Short-term forecasting with dynamic factor models

Rytis Bagdžiūnas, Aspirant FNRS
CORE, Université catholique de Louvain, Belgium

Initial problem

Predict economic growth (**GDP**), **inflation** and other summary variables in the short-term. Construct an index representing current state of the economy. Some challenges to take into account:

- integrate large number of time series
- asynchronous data arrival
- deal with NAs
- mixed-frequency nature of data
- possible regime changes, e.g. boom and bust cycles
- identification and economic constraints

Getting data with `rsdmx`

Statistical data and metadata exchange (SDMX) standard has been sponsored by ECB, Eurostat, IMF, OECD, etc.

`rsdmx` package aims to fully implement SDMX-ML standard in R:

- `readSDMX` downloads and parses XML data and metadata
- `as.data.frame` converts `SDMXData` object to a data.frame

`rsdmx` is still under development but already very useful. In the future, it seeks to implement `writeSDMX`, automate timestamp treatment and assist in query construction.

Example: 767 monthly business and consumer surveys for OECD and partner countries are downloaded in seconds.

Sample plot

```
library(rsdmx)
url <- "http://stats.oecd.org/restsdmx/sdmx.ashx/GetData/"
key <- "MEI_BTS_COS/..BLSA.M/all?startTime=2014-Q4"
xml_data <- readSDMX(paste0(url,key))
data <- as.data.frame(xml_data)
```

Our objectives

⇒ Package for R that facilitates the estimation part: **dynfactoR**

Why? Current "libraries" for dynamic factor model estimation are scattered and mostly Matlab-only. State-space modelling is already available for R but exploiting dynamic factor model structure will lead to gains in efficiency.

The following snippet estimates a dynamic factor model with a single regime where 2-dimensional factor $\mathbf{X}_t$ is assumed to follow VAR(2). This model is statistically identified since $\mathbf{Q} = \mathrm{Id}$ and $\mathbf{F}$ is partially upper triangular. See Bai & Wang (2012).

```
library(dynfactoR)
DFMfit <- dfm(data, f=2, p=2, rF='upper', rQ='identity')
summary(DFMfit, plot=TRUE)
```

Result

- Restrictions on $\mathbf{F}$ can be specified to deal with mixed-frequency, identification and other constraints.
- Kalman filtering supports missing data and asynchronous updates

References

1. Stock & Watson (2002). *Forecasting using principal components from a large number of predictors.*
2. Doz, Gianone & Reichlin (2006). *A quasi-maximum likelihood approach for large approximate dynamic factor models.*
3. Bai & Wang (2012). *Identification and estimation of dynamic factor models.*
4. Banbura & Modugno (2014). *Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data.*

