Generalitat de Catalunya
Institut d'Estadística de Catalunya

# Dynamic Factor Models for GDP nowcasting: An application for the Catalan economy

**Mònica Gasulla, Jonatan Jorba, Jesús Muñoz, Sergi Plaza**

Barcelona, 21st October 2016

# CONTENTS

# INTRODUCTION

- ✓ The **Gross Domestic Product (GDP)** is the most frequently used synthetic indicator for evaluating the economic evolution in a territory.

- ✓ **Eurostat estimates** the GDP flash for the EU aggregates in t+45 days time.

- ✓ **Eurostat's declared aim** is to cut the availability period of the GDP flash for the EU aggregates to t+30 days. This objective allows **the homogeneization with the equivalent USA estimations**.

- ✓ With this aim, Eurostat has encouraged diverse **methodological works**, all of them confirming the existence of **a great diversity of approaches** among EU member states.

# INTRODUCTION

**The approaches followed by the EU member states** can be classified in two main groups:

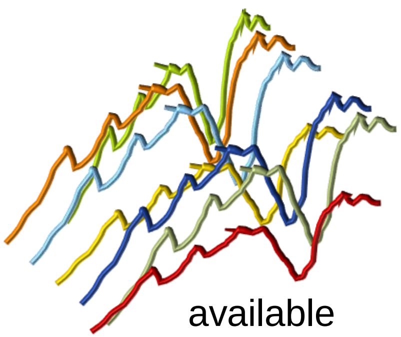
Direct Methods:
- Autorregresive Models (ADL & VAR)
- Dynamic Factor Models (DFM)
- Bridge Models

Indirect Methods:
- Denton Model
- Chow Lin Models
- Multivariate Models

Taking into account the previous experience of the diverse Federal Reserve Banks and the AIReF in Spain, **Idescat** has opted for developing the **DFM methodology** given the robustness, coherence and predictive power of the published works.

# INTRODUCTION

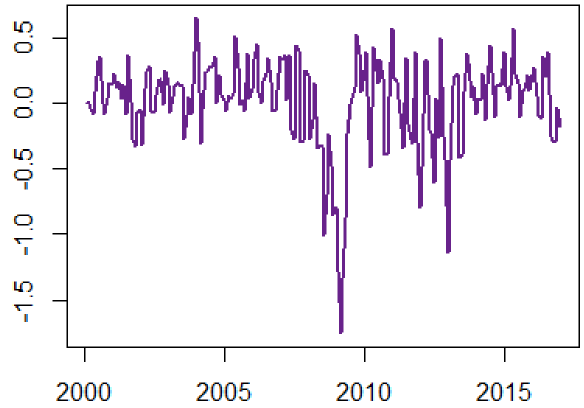
## DFM APPROACH



available
macroeconomic
indicators
(current quarter)

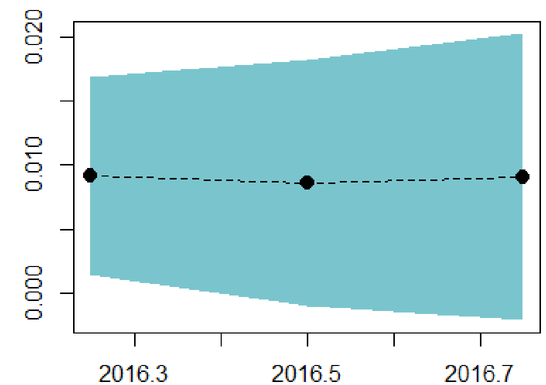
DFM

**common factor**

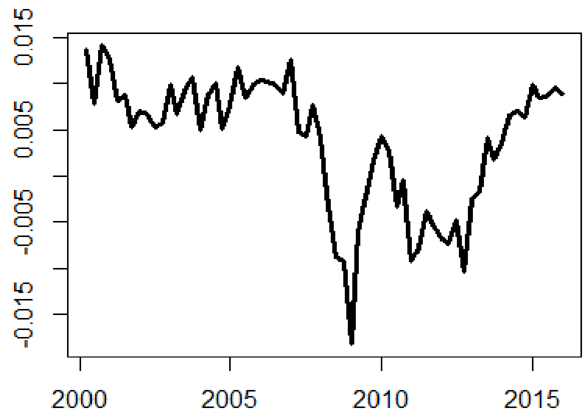



TRANSFER
FUNCTION

**GDPgr: nowcast and forecasts**




**GDP (grothw rates)**




GDP
(up to previous quarter, if possible)

# DYNAMIC FACTOR MODELS (DFM)

Given a period t, t=1,…,T, the observation of each stationary indicator $Z_i$ (t) i=1,..,n is descomposed as a lineal combination of a set of common factors ($f_1$(t),…, $f_r$(t)) following a VAR(p) structure and an idiosyncratic white noise or AR(1) term $e_{it}$

