FORECASTING DISAGGREGATES AND AGGREGATE WITH COMMON FEATURES

Iván Mayo and Antoni Espasa

Seminar at the Einaudi Institute for Economics and Finance

12th March 2009

Handout

1.-INTRODUCTION AND MOTIVATION

A growing flow of information

- One important fact in economic forecasting is the steadily growing flow of information available;
- Data is becoming increasingly available at a higher degree of disaggregation,
  - at temporal
  - sector and
  - regional levels.

Importance of the disaggregated analysis by sectors and regions

- 1.-The characteristics of the aggregate are not properly understood unless disaggregated data is analyzed.

The **role of a trend** in the aggregate is not the same if it mainly comes from

- a common trend in many components,
- many different unrelated trends in the components or
- the trends in a minority of components.

Negative inflation rates in US in 2009 do not have the same implications if they are generalized or if they come only from prices of energy and durable goods.

2.- Economic policy

Information and forecasts at the disaggregated level by sectors and regions are essential for economic policy.

Then the forecasts of components and aggregate are required and must be consistent.

Relative analysis through sectors and regions are of great importance.

3.- Forecasting the aggregate

A way to improve forecasting accuracy is to disaggregate as much as possible the macroeconomic variable in question and

use all the **relevant information** in the extended data set.

For data analysis, and forecasting

we should not ignore relevant information and the components can contain –**mainly if they are interrelated-** valuable information about the behaviour of the aggregate.

For policy recommendations

results on the components are of direct interest.

AGGREGATE AND DISAGGREGATE INFORMATION

1.- Try to understand the behaviour of the components and from it the behaviour of the aggregate

2.- Try to obtain a full picture of the forecasts of the components and from it of the aggregate (indirect forecast).

- 3.- Try to validate these forecasts showing that the indirect forecast of the aggregate is more accurate than the direct forecast.

Advantages of disaggregated models in forecasting the aggregate

- The advantages of using models based on relevant disaggregated information (indirect forecasting) relative to aggregate information (direct forecasting) are potentially large in forecasting the aggregate because
  - we have more information
  - we can introduce exogenous information more efficiently.
- But all this requires the possibility of **building adequate models.**

Theoretical result

- **The forecast of a stationary macroeconomic aggregate** variable with a disaggregate information set that includes the past information on the components should be at least
- as precise as an aggregate univariate forecast, which uses only past information of this aggregate (for more details, see Tiao and Guzman(1980) and Lütkepohl (1984).)

Problem in empirical applications

- This theoretical result requires that the model be known.
- From an empirical point of view, when the model specification and the parameters are unknown, they must be estimated from the data, and
- The estimation uncertainty could do that the expected increase in forecast accuracy from disaggregation may not be found.

The problem of estimation uncertainty

- Models based on disaggregate data can be much more difficult to built and
- Will include much more parameters increasing the estimation uncertainty.

- The increase of estimation uncertainty could spoil the advantages of disaggregation.

We illustrate how the theoretical results on indirect forecasting extend for the case of an I(1) aggregate

when the process is known and

there is full cointegration among several components.

Full cointegration: in m time series there are (m-1) cointegration relationships.

The convenience of disaggregating in presence of common trends

The direct modelling of the aggregate ignores a part of the disaggregate information about stochastic trends.

- The disaggregation is specially informative when there are cross-restrictions between the components.
- The paper justifies the convenience of **disaggregating in presence of stable common trends** if this is done **in a way in which the trend restrictions are maintained.**

MOTIVATION FOR FULL DISAGGREGATION

- We call basic components (BC) to the components of the aggregate at the full disaggregation level.
- Motivation for it:

To understand the properties of the data set and maintain **the most important restrictions** in the BC in intermediate sub-aggregates.

To approach the **efficiency** gain due to disaggregation

2.- DIRECT AND INDIRECT FORECAST OF AN AGGREGATE:

Theoretical results and empirical applications.

DIRECT AND INDIRECT FORECAST OF AN AGGREGATE

- Theoretical result for a stationary system which is known.

In this case, if the mean square error h-step forecasts of the aggregate derived from both processes are $W'\Sigma_C(h)W$ and $\sigma_y^2(h)$, respectively, then

$$\Delta(h)=\sigma_y^2(h)-W'\Sigma_C(h)W$$

is a positive scalar.

