14th February 2011

FORECASTING AGGREGATE AND DISAGGREGATES WITH COMMON FEATURES

Preliminary version

Iván Mayo and Antoni Espasa[1],

ABSTRACT

The debate about forecasting an aggregate variable, directly or indirectly, is usually centred only on the forecasting accuracy of the aggregate. In contrast, the starting point of this paper is that all data matter -aggregate and components. The paper is focused on providing joint consistent forecasts for the aggregate and its components and in showing that the indirect forecast of the aggregate is at least as accurate as the direct one. The procedure developed in the paper is a disaggregated approach based on single-equation models for the components, which take into account common stable features which some components share between them. The procedure can be easily extended to include exogenous variables.

The procedure is applied to forecasting euro area, UK and US inflation and it is shown that its forecasts are significantly more accurate than the ones obtained by the direct forecast of the aggregate or by dynamic factor models.

A by-product of the procedure is the classification of a large number of components by restrictions shared between them, which could be also useful in other respects, as the application of dynamic factors, the definition of intermediate aggregates or the formulation of models with unobserved components.

---

[1] Espasa, Dept. Estadística, Univ. Carlos III de Madrid 28903, Madrid, Spain, e-mail:antoni.espasa@uc3m.es; Mayo, Dept. Estadística y Econometría, Univ. Carlos III de Madrid 28903, Madrid, Spain, e-mail:imayo@eco.uc3m.es

This paper was originally distributed as "Forecasting A Macroeconomic Aggregate Which Includes Components With Common Trends". We appreciate the comments and suggestions by Katerina Juselius. We are also grateful for the comments of the participants at the Econometric Seminars of the Nuffield College and the European University Institute. Antoni Espasa acknowledges financial support from Ministerio de Educación y Ciencia, proyect ECO 2009-08100. Any opinions are solely those of the authors.

The disaggregated inflation forecasts could show market differences, providing then, hopefully, some clues about the main factors causing inflation. But the procedure presented in this paper does not include causal economic variables to explain inflation, because in general the results from the economic theory will refer to the aggregate but not to the basic components. If forecasts for the aggregate are available from a causal model, they can be linked to the ones, usually more accurate, from the disaggregation approach by means of a regression, obtaining then the most accurate forecast with an econometric explanation.

Key Words: common trends, common serial correlation, inflation, euro area, UK, US, cointegration, single-equation econometric models.
JEL codes: C18,22,38,43,51, 53 and E17,31, and 37.

I.- Aim and Motivation of the paper.

The demand for macroeconomic forecasts has increased considerably in the last ten year or so and with it the requests of having quicker releases for the official data. To palliate the delays in official data more soft economic indicators have appeared and for their use dynamic factors techniques are extensively employed. Questions as flash indicators and nowcasting have being considered with some attention in the literature, and it appears that they are more relevant for GDP than for other macro-variables, especially inflation which is the variable of main interest in this paper.

In this context of increasing relevance of economic forecasts, one important fact in is the steadily growing flow of information available to the forecasters; in particular data is becoming increasingly available at a higher degree of disaggregation, at the regional, temporal and sector levels. Thus, the debate about forecasting an aggregate variable directly or indirectly by aggregating the forecasts of its components, which is an old one, has recently received considerable attention, certainly within the euro area. This topic is also of great interest in nowcasting and flash indicators, thus even when the paper is applied to inflation its results are of interest for many more macro-variables. Usually, the discussion about direct or indirect forecasting is centred only in the forecasting accuracy of the aggregate. By contrast, the starting point of this paper is that all data matter -aggregate and components-, both for a full understanding of the aggregate and for the formulation of useful economic policies.

Behind an aggregate, there is a great amount of data which cannot be ignored in generating the forecasting results which economic agents need for the designing of economic policy measures, investment decisions, etc. For instance, the component prices in an aggregate price index. The aggregate trend behaviour could be mainly due to the effects of (a) monetary policy, (b) the incorporation of technical changes, (c) changes in consumer preferences or (d) supply constraints in different markets. If prices are analyzed at the full disaggregation level it is common to find that several prices share some features, but others not, because they, for instance, are affected by technological changes in a particular way or because they are affected in a different form by changes in preferences. The idea is that in examples like this one we can take as a valid hypothesis that certain set of components of an aggregate share a common feature but others do not. Consequently, it seems convenient to use disaggregated information in econometric modelling and provide the decision makers with an estimation of the future which refers to the aggregate and its components. A forecast of 1.5% headline inflation next year does not mean the same if all the price components are going to grow around this rate or if the rate of growth of energy prices is forecast at 15% and the other prices at minus 0.5%. Thus our approach advocates, at least at the initial level of a study, for the consideration of all the components of an aggregate, say n, which we called basic components, and the paper is focused on providing joint consistent forecasts for the aggregate and its components and possibly for some useful intermediate aggregates. In order to validate our proposal, one would have to show that the indirect forecast of the aggregate is at least as accurate as the direct one. When this happens it provides a third reason for disaggregated analysis.

The literature in this area of research considers three approaches. I) The direct one, which only works with the time series for the aggregate, II) the multivariate disaggregated approach and III) the disaggregate procedure based on univariate models for each one of the basic components. A fourth alternative, IV), developed in this paper is a disaggregated approach based on single-equation models for the components, which take into account common stable features which some components share between them.

Theory shows that when the data generation process (DGP) is known the forecasting accuracy from II) is better or equal than the forecasting accuracy of the other procedures. Nevertheless, if the number of components is large, as it is usually the case when one works with basic components, then approach II) is unfeasible, and in any case it would be very much affected by the problem of estimation uncertainty, which we comment below. On the other hand approach III) can be better or worse than I) depending on the properties of the data. Since restrictions between the components is one of the main reasons why the disaggregated approach could matter, in this paper we develop an intermediate approach between I), II) and III) based on single-equation models which take into account, when this is the case, the important restrictions between the components coming from the fact that some of them share common features. The approach is kept simple because identification of a unique common feature in a sub-set of components can be done by bivariate methods.

Our procedure differs from the dynamic factor literature by the fact that we always consider the possibility of common trends and look for common features (factors) analyzing the behaviour of each one of the variables, the basic components of an aggregate, and estimating only common features –estimation restriction- between the components which really share them. Then each factor is used as a restriction only in modelling the behaviour of the components which have the corresponding common features –forecasting restriction. At the same time the procedure is very restrictive in applying statistical tests for identifying common features and requires that they are stable. In dynamic factors literature all elements are considered to estimate a common factor without the above estimated restriction, leaving for the estimation process the finding of which components enter with practically zero weights. If the estimation restriction is right, the common factors (features) in our procedure could be estimated more precisely in small samples and could hopefully have also a more direct economic interpretation.

Recently Hendry and Hubrich (2006 and 2010), in what follows HH, have proposed a procedure to forecast the aggregate by means of a model for the aggregate which include as regressors its own lags and lags of the components. This procedure is especially relevant because in building the model the general-to-specific approach can be used. Thus, the selection of lags is done by applying the Autometrics procedure and in order to forecast at different horizons the model is specified and estimated by minimizing the sum of squares of forecasting errors at horizons 1, 2, etc. Our procedure differs from the HH procedure mainly in two respects. One refers to the fact it includes specific identified restrictions between the components in forecasting the aggregate, and HH incorporate some unknown restrictions, by including some but not all the components in the equation for the aggregate. A second difference consists in that our procedure also provides forecasts for the components, which is considered of interest since they could be required by decision makers. On the other hand, in building the single-equation models proposed in this paper the general-to-specific approach can also be applied.

The remaining of the paper is as follows. In section 2 we comment about the theoretical efficiency, estimation uncertainty and relevant restrictions in forecasting an aggregate. In section 3 we present our procedure based on single-equation models. In section 4 we apply our procedure to forecast inflation in the Euro Area, UK and US and we compare these results with those obtained from a direct forecast. In section 5 a comparison is done with results from models based on dynamic factors using the same sample and with the HH procedure using different samples. Section 6 faces the question of linking these forecasts with forecasts for the aggregate based on causal models and section 7 concludes.

2.- Theoretical efficiency, estimation uncertainty and relevant restrictions.

As mentioned above, theoretical results -for more details, see Rose (1977), Tiao and Guzman(1980), Wei and Abraham (1981), Kohn (1982) and Lütkepohl (1984) between others- for stationary variables have shown that, in general, approach II) will provide more accurate forecasts of the aggregate. Only under special conditions –Conditions for Efficiency of the Direct Forecast, CEDF- of the data the direct approach is efficient. Lütkepohl (1987) indicates that this happens for instance when the components are uncorrelated and have identical stochastic structures. This can be taken as a hint that when components have different distributions -for instance some are conditional heteroskedastic or have a conditional mean with a nonlinear structure- or there are cross restrictions between them, disaggregation could be important. In this paper we limit ourselves to considering the case in which there are restrictions between the components. This does not mean that we believe that distributional differences are less important, but merely that we start to study the problem in a way which is easier to solve in a general framework.