**INPUT** $Z_1$ $Z_2$ … $Z_i$ … $Z_n$

$$Z_i(t) = \lambda_i f(t) + e_i(t)$$

**UNOBSERVED**

$f_1$ … $f_r$ **VAR(p)**

$$(I_r - \psi(L))f(t) = \eta(t)$$

$e_1$ **N(0,$\sigma_1$) [AR(1)]**

$e_2$ **N(0,$\sigma_2$) [AR(1)]**

…

$e_i$ **N(0,$\sigma_i$) [AR(1)]**

…

$e_n$ **N(0,$\sigma_n$) [AR(1)]**

$$e_i(t) \sim N\,(0,\sigma_i)$$

o

$$e_i(t) \sim AR(1)$$

$$(1-\varphi_i(L))e_{it} = \eta_t$$

Static version of the model: $Z_{it} = \lambda_i f_t + e_{it}$ i=1,…,n t=1,…,T

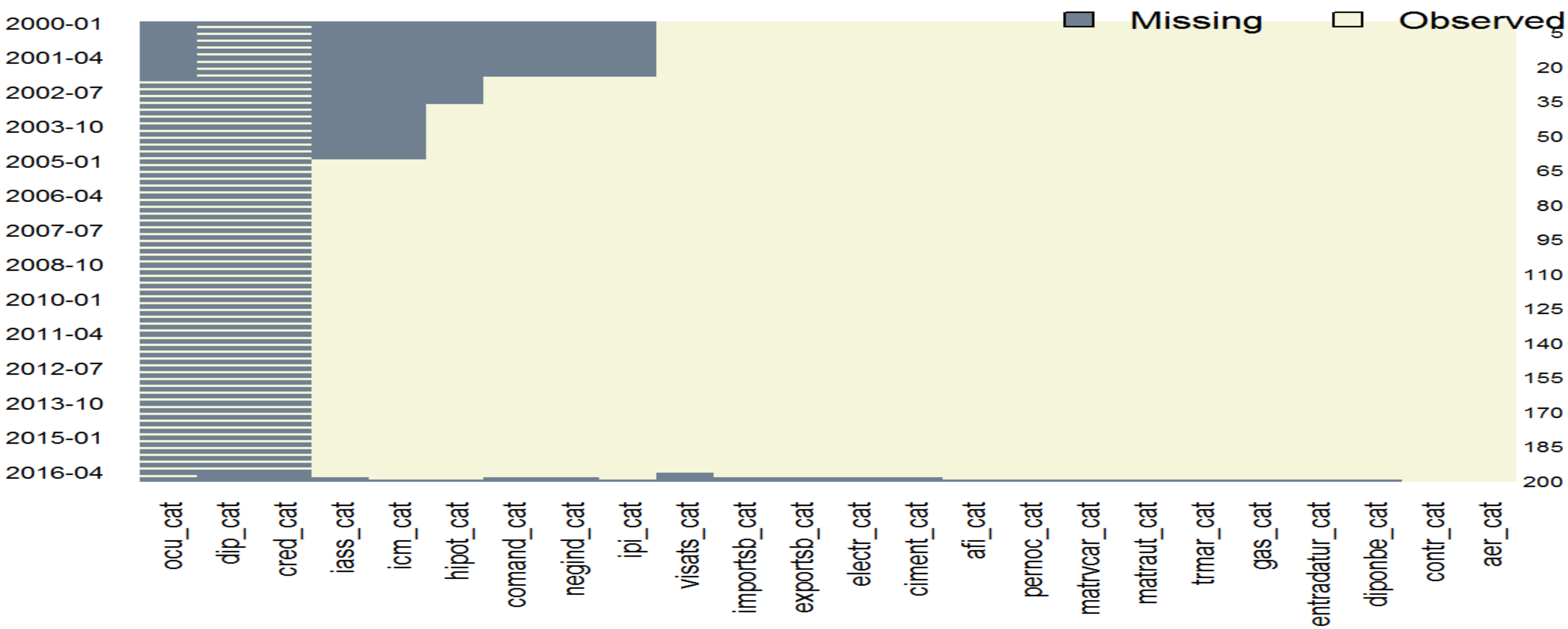
# DYNAMIC FACTOR MODELS (DFM)

Drawbacks:

- Starting point dates are different →
- Release calendars are diferent (*ragged edge*) →

**ITERATIVE ALGORITHMS (EM)**
**IMPUTATION METHODS**
**KALMAN FILTER**

- Mixed frequencies → **TEMPORAL DISAGGREGATION**



# DYNAMIC FACTOR MODELS (DFM)

**EM ALGORITHM: Factor and parameter estimation with missing data**

| Fed ESTIMATION | AIReF ESTIMATION |
|---|---|
| Imputed dataset | Trimmed dataset |
| | Static estimation of factor/s |
| | Original dataset imputation by factor-indicator regressions |

**Static estimation of factor/s**

| | |
|---|---|
| $\Lambda$ estimation by regression over the original dataset | $\Lambda$ estimation by regression over the imputed dataset |

**$\psi$ (and $\varphi_i$) estimation by autoregressions**

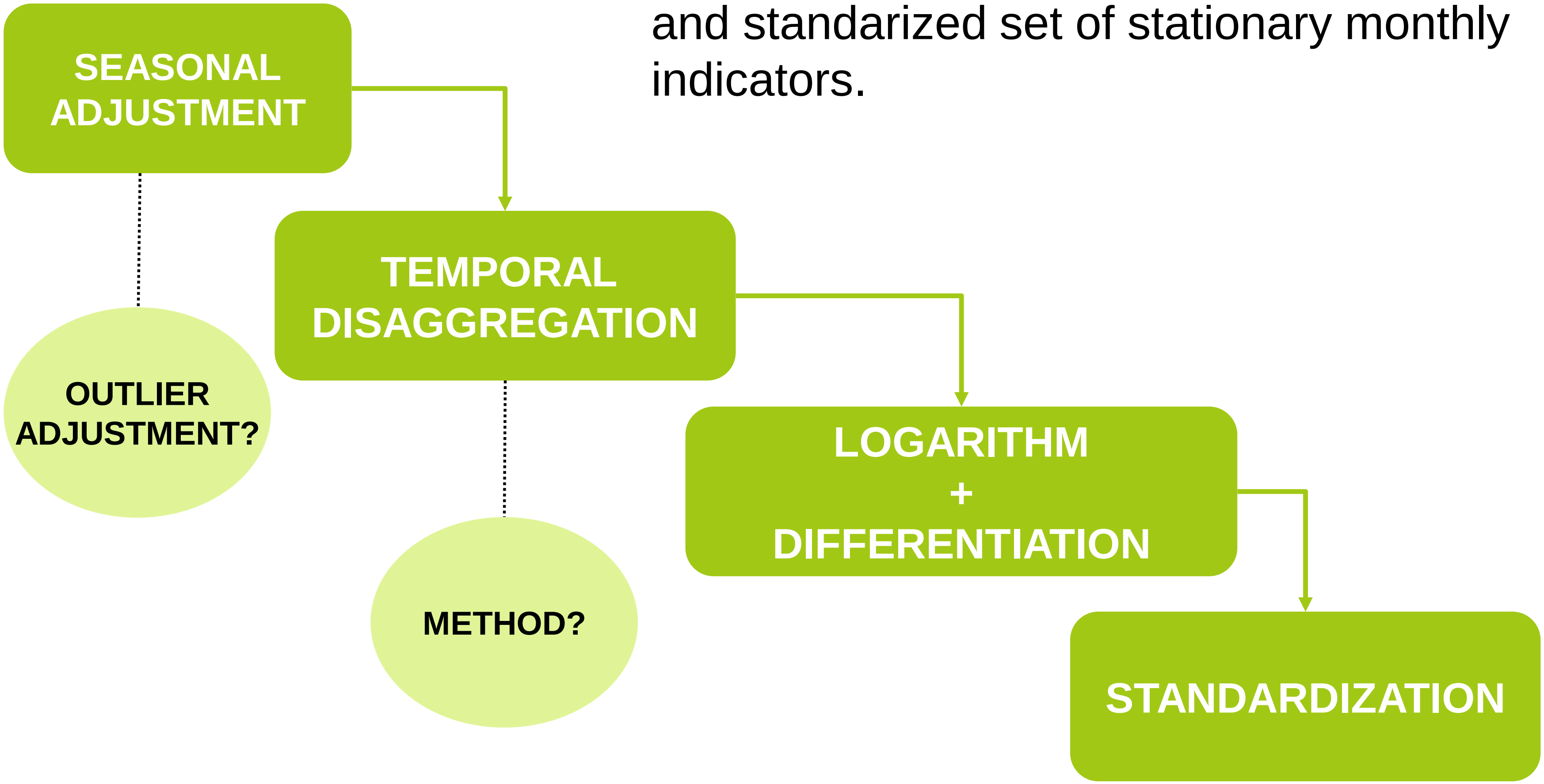
**Kalman filter:**

- **Factor estimation and projection**
- **Likelihood computation and evaluation**