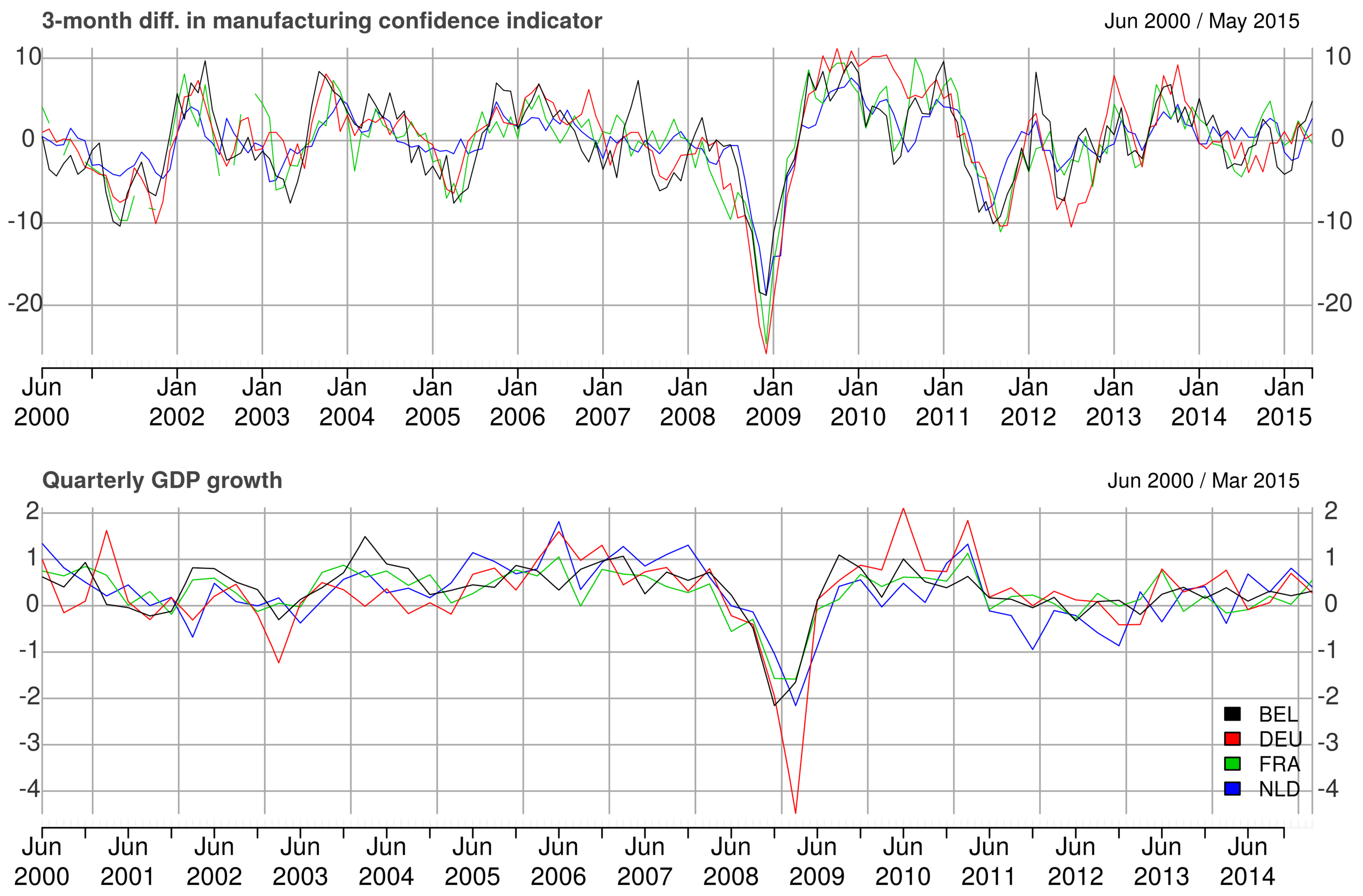

Figure: **Top**: monthly manufacturing survey data in Belgium, Germany, France and the Netherlands. **Bottom**: quarterly GDP growth. It looks as if there was a common underlying pattern.

Source: OECD

Modelling solution: dynamic factor models

Let $\mathbf{Y}_t$ be a high-dimensional dataset and $\mathbf{X}_t$ an unobservable latent low-dimensional process such that

$$\mathbf{Y}_t = \mathbf{F}_{s_t}\mathbf{X}_t + \varepsilon_t$$
$$\mathbf{X}_t = \mathbf{A}_{1,s_t}\mathbf{X}_{t-1} + \cdots + \mathbf{A}_{p,s_t}\mathbf{X}_{t-p} + \mathbf{u}_t$$

with $\varepsilon_t \sim \mathcal{N}(0, \mathbf{R})$ and $\mathbf{u}_t \sim \mathcal{N}(0, \mathbf{Q})$. $(\mathbf{F}_{s_t}, \mathbf{A}_{1:p,s_t})$ may also switch regime $s_t \in \{0, 1\}$ with s_t following a Markov-switching process.

This model is **flexible** enough to deal with most issues stated initially.