In the appendix we show that for aggregates I(1) with cointegrated components and with components with a common serial correlation factor in their stationary transformations, CEDF applies only under very specific weighting vectors in the aggregation scheme. When the CEDF does not apply the direct forecasting approach imposes invalid restrictions in the DGP -all the available data, the basic components. Consequently, we can add that in order to avoid invalid restrictions one needs to start working from the basic components. These results confirm the intuition that common trends, and in general common features, in disaggregates are types of restrictions which could be worth exploiting in indirect forecasting. But in practice, the DGP is unknown and the efficiency result depends then on the specification and estimation problems. Large disaggregation levels (basic components) include a vast number of components and increase the estimation uncertainty and a loss of forecast accuracy in terms of MSFE is likely. Thus, a large disaggregation may not constitute, at least with the current econometric techniques, a realistic alternative to the direct forecast of an aggregate. But full disaggregation information provides a deeper understanding of the more important statistical properties of the DGP, as common features between the components, which could lead to the design of an indirect forecast strategy based on a disaggregation map which could include some basic components and intermediate aggregates. One could expect that this indirect forecast would not perform worse than the direct one, since it captures relevant restrictions in the data and keeps the procedure relatively simple using, as it will explain later, single-equation models and, possibly, intermediate aggregates.

3.- Our procedure.-

Bearing in mind the above considerations, we propose a simple procedure, which starting from the set of basic components, configures a disaggregation map by using selected common features to group the basic components in subsets such that all the basic components of a subset share a common feature. Usually the common features will be a common trend or a common serial correlation factor (CSCF), but this can be enlarged to consider two common trends, as it will become clear later. In order to keep the procedure simple it is important to formulate the strategy for detecting common features by using only bivariate methods. Also a residual group, R, is formulated with all the basic components which do not share any of the selected properties.

The disaggregation map can be used for constructing forecasting models of dimension smaller than n by using some intermediate sub-aggregates. The forecasting models will be formulated as dynamic single-equation models for each one of the components of a group of basic components in the disaggregation map. A specific model will include transitory dynamics in terms of lags of the variables and, when it is the case, the restriction coming from the common features. Alternatively, these single-equation models can be built for the sub-aggregate formed with the basic elements of a group in the disaggregation map. Which alternative to follow is something which can be determined empirically.

The above approach is not popular because often there are over a hundred basic components. This is why many papers dealing with disaggregation have been limited to considering an official breakdown of the aggregate into just a few components, usually five in the case of euro area inflation as proposed in Espasa et al.(1987). 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 is huge but it can be done by the 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.

In our procedure we distinguish three phases. 1) selection of relevant constraints; 2) identification of a grouping map with non-overlapping subsets of basic components with different restricting definitions in each subset; and 3) building forecasting single-equation models for the elements of the grouping map, working with each one of the basic elements in each group or with the sub-aggregate obtained with them. The elements of each group could have a common trend and a CSCF, therefore the general structure for the single-equation model of a basic component x_{it} is:

- (1) A constant and possible deterministic terms mainly related to seasonality.
- (2) A cointegration restriction between x_{it} and a common trend.
- (3) The CSCF.
- (4) Lags of the first differences of x_{it}, and of relevant intermediate sub-aggregates.

The terms (1) to (3) are only present when they apply. In what follows the sub-sets of basic components which could have overlapping elements will be denoted by a term starting by S and by a term starting by B when they do not have overlapping components. In both cases the intermediate sub-aggregate obtained from the aggregation of the elements of the corresponding subset will be denoted by adding an A at the beginning of the name of the subset.

Thus the procedure operates as follows: 1) select the types of the possibly most relevant constraints in the components which, in each case, initially involve just a single factor in a subset of components. In the paper we propose two types of constraints: (PT) the presence of a common trend between a subset of basic components and (PC) the presence of a common serial correlation factor 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 CSCF. The elements in both subsets must be identified, verifying that the constraint in question is stable over time for each element.

To identify the elements in S1 we start by testing for cointegration in all possible pairs of elements coming from the basic components and selecting the largest set, say S0, of basic components in which all pairs are cointegrated. The cointegration tests are performed using a very restrictive criterion for ending up with the presence of bivariate 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 cointegreted and zero otherwise. The elements of this matrix can be arranged in such a way that in the upper left corner of the matrix we find the largest m0 x m0 submatrix full of ones, meaning that in these m0 basic components there exist (m0-1) cointegration restrictions – full cointegration- , or equivalently that these components share the same stochastic trend, see charts 1,2 and 3.

Furthermore, the identification of the elements must be done checking that for each element the restriction in question is stable along the sample. For this purpose a sub-aggregate, AS0, with all the elements of S0 is built. Each element of S0 must be cointegrated with AS0 and the stability of this restriction is studied by estimating and testing cointegration across the sample by means of a rolling window. The elements of S0 which do not pass the stability test are removed from S0 and the resulting set is denoted by $S0^1$.

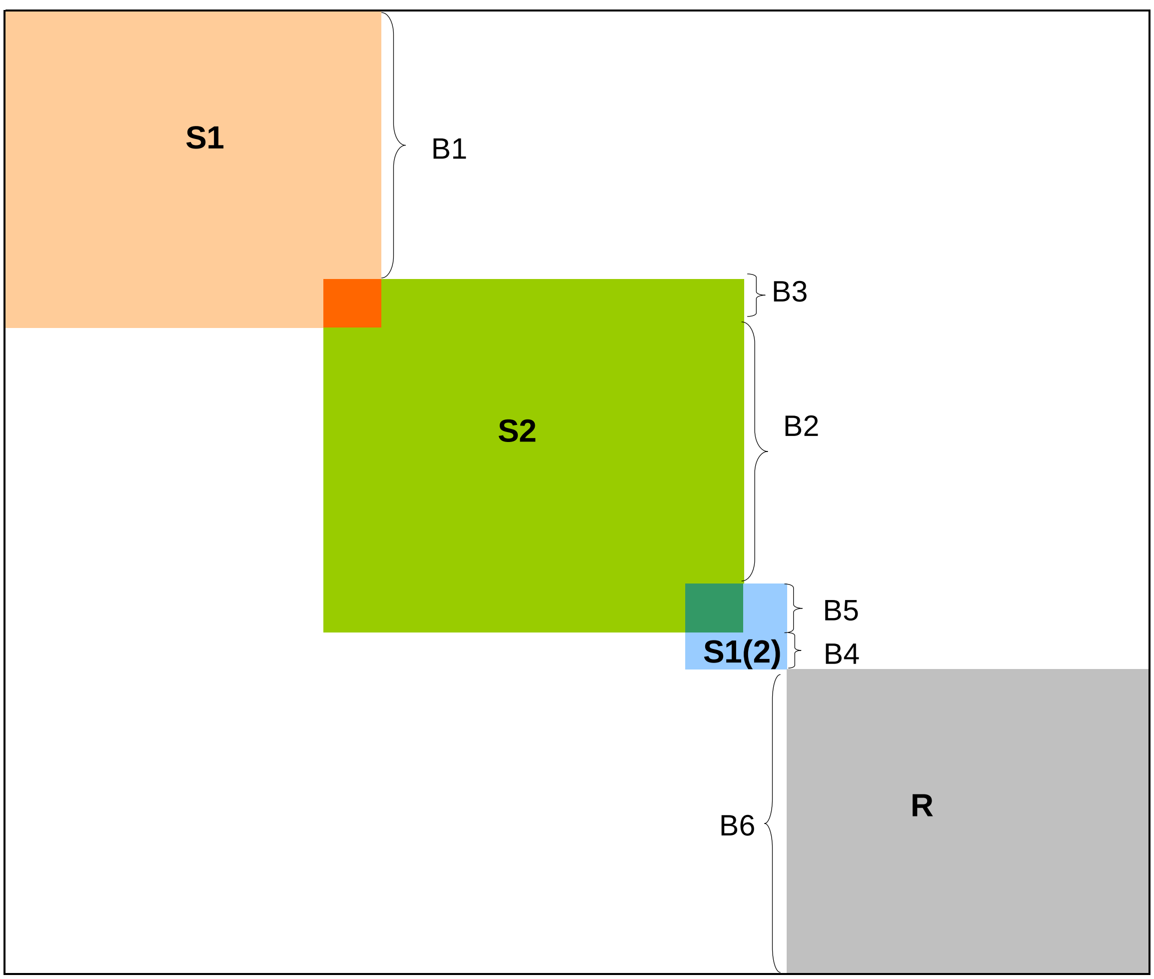
In order to explore a possible enlargement of $S0^1$ we considered the rest of elements out of S0 as potential candidates and a bivariate cointegration test between each of them and the new aggregate $AS0^1$ is performed. If some elements are cointegrated with this $AS0^1$ then $S0^1$ is enlarged with them and we call $S1^0$ to this new set and with its elements a sub-aggregate $AS1^0$ is constructed. Finally, bi-variant cointegration and stability tests are performed between each component of $S1^0$ and $AS1^0$ and those components which do not pass the test are removed from $S1^0$ resulting a new set denoted S1 and from it the sub-aggregate AS1 is defined. At the end of the process we have a set of, say m, basic components which are fully cointegrated (they share the same trend) in a stable sense along the sample and a set of the remaining components. The cointegration tests are performed using the procedure in Engle and Granger (1987).