***ITERATE UNTIL LIKELIHOOD CONVERGENCE***

# DATA PROCESSING

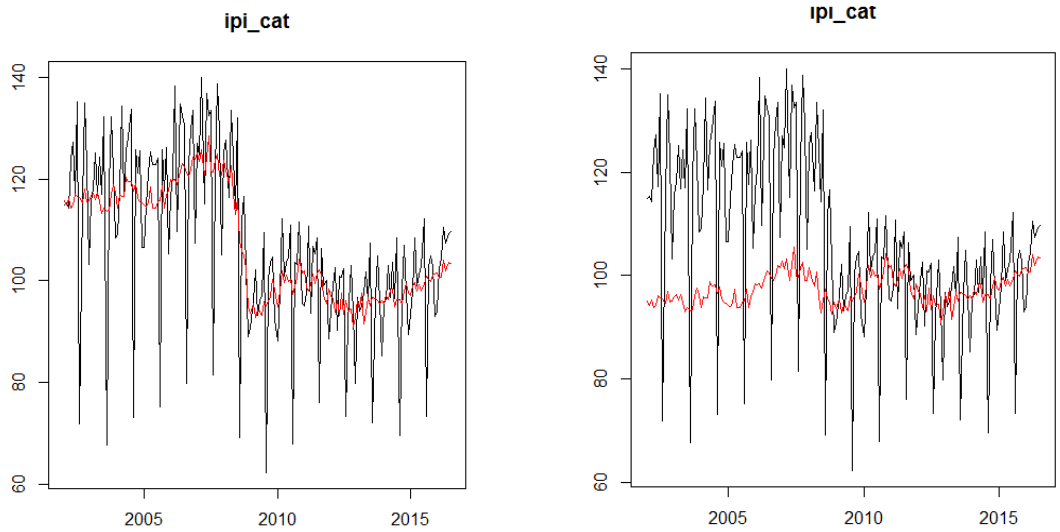
Transformations to obtain a non-seasonal and standarized set of stationary monthly indicators.



