .., y_1\right)$$

- One can use the model for forecasting purpose: mean-wise h horizon forecast $E\left(y_{t+h}|I_t\right)$ with $I_t = \{Y_1, Y_2, ..., Y_t\}$. In practice, computation of $E\left(y_{t+h}|I_t\right)$ numerically costly (e.g. monte-carlo)...
- ... but under certains conditions (stationarity, gaussian errors, linearity) the computation of $E\left(y_{t+h}|I_t\right)$ is **straightforward**.
- In this class, our focus will be only on **linear gaussian time series models**, which imply to be careful on the features of the sample we utilize.

Core assumption in conventional models

- One can assume a specific **Data Generating Process** (i.e. a time series model) with specific distribution for prediction errors. These models typically relate the present value of a series to past values (AR model) and past prediction errors (MA):

$$y_t = f\left(y_{t-1}, y_{t-2}, .., y_1\right)$$

- One can use the model for forecasting purpose: mean-wise h horizon forecast $E\left(y_{t+h}|I_t\right)$ with $I_t = \{Y_1, Y_2, ..., Y_t\}$. In practice, computation of $E\left(y_{t+h}|I_t\right)$ numerically costly (e.g. monte-carlo)...
- ... but under certains conditions (stationarity, gaussian errors, linearity) the computation of $E\left(y_{t+h}|I_t\right)$ is **straightforward**.
- In this class, our focus will be only on **linear gaussian time series models**, which imply to be careful on the features of the sample we utilize.

Core assumption in conventional models

But which kind of features should be look for?

- **Stationarity** (weak definition) typically allows the mean function $E(y_t)$ and auto-covariance function $E[(y_s - E(y_s))(y_t - E(y_t))]$ to be time independent.

- One can estimate the parameters of the time series model $\hat{f}(\cdot)$, and provide out-of-sample forecast assuming distribution of y_t **remains unchanged** over the forecasting horizon (very convenient).

- Suppose our time series model is $y_t = \mu + \phi y_{t-1} + \varepsilon_t$ with $\varepsilon_t \sim N(0, \sigma^2)$. One can estimate $\hat{\mu}$, $\hat{\phi}$ and $\hat{\sigma}$, and compute the mean-wise forecast $E(y_{t+1}|I_t) = \hat{\mu} + \hat{\phi} Y_t$ and quantify uncertainty: $E(y_{t+1}|I_t) \pm 1.96 \cdot \hat{\sigma}$.

Core assumption in conventional models

But which kind of features should be look for?

- **Stationarity** (weak definition) typically allows the mean function $E(y_t)$ and auto-covariance function $E[(y_s - E(y_s))(y_t - E(y_t))]$ to be time independent.
- One can estimate the parameters of the time series model $\hat{f}(\cdot)$, and provide out-of-sample forecast assuming distribution of y_t **remains unchanged** over the forecasting horizon (very convenient).
- Suppose our time series model is $y_t = \mu + \phi y_{t-1} + \varepsilon_t$ with $\varepsilon_t \sim N(0, \sigma^2)$. One can estimate $\hat{\mu}$, $\hat{\phi}$ and $\hat{\sigma}$, and compute the mean-wise forecast $E(y_{t+1}|I_t) = \hat{\mu} + \hat{\phi} Y_t$ and quantify uncertainty: $E(y_{t+1}|I_t) \pm 1.96 \cdot \hat{\sigma}$.

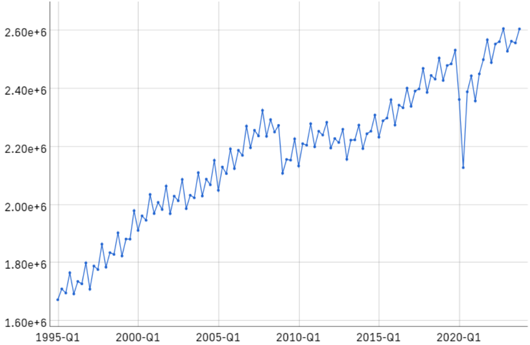
Core assumption in conventional models

But which kind of features should be look for?

- **Stationarity** (weak definition) typically allows the mean function $E\left(y_t\right)$ and auto-covariance function $E\left[\left(y_s - E\left(y_s\right)\right)\left(y_t - E\left(y_t\right)\right)\right]$ to be time independent.
- One can estimate the parameters of the time series model $\hat{f}\left(\cdot\right)$, and provide out-of-sample forecast assuming distribution of y_t **remains unchanged** over the forecasting horizon (very convenient).
- Suppose our time series model is $y_t = \mu + \phi y_{t-1} + \varepsilon_t$ with $\varepsilon_t \sim N(0, \sigma^2)$. One can estimate $\hat{\mu}$, $\hat{\phi}$ and $\hat{\sigma}$, and compute the mean-wise forecast $E\left(y_{t+1}|I_t\right) = \hat{\mu} + \hat{\phi} Y_t$ and quantify uncertainty: $E\left(y_{t+1}|I_t\right) \pm 1.96 \cdot \hat{\sigma}$.