The elements of S2 can be identified in a similar way but testing for CSCF following Engle and Kozicki (1993), and with them the sub-aggregate AS2 is formed. $CSCF_t = (\sum_{j=1}^{m} \rho_j \Delta AS2_{t-j})$, where the right hand side term is the fitted value in an ARIMA model for $\Delta AS2_t$.

In what follows we consider that we only have subsets S1 and S2. Obviously the procedure could be extended to consider the largest subset of basic elements outside S1 which share a common trend and, similarly, the largest subset of basic elements outside S2 which share a CSCF. But in the applications done for this paper the sizes of theses subsets are very small and we do not considered them in this paper.

The procedure can be extended to identify another new subset of basic components outside S1 which share two common trends – denoted as S1(2). As a proxy for one of the two common trends we use AS1 and then each element outside S1 can be enlisted by testing for cointegration in triplets formed from by AS1 and all possible pairs of elements coming from the basic components outside S1. The procedure keeps going in a similar way to the one followed to detect the stable single common trend in S1. Thus we end up with a subset S1(2) of basic components which share two common trends and from it the corresponding subaggregate AS1(2) can be obtained. By construction there is no intersection between S1 and S1(2). For the common trend in S1 we use as a proxy AS1. The subsets S1, S1(2) and S2 can have overlapping elements. To avoid that, we define the following subsets. B1, defined by the above elements which only share a common trend; B2, in which the elements are those which share just a CSCF; B3, formed by the elements which share a common trend and a CSCF; B4 defined by the elements which share the two common trends; B5 defined by the elements which share the two common trends and CSCF and R which includes the basic elements outside the union of the previous subsets. Then aggregating all the elements in the respective subset, B1, B2, B3, B4, B5 and R and we obtain sub-aggregates AB1, AB2, AB3, AB4, AB5 and AR. In chart 1 it can be seen the estimated composition of set S and B for the euro area.

Chart 1: Classification of the basic components in the euro area.



3) Forecast models for the basic components or intermediate sub-aggregates from B1, B2, B3, B4 and B5.

The elements of B1 could be forecast by single-equation equilibrium correction models, which could include lags of the differences of the element in question, of AB1 and of AR, and perhaps of the other intermediate sub-aggregates, and the equilibrium correction term with respect AB1.

$$\Delta x_{it} = \mu + \sum_{j=1}^{12} \delta_{ij} D_{jt} + \sum_{j=1}^{s} \left(\lambda_{i,j} \Delta x_{i,t-j} \right) + \alpha_i (x_{i,t-1} - \beta_i AS1_{t-1}) + + \sum_{j=1}^{p} \rho_{i,j} \Delta AS1_{t-j} + \sum_{j=1}^{q} \gamma_{i,j} \Delta AR^{*}_{t-j} + e_{it} \qquad \forall i \in B1 \tag{1}$$

Where μ, D_{jt} are deterministic terms and $\lambda_{ij}, \alpha_i, \rho_{i,j}$ and $\gamma_{i,j}$ are coefficients.

For an element of B2, we use a model with the CSCF. To estimate CSCF we fit an ARIMA model to ΔAS2 and we approximate CSCF by $CSCF_t = (\Delta AS2 - a_t)$, where a_t are the residuals of the ARIMA model, and we denote the approximation by ΔAS2F. The models for the elements of B2 include ΔAS2F and lags of the differences of the dependent variable and of the intermediate sub-aggregates as above.

$$\Delta x_{it} = \mu + \sum_{j=1}^{12} \delta_{ij} D_{jt} + \alpha_{1,i} (CSC_t) + \sum_{j=1}^{s} \left(\lambda_{i,j} \Delta x_{i,t-j} \right) + \sum_{j=1}^{p} \rho_{i,j} \Delta AS2_{t-j} + + \sum_{j=1}^{q} \gamma_{i,j} \Delta AR^{*}_{t-j} + e_{it} \qquad \forall i \in B2 \tag{2}$$

For the elements of B3 and B5, we use a model like (2) but adding the corresponding equilibrium correction term, which will be with respect AS1 for the elements of B3, see equation (3) and with respect AS1 and AS1(2) for the elements of B5, see equation (5).

$$\Delta x_{it} = \mu + \sum_{j=1}^{12} \delta_{ij} D_{jt} + \alpha_{1,i}(CSC_t) + \alpha_{2,i}(x_{i,t-1} - \beta_i AS1_{t-1}) + \sum_{j=1}^{s} \left(\phi_{i,j} \Delta x_{i,t-j}\right) +$$

$$+ \sum_{j=1}^{r} \rho_{i,j} \Delta AS1_{t-j} + \sum_{j=1}^{p} \lambda_{i,j} \Delta AS2_{t-j} + + \sum_{j=1}^{q} \gamma_{i,j} \Delta AR^{*}_{t-j} + e_{it} \qquad \forall i \in B3 \qquad (3)$$

The elements of B4 are forecast by single-equation equilibrium correction models which could include lags in differences of the element in question, of the intermediate sub-aggregates.

$$\Delta x_{it} = \mu + \sum_{j=1}^{12} \delta_{ij} D_{jt} + \alpha_{1i}(x_{i,t-1} - \beta_{2i} AS1_{t-1} - \beta_{3i} AS1(2)_{t-1}) + \sum_{j=1}^{s} \left(\phi_{i,j} \Delta x_{i,t-j}\right) +$$

$$+ \sum_{j=1}^{r} \rho_{i,j} \Delta AS1_{t-j} + \sum_{j=1}^{s} \tau_{i,j} \Delta AS1(2)_{t-j} + \sum_{j=1}^{q} \gamma_{i,j} \Delta AR^{*}_{t-j} + e_{it} \qquad \forall i \in B4 \qquad (4)$$

For the elements of B5 the models are formulated as:

$$\Delta x_{it} = \mu + \sum_{j=1}^{12} \delta_{ij} D_{jt} + \alpha_{1,i}(CSC_t) + \alpha_{2i}(x_{i,t-1} - \beta_{2i} AS1_{t-1} - \beta_{3i} AS1(2)_{t-1}) +$$

$$+ \sum_{j=1}^{s} \left(\phi_{i,j} \Delta x_{i,t-j}\right) + \sum_{j=1}^{r} \rho_{i,j} \Delta AS1_{t-j} + \sum_{j=1}^{p} \lambda_{i,j} \Delta AS2_{t-j} + \sum_{j=1}^{s} \tau_{i,j} \Delta AS1(2)_{t-j} +$$

$$+ \sum_{j=1}^{q} \gamma_{i,j} \Delta AR^{*}_{t-j} + e_{it} \qquad \forall i \in B5 \qquad (5)$$

In order to forecast the elements in R we use ARMA models for ΔR enlarged with lags of the first differences of AS1, AS2 and AS1(2).

$$\Delta AR_{it} = \mu + \sum_{j=1}^{12} \delta_{ij} D_{jt} + \sum_{j=1}^{r} \rho_{i,j} \Delta AB1_{t-j} + \sum_{j=1}^{s} \tau_{i,j} \Delta AS1(2)_{t-j} + \sum_{j=1}^{p} \lambda_{i,j} \Delta AS2_{t-j} +$$

$$\sum_{j=1}^{q} \gamma_{i,j} \Delta AR_{t-j} + e_{it} \qquad \forall i \in R \qquad (6)$$

In this case is especially important to check how AR is forecast best, by the direct or indirect procedure and used the preferred one. If the former is best, it implies that in this case working with an intermediate aggregate, AR, is not worthy. This could happen often since the elements in R are not related by the constraints considered in this paper. In the other subsets B1, B2, B3, B4 and B5 it could also be useful to test if the corresponding sub-aggregates are forecast best directly or indirectly and proceed consequently. In all the cases in which a sub-aggregate is forecast best directly and forecasts of the corresponding elements are also required, then the individual elements should be forecast using the above models and imposing the restriction derived from the direct forecast.

With the above models we obtain forecasts of all the basic components and evidence of which sub-aggregates if any are forecast best directly. Thus the forecast for the overall aggregate is obtained aggregating the forecasts of the sub-aggregates which are forecast best directly and the forecasts of the remaining basic components.

With information up to time n, the forecasts are calculated sequentially for a path of h data points. The models are re-estimated each time that the base of the forecasts changes.

4.- Forecasting results for inflation in the euro area, US and UK.