# DATA PROCESSING

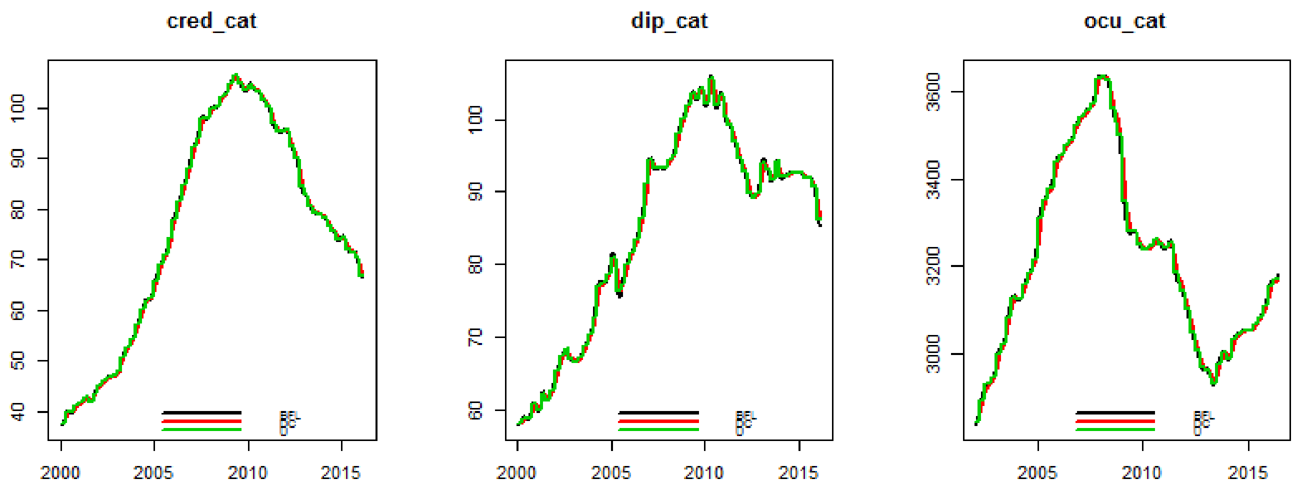
**SEASONAL ADJUSTMENT**

Filtering outliers
Non-filtering




**TEMPORAL DISAGGREGATION**

Boot, Feibes & Lisman
Denton-Cholette
Uniform



# PRELIMINARY ANALYSIS

**Number of factors**
- Visual criteria
- Numerical criteria

**MISSING DATA IMPUTATION**
- 0, MEANS
- ARIMA, LINEAR REG
- REGULARIZED PCA

**STARTING DATASET**
- ORIGINAL
- LONGITUDINAL
- CROSS-SECTIONAL
- OPTIMIZED (AIReF)

**STATIC FACTOR MODELS**
- PCA-OLS
- PCA-WLS / AIReF
- PCA-GLS

**VAR (p): p DETERMINATION**
- Correlogram inspection
- BIC, AICC,..

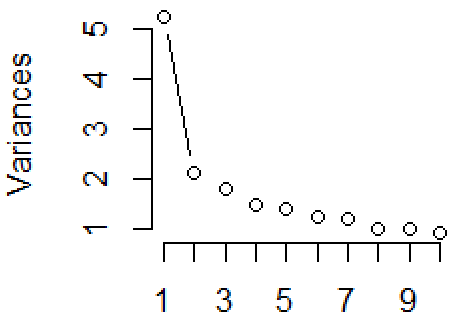
# PRELIMINARY ANALYSIS

## NUMBER OF FACTORS

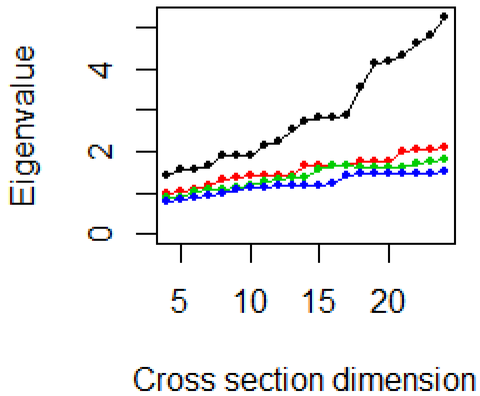
**GRAPHICAL CRITERIA**
SCREEPLOT
CROSS-SECTION SUCCESSIONS

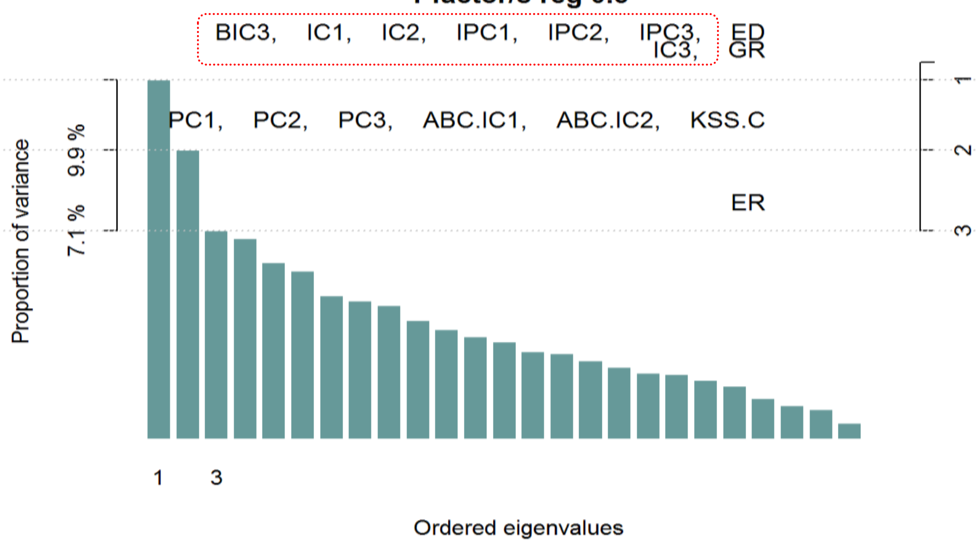
**QUANTITATIVE CRITERIA**
BIC3,
IC1, IC2, IC3,
IPC1, IPC2 e IPC3
Bai & Ng Criteria

SCREE PLOT - ALL




FIRST 4 EIGENVALUES



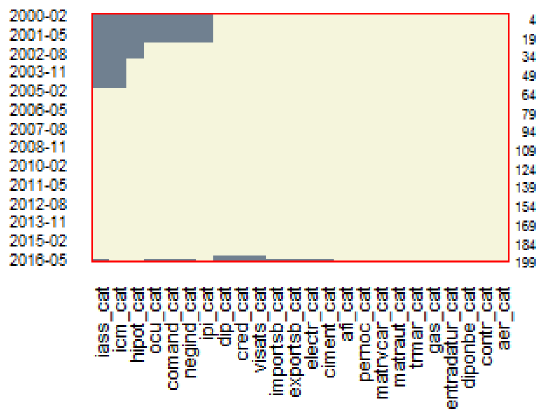



# PRELIMINARY ANALYSIS

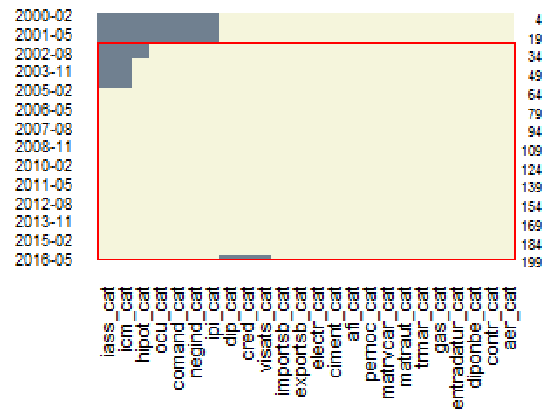
## STARTING DATASET

ORIGINAL
CROSS-SECTION
LONGITUDINAL
OPTIMIZED(AIReF)

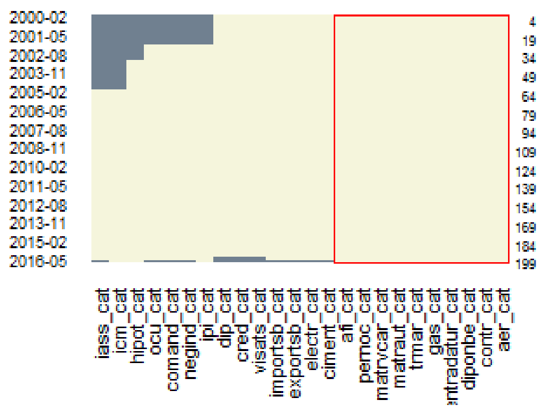
### CAT 24



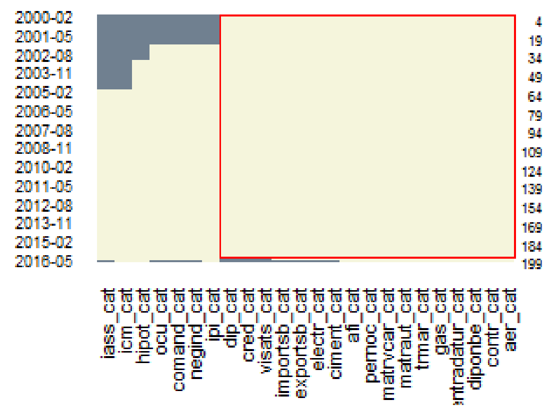

### CROSS-SECTION 80 %




### LONG 99 %




### CAT 24 AIReF



# PRELIMINARY ANALYSIS

## IMPUTATION

**PREDICTIVE MEAN MATCHING**

$\mathbf{X_j}$ incomplete, **Y** complete

$\hat{X}_i(t) = a + b \cdot Y(t)$ [ a, b computed with OLS ]

**Y** can be an indicator or a common factor.

**ITERATIVE PCA IMPUTATION (Stock & Watson, Josse & Husson)**

1. Get a balanced dataset through simple imputation or longitudinal subsetting.
2. Compute a set of factors with PCA.
3. Impute missing values using predictive mean matching with the computed factors.
4. Iterate until convergence.