Only under the special condition, that $W'\Psi(L)=\theta(L)W'$ the direct forecast is as efficient as the indirect forecast.

It happens, for instance, when the components are uncorrelated and have the same stochastic structure.

This would point to a similar but independent behaviour in components data set. In the following, we will denote the former condition as efficient-direct-forecasting condition (EDFC.)

When the process is unknown

- it must be specified and estimated on the basis of sample data, then the results may change.
- Given the current tools of multiple time series model building, misspecifications in high-dimensional time series models are highly possible.

The interesting elements in a breakdown of the aggregate

- The ECDF is fulfilled when the components are:

  uncorrelated and

  have the same stochastic structure.

Hint:

Single out components which

are inter-related and/or

have different stochastic structure.

Inter-related components

- To consider **all type of inter-relationships** between the components would require complex systems and a huge estimation uncertainty associate to them will spoil the advantages of disaggregation.
- Therefore look for a type (or more) of **common feature** – initially in this paper a common trend-between a subgroup, B, of components, to simplify the specification and estimation of the model.

OUR PROCEDURE

- Initially we consider that if a common trend exist between some of the **basic components**,
- this is the most important property to explode in data analysis and forecasting.

Our Approach

One aggregate X_t

$\{X_t = X_{1t}, X_{2t},\ldots,X_{nt}\}$ can have several hundreds components.

Principle: relevant information in $\{X_{jt}\}$ can not be ignored.

Relevant information implies restrictions between the components and different distributions.

Base for the Procedure: Select one important restriction.

STEPS IN THE PROCEDURE

1. Chech that this property is stable along time
2. Single out these components in set B

   $\{B_t\}=\{X_{it},\ldots,X_{mt}\}$

4. and the rest of the components in X_t

   $X_{m+1,t}$, $X_{m+2,t},\ldots,X_{nt}$

in set R. Call SR_t the aggregate of those elements.

5. Break down Xt in

X1t, X2t,…,Xmt and SRt

6. Forecast Xt by aggregating the forecasts of

X1t, X2t,…,Xmt and SRt

7. The components Xit, …, Xmt are forecast taken into consideration the common feature restriction.

8. Check if the indirect forecast of SRt is better than the direct one. If not forecast the components using the restriction of the direct forecast.

Generalization of the procedure

Base the proceduce in selecting more than one important restriction, which could include specific type of distribution. For instance, components with conditional heteroskedasticity.

Ilustration: suppose that two important restrictions P1 and P2 have been selected.

Then we look for sets of components as:

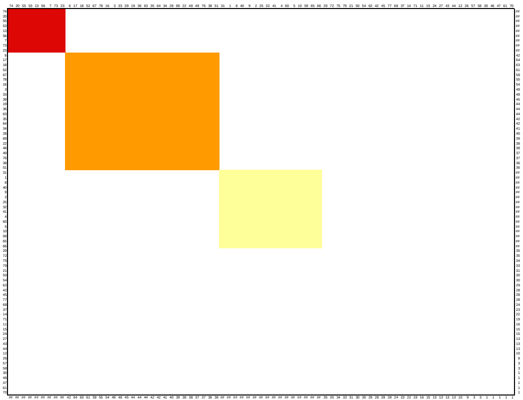
{$B1_t$} = Includes all components sharing restrictions P1 and P2

{$B2_t$} = Includes all components sharing only restriction P1

{$B3_t$} = Includes all components sharing only restriction P2

{R_t} = The reemaning components

Summary of common trend and commovements



Forecast of the aggregate in this general case

a. Forecast all the elements in {$B1_t$} taken into consideration the both restrictions which they share.
b. Forecast all the elements in {$B2_t$} taken into consideration the common restriction.
c. Proceede similarly whith the elemens in $B3_t$
d. Aggregate the elements in {R_t} in SRt and forecast SR_t
e. Aggregate all previous forecasts as the forecast of X_t

We approach the problem of designing an indirect forecast based on a disaggregation scheme which is

- informative about relevant restrictions between the components –in this case is more difficult that the EDFC could hold- and
- for which the specification and estimation problems are relatively simple: pairwise comparisons and single-equation models.

Empirical literature

- Although there is a much empirical literature on forecasting inflation in the context of the euro-area,
- the typical focus is to consider that the HICP, can be written as a function of a weighted sum of a small number of components, traditionally some official sub-aggregates.
- They use a disaggregation shceme far away from the full one.