In this section we forecast inflation, defined in terms of consumer prices, CPI, for US and the Harmonized Index of Consumer Prices (HICP) for the euro area and UK.

The official breakdown of HICP mentioned in section 3 contains five categories: energy (ENE), Non-processed food (NPF), Processed food (PF), Non-energy industrial goods (MAN) and Services (SERV) and in reporting the results for the basic components for the euro area, US and UK we classify them using the mentioned categories, even when in some cases the correspondence is not perfect.

The forecasting exercise for each area compares the performance of different procedures against the performance of the direct procedure using an ARIMA model. For subset R the direct forecast of the sub-aggregate AR usually performs better than using an ARIMA indirect approach. Consequently we report the results for the three areas using a direct ARIMA model for AR, possibly with lags from the other sub-aggregates . The forecast exercises have been conducted on the year-on-year rate of growth of the price indexes.

The forecasting procedures considered with the type of models used in each case are the following ones (all models include the appropriate sets of seasonal dummies if they are required):

Direct procedure

P1 ARIMA model for the aggregate

Indirect procedures based on intermediate disaggregations considered in this paper

P2 Procedure using just a common trend restriction.
Model (1) for the elements of S1 and model (6) for the sub-aggregate formed with the remaining elements.

P3 Procedure using just a CSCF restriction.
Model (2) for the elements of S2 and model (6) for the sub-aggregate formed with the remaining elements.

P4 Procedure using a common trend and a CSCF restrictions.
Model (1) for the elements of B1, model (3) for the elements of B3, model (2) for the remaining elements of S2 and model (6) for the sub-aggregate formed with the elements of HICP outside S1 and S2.

P5 Procedure using one and two common trends restrictions and a CSCF.
Model (j) for the elements in Bj, for j=1,..., 5; and model (6) for the sub-aggregate formed with the elements of HICP outside S1,S2 and S1(2).
.

P5AG Like P5 but replacing the forecasts of the elements of a given subset Bj, j=1,...,5 by the direct forecast of the corresponding intermediate sub-aggregate.

Indirect procedure based on ARIMA models

P6 ARIMA models for each basic component.

The samples used go from 1995:01 to 2009:12 for the euro area and UK and from 1999:01 to 2009:12 for US. The first years till 2003 are used to estimate the models and the following six years to evaluate their forecasting performance, estimating the models with an enlarged sample each time that the base of the forecasts changes. The results are reported in different tables below given the RMSFE for the direct procedure in the first column of results and the ratio between the RMSFE of this procedure and the RMSFE of each one of the other specific procedures in the remaining columns. In these tables one star indicates that the difference in RMSFE's between the direct procedure and the one of the corresponding column is significantly significant at 5% level following the Diebold-Mariano test and two stars that such difference is significant at 1% level.

4.1 Euro area.

The data used corresponds to a breakdown of the HICP in 79 basic components as reported in appendix 3. For greater breakdowns the sample size available is too small.
The number of basic components with a common trend (subset S1) is 26 and except price for vegetables all of them belong to MAN and SERV. S1 weights 39.5% in HICP, see table 1. Given the nature of products whose prices are in S1 it seems that the underlying common trend could be associated mainly to monetary policy. The number of prices with a CSCF (a common cycle) is 28, the corresponding set S2 weights 26.5% in HICP and its elements are more distributed along the official categories and among the products, see table 2. Set S2 includes prices of food, fuels, heat energy, services related with transport and tourism, nondurable household goods, equipment for sport and goods related with new technologies. The number of elements with two common trends, set S1(2), is seven and the set weights 4.12% in the HICP. Its elements refer mainly to fuels and transport. The set R of elements which do not share any of the common features considered has 25 elements and weights 34.18% in the HICP. A summary of these results are in table 1, see also chart 1 above, and table 2 shows the results by official categories.

Table 1: Composition of subsets of basic components in the euro area taken into account common trends and a CSCF.

	Number of Components	Weight in the official category Index
One common trend	26	39.51%
Only common cycle	21	22.18%
Common cycle and a common trend	(3)	(2.16%)
Common cycle and two common trends	(4)	(2.18%)
Two common trends	7	4.12%
Residual	25	34.18%
Total	79	100.00%

Table 2:

Classification by official categories of the basic components with a common trend (subset S1) in the euro area.

	NPF	ENE	PF	MAN	SERV	TOTAL
Number of Components	1	0	0	11	14	26
Weight in the subset S1	3.9%	0.0%	0.0%	27.7%	68.4%	100.0%
Weight in the official category Index	1.6%	0.0%	0.0%	10.9%	27.0%	39.5%

Classification by official categories of the basic components with a common serial correlation factor (subset S2) in the euro area.

	NPF	ENE	PF	MAN	SERV	TOTAL
Number of Components	2	4	8	8	6	28
Weight in the subset S2	17.9%	18.4%	32.8%	18.1%	12.8%	100.0%
Weight in the official category Index	4.7%	4.9%	8.7%	4.8%	3.4%	26.5%

Classification by official categories of the basic components with two common trends (subset S1(2)) in the euro area.

	NPF	ENE	PF	MAN	SERV	TOTAL
Number of Components	1	2	0	2	2	7
Weight in the set S1(2)	26.9%	16.7%	0.0%	33.6%	22.8%	100.0%
Weight in the official category Index	1.1%	0.7%	0.0%	1.4%	0.9%	4.1%

A simple preliminary analysis shows, as pointed out in Espasa and Albacete (2007), two different seasonal regimes during the sample period. The first lasts till December 2000, and the second one from January 2001 onwards. We model the break in the seasonal structure by including two sets of seasonal dummies, one for each regime, when they are required.

The RMSFE for the direct procedure (1) is 0.17% for horizon one and 0.81% for horizon twelve (see column 1 on table 3). The procedure P5 proposed in this paper outperforms the direct ARIMA forecasts from horizon three onwards and the differences are statistically significant, using the Dieblod and Mariano test, from horizon fourth onwards. Also the importance of using procedure P5 grows with the horizon of the forecast. Thus the ratio of the RMSFE's from the ARIMA direct model and P5 goes from 1.00 for horizon one to 0.82 for horizon twelve. In this case the individual modelling of the components of the subset S2 with a CSCF restriction (compare columns P2 and P3 in table 3) and of the components of subset S1(2), which singles out some energy and transport prices as having two common factors (compare columns P4 and P5 in table 3) turn to be important.

For the euro area the indirect ARIMA procedure, P6, performs worse than the direct approach from horizons 6 onwards, pointing the importance of using restrictions between the basic elements in indirect procedures.

Table 3: RMSFE of the direct procedure P1 and ratios of the RMSFE`s between P1 and each one of the other forecasting procedures.
Euro area, year-on-year inflation rate.

	DIRECT PROCEDURE	INDIRECT PROCEDURES BASED ON INTERMEDIATE DISAGGREGATION CONSIDERED IN THE PAPER									INDIRECT ARIMA MODELS
Periods	**P1**	**P2**		**P3**		**P4**		**P5**		**P5AG**	**P6**
1	0.17	0.99		1.00		1.00		1.00		1.47	0.89
2	0.27	0.97		0.99		0.94	*	1.00		1.35	0.93
3	0.34	0.94	*	0.96	*	0.96	*	0.98		1.25	0.95
4	0.40	0.95	*	0.91	*	0.99	*	0.94	*	1.12	0.98
5	0.46	0.96		0.90	**	0.93	**	0.90	**	1.05	0.99
6	0.50	0.97		0.91	**	0.94	**	0.90	**	1.00	1.02
7	0.55	0.98		0.89	**	0.93	**	0.87	**	1.03	1.03
8	0.60	0.99		0.87	**	0.93	**	0.86	**	1.04	1.04
9	0.65	1.00		0.87	**	0.95	**	0.86	**	1.02	1.06
10	0.70	1.00		0.85	**	0.94	**	0.85	**	0.97	1.08
11	0.76	1.00		0.83		0.94	*	0.83	*	0.93	1.07
12	0.81	1.00		0.84		0.94	*	0.82	*	0.91	1.08

Sample: 1995/01-2009/12

Forecast Sample: 2004/01-2009/12

* 95%
** 99%

Table 4 Diebold–Mariano test using a multivariate loss function for the path forecast between two procedures (Capistrán 2006).
Euro area results

	DIRECT PROCEDURE	INDIRECT PROCEDURES BASED ON INTERMEDIATE DISAGGREGATION CONSIDERED IN THE PAPER					DYNAMIC FACTORS			INDIRECT ARIMA MODELS
	P1	P2	P3	P4	P5	P5AG	DF1S	DF1NS	DF2	P6
P1			**		**					
P2			**		**					
P3										
P4			**		**					
P5										
P5AG			**		**				**	
DF1S	**	**	**		**				*	
DF1NS	**	**	**	*	**	**	**		**	**
DF2			**		**					
P6		**	**		**					

* 95%
** 99%

In table 4 we report the Diebold–Mariano test using a multivariate loss function proposed by Capistrán (2006) to test jointly the forecast accuracy between two procedures along the 12 horizons. This table also contains results corresponding to procedures based on dynamic factors which will be discussed below. The table shows that our procedure P5 outperforms significantly all others except procedure P3 which is also outperformed but not in a significant way. Procedure P3 is restricted to consider only individual forecasts for price indexes sharing a CSCF and the results in table 4 show the importance of these restrictions when forecasting inflation in the euro area.