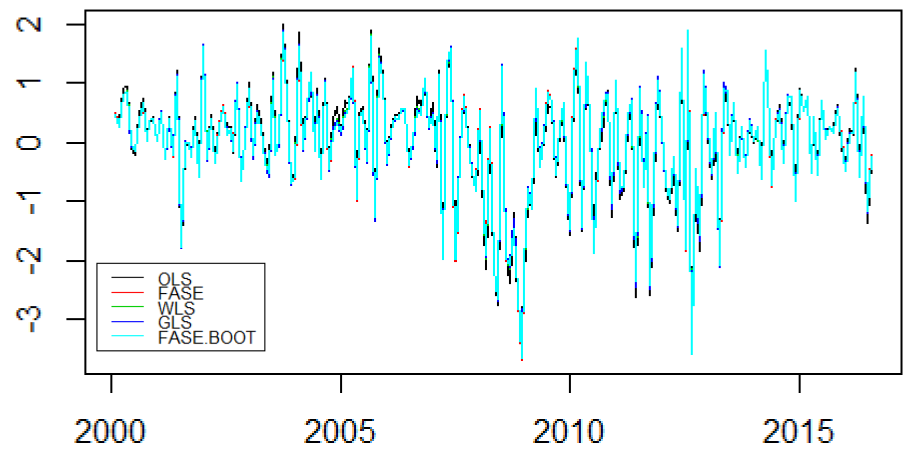
After convergence, factor/s can be directly introduced in the transfer function?

J&H: Regularized version of this EM algorithm

# PRELIMINARY ANALYSIS

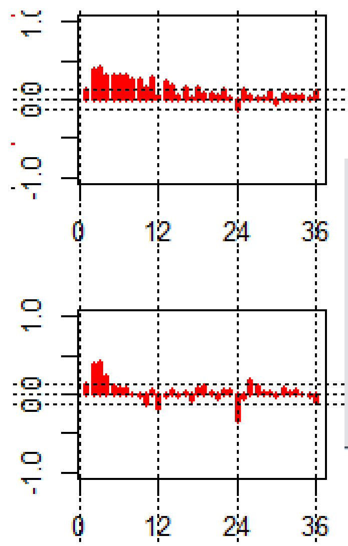
## STATIC FACTOR MODELS

PCA-OLS
PCA-WLS
PCA-GLS




## p DETERMINATION

VISUAL INSPECTION
BIC



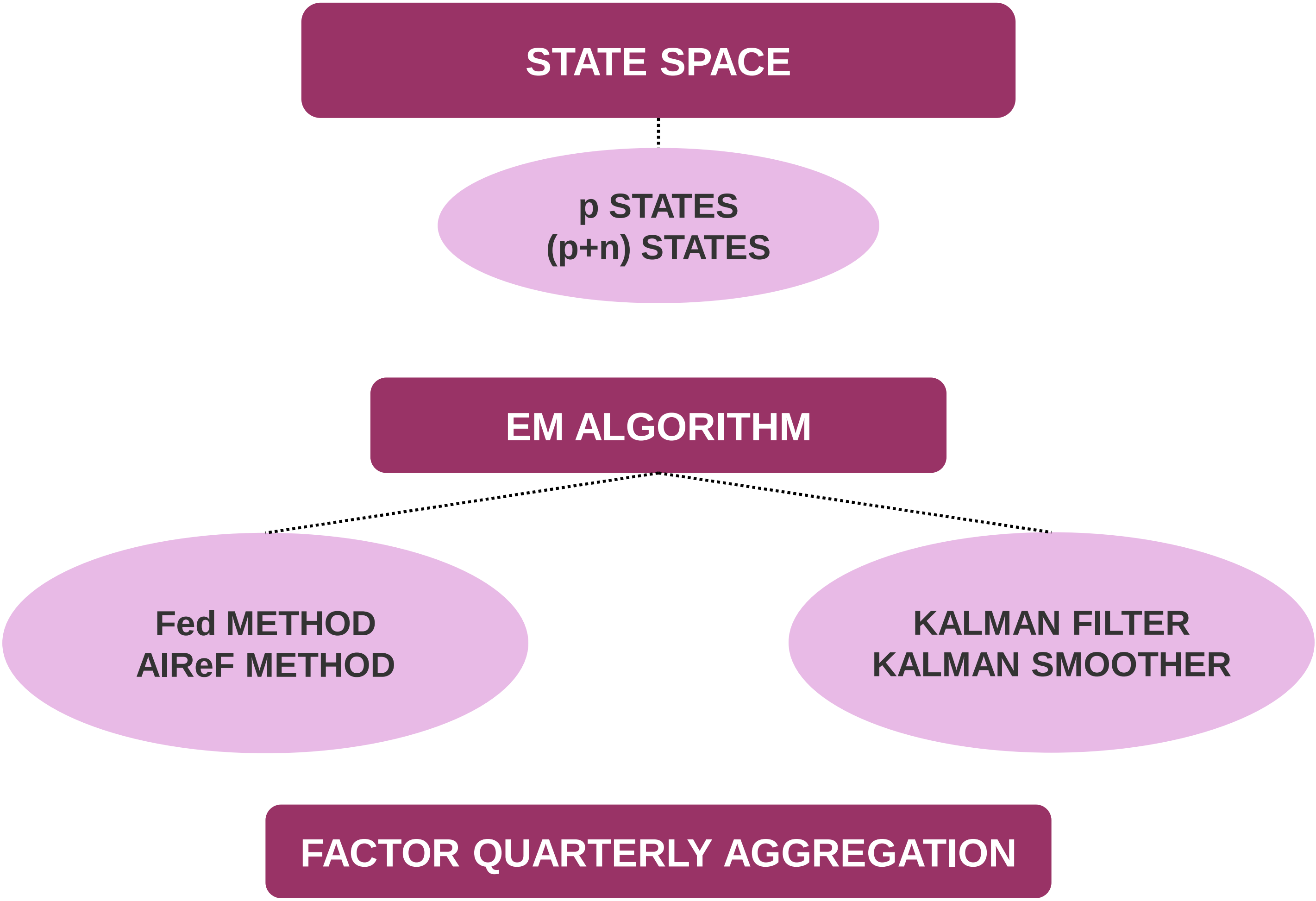

```
$WLS

Call:
ar.ols(x = scale(f.object$factor), demean = F)

Coefficients:
      1        2        3        4
-0.1889   0.2717   0.4398   0.2220

Order selected 4  sigma^2 estimated as  0.6607
```