3.-COMPONENTS WITH A COMMON TREND.

THE EDF CONDITION

Estimating common trends in the n basic components

- allows us, if this is the case, to define two sets of components,

SETS B AND R

- One, B, which includes all the components, say

  X1,X2,…,Xm,

  which share a common trend – this is the property which we have selected to simplify the disaggregation -, and
- Another, R, formed by the r remaining components.

INTERMEDIATE DISAGGREGATION

- Aggregating all the elements in R in a sub-aggregate denoted SR,
- For indirect forecasting we will forecast the elements

  X1 to Xm

taking into account the restriction between them and

the elements of R, directly or indirectly as it is more accurate.

A parsimonious breakdown of the aggregate

- This approach enables us to obtain a parsimonious breakdown of the data in m+1 components

  $X1_t$, $X2_t$,…,Xm_t and SR_t

- which maintain the selected restriction existing in the full disaggregation.

JUSTIFICATION OF THE PROCEDURE

- IN THE PRESENCE OF A COMMON TREND IN THE COMPONENTS, THE EDF CONDITION IS VERY SPECIFIC.

Let us illustrate the case in which the components are fully cointegrated and the process is known. Our interest is to forecast the year-on-year rate of growth. The EDF condition can not be satisfied for an arbitrary W vector when there exits full cointegration among n the components.

We show the above proposition for the case in which the full disaggregation implies n random walks with a common unit root.

Assumption 1: We have an aggregate variable composed by a *n* dimensional non-stationary vector of basic components, $C_t = (C_{1t}, ..., C_{nt})'$.

contains n pure random walks generated from a single unit root (a common stochastic trend.) The aggregate is calculated as

$$y_t = W' C_t = \sum_{i=1}^{n} w_i C_{it}, \quad t = 1, ..., T$$

Proposition 1: For an arbitrary weighting vector, it is better to forecast the disaggregated components and aggregate them (indirect forecast) than forecasting the aggregate directly (direct forecast), since the MSE (Indirect forecast) < MSE (direct forecast).

Forecasting Aggregates with a Common Trend

Equations (4) and (6) are true at the sometime if:

$\theta - (1 + \beta(\alpha_1 - \alpha_2)) = 0$ **and** $\theta - (1 + (\alpha_1 - \alpha_2)) = 0$ **(7)**

Then, we get the same forecast error for both models, aggregate and disaggregate.

Equations (7) require, $\beta = 1$ **or** $\alpha_1 = \alpha_2$.

Conclusion:

If the process in a CI(1,1) disaggregate DGP is known, then, in general, the mean square error of the disaggregate process is lower than the aggregate one.

If the components are non-cointegrated the MSE of both models is equal.

Efficiency of the direct forecast in the two-components case:

(1) If the cointregration vector is (1,-1) and

(2) If the adjustment parameters to the disequilibrium are equal.

4.-Specification of the components with a common trend, set B

Specification of the components of set B

- The components in B are restricted by (m – 1) cointegration restrictions and we say that they are **fully cointegrated**.
- The m components sharing a common trend are identified by a **testing procedure carried over each one of all the possible pairs of components** from the full disaggregation.
-

Cointegrated and cotrending componponents

- **In this paper a pair of components is considered cointegrated if and only if there exists a linear relationship between them in which both the stochastic and possible deterministic trends of the components cancel out.**

STABILITY IN THE COINTEGRATION RESTRICTIONS

- **A key aspect for the proposed reduction in dimensionality from n to (m+1) relies on the idea that**
- **the full cointegration property between the elements of set B is stable over time.**

Test for cointegration stability

- **Once set B has been inicially defined by the previous pairwise cointegration tests,**
- **a sub-aggregate SB can be constructed with the components of B.**
- **This sub-aggregate can be taken as a proxy for the common trend.**

- **Then for all components in B we test that the cointegration relationship of each one of them with SB is stable along time.**
- **The components for which the stability hypothesis is rejected are excluded from B and the remaining ones define the final set B.**

Methodology

- The first step of our method is to check the order of integration of the basic component series.
- The standard analysis using the Augmented Dickey-Fuller test suggests that Euro area HICP and USA CPI should be treated as non-stationary series.
- Whenever it seems appropriate the HEGY test in the monthly version is performed for the HICP due to the complexity of the seasonal scheme of the index.