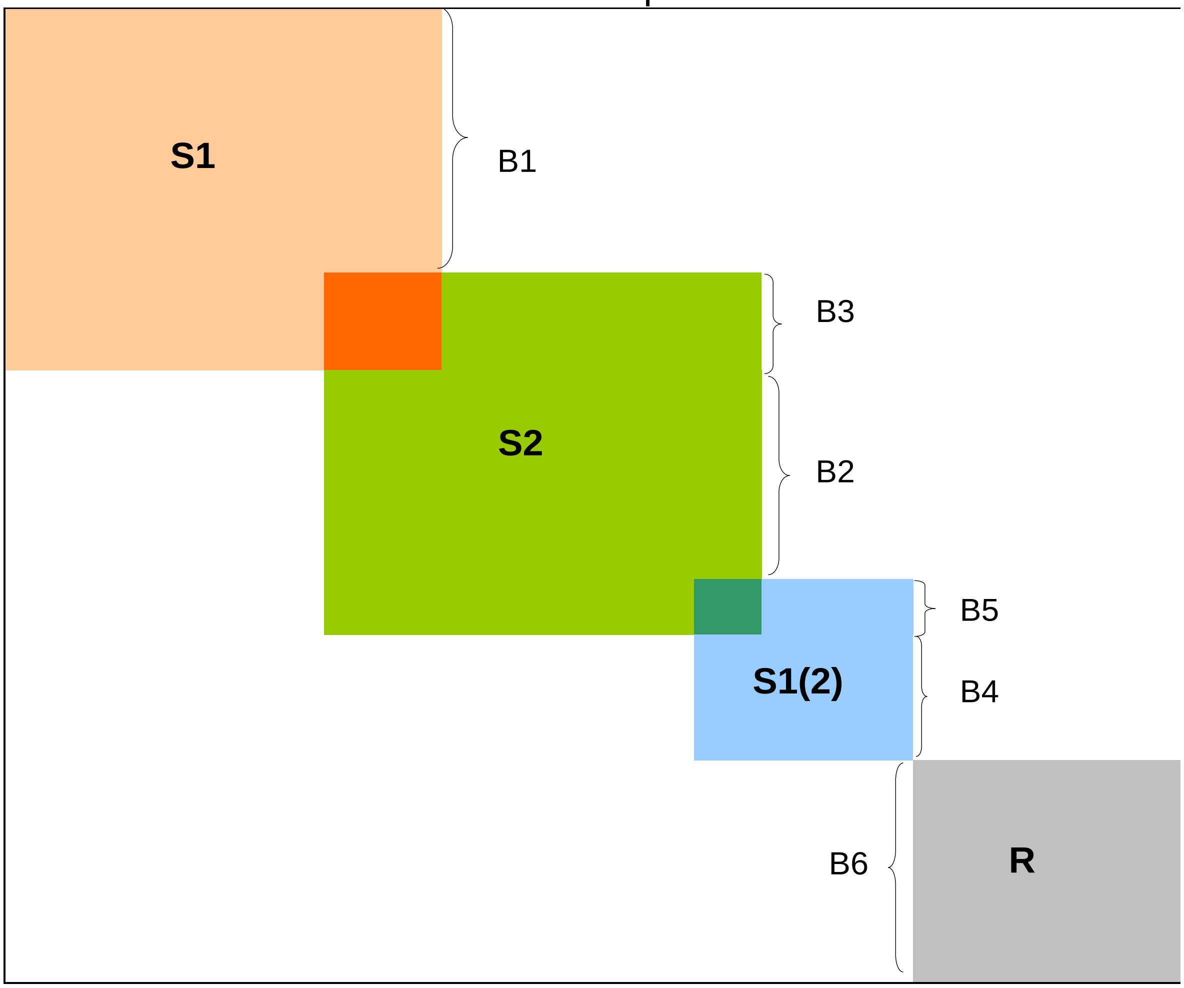
4.2 UK

In this case the data only require one set of seasonal dummies. The number of basic elements is 70. Searching for common features we found that 26 basic components (39.3%) make up the largest set of basic components with a common trend, 24 basic components (36.5%) make up the largest set of elements with a CSCF, 14 basic components (24.30%) share two common trends and 18 (22.60%) do not share any of the above common features. A summary of these results are in table 5 and chart 2 and table 6 shows the results by official categories.

Table 5 Composition of subsets of basic components in UK taken into account common trends and a CSCF.

	Number of Components	Weight in the official category Index
One common trend	26	39.30%
Only common cycle	12	13.80%
Common cycle and a common trend	(5)	(6.10%)
Common cycle and two common trends	(7)	(16.60%)
Two common trends	14	24.30%
Residual	18	22.60%
Total	70	100.00%

Chart 2: Classification of the basic components in the UK.



The RMSFE from the aggregate ARIMA model is 0.26% for horizon one and 1.64% for horizon twelve (see column 1 on table 7), therefore the forecasting uncertainty in this case is considerably greater than in the euro area. The forecasting procedure P5 outperforms P1 from horizon one onwards with statistically significant differences in all of them but lags 2, 3 and 4. In this case P5 performs worse than P4, which do not single out the basic components identified as having two common trends, and also worse than P3 which only considered individually the basic components with a CSCF. Nevertheless the classification of the basic components used in P5 seems relevant, because based on it the procedure P5AG, which uses the sub-aggregates B1 to B5 and R to forecast the aggregate, turns to be best one from horizon 3. For UK procedure P4 also performs very well, but better than P5AG only in the first two horizons. In the tests for the global forecasting path given in table 8, P5AG outperforms significantly all others except procedure P4, while P4 only outperforms P1 and P6.

Table 6:

Classification by official categories of the basic components with a common trend (subset S1) in the UK.

	NPF	ENE	PF	MAN	SERV	TOTAL
Number of Components	1	1	1	15	8	26
Weight in the subset S1	4.07%	11.70%	3.31%	51.40%	29.52%	100%
Weight in the official category Index	1.60%	4.60%	1.30%	20.20%	11.60%	39.3%

Classification by official categories of the basic components with a common serial correlation factor (subset S2) in the UK.

	NPF	ENE	PF	MAN	SERV	TOTAL
Number of Components	1	1	5	10	7	24
Weight in the subset S2	2.7%	9.3%	15.6%	26.0%	46.3%	100%
Weight in the official category Index	1.0%	3.4%	5.7%	9.5%	16.9%	36.5%

Classification by official categories of the basic components with two common trends (subset S1(2)) in the UK.

	NPF	ENE	PF	MAN	SERV	TOTAL
Number of Components	1	1	2	5	5	14
Weight in the subset S1(2)	4.12%	13.99%	3.29%	13.58%	65.02%	100%
Weight in the official category Index	1.00%	3.40%	0.80%	3.30%	15.80%	24.3%

Table 7:RMSFE of the direct procedure P1 and ratios of the RMSFE`s between P1 and each one of the other forecasting procedures.
…………..United Kingdom, year-on-year inflation rate.

	DIRECT PROCEDURE	INDIRECT PROCEDURES BASED ON INTERMEDIATE DISAGGREGATION CONSIDERED IN THE PAPER										INDIRECT ARIMA MODELS
Periods	**P1**	**P2**		**P3**		**P4**		**P5**		**P5AG**		**P6**
1	0.26	0.94		0.92	*	0.93	*	0.90	*	1.25		0.97
2	0.41	0.93		0.89		0.91	**	0.89		1.02		0.95
3	0.55	0.93		0.91		0.92	**	0.93		0.89		1.07
4	0.69	0.90		0.91		0.91	**	0.93		0.85		1.06
5	0.83	0.88	**	0.88	*	0.88	*	0.90	**	0.83		1.03
6	0.96	0.88	**	0.88		0.86	*	0.90	**	0.81		1.04
7	1.08	0.89	**	0.89		0.86	**	0.92	**	0.81		1.03
8	1.19	0.89	**	0.89	*	0.87	**	0.92	**	0.81	*	1.03
9	1.31	0.89	**	0.89	**	0.87	*	0.92	**	0.80	**	1.02
10	1.43	0.89	**	0.88	**	0.86	*	0.92	**	0.80	**	1.02
11	1.54	0.89	**	0.88	**	0.86	*	0.92	**	0.80	**	1.01
12	1.64	0.89	**	0.88	**	0.85	*	0.92	**	0.80	**	0.97

Sample: 1995/01-2009/12 Forecast Sample: 2004/01-2009/12 * 95% ** 99%

Table 8 Diebold–Mariano test using a multivariate loss function for the path forecast between two procedures.