# DFM ESTIMATION



# DFM ESTIMATION

## DYNAMIC FACTOR MODEL

$Z_{it} = \lambda_i f_t + e_{it}$ $\quad$ $i=1,\dots,n$

$(I_r - \psi(L))f_t = \eta_t$ $\quad$ $t=1,\dots,T$

$(1-\varphi_i L)\, e_{it} = u_t$

## STATE SPACE MODEL

$Z_{it} = HX_t + W_t$ $\quad$ **Measurement**

$X_t = GX_{t-1} + V_t$ $\quad$ **Transition**

$W_t \sim N(0,R)$ $\quad$ $V_t \sim N(0,Q)$

### IDIOSYNCRATIC FACTORS $N(0,\sigma_i^2)$
### Number of states = p

$$X_t = \begin{pmatrix} f_t \\ f_{t-1} \\ \vdots \\ f_{t-p+1} \end{pmatrix} \qquad H = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ \lambda_n & 0 & \dots & 0 \end{pmatrix}$$

$$R = diag(\sigma_i^2)_{i=1,\dots,n}$$

$$G = \begin{pmatrix} \Psi_1 & \dots & \Psi_{p-1} & \Psi_p \\ 1 & & 0 & 0 \\ & \ddots & & \vdots \\ 0 & & 1 & 0 \end{pmatrix} \qquad Q = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & & \\ \vdots & & \ddots & \\ 0 & & & 0 \end{pmatrix}$$

### IDIOSYNCRATIC FACTORS AR(1)
### Number of states = p+n

$$X_t = \begin{pmatrix} f_t \\ f_{t-1} \\ \vdots \\ f_{t-p+1} \\ e_{1t} \\ \vdots \\ e_{nt} \end{pmatrix} \qquad H = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 & 1 & & 0 \\ \vdots & \vdots & & \vdots & & \ddots & \\ \lambda_n & 0 & \dots & 0 & 0 & & 1 \end{pmatrix}$$

$$R = (0) \qquad G = \begin{pmatrix} \Psi_1 & \dots & \Psi_{p-1} & \Psi_p & 0 & \dots & 0 \\ 1 & & 0 & 0 & 0 & & 0 \\ & \ddots & & \vdots & \vdots & & \vdots \\ 0 & & 1 & 0 & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & \varphi_1 & & 0 \\ \vdots & \ddots & & \vdots & & \ddots & \\ 0 & \dots & & 0 & 0 & & \varphi_n \end{pmatrix}$$

# DFM ESTIMATION

## KALMAN FILTER

**Recursive estimation:**

$E[X_t \mid X_{t-1}, X_{t-2}, \ldots, X_1]$

$Var(X_t \mid X_{t-1}, X_{t-2}, \ldots, X_1)$

**Computational complexity increases with size of G**

**Aspects of initialisation**

## KALMAN SMOOTHER

**Two-step estimation:**

1. **Kalman filter.**
2. **Backward estimation:**

$E[X_t \mid X_T, X_{t-2}, \ldots, X_1]$

$Var(X_t \mid X_T, X_{t-2}, \ldots, X_1)$

**Inverse of $Var(X_t \mid X_{t-1}, X_{t-2}, \ldots, X_1)$ → often ill-conditioned**

# NOWCASTING

**PIB in DFM**

$PIB^*_t = \lambda_{PIB^*} f_t + e_{PIB^*t}$ PIB* = transformed PIB

Simple relation, but undoing the transformation can be hard!

**PIB out of DFM**

## ADL MODEL

PIB depends on its own history and factors' histories

$$PIB_t = a + b_1 f_t + b_2 f_{t-1} + b_3 PIB_{t-1} + u_t$$

## TRANSFER FUNCTION

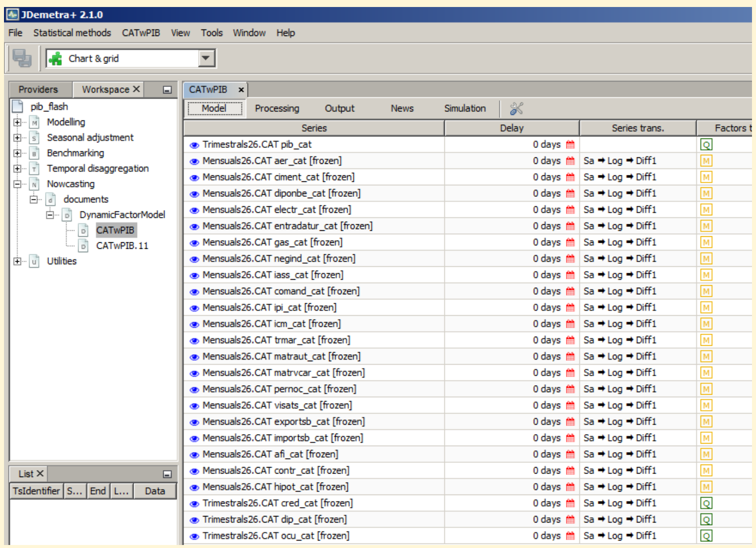
PIB depends on its lags and shocks + factors' lags and shocks

$$PIB_t = c + \frac{\omega_s(L)L^d}{1-\delta_r(L)} f_t + \frac{\theta_q(L)}{\varphi_p(L)} u_t$$

# TOOLS

## WITH GRAPHICAL INTERFACE

JDemetra+ for Fed methodology with idiosyncratic factors N(0,σ)




## WITHOUT GRAPHICAL INTERFACE

R for all the models.

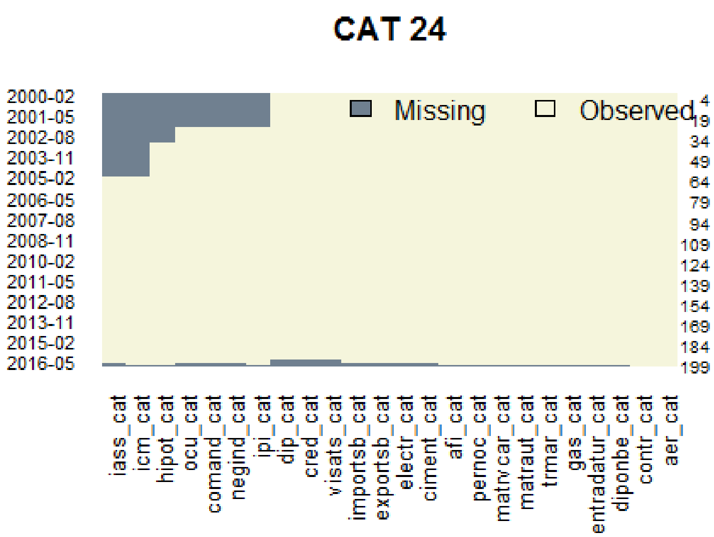
Some issues are hard to develop (time-dependent state space matrices, for example).

Packages: JDemetra+/R interface, forecast, missMDA, phtt, MTS, tempdisagg, signal, KFAS, nlme, Amelia, …

Shiny allows to develop GUIs.