SEASONALITY

- In the HICP variable, the hypothesis of seasonal unit roots was tested on all basic components and it was rejected for all, but the hypothesis of regular unit root was not.
-

The time series of the basic components of the both HICP and USA CPI are modelled

- taking the first differences of each variable and
- considering a deterministic seasonal structure capture by dummies for the seasonality.

In a second step

- We test for cointegration in all possible pairs which can be formed with the basic components.
- This test is performed using a very restrictive criterion for ending up with the presence of bi-variately cointegration.

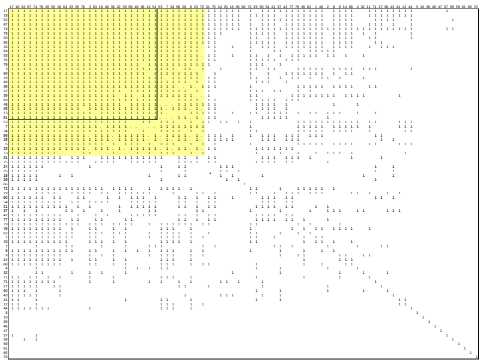
- With these results a binary nxn matrix, M, which resumes the test results is constructed,
- putting a digit one in cell (i,j) if the corresponding two components are cointegrated and zero otherwise.
- Each pair is tested in both directions and we assume cointegration only if it is not rejected in both cases.

- The elements of M are arranged in such a way that in the upper left corner of the matrix we put the largest m_0 x m_0 sub-matrix full of ones,
- meaning that in these m_0 basic components there exist (m_0-1) cointegration restrictions – full cointegration- , or equivalently that these components share the same stochastic trend.

- We denote B_0 the set whose elements are these m_0 basic components and
- with them a sub-aggregate, SB_0 , is built.

Tables and Figures

Matrix 1. Summary of binary cointegration relationship between Euro Area HICP components:



- Then for elements outside B_0 we considered the row of M with more ones corresponding to the columns of the components in B_0 .
- Denoting $C(m_0 +1)$ to this component, a bi-variety cointegration test between $C(m_0 +1)$ and SB_0 is performed.

- If they are not cointegrated $C(m_0 + 1)$ is assigned to a set R of basic components which are not full cointegrated.
- If $C(m_0 + 1)$ is cointegrated with SB_0 then B_0 is enlarged with $C(m_0 + 1)$, and
- we call B1 to this set and
- with its elements a sub-aggregate SB1 is constructed.

- Then for the elements outside B_1 the component, say $C(m_0 + 2)$, with more bi-variately cointegration relationships with the elements of B1 is selected and
- a cointegration test between $C(m_0 + 2)$ and SB_1 is done and,
-

- as before, if no cointegration is found C(m0 +2) is assigned to R and
- otherwise is used to enlarge B1 forming a new set B2 and
- from it a new sub-aggregate SB2.

- Now, before to pass to consider a new basic component,
- we test if SB1 and SB2 are cointegrated.
- If they are not we remove C(m0 +2) from B2 and SB2 and assigned it to R..

- Then we continue with the remaining basic components proceeding in a same way as with C(m0 +2).
- At the end of the process we have a set of, say m, basic components which are fully cointegrated and a set of the remaining components, which define the set R.
- From R the sub-aggregate SR is constructed

- The disaggregation proposed to forecast the aggregate yt is the one which its elements are C(1), C(2),…, C(m) and R.
- In other words we use a breakdown of the aggregate that only singles out the basic components which are fully cointegrated between them.
- The test used is the one proposed byEngle-Granger (1987)

FORECASTING SR_t AND ITS COMPONENTS

- Check which procedure, direct or indirect, is more accurate.
- If the direct one, forecast the components taking into account the restriction implied by the direct forecast of the aggregate.

Set B in EU and US

- For the HICP in the Euro area, its full disaggregation implies 79 basic components, and it is found that a unique common trend is shared by 33 components with a weight in the index of 37%.
- In USA inflation, a unique common trend is shared by 63 components from 161 basic components, with a weight of 43% in the total index.

Results from components with and without a common trend could be useful for study differences between demand and supply in different sectors. EU

- Food prices which do not share a common trend.

  Bread and cereals, milk, cheese, eggs, oils and fats, sugar ,honey, jam, chocolate, coffee, tea and cocoa, **mineral waters, soft drinks**, fruit and vegetables juices, beer, spirits and tobacco, and meat.