But in practice...

Most of time series are affected by trend and seasonal components that do **not** imply (weak) stationarity. See GDP time series:



Gross domestic product at market prices – Euro area

What we should do?

- So one needs to be careful:

1. **Check for seasonality** (visual screening and description from the data provider), and if necessary, use seasonal filters (most common: X-13ARIMA-SEATS, available at https://www.census.gov/data/software/x13as.html).
2. **Check for stationarity** (through visual screening and stationarity tests), and if necessary, impose stationarity:
   - Use first difference filtering (e.g., express in growth rates)
   - Use a business cycle filter (HP, bandpass, etc)
   - Eventually divide by working age population to get rid off low frequency component

What we should do?

- So one needs to be careful:

1. **Check for seasonality** (visual screening and description from the data provider), and if necessary, use seasonal filters (most common: X-13ARIMA-SEATS, available at https://www.census.gov/data/software/x13as.html).
2. **Check for stationarity** (through visual screening and stationarity tests), and if necessary, impose stationarity:
   - Use first difference filtering (e.g., express in growth rates)
   - Use a business cycle filter (HP, bandpass, etc)
   - Eventually divide by working age population to get rid off low frequency component

What we should do?

- So one needs to be careful:

1. **Check for seasonality** (visual screening and description from the data provider), and if necessary, use seasonal filters (most common: X-13ARIMA-SEATS, available at https://www.census.gov/data/software/x13as.html).
2. **Check for stationarity** (through visual screening and stationarity tests), and if necessary, impose stationarity:
   - Use first difference filtering (e.g., express in growth rates)
   - Use a business cycle filter (HP, bandpass, etc)
   - Eventually divide by working age population to get rid off low frequency component

What we should do?

- So one needs to be careful:

1. **Check for seasonality** (visual screening and description from the data provider), and if necessary, use seasonal filters (most common: X-13ARIMA-SEATS, available at https://www.census.gov/data/software/x13as.html).
2. **Check for stationarity** (through visual screening and stationarity tests), and if necessary, impose stationarity:
   - Use first difference filtering (e.g., express in growth rates)
   - Use a business cycle filter (HP, bandpass, etc)
   - Eventually divide by working age population to get rid off low frequency component

What we should do?

- So one needs to be careful:

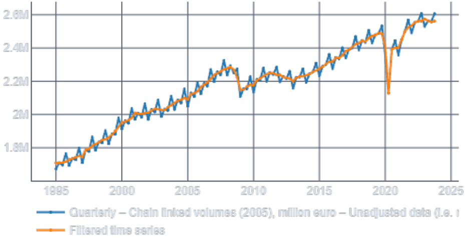
1. **Check for seasonality** (visual screening and description from the data provider), and if necessary, use seasonal filters (most common: X-13ARIMA-SEATS, available at https://www.census.gov/data/software/x13as.html).
2. **Check for stationarity** (through visual screening and stationarity tests), and if necessary, impose stationarity:
   - Use first difference filtering (e.g., express in growth rates)
   - Use a business cycle filter (HP, bandpass, etc)
   - Eventually divide by working age population to get rid off low frequency component

What we should do?

- So one needs to be careful:

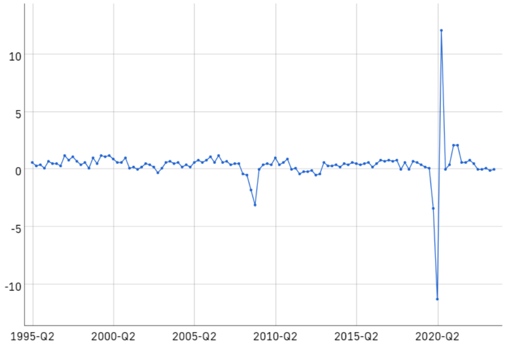
1. **Check for seasonality** (visual screening and description from the data provider), and if necessary, use seasonal filters (most common: X-13ARIMA-SEATS, available at https://www.census.gov/data/software/x13as.html).
2. **Check for stationarity** (through visual screening and stationarity tests), and if necessary, impose stationarity:
   - Use first difference filtering (e.g., express in growth rates)
   - Use a business cycle filter (HP, bandpass, etc)
   - Eventually divide by working age population to get rid off low frequency component

First deseasonalize:



Deseasonalized gross domestic product at market prices – Euro area

Then impose stationarity $\Delta \log Y_t$



Deseasonalized GDP growth at market prices – Euro area

Plan

DBnomics: Real-Time Data for Forecasting

What is dbnomics?

dbnomics is a platform that aggregates economic datasets from various public sources.

Why is it useful for forecasting?

- Provides access to a vast array of real-time economic data.
- Allows economists and analysts to track economic indicators as they are released.
- Enables more accurate and timely forecasting models by incorporating the latest data.

DBnomics: Real-Time Data for Forecasting

Scenario: INSEE French GDP Data Release

- INSEE releases French GDP data every first Monday of the month at 10:00 AM.
- As a forecaster, you only need to rerun your code at 10:01 AM to access the latest data.
- This allows you to quickly generate graphs and draft reports for management.

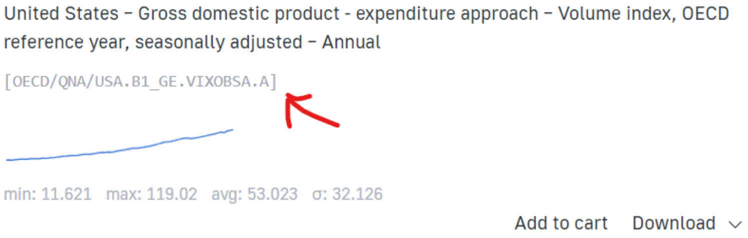
Using dbnomics

- Access the latest French GDP data from INSEE via dbnomics API.
- Integrate the new data into your forecasting model.
- Generate updated graphs and reports for management with minimal delay.

DBnomics: Real-Time Data for Forecasting

How to use dbnomics?

1. Go to db.nomics.world
2. Browse among all the available databases, find your time series, and pick its token (a unique identifier for each series):



DBnomics: Real-Time Data for Forecasting

How to use dbnomics?

3. Download DBnomics MATLAB pluging call_dbnomics.m
4. Query dbnomics on real time (assuming you have internet connection):

```
[output_mat,output_table,dates_nb] =  call_dbnomics(varargin);
```

Input:

`varargin`: string tokens list ('token1','token2',...)

Outputs:

`output_mat`: $T \times (N+1)$ matrix of extracted data, 1st column is date vector in MATLAB format.

`output_table`: same matrix but in table format.

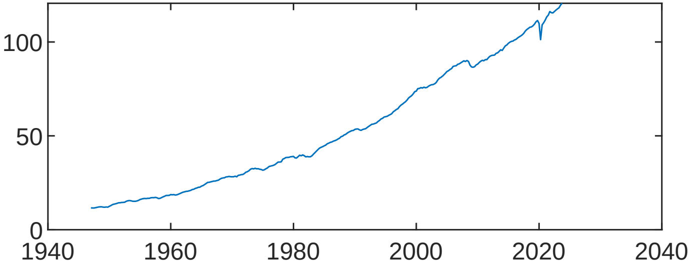
`dates_nb`: vector of dates in numeric values.

DBnomics: Real-Time Data for Forecasting

How to use dbnomics?

Example:

```
[output_mat,output_table,dates_nb] =  call_dbnomics('OECD/QNA/USA.B1_GE.VI
figure;
plot(dates_nb,output_mat(:,2))
```



DBnomics: Real-Time Data for Forecasting

Building my sample for the US

Suppose a New Keynesian model comprising 3 core variables $[y_t, \pi_t, r_t]$. Our gross data from dbnomics comprise a price data $\mathbb{P}_t$, GDP (volume) $\mathbb{Y}_t$ and nominal rate $\mathbb{R}_t$. One can simply impose for the two stationary variables the following transformation:

$$
\begin{aligned}
G_{\mathbb{Y}t} &= \log(\mathbb{Y}_t) - \log(\mathbb{Y}_{t--1}) \\
G_{\mathbb{P}t} &= \log(\mathbb{P}_t) - \log(\mathbb{P}_{t--1})
\end{aligned}
$$

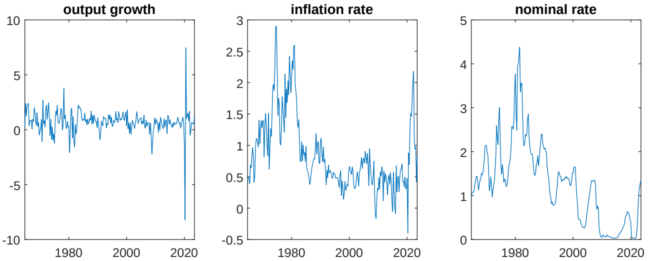
One can next map the model and the data as follows:

$$
\begin{aligned}
y_t - y_{t-1} &= G_{\mathbb{Y}t} \\
\pi_t &= G_{\mathbb{P}t} \\
\underbrace{r_t}_{\text{model}} &= \underbrace{\mathbb{R}_t}_{\text{data}}
\end{aligned}
$$

DBnomics: Real-Time Data for Forecasting

Building my sample for the US

Run `my_db_US.m` to get:



A quick recap for your report:

- We will employ linearized and stationary macro models, which implies to have a sample distribution consistent with our model.
- Check first that your data do not exhibit seasonal component $\rightarrow$ use deseasonalized series or applies methods to remove seasonal component (e.g. X13 method).
- Check your sample is stationary (visual screening to assess the data have no upward trend as for GDP). If so, employ first order difference.
- If stationary and no seasonal features, your sample is ready for applied work.

gauthier@vermandel.fr

Nelson, C. R. (1972). The prediction performance of the frb-mit-penn model of the us economy. *The American Economic Review*, 62(5):902–917.