# RESULTS

- 24 INDICATORS
- 1 FACTOR, p = 4
- JANUARY 2000–AUGUST 2016
- 200 MONTHLY VALUES

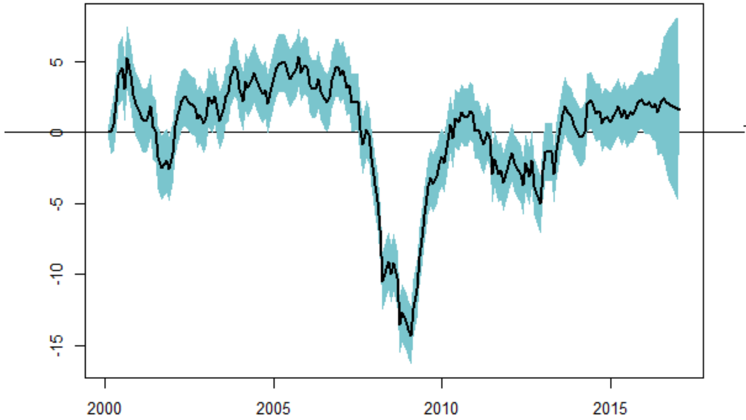



| VARIABLE | YEAR.I | MONTH.I | FREQUENCY | TYPE | STOCK | ON | M1 | M2 | M3 |
|---|---|---|---|---|---|---|---|---|---|
| aer_cat | 2000 | 1 | 12 | 0 | 0 | 1 | -18 | 13 | 43 |
| ciment_cat | 2000 | 1 | 12 | 0 | 0 | 1 | -14 | 17 | 48 |
| diponbe_cat | 2000 | 1 | 12 | 0 | 0 | 1 | -15 | 16 | 46 |
| electr_cat | 2000 | 1 | 12 | 0 | 0 | 1 | -19 | 14 | 46 |
| entradatur_cat | 2000 | 1 | 12 | 0 | 0 | 1 | -5 | 27 | 57 |
| gas_cat | 2000 | 1 | 12 | 0 | 0 | 1 | -21 | 10 | 40 |
| negind_cat | 2002 | 1 | 12 | 0 | 0 | 1 | -14 | 15 | 48 |
| iass_cat | 2005 | 1 | 12 | 0 | 0 | 1 | -14 | 15 | 48 |
| comand_cat | 2002 | 1 | 12 | 0 | 0 | 1 | -14 | 15 | 48 |
| ipi_cat | 2002 | 1 | 12 | 0 | 0 | 1 | -26 | 3 | 34 |
| icm_cat | 2005 | 1 | 12 | 0 | 0 | 1 | -6 | 27 | 56 |
| trmar_cat | 2000 | 1 | 12 | 0 | 0 | 1 | -31 | 0 | 30 |
| matraut_cat | 2000 | 1 | 12 | 0 | 0 | 1 | -19 | 12 | 42 |
| matrvcar_cat | 2000 | 1 | 12 | 0 | 0 | 1 | -19 | 10 | 42 |
| pernoc_cat | 2000 | 1 | 12 | 0 | 0 | 1 | -12 | 20 | 50 |
| visats_cat | 2000 | 2 | 12 | 1 | 0 | 1 | -5 | 24 | 57 |
| exportsb_cat | 2000 | 1 | 12 | 0 | 0 | 1 | -15 | 16 | 46 |
| importsb_cat | 2000 | 1 | 12 | 0 | 0 | 1 | -15 | 16 | 46 |
| afi_cat | 2000 | 1 | 12 | 0 | 0 | 1 | -30 | -1 | 29 |
| contr_cat | 2000 | 1 | 12 | 0 | 0 | 1 | -32 | -2 | 28 |
| hipot_cat | 2003 | 1 | 12 | 0 | 0 | 1 | -7 | 24 | 54 |
| cred_cat | 2000 | 1 | 4 | 0 | 1 | 1 | | | 45 |
| dip_cat | 2000 | 1 | 4 | 0 | 1 | 1 | | | 45 |
| ocu_cat | 2002 | 1 | 4 | 0 | 1 | 1 | | | -6 |
| pib_cat | 2000 | 1 | 4 | 0 | 1 | 1 | | | 43 |