- Food prices which share a common trend.

  Wine, **fish and sea food**, **fruit and vegetables**.

•Forecasting Strategy and results.

The proposed forecasting strategy for the aggregate consists of

- forecasting each one of the components in B, taking into account the common feature they share, and

- forecasting independently the sub-aggregate SR, directly or indirectly, and

- then aggregating all these forecasts.

All the models include lags of the endogenous variable and of SB and SR.

Model with common trends

- The disaggregate model:

$$\Delta X_t = \mu + \Psi\ S_t + \alpha(SB_{t-1} - \beta X_{t-1}) +$$

$$+ \rho(L)\Delta SB_t + \lambda(L)\Delta SR_t` + \gamma(L)\Delta X_t + a_t$$

$$\Delta SR = \mu + \Gamma S_t + \theta(L)\Delta SR_t + e_t$$

$$SB_t = \sum_{j\in B} w_j C_j$$

$$SR_t = \sum_{j\in R} w_j C_j$$

Forecast evaluation

The models are estimated every period.

- The main criterion for the comparison of the forecasts is the root mean square forecast error statistic (RMSFE),
- the prediction accuracy is evaluated at any given forecast horizon h from 1 to 12.
- The forecast is evaluated in terms of year-on-year inflation rate in percentages.

The D&M test is implemented to test if the relative accuracy is statistically significant.

- The tables also report the results in relative terms, using a coefficient defined as the ratio between the RMSFE of the proposed model and the RMSFE of the univariate model for the aggregate.
- Numbers smaller than one indicate better performances than the univariate aggregate model.

FORECASTING EURO AREA INFLATION

- . The euro area HICP s extends from 1995:01 to 2006:12,
- the first seven years were used to estimate the models, and
- the following five years, to evaluate the forecasting performance.

- In the HICP case, a simple preliminary analysis shows, as pointed out in Espasa and Albacete (2007), two different seasonality patterns during the sample period.
- The first lasts till December 2000, and the second one from the next month onwards.

- the best univariate model for the aggregate variable using an out-of-sample experiment for a battery of models. The models considered are three. Model 1: . Model 2:. Model 3: . Among these model ones, best performance is realised by the model 2 (see table 1),

Univariant Models for the HICP

- Summary of Univariate models for the aggregate Euro area HICP
- Model 1:

$$\Delta Y_t = \hat{c} + \sum_{i=1}^{11} \hat{\alpha}_i \Delta d_{it} + \sum_{i=1}^{11} \hat{\delta}_i \Delta D_{it} + \hat{\phi}(L)\hat{\varepsilon}_t$$

- Model 2:

$$\Delta Y_t = \hat{c} + \sum_{i=1}^{11} \hat{\alpha}_i \Delta d_{it} + \sum_{i=1}^{11} \hat{\delta}_i \Delta D_{it} + \hat{\varepsilon}_t$$

- Model 3:

$$\Delta^{12} \Delta Y_t = \hat{\phi}(L)\hat{\varepsilon}_t$$

Empirical Results

- Aggregate alternative in the Euro area:

Table 1: Forecast results for the annual growth rate of the aggregate variable of alternative univariate models for the aggregate

	AGGREGATED VARIABLE			RMSFE RATIOS (RMSFE(Model i)/RMSFE(Model 1))		
Forecast horizons	Univariante Model 1	Univariante Model 2	Univariante Model 3	Univariante Model 1	Univariante Model 2	Univariante Model 3
1	0.18	0.17	0.18	1.00	0.98	0.99
2	0.27	0.27	0.25	1.00	0.98	0.91
3	0.35	0.34	0.31	1.00	0.96	0.88
4	0.37	0.36	0.33	1.00	0.96	0.88
5	0.37	0.35	0.33	1.00	0.95	0.90
6	0.36	0.34	0.33	1.00	0.94	0.92
7	0.35	0.33	0.33	1.00	0.94	0.94
8	0.34	0.31	0.33	1.00	0.93	0.98
9	0.31	0.29	0.32	1.00	0.94	1.03
10	0.28	0.26	0.33	1.00	0.94	1.16
11	0.25	0.24	0.33	1.00	0.95	1.34
12	0.25	0.24	0.37	1.00	0.95	1.45

Empirical Results

- Disaggregate forecasting procedure for the Euro area:

