FORECASTING DISAGGREGATES AND AGGREGATE WITH COMMON FEATURES

Iván Mayo and Antoni Espasa

First Macroeconomic Forecasting Conference

Rome, 27th March 2009

(Previously distributed as Forecasting a Macrovariable which includes components with common trends)

The discussion on forecasting an aggregate variable

directly or indirectly, by aggregating the forecasts of its components, is an old one, and

it has received much more attention recently with the establishment of the euro area.

Usually this discussion only pays attention to the forecast of the aggregate.

By contrast, the starting point of this paper

- is that all data, for the aggregate as a whole and for its components, matter for
- (I) a proper full understanding of the aggregate and
- (II) certainly for the formulation of useful economic policies.

Basic Components of an aggregate

This approach eventually leads to the consideration of all the components of an aggregate, which we called basic components.

In this context we focus the paper in providing joint consistent forecasts for the aggregate and its components.

In the **process of validation of our results** it would be necessary to show that the indirect forecast of the aggregate is at least as accurate as the direct one.

When this happens it provides **a (III) third reason for the disaggregated analysis.**

This is a non-popular approach

because the number of basic components commonly is **over one hundred** and considering disaggregation by sectors and regions easily goes over one thousand.

This is the reason why many papers dealing with disaggregation have been constrained to consider some official breakdown of the aggregate **in a few components, usually five in the case of the euro area inflation.**

The second author has published different papers with this orientation, but we will show that such an approach is not satisfactory.

The computation required

in this general context **could be huge** but

it can be done by computer without consuming much of the researcher’s time.

- Researchers would, however, have to spend considerable time trying to
- find a general econometric framework for the problem, and
- understanding the economic implications of the results.
- But in both cases, it would be time well spent.

Theoretical results for stationary variables

have shown that in general **the indirect approach will provide more accurate forecasts of the aggregate**,

but when special conditions are fulfilled the direct approach is efficient. We call them Efficient-Direct-Forecasting Conditions (EDFC).

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.

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.

The case in which there are restrictions between the components

In this paper we constrain ourselves to that case.

This does not means that we consider that the **distributional differences** are not or less important,

it simply means that we start to study the problem from a way which turns to be easier to solve in a general framework.

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.

A simpler approach based on bivariate analyses

- Therefore look for types of **common features, P_j ,**between the elements of subset, B_j, of basic components.
- Then only **bivariate analyses** are required to simplify the specification and estimation of the model.
- And at the same time this approach allows us to **maintaining the important restrictions** between the components, which is one of the main reasons why the disaggregated approach matters.

I(1) Aggregates

In the paper we show that for aggregates I(1) with cointegregated components

the EDFC applies only under very specific weighting vectors in the aggregation scheme.

This confirms the intuition that common trends in the components is a type of restriction which is worthwhile to exploit in indirect forecasting.

Bivariate methods

In order to be able to run the analysis only by bivariate methods we look for situations in which a subset of components share just one common trend.

The common trend procedure can also be similarly applied to situations in which a **subset of components shares one common serial correlation factor (CSCF).**

OUR PROCEDURE

1). Select the types of the possibly most relevant constraints in the components which, in each case, involve just a single factor in a subset of components.

In the paper we propose two types of restrictions:

(PA) the presence of **a common trend** between a subset of basic components and

(PB) the presence of a **common serial correlation** between a subset of basic components.

- 2) Look for the largest subset, say S1, of basic components with a common trend and

  also for the largest subset, say S2, of basic components with a common serial correlation factor (CSCF).

  The identification of the elements in both subsets must be done checking that for each element the restriction in question is stable along time.

- From the elements in S1 and S2 define the following subsets.
- B1, formed by the elements which share a common trend and a CSCF ;
- B2, defined by the elements which only share a common trend;
- B3, whose elements are those which share just a CSCF and
- R which includes the basic elements outside S1 and S2.

- Then aggregating in each case all the elements in the corresponding subset, B1, B2, B3 and R,
- obtained the subaggregates, SB1, SB2, SB3 and SR.

- 3) Forecast the basic components in B1, B2 and B3 in the following way.
- **The elements of B1 by single-equation equilibrium correction models** which

  includes the equilibrium correction term with respect the common trend

  and the CSCF

  and could include lags in differences of the element in question, of SB2 and SR.

- The common trend can be approximated by SS1, the aggregate obtained with all the elements of S1.
- SS1 can also be used to test the stability of the common trend constraint, by testing whether the cointegration relationships between each element in S1 and SS1 are stable.