Useful to know beforehand

- **Pre-processing**
  Normalize and deseasonalize $\mathbf{Y}_t$, e.g. `x12` package in R can do batch processing with `X13-ARIMA-SEATS` methodology
- **Identification**
  If statistical identification is an issue, impose constraints on $(\mathbf{F}, \mathbf{R}, \mathbf{Q})$.
- **Estimation methods**
  Based on Kalman filter and Kim filter (still experimental!) for Markov-switching case
  - Maximizing likelihood with non-linear optimization procedures
  - EM-algorithm

First dynamic factor and trend for 2 months-ahead

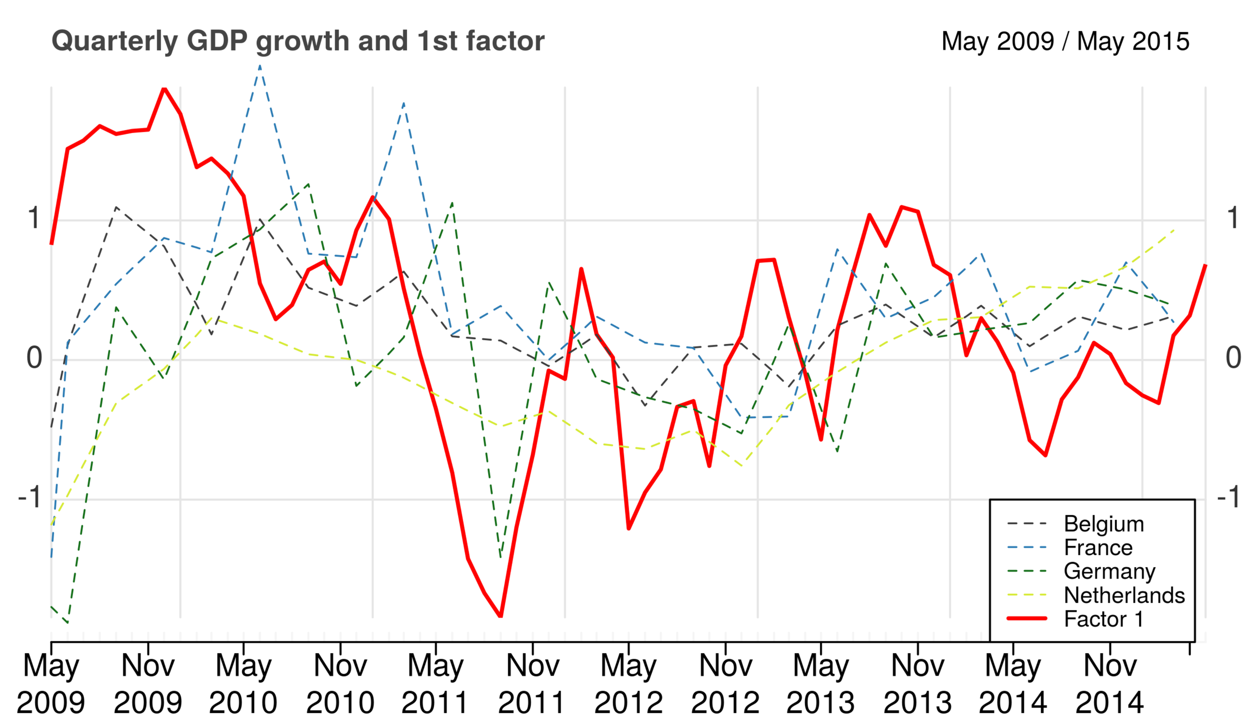

Figure: Quarterly GDP growth in Belgium, France, Germany and the Netherlands. First factor often precedes trend variations in GDP data.

- Source for this poster and other projects: rbagd@github
- Tutorial on using SDMX from R: rbagd.eu
- `rsdmx` is available on CRAN: its developer Emmanuel Blondel would appreciate any programming and financial assistance!

UCL Université catholique de Louvain
fnrs