United Kingdom results

	DIRECT PROCEDURE	INDIRECT PROCEDURES BASED ON INTERMEDIATE DISAGGREGATION CONSIDERED IN THE PAPER					INDIRECT PROCEDURES BASED ON DYNAMIC FACTORS			INDIRECT ARIMA MODELS
	P1	P2	P3	P4	P5	P5AG	DF1S	DF1NS	DF2	P6
P1		**	**	**	*	**				
P2						**				
P3						**				
P4										
P5			**			**				
P5AG										
DF1S								**	**	
DF1NS										
DF2										
P6		**	**	**	*	**				

* 95%
** 99%

4.3 United States

For US, we were able to work with 160 basic components. We find that 30 of them (7.66%) make up the largest set of basic components with a common trend, 61 (68.7%) make up the largest set of elements with a CSCF, 32 (13.98%) share to common trends and 65 (22.96%) do not share any common feature. A summary of these results are in table 9 and chart 3 and table 10 shows the results by official categories.

Table 9 Composition of subsets of basic components in US taken into account common trends and a CSCF.

	Number of Components	Weight in the official category Index
One common trend	30	7.66%
Only common cycle	33	55.44%
Common cycle and a common trend	(16)	(6.08%)
Common cycle and two common trends	(12)	(7.18%)
Two common trends	32	13.98%
Residual	65	22.96%
Total	160	100.0%

Chart 3: Classification of the basic components in the US.

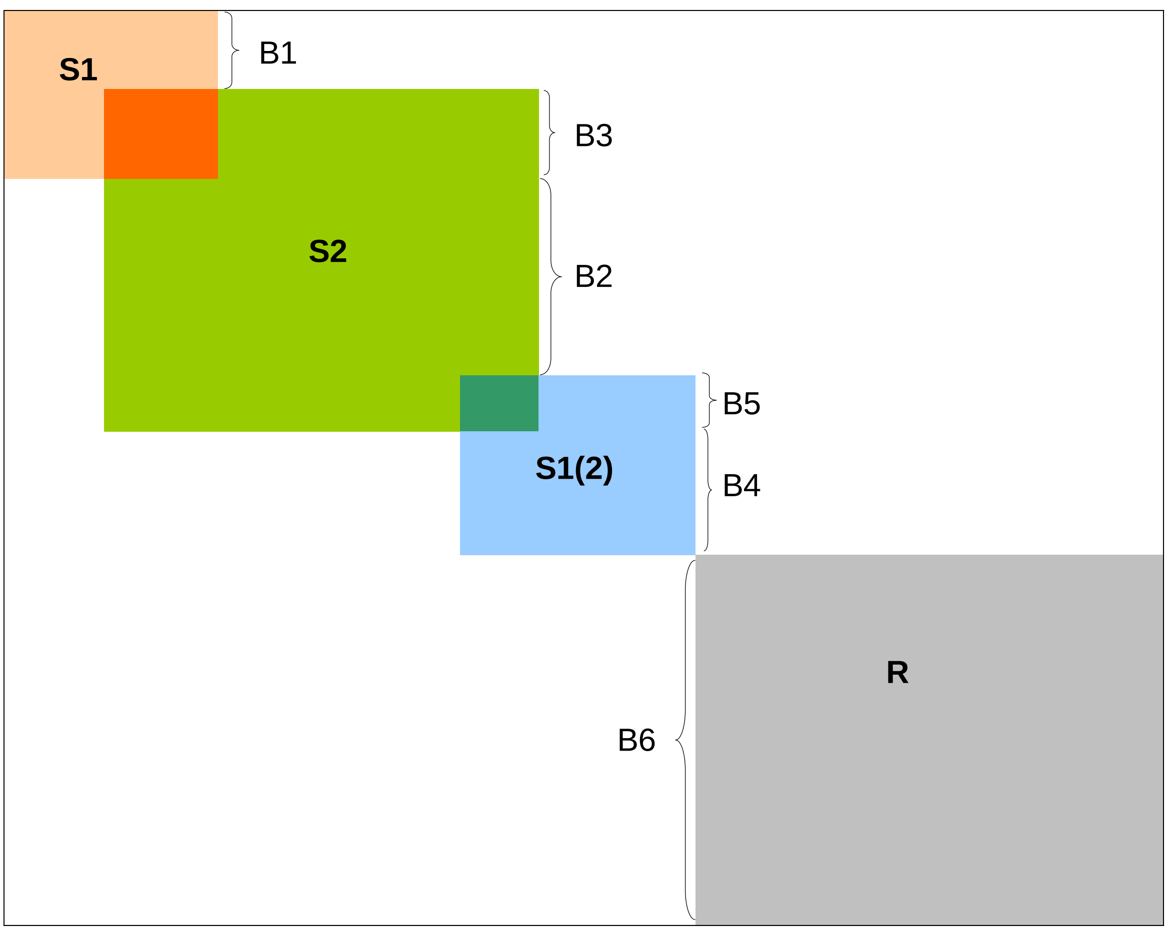


Table 10:

Classification by official categories of the basic components with a common trend (subset S1) in the US.

	NPF	ENE	PF	MAN	SERV	TOTAL
Number of Components	5	2	4	15	4	30
Weight in the subset S1	6.96%	17.56%	9.49%	37.91%	28.08%	100%
Weight in the official category Index	0.53%	1.35%	0.73%	2.90%	2.15%	7.7%

Classification by official categories of the basic components with a common serial correlation factor (subset S2) in the US.

	NPF	ENE	PF	MAN	SERV	TOTAL
Number of Components	8	5	12	21	15	61
Weight in the subset S2	1.84%	10.80%	4.24%	20.10%	63.03%	100%
Weight in the official category Index	1.26%	7.42%	2.91%	13.81%	43.30%	68.7%

Classification by official categories of the basic components with two common trends (subset S1(2)) in the US.

	NPF	ENE	PF	MAN	SERV	TOTAL
Number of Components	7	0	10	10	5	32
Weight in the subset S1(2)	4.83%	0.00%	10.94%	30.09%	54.14%	100%
Weight in the official category Index	0.68%	0.00%	1.53%	4.21%	7.57%	14%

Table 11: RMSFE of the direct procedure P1 and ratios of the RMSFE`s between P1 and each one of the other forecasting procedures.

United States, year-on-year inflation rate.

	DIRECT PROCEDURE	INDIRECT PROCEDURES BASED ON INTERMEDIATE DISAGGREGATION CONSIDERED IN THE PAPER										INDIRECT ARIMA MODELS
Periods	P1	P2		P3		P4		P5		P5AG		P6
1	0.57	1.03		1.03		0.84	**	1.00		0.99		1.04
2	1.13	0.98		1.05		0.88	**	1.03		1.06		1.00
3	1.49	0.91	**	0.96	*	0.91	**	0.94	*	1.04		1.01
4	1.71	0.90	**	0.94	**	0.94	*	0.93	**	1.09		1.01
5	1.90	0.89	**	0.93	**	0.91	**	0.92	**	1.10		1.01
6	2.03	0.89	**	0.93	**	0.91	**	0.92	**	1.10		1.01
7	2.10	0.90	**	0.93	**	0.91	**	0.92	**	1.09		1.01
8	2.18	0.89	**	0.93	**	0.91	**	0.92	**	1.09		1.01
9	2.26	0.87	**	0.89	**	0.88	**	0.89	**	1.07		1.01
10	2.32	0.89	**	0.91	**	0.86	**	0.91	**	1.07		1.01
11	2.35	0.91	**	0.90	**	0.84	**	0.92	**	1.08		1.03
12	2.40	0.93	**	0.92	**	0.82	**	0.93	**	1.09		1.02

Sample: 1999/01-2009/12 Forecast Sample: 2004/01-2009/12 * 95% ** 99%

Table 12 Diebold–Mariano test using a multivariate loss function for the path forecast between two procedures (Capistrán 2006).
United States results

	DIRECT PROCEDURE	INDIRECT PROCEDURES BASED ON INTERMEDIATE DISAGGREGATION CONSIDERED IN THE PAPER					INDIRECT PROCEDURES BASED ON DYNAMIC FACTORS			INDIRECT ARIMA MODELS
	P1	P2	P3	P4	P5	P5AG	DF1S	DF1NS	DF2	P6
P1		**	*	**	**					
P2										
P3										
P4										
P5										
P5AG	*	**	**	**	**					
DF1S	**	**	**	**	*					
DF1NS	**	**	**	**	*					*
DF2	**	**	**	**	**					*
P6		**	**	**	*					

* 95%
** 99%

The forecast of the US CPI is done by the same procedures than in the previous cases. The RMSFE from the direct ARIMA model is 0.57% for horizon one and 2.40% for horizon twelve (see column 1 on table 11). The procedure P5 outperforms significantly the direct procedure from horizon three but performs worse than P4, which does not single out the basic components identified as having two common trends. In this case then the method P4 is the recommended method.