Table 3: RMSFE summary of Euro area year-on-year inflation in percentage points

Forecast Horizons	ARIMA Model	Forecasting Procedure	D&M Test	RMSFE(FP)/ RMSFE(AM)
1	0.17	0.16		0.95
2	0.27	0.25		0.94
3	0.34	0.30	**	0.90
4	0.36	0.33	**	0.92
5	0.35	0.33	**	0.93
6	0.34	0.31	**	0.92
7	0.33	0.29	**	0.90
8	0.31	0.29	*	0.92
9	0.29	0.26	**	0.91
10	0.26	0.24	*	0.91
11	0.24	0.21	*	0.90
12	0.24	0.21	*	0.86

* 95%
** 99%

FORECASTING US INFLATION

- The results show that the gains with our procedure with respect an aggregate univariate model are greater than in the euro area
- with the ratio of RMSE going from 0.75 for h=2 to 0.86 for h=12.

Main empirical conclusions:

- The above applications show that cointegration relationships from the full disaggregation turn to be stable and improve the accuracy of the aggregate forecast at longer horizons.
- The results are different to the previous literature exercises, where the long term performance were less encouraging.

Other related literature with relevance for this paper

- is the one about dynamic factor models.
- These models can collect the relevant information of large number of exogenous variables in a few variables and use them for forecasting the variable of interest.
- Usually they work with the stationary transformation of the data.

Main differences of our procedure with the dynamic factor model.

- We use all the information with the aim of: - understanding it
  - to provide a full picture of the properties of the data and
  - of the forecasts of all variables.
- The common features in our case usually include a non-stationary one.

Main differences of our procedure with the dynamic factor model.

- It is obtained from the breakdown of the aggregate.
- The most important factor is the common trend of the largest group of prices sharing a common trend.
- It can be approximated by a sub-aggregate which could be interpreted in economic terms.

Main differences of our procedure with the dynamic factor model.

- Divides the basic components of the aggregate in sets, B_j, whose components share a common feature, and R, whose components do not share any. This could be useful for economic policy and for further economic analysis.
- The procedure insures stability of the common features.
- Provides an endogenous design for breaking down the aggregate.

Rank of Models for Euro area HICP (RMSE)

Models:	Univariant Model ARIMA	Dynamic Factor Model	Forecasting Procedure from Paper	Dynamic Factor Model	Dynamic Factor Model	Univariant Model ARIMA
Forecast Equations	1	1	33+1	33+1	79	79
Style:	*Aggregate*	*Aggregate*	*Disaggregate*	*Disaggregate*	*Disaggregate*	*Disaggregate*
1	0.17	0.19	0.16	0.17	0.18	0.17
2	0.27	0.29	0.25	0.26	0.27	0.28
3	0.34	0.36	0.30	0.32	0.33	0.34
4	0.36	0.40	0.33	0.35	0.38	0.38
5	0.35	0.41	0.33	0.36	0.39	0.38
6	0.34	0.40	0.31	0.35	0.41	0.40
7	0.33	0.39	0.29	0.34	0.43	0.41
8	0.31	0.38	0.29	0.32	0.45	0.43
9	0.29	0.36	0.26	0.28	0.44	0.42
10	0.26	0.34	0.24	0.24	0.41	0.41
11	0.24	0.33	0.21	0.21	0.41	0.39
12	0.24	0.35	0.21	0.22	0.41	0.39

Rank of Models for Euro area HICP (% RMSE over Aggregate Benchmark)

Models:	Univariant Model ARIMA	Dynamic Factor Model	Forecasting Procedure from Paper	Dynamic Factor Model	Dynamic Factor Model	Univariant Model ARIMA
Forecast Equations	1	1	33+1	33+1	79	79
Style:	*Aggregate*	*Aggregate*	*Disaggregate*	*Disaggregate*	*Disaggregate*	*Disaggregate*
1	100%	107%	95%	100%	103%	97%
2	100%	108%	94%	98%	100%	104%
3	100%	108%	90%	94%	99%	101%
4	100%	112%	92%	98%	106%	106%
5	100%	116%	93%	102%	112%	108%
6	100%	118%	92%	104%	121%	117%
7	100%	119%	90%	104%	130%	124%
8	100%	120%	92%	102%	144%	136%
9	100%	124%	91%	98%	152%	144%
10	100%	130%	91%	92%	157%	154%
11	100%	141%	90%	90%	174%	164%
12	100%	144%	86%	92%	168%	161%