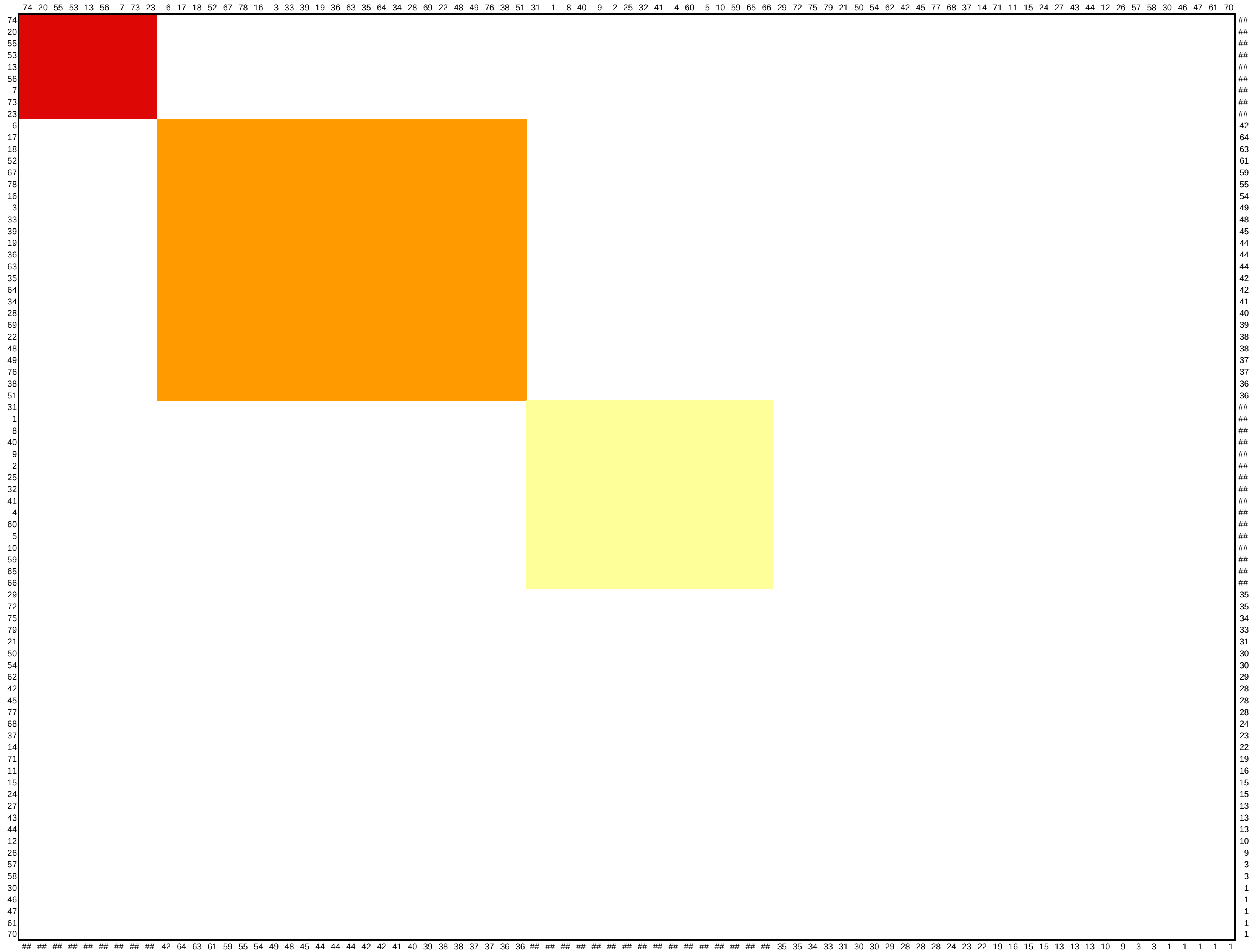
- The elements of B2 and B3 by similar models that in the previous case, but

  - omitting the CSCF for the elements of B2 and

  - omitting the equilibrium correction term in the models for elements in B3.

- 4) In order to forecast the elements in R, check how SR is forecast best, by the direct or indirect procedure.
- If the latter is best, it should be used,
- otherwise we should forecast the elements in R imposing the constraint from the direct forecast of SR.
- 5) Forecast the aggregate aggregating the forecasts of the components.

Summary of common trend and comovements



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.

With the above procedure we obtain forecasts of all the basic components and aggregating them we have the forecast of the aggregate.

Since in the indirect forecast we have taken **into consideration the main stable restrictions** between the components,

it is reasonable to expect that this forecast is more accurate than the direct one.

This can be expected in spite of the fact that the estimation uncertainty is greater in the former than in the latter.

Forecasting inflation in the euro area and in US

In the **euro area** the number of basic components for which a reasonable large time series exists is **79** and

we find that 33 of them (37%) make up the largest set of basic components with a common trend.

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 Procedure with indirect forecast of SR

Models:	Univariant Model ARIMA	Forecasting Procedure from Paper Forecasting Disaggregate SR	Forecasting Procedure from Paper
Forecast Equations	**1**	**79**	**33+1**
Style:	***Aggregate***	***Aggregate***	***Disaggregate***
1	0.17	0.17	0.16
2	0.27	0.26	0.25
3	0.34	0.31	0.30
4	0.36	0.34	0.33
5	0.35	0.34	0.33
6	0.34	0.35	0.31
7	0.33	0.35	0.29
8	0.31	0.37	0.29
9	0.29	0.37	0.26
10	0.26	0.35	0.24
11	0.24	0.35	0.21
12	0.24	0.36	0.21

The proposed forecasting procedure

improves the accuracy of the direct forecast for all the horizons which have been considered,1 to 12, and

the improvement is **significant** from horizon three onwards.

In all models a break in the seasonal structure from 2001 has been included in all models, when required.

For the US

it has been possible to work with **161** basic components and 63 of them (43%) are in the largest set with a common trend.

The indirect method is significantly superior from all horizons.

Empirical Results

- Disaggregate forecasting procedure for the USA:

Table 4: RMSFE summary of USA year-on-year inflation in percentage points

Forecast Horizons	ARIMA Model 1	Forecasting Procedure	D&M Test	RMSE(FP)/ RMSE(AM1)
1	0.37	0.31	**	0.85
2	0.59	0.44	**	0.75
3	0.71	0.55	**	0.78
4	0.91	0.74	**	0.81
5	1.06	0.88	**	0.83
6	1.18	0.98	**	0.83
7	1.24	1.04	**	0.84
8	1.28	1.06	**	0.83
9	1.35	1.15	**	0.85
10	1.45	1.22	**	0.84
11	1.54	1.31	**	0.85
12	1.63	1.40	**	0.86

* 95%

** 99%

An empirical conclusion:

- The above applications show that, for the examples considered, stable cointegration relationships from the full disaggregation improve the accuracy of the aggregate forecast at all horizons (1 to 12).
- These results differ from the previous literature exercises, where the medium term performance was 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 factors and use them for forecasting the variable of interest.
- In this case the large group of variables are the components.

Main differences of our procedure with the dynamic factor model.

- (a) We use all the information with the ambitious aim of: - understanding it

  - to provide a full picture of the properties of the aggregate and

  - of the forecasts of all variables,

which could be **useful for policy**.

- (b) The common features are obtained from a detail analysis of the restrictions between the component series.
- (c) Our procedure always looks for a common trend.
- (d) The common trend can be approximated by a sub-aggregate of the components and can be interpreted in economic terms.

- (e) A by-product of our procedure is the classification of the components in sets whose elements share a common trend and/or a common serial correlation factor.

This could be **useful for further economic analysis and for economic policy.**

- In this respect, it would be interesting, for example, to provide an economic explanation of
- why prices of **soft drinks in the euro area share a common trend** with several other prices and
- the prices of alcoholic drinks do not.

- (f) Our procedure ensures common factor stability.
- (g) It also shows that the definition of intermediate aggregates is an endogenous question and **provides strong statistical results for their formulation.**

- (h) Another by-product of our method is that it is easy to test whether there are leading indicators between the components.
- (i) Finally, the paper shows that, in both the above applications, our procedure performs significantly better than dynamic factor analysis.

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%

Euro Area: More factors do not improve the results

Models:	**Univariant Model ARIMA**	**Dynamic Factor Model**	**Dynamic Factor Model**	**Dynamic Factor Model**
Forecast		**1FACTOR**	**1FACTOR**	**4FACTOR**
Equations	**1**	**1**	**79**	**79**
Style:	***Aggregate***	***Aggregate***	***Disaggregate***	***Disaggregate***
1	0.17	0.19	0.18	0.17
2	0.27	0.29	0.27	0.27
3	0.34	0.36	0.33	0.36
4	0.36	0.40	0.38	0.40
5	0.35	0.41	0.39	0.42
6	0.34	0.40	0.41	0.41
7	0.33	0.39	0.43	0.41
8	0.31	0.38	0.45	0.42
9	0.29	0.36	0.44	0.43
10	0.26	0.34	0.41	0.43
11	0.24	0.33	0.41	0.43
12	0.24	0.35	0.41	0.44

Euro Area: Common trend and CSCF

Models:	Univariant Model ARIMA	Forecasting Procedure from Paper (only CT)	Forecasting Procedure from Paper (CT&CM)	Forecasting Procedure from Paper Stability CM (CT&CM)
Forecast Equations	**1**	**33+1**	**49+1**	**42+1**
Style:	***Aggregate***	***Disaggregate***	***Disaggregate***	***Disaggregate***
1	0.17	0.16	0.16	0.17
2	0.27	0.25	0.24	0.25
3	0.34	0.30	0.30	0.31
4	0.36	0.33	0.32	0.33
5	0.35	0.33	0.32	0.33
6	0.34	0.31	0.32	0.32
7	0.33	0.29	0.32	0.31
8	0.31	0.29	0.32	0.29
9	0.29	0.26	0.31	0.27
10	0.26	0.24	0.30	0.25
11	0.24	0.21	0.26	0.23
12	0.24	0.21	0.25	0.23

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 special treatment for the non-normal components.

Thank you