In table 12 we report the tests for the forecast accuracy between pairs of procedures along the 12 horizons. The table shows that all the procedures P2 to P5 outperform significantly P1.

4.4 Main conclusions from the above applications.

1.- The indirect forecast based on ARIMA models for all the basic components (P6) cannot be considered a better procedure than the direct one. In fact, it performs worse than the direct procedure for a majority of horizons in the euro area and UK, and in all of them in US.

2.- On the contrary, indirect forecasting based on identifying subsets of basic components which share common trends and/or CSCF improves significantly the performance of the direct method. Thus, any of the indirect procedures, P3 to P5, based on the disaggregation map considered in this paper performs better than the direct method in the euro area and UK. For P2 this is also true for UK and US (tables 8 and 12). For the euro area P2 performs better than P1 in 8 horizons and in two of them significantly better. In the other four horizons its RMSFE is the same than in the direct procedure and for the whole path (table 4) P2 is not significantly better than P1.

3.- P5 is not the best method for UK and US, suggesting that in these cases the subset S1(2) could not be well identified to make profit of it in indirect forecasting.

4.- The paper provides evidence that it matters to distinguish basic components with a common trend or a CSCF from the rest, but it is not clear what matters more, the common trend or the CSCF. The consideration of only a subset with a CSCF and another for the rest of basic elements, procedure P3, is more relevant than the incorporation of only a subset with a common trend, procedure P2, for the case of the euro area, but for UK procedures P2 and P3 have similar performances and for US P2 performs better than P3 in nine of the twelve horizons considered.

5.- Our proposal in this paper of identifying six non-overlapping subsets of basic components depending on if they share or not common trends and a CSCF provides a good framework to propose an indirect forecasting strategy. It could be formulated as follows. Try the forecasting procedures P2 and P3 and test if they forecast significantly better than P1 applying Diebold-Mariano tests. If both of them do, as it is the case in the three applications above, consider procedures P4 and P5 and P5AG and choose the best one. If one of them does not improve significantly the direct method, the recommendation could be to use the other procedure.

6.- The disaggregation in six subsets leads to procedures which perform better than those just based in one subset with a common feature and the rest of components. The best results are obtained with P5 in the euro area, with P5AG for UK and with P4 for US.

7.- Our results show that between the direct procedure and the theoretically efficient one based on a vector model for all basic components –which usually is unfeasible and very often unreliable- the indirect procedure based on single-equations models for the basic elements which share some common features is an intermediate alternative which has been proved successful in forecasting inflation in three different economies. The key point for it seems to be the fact that the procedure incorporates important restrictions between the basic components. This in turns suggest that the question of how to use the disaggregate information to forecast the aggregate requires to consider the restrictions present in the disaggregate information.
In the case of UK the procedure P5 is outperformed by a method, P5AG, which, using the recommended disaggregation in six subsets of basic elements, forecast the aggregate by forecasting the corresponding sub-aggregates formed from those subsets. This results points out that in order to break down an aggregate in a small number of intermediate sub-aggregates it is important to construct them taking into account the main characteristics of the basic components. But the forecasting approach using intermediate sub-aggregates based on the proposed grouping map is not in general a good procedure as can be seen from the results corresponding to procedure P5AG for the euro area and US in tables 3 and 11.

Thus the analysis of the basic components turns to be important even when in occasions at the end we could not use them in forecasting the aggregate, but in defining useful intermediate sub-aggregates. In fact, a by-product of the procedure is the classification of a large number of components by restrictions shared between them, which could be also useful in other respects, as the application of dynamic factors, the definition of useful intermediate sub-aggregates or the formulation of models with unobserved components.

In this paper we have put the emphasis in trend and cyclical restrictions, but the consideration of subsets according the distributional properties of the basic components seems very relevant and could also have an important contribution in interval forecasting.

5 **Comparison with other procedures.**

5.1 Procedures based on dynamic factors.

The procedures developed in this paper are related to the literature on dynamic factors, which using jointly all the information set look for a reduced number of factors which could summarise it. Our procedure differs from this literature by the fact that we look for common features (factors) analyzing, by pairs, the behaviour of each one of the variables in the information set (basic components) with all the others, and by using each factor as a restriction only in the behaviour of the components which share it. In this way the common factors (features) could hopefully have also a more direct economic interpretation.
Thus the important differences between these two approaches can be described as follows. (a) In our approach one is interested in forecasting the aggregate and all its components –or at least some relevant intermediate sub-aggregates- in order to provide a full picture of the aggregate. (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 and a common cycle and, if convenient, for a second common trend. (d) The common features can be approximated by a sub-aggregate of the components and it facilitates the interpretation of the factors in economic terms. (e) As mentioned above 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 in dynamic factor applications, trying to get factors from each specific subset of basic components instead of deriving them from the whole data set. (f)

The classification of the basic components could also be useful for further economic analysis and for the definition of relevant intermediate sub-aggregates for economic policy. In fact, the paper suggests that the definition of intermediate aggregates is an endogenous question and provides statistical hints for their formulation. (g) Our procedure ensures common factor stability. (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, as discussed below, our procedure performs significantly better than dynamic factor analysis.

We have run a forecasting experiment using dynamic factors in the following way. The factors are obtained from the set of basic components and we obtain non-stationary dynamic factors from the logarithmic transformation of the data and stationary ones from the data on first differences of the logs. For each basic component in first differences we formulate a model which includes as explanatory variables its on lags, lags of the dynamic factors and seasonal dummies as in our models. In the experiment we have used up to four factors of each type, but the best results are obtained with just one factor. We denote by DF1NS the procedure in which the forecasting models only include a non-stationary dynamic factor, DF1S when the models include just one stationary factor and DF2 when they include one factor of each type. This strategy is similar to the one used by Duarte and Rua (2007). It should be mentioned that a direct forecasting method in which the dynamic factors are added to the ARIMA model for the aggregate, forecast in all the three areas worse than the models reported bellow with dynamic factors applied to the basic components.

The results using basic components and dynamic factors for the euro area are in table 13 and they show that our procedure P5 clearly outperforms the results from models with dynamic factors and table 4 shows that these differences are significant for the forecasting path as a whole. In fact the results with dynamic factors do worse than the direct method.

In the case of UK, table 14, dynamic factor procedures perform significantly better than the direct model for many horizons, but our procedure P5AG outperforms them from horizon three.

Finally for US, table 15, the procedures based on dynamic factors does not performed better than the direct method at any horizon, and certainly they perform clearly worse than procedures P2 to P5 and table 12 shows that the differences are significant for the forecasting path as a whole.

Table 13: RMSFE of the direct procedure P1 and ratios of the RMSFE`s between P1 and each one of the other forecasting procedures.
EA year-on-year inflation

Euro area results

	DIRECT PROCEDURE	INDIRECT PROCEDURES BASED ON INTERMEDIATE DISAGGREGATION CONSIDERED IN THE PAPER		INDIRECT PROCEDURES BASED ON DYNAMIC FACTORS			INDIRECT ARIMA MODELS
Periods	P1	P5		DF1NS	DF1S	DF2	P6
1	0.17	1.00		1.01	1.01	1.00	0.89
2	0.27	1.00		1.00	1.00	1.00	0.93
3	0.34	0.98		1.05	0.99	1.02	0.95
4	0.40	0.94	*	1.06	1.00	1.03	0.98
5	0.46	0.90	**	1.07	1.01	1.04	0.99
6	0.50	0.90	**	1.07	1.03	1.05	1.02
7	0.55	0.87	**	1.07	1.04	1.05	1.03
8	0.60	0.86	**	1.07	1.05	1.05	1.04
9	0.65	0.86	**	1.07	1.06	1.05	1.06
10	0.70	0.85	**	1.09	1.07	1.08	1.08
11	0.76	0.83	*	1.10	1.07	1.08	1.07
12	0.81	0.82	*	1.10	1.07	1.09	1.08

Sample: 1995/01-2009/12 Forecast Sample: 2004/01-2009/12 * 95% ** 99%

Table 14: RMSFE of the direct procedure P1 and ratios of the RMSFE`s between P1 and each one of the other forecasting procedures.