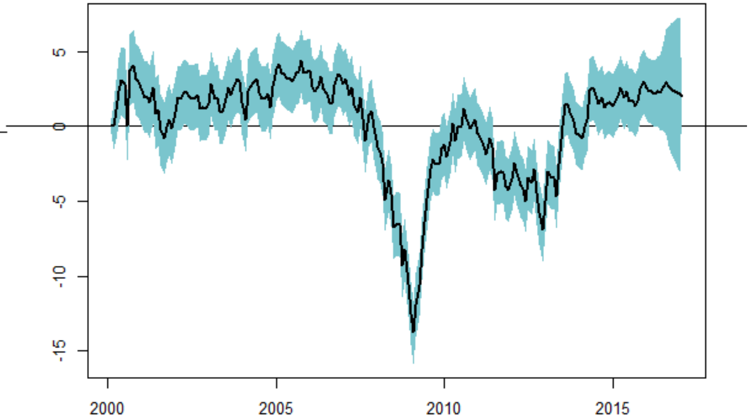
# RESULTS

## Fed

A. FACTOR with WN idio

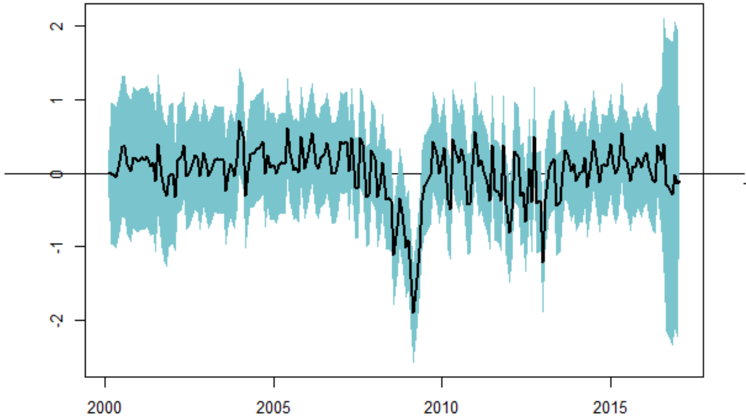



A. FACTOR with AR(1) idio

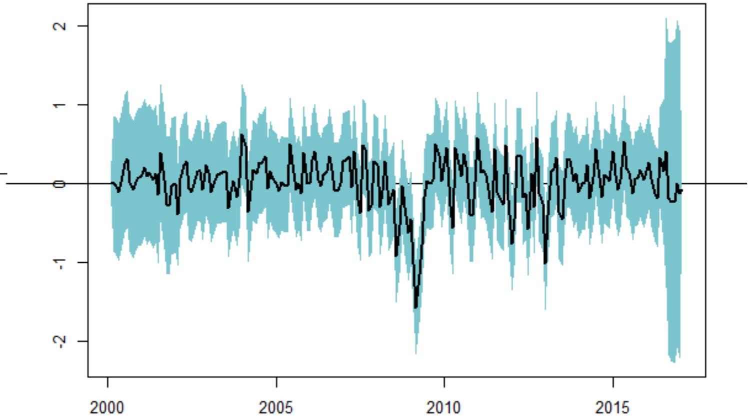



## AIReF

B. FACTOR with WN idio




B. FACTOR with AR(1) idio



# RESULTS

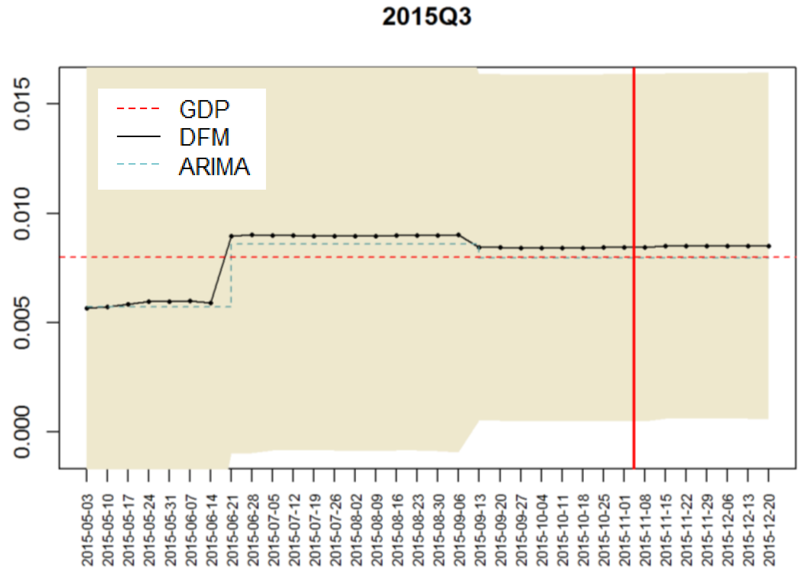
- 2014-2016 simulation
- First provisional results
- RMSE 2014-2016

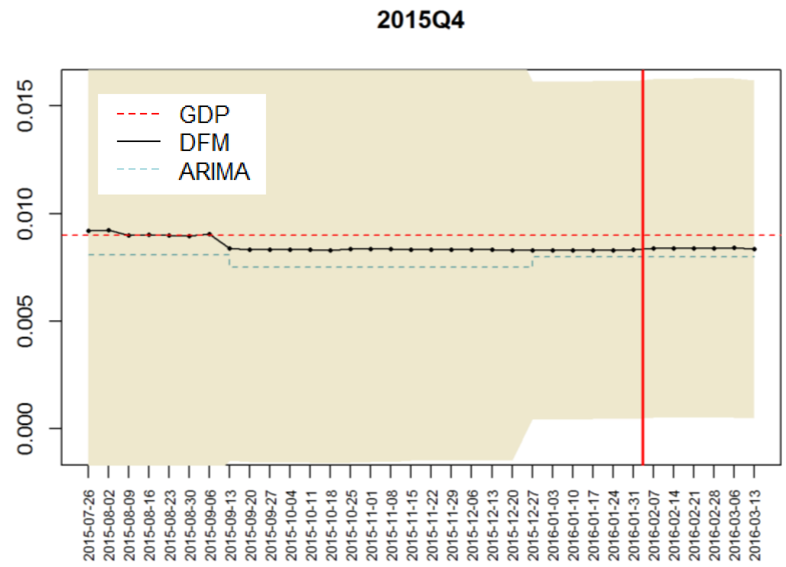
| Fed, AR | OA | NOA |
|---|---|---|
| KF | 0.7607 | 0.760865 |
| KS | 0.7609044 | 0.760486 |

| AIReF, AR | OA | NOA |
|---|---|---|
| KF | 0.761024 | 0.761039 |
| KS | 0.761983 | 0.761854 |

| Fed, WN | OA | NOA |
|---|---|---|
| KF | 0.76116 | 0.7611596 |
| KS | 0.761347 | 0.761093 |

| AIReF, WN | OA | NOA |
|---|---|---|
| KF | 0.760909 | 0.7609691 |
| KS | 0.762122 | 0.7617329 |



# RESULTS: FUTURE WORK

These results are still **provisional** and will be analyzed in depth.

**Future work**:

- Determine the optimal set of indicators to include in the Dynamic Factor Model.
- Improvement of the Kalman smoother implementation.
- Simulate the results for unstable periods (recessive cycles, for example).
- Improvement of the residual heterocedasticity using Markov Regime-Switching models. These models will also allow us to predict the state of the economic cycle among a list of pre-defined regimes.

# CONCLUDING REMARKS

- The development of an estimation methodology for the GDP flash through Dynamic Factor Models **allows to obtain robust forecasts of the Catalan GDP in compliance with the availability of the statistical information** of reference

- The method employed for estimating the GDP flash allows **to generalize the results obtained to other time frequencies** and consequently to callibrate the model in real time

- The process of validation and selection of economic indicators to include them in the Dynamic Factor Model allows to determine **their relevance as underlying measures of the economic activity**

# REFERENCES


- **STATE OF THE ART**
  - Eurostat (2016). Overview of GDP flash estimation methods.
  - Stock, J. H., & Watson, M. W. (2010). Dynamic Factor Models.
  - Stock, J. H., & Watson, M. W. (2016). Factor Models and Structural Vector Autoregressions in Macroeconomics.
  - Reichlin, L., Giannone, D., Banbura, M., Modugno, M. (2013). Now-Casting and the Real-Time Data-Flow.
- **AIReF**
  - Cuevas A, Quilis E (2009) A factor analysis for the Spanish economy (FASE). Mineco, Ministerio de Economía y Hacienda, Madrid

# REFERENCES

- **FED/ECB**


- Banbura & Modugno, 2010, “Maximum Likelihood Estimation of Factor Models on Data Sets with Arbitrary Pattern of Missing Data" Working Paper Series 1189, European Central Bank
- Doz C, Giannone D, Reichlin L, (2009) A two-step estimator for large approximate dynamic factor models based on Kalman filtering.
- Giannone D, Reichlin L, Small D (2008) Nowcasting: the real-time informational content of macroeconomic data. J Monet Econ 55:665–676.


**JDEMETRA+**

JDemetra+ on Github
https://github.com/jdemetra/jdemetra-app/releases

JDemetra+ documentation
https://ec.europa.eu/eurostat/cros/content/documentation_en

R/JDemetra+ interface
https://github.com/nbbrd/jdemetra-R

“Nowcasting” pluggin for JDemetra+
https://github.com/nbbrd/jdemetra-nowcasting/releases