Rank of Models for USA CPI (RMSE)

Models:	Univariant Model 1 ARIMA	Univariant Model 2 ARIMA	Dynamic Factor Model	Forecasting Procedure from Paper	Dynamic Factor Model	Univariant Model ARIMA
Forecast Equations	**1**	**1**	**1**	**63+1**	**63+1**	**161**
Style:	*Aggregate*	*Aggregate*	*Aggregate*	*Disaggregate*	*Disaggregate*	*Disaggregate*
1	0.37	0.29	0.27	0.31	0.32	0.40
2	0.59	0.51	0.46	0.44	0.56	0.60
3	0.71	0.69	0.58	0.55	0.72	0.80
4	0.91	0.97	0.76	0.74	0.96	1.07
5	1.06	1.21	0.94	0.88	1.18	1.30
6	1.18	1.40	1.07	0.98	1.34	1.42
7	1.24	1.52	1.14	1.04	1.41	1.46
8	1.28	1.60	1.19	1.06	1.50	1.53
9	1.35	1.75	1.28	1.15	1.62	1.63
10	1.45	1.92	1.36	1.22	1.74	1.70
11	1.54	2.07	1.46	1.31	1.84	1.66
12	1.63	2.32	1.59	1.40	1.98	1.97

Rank of Models for USA CPI (% RMSE over Aggregate Benchmark)

Models:	Univariant Model 1 ARIMA	Univariant Model 2 ARIMA	Dynamic Factor Model	Forecasting Procedure from Paper	Dynamic Factor Model	Univariant Model ARIMA
Forecast Equations	**1**	**1**	**1**	**63+1**	**63+1**	**161**
Style:	*Aggregate*	*Aggregate*	*Aggregate*	*Disaggregate*	*Disaggregate*	*Disaggregate*
1	1.00	0.79	0.73	0.85	0.88	1.08
2	1.00	0.87	0.77	0.75	0.95	1.02
3	1.00	0.97	0.82	0.78	1.01	1.12
4	1.00	1.06	0.83	0.81	1.05	1.17
5	1.00	1.14	0.89	0.83	1.11	1.23
6	1.00	1.19	0.91	0.83	1.14	1.20
7	1.00	1.22	0.92	0.84	1.14	1.18
8	1.00	1.25	0.93	0.83	1.17	1.20
9	1.00	1.30	0.95	0.85	1.20	1.21
10	1.00	1.32	0.94	0.84	1.20	1.18
11	1.00	1.35	0.95	0.85	1.20	1.08
12	1.00	1.42	0.97	0.86	1.21	1.21

D&M Test

Table 1: D&M test respect to the model 1 for the aggregate

- The D&M test shows a significative forecast accuracy improvement respect to the aggregate benchmark in both procedures.
- Our Forecasting Procedure outperforms the aggregate benchmark significatively for all periods.

	D&M Test	D&M Test
Models:	Forecasting Procedure from Paper	Dynamic Factor Model
Forecast Equations	**63+1**	**1**
Style:	*Disaggregate*	*Aggregate*
1	**	**
2	**	**
3	**	**
4	**	**
5	**	**
6	**	*
7	**	*
8	**	*
9	**	
10	**	
11	**	
12	**	

D&M Test

Table 2: D&M test respect to the Forecasting Procedure from paper

- Our Forecasting Procedure outperforms the Dynamic Factor Model for the aggregate according the D&M test.

	D&M Test
Models:	Dynamic Factor Model
Forecast Equations	**1**
Style:	*Aggregate*
1	
2	*
3	**
4	*
5	**
6	**
7	**
8	**
9	**
10	**
11	**
12	**

Looking for leading indicators in the set B

- Testing for strong exogeneity within the components of set B we could detect the presence of a leading price.
- This is not the case for the CPI in US and the HICP in the euro area.

Triplets of basic components with two common trends.

- For all possible pairs of basic components in set R we could test if there is a cointegration relationship between the triplet composed by the corresponding pair from R and the aggregate SB. Then a set C could be defined with all pairs in R with share this cointegration restriction. This would imply that prices in C are driven by two common trends
- This type of analysis did not show positive results with the inflation data of this paper.

Considerations for future work

- Make a rigorous outlier detection of the data before the cointegration analysis.
- Test for normality in all basic components and consider specias1 treatment for the non-normal components.

Thank you.