United Kingdom results

	DIRECT PROCEDURE	INDIRECT PROCEDURES BASED ON INTERMEDIATE DISAGGREGATION CONSIDERED IN THE PAPER				INDIRECT PROCEDURES BASED ON DYNAMIC FACTORS						INDIRECT ARIMA MODELS
Periods	**P1**	**P5**		**P5AG**		**DF1NS**		**DF1S**		**DF2**		**P6**
1	0.26	0.90	*	1.29		0.92	*	0.91	*	0.92	*	0.97
2	0.41	0.89		1.08		0.91		0.89		0.90	0.00	0.95
3	0.55	0.93		0.83		0.91		0.91		0.90	0.00	1.07
4	0.69	0.93		0.80		0.92		0.91		0.91	0.00	1.06
5	0.83	0.90	**	0.80		0.92	*	0.91	*	0.91	*	1.03
6	0.96	0.90	**	0.78		0.92	*	0.91	*	0.91	*	1.04
7	1.08	0.92	**	0.78		0.93	*	0.92	**	0.92	**	1.03
8	1.19	0.92	**	0.78	*	0.94	*	0.93	*	0.93	*	1.03
9	1.31	0.92	**	0.78	**	0.94		0.93	*	0.93	*	1.02
10	1.43	0.92	**	0.78	**	0.93	*	0.92	*	0.92	*	1.02
11	1.54	0.92	**	0.79	**	0.93	*	0.92	*	0.92	*	1.01
12	1.64	0.92	**	0.82	**	0.93	*	0.92	**	0.92	**	0.97

Sample: 1995/01-2009/12

Forecast Sample: 2004/01-2009/12

* 95%

** 99%

Table 15: RMSFE of the direct procedure P1 and ratios of the RMSFE`s between P1 and each one of the other forecasting procedures.

United States results

	DIRECT PROCEDURE	INDIRECT PROCEDURES BASED ON INTERMEDIATE DISAGGREGATION CONSIDERED IN THE PAPER		INDIRECT PROCEDURES BASED ON DYNAMIC FACTORS			INDIRECT ARIMA MODELS
Periods	P1	P4		TF1	ST1	T&SF1	P6
1	0.57	0.84	**	1.09	1.1	1.10	1.04
2	1.13	0.88	**	1.00	1.01	1.00	1
3	1.49	0.91	**	1.01	1.02	1.01	1.01
4	1.71	0.94	*	1.03	1.05	1.04	1.01
5	1.90	0.91	**	1.04	1.07	1.07	1.01
6	2.03	0.91	**	1.08	1.09	1.07	1.01
7	2.10	0.91	**	1.07	1.09	1.11	1.01
8	2.18	0.91	**	1.06	1.06	1.06	1.01
9	2.26	0.88	**	1.06	1.06	1.06	1.01
10	2.32	0.86	**	1.06	1.07	1.11	1.01
11	2.35	0.84	**	1.10	1.10	1.09	1.03
12	2.40	0.82	**	1.10	1.11	1.14	1.02

Sample: 1999/01-2009/12 Forecast Sample: 2004/01-2009/12 * 95% ** 99%

5.2 Hendry-Huldrich method.

We have not implemented with our data the Hendry-Huldrich method, therefore a proper comparison between procedures is not done in this paper. We only can report Hendry-Huldrich results which correspond to different samples than ours. These authors applied their procedure for US data for different sample periods. The most recent one, which is the period for which they get better results, goes from 1984 to 2004. In this case their best method reduces the RMSFE of the direct procedure (0.19) by less than one percent at horizon one. At horizon 12, where the RMSFE of the direct method is 1.30, the reduction is 13%. In our case the sample for forecast evaluation goes from 2004 to 2009 and our procedure P4 reduces the RMSFE of the direct method

(0.57) by 16% at horizon one and 18% at horizon 12. Nevertheless, the samples are very different and no conclusion can be obtained with this comparison. In fact, the unpredictability measured by the RMSFE of the direct method is one third of ours in the sample used by HH at horizon one and almost one half at horizon twelve.

6.- Exogenous variables y causal models.

Our procedure can be easily extended to include exogenous variables in the different models, (1) to (6), for the basic components or intermediate aggregates. In fact, we can classified the candidates for exogenous explanatory variables in groups in such a way that for a particular basic component some groups could be relevant and the others no, and in each case test for the inclusion of regressors from the relevant groups. For instance, one group of exogenous candidates could be "international prices for energy goods". With no doubt the inclusion of regressors from that group in explaining the basic components referred to domestic consumer prices of energy products will improve the forecasting accuracy for short horizons. In other cases wholesale and import prices, demand indicators, etc could be also useful for different basic components.
Causal models are very important in forecasting because the economic agents *are* not only interested in forecasts but also on an explanation of how these forecasts are determined. The disaggregated forecasts in our procedure could show market differences and provide some clues about the main factors causing inflation. For instance, they could show that inflation is coming mainly from basic components related to energy and unprocessed food, suggesting that the inflation problem could be due more to imported inflation than to pressure of the domestic demand. But our procedure does not currently include causal economic variables to explain inflation. The consideration of causal models in our approach is more complex, because in general the results from the economic theory will refer to the aggregate or intermediate aggregates but not to the basic components. Also, for the basic components we could have very often monthly data but for the explanatory variables referred to the whole economic usually only quarterly information is available. In these circumstances a one could

proceed, for example, by building quarterly causal models for the aggregate. In this case it would not be strange that our procedure provides more accurate quarterly forecasts than the causal model, but the latter provides an explanation for the forecasts which should be taken into consideration. This can be done by linking both results by means of a regression which could be used to provide the most accurate forecast with an econometric explanation.

In order to obtain a causal explanation for headline inflation forecasts a simple regression between the forecasts of our procedure aggregated at quarterly level, y_t, and the forecasts from the causal model, x_t, can be run. In this regression:

$$y_t = c + bx_t + r_t$$

we can test the null that c equals zero and b equals one. If the null is not rejected, we can substitute the x_t forecasts in the above regression for their composition in terms of the explanatory variables used to calculate them. Thus we end up with a causal explanation for the inflation forecasts (y_t) coming from our approach. The component r_t (the part of the disaggregated forecasts which is not explained by the econometric forecasts) can be interpreted as the impact on total inflation of the heterogeneous inflation situation of different markets or the bias in the econometric model's causal explanation derived from not contemplating specific market effects. An application of this way of linking time series and econometric forecasts can be seen in European Forecasting Network (2003), chapter 1, box 4. In this example referred to the inflation in the euro area the linking regression shows, in the words of the authors, "that the amount of money in relation to output is pushing inflation up, that there are three factors pushing in the opposite direction and that another two currently have a practically insignificant effect. The heterogeneous inflation situation on different markets is favouring lower inflation rates". Based on these results the report concludes that a loose monetary policy can continue.

7.- Conclusions.

The paper, as the literature on dynamic factors, faces the question of dealing with a large number of variables, 160 in the case of the CPI in US, and departing from it proposes the analysis of all variables individually considered to detect common features which are shared by some variables but not for all. This leads to a classification of the variables in different non-overlapping subsets which turns to be very useful for building econometric models for each variable which in each case includes the corresponding common features.

The cases of large number of variables contemplated in this paper correspond to all the components of an aggregate, because a pivotal point in the paper is that for a full understanding of the aggregates and for the formulation of economic policies, it is useful to consider all the variables comprising a macro-magnitude, aggregate and components, and in particular the forecasts of all of them. The paper develops a procedure based on bivariate methods which groups the components of an aggregate in a few subsets characterized by the fact that their elements share a common trend and/or a common stationary factor and a remaining subset whose elements do not have such restrictions. It seems that in occasions it is useful a step further, removing from this residual subset the largest subset of components which share two common trends. Applying the procedure in a restrictive way to study consumer price indexes we found that components which weight between 65-77% in those indexes for the euro area, UK and US are restricted in the above form. For forecasting inflation the paper shows that in the euro area and UK it is useful to consider the subset of elements which share two common trends but this is not the case for US.

For policy purposes one would be interested in working with the observed and forecasts values of an aggregate and its components at the same time, thus we need to insure that an indirect forecast of the aggregate is more accurate than (I) a direct one. Discarding as unfeasible (II) the indirect forecast based on a vector model, an indirect alternative easily to implement is (III) the one based on univariate models for the components, but it ignores restrictions between the components and is exposed to a great deal of estimation uncertainty. Thus univariate indirect forecast can perform worse than the direct method. In fact, we show that this is the case for the data studied in this paper.

An intermediate alternative (IV) between the direct forecast and the indirect forecast based on vector modelling is an indirect one based on single-equation models with explanatory variables. Having obtained the grouping of components described above, the forecasting procedure proposed in this paper consists in forecasting each component by a single-equation model which incorporates the common trend and the CSCF affecting the component in question. This procedure, denoted above as P4, performs significantly better than the direct method for all horizons in UK and US and from horizon two for the euro area. In the cases of the euro area and UK the procedure can be improved considering also the largest subset of basic elements, S1(2), which share two common trends. For the euro area this is obtained by individual forecasts of the elements of S1(2) and for UK by the direct forecast of the sub-aggregate, AS1(2), of this subset.

Our procedure outperforms the procedure based on dynamic factors in the three economies and in a significant way for the whole path of forecasts in the euro area and in US.

The classification, proposed in this paper can be of interest for further economic and econometric analysis and for the formulation of intermediate sub-aggregates.

The forecasting procedure proