Modelling with `dynfactoR`

Rytis Bagdžiūnas
CORE, Université catholique de Louvain

June 26, 2015

`dynfactoR` package is aimed at researchers dealing with economic modelling. In particular, it is useful at extracting underlying trends from a potentially large number of economic time series.

Dynamic factor model is a special case of a state space equation. In its general form it can be written as

$$\begin{aligned} \mathbf{X}_t &= \mathbf{C}\mathbf{f}_t + \boldsymbol{\varepsilon}_t, \quad & \boldsymbol{\varepsilon}_t &\sim \mathcal{N}(\mathbf{0}, \mathbf{R}) \\ \mathbf{f}_t &= \mathbf{A}\mathbf{f}_{t-1} + \mathbf{u}_t, \quad & u_t &\sim \mathcal{N}(\mathbf{0}, \mathbf{Q}) \end{aligned} \tag{1}$$

where $\mathbf{X}_t$ is a vector of observable data which might contain missing data. It is assumed that observable data is linearly driven by a low-dimensional unobservable process $\mathbf{f}_t$.

Use case

A typical use case for such a model is nowcasting gross domestic product (GDP). GDP data is published every 3 months, so that economic growth from January to March would only be known in early April. However, many business and consumer surveys are published monthly and are often correlated with GDP growth. Therefore, predictions of economic growth for first quarter should be different depending on whether they are done in December or in February (hence the term *nowcasting*). Dynamic factor model is one way to do that by extracting an underlying trend which often follows economic growth pattern. Besides, if restrictions are imposed on $\mathbf{C}$, economic growth can be modelled explicitly by factors.

1 State space equations

It is assumed that $\mathbf{A}(L)\mathbf{f}_t = \mathbf{u}_t$ where $\mathbf{u}_t \sim N(\mathbf{0}, \mathrm{I}_r)$ and $\mathbf{A}(L)$ is a lag polynomial with roots inside the unit circle, that is $\mathbf{f}_t$ is a covariance-stationary process.

Let $\mathbf{f}_t$ follow a p-th order autoregressive process, so that

$$\mathbf{f}_t = \mathbf{A}_1 \mathbf{f}_{t-1} + \cdots + \mathbf{A}_p \mathbf{f}_{t-p} + \mathbf{u}_t \tag{2}$$

Suppose that there are r-factors, that is $\mathbf{f}_t$ and $\mathbf{u}_t$ are r-variate processes. If $\mathbf{f}_t$ were observed, it would be easy to estimate $\{\mathbf{A}_1, \ldots, \mathbf{A}_p\}$ with a simple OLS procedure in the context of VAR(p) model. This estimator would also be a maximum likelihood estimator.

Let $\mathbf{f}_t = (f_{1,t}, \ldots, f_{r,t})$ and $\mathbf{u}_t = (u_{1,t}, \ldots, u_{r,t})$. Then (2) can be rewritten in state space equation form of (1) as

$$\begin{pmatrix} f_{1,t} \\ f_{2,t} \\ \cdots \\ f_{r,t} \\ f_{1,t-1} \\ \cdots \\ f_{r,t-p+1} \end{pmatrix} = \begin{bmatrix} \mathbf{A}_1 & \cdots & \cdots & \mathbf{A}_p \\ \mathrm{I}_r & 0 & 0 & \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & 0 & \mathrm{I}_r & 0 \end{bmatrix} \begin{pmatrix} f_{1,t-1} \\ f_{2,t-1} \\ \cdots \\ f_{r,t-1} \\ f_{1,t-2} \\ \cdots \\ f_{r,t-p} \end{pmatrix} + \begin{pmatrix} u_{1,t} \\ \cdots \\ u_{r,t} \\ 0 \\ \cdots \\ 0 \end{pmatrix} \tag{3}$$

Evidently, state space covariance matrix has to be extended by zeroes. With this in mind, measurement equation then becomes

$$\mathbf{X}_t = \begin{bmatrix} \mathbf{C} & \mathbf{0}_{n\times r} & \ldots & \mathbf{0}_{n\times r} \end{bmatrix} \begin{pmatrix} f_{1,t} \\ f_{2,t} \\ \cdots \\ f_{r,t} \\ f_{1,t-1} \\ \cdots \\ f_{r,t-p+1} \end{pmatrix} + \boldsymbol{\varepsilon}_t \tag{4}$$

where $\mathbf{0}_{n\times r}$ is zero-matrix with n rows and r columns.

2 Estimators

- PCA estimator where factors are estimated as principal components of observable data. Its properties are well discussed in Stock & Watson (2002). If missing data is present, other estimators should be used.
- Two-step estimator by Doz, Gianone and Reichlin (2011). This methods uses PCA estimates and runs them through Kalman filter.
- QML estimator by Doz, Gianone and Reichlin (2012). Similar to two-step estimator, however Kalman filtering procedure is iterated until EM convergence.

3 Forecasting

Once a dynamic factor model is estimated, $\left(\mathbf{X}_{1:T}, \hat{\mathbf{A}}, \hat{\mathbf{C}}\right)$ are available. Predicting $(\mathbf{X}_{f,T+h}, \mathbf{f}_{T+h})$ for some horizon $h \in \{1, \dots\}$ consists in three steps:

1. predict $\mathbf{f}_{f,T+1}$ given estimate of $\mathbf{f}_T$ and VAR(p) process, i.e.

$$\mathbf{f}_{f,T+1} = \hat{\mathbf{A}}_1 \mathbf{f}_T + \cdots + \hat{\mathbf{A}}_p \mathbf{f}_{T-p}$$

2. predict $\mathbf{X}_{f,T+1}$ based on factor loading and factor estimates, i.e.

$$\mathbf{X}_{f,T+1} = \hat{\mathbf{C}} \mathbf{f}_{f,T+1}$$

3. repeat steps 1 and 2 until desired horizon level is reached

4 Future and ongoing work

Some of the ideas for dynamic factor modelling are not yet implemented completely or even at all. This section gathers future ideas for the package.

1. Introduce Markov-switching possibility so that, for example, $\mathbf{C}$ or $\mathbf{A}$ could be state dependent. It would be in particular interesting if this could capture business cycle growth variations. In case of this model, approximation to Kalman filter can be used, e.g. Kim filter, in order to avoid exponentially growing number of calculations.

2. Analyze and propose some defaults for restrictions on $\mathbf{C}$. Typically, this includes identification constraints and currently, only upper triangular $\mathbf{C}$ is implemented. Constraints can also be specified based on economic reasoning and render model hierarchical.

3. Kalman and Kim filters are slow in `R` due to their recursive nature. They should be reimplemented in `C` or call another package such as `FKF`. Kim filter will probably need to be rewritten in `C` anyway.

4. As presented, dynamic factor model is only dynamic in the state equation. It can be generalized to have dynamics in the measurement equation as well, i.e. $\mathbf{X}_t$ depending on current and past values of $\mathbf{f}_t$. This should not be difficult to implement as such model would be eventually reduced to (1).

5. Currently, there is no automated testing for the package. This should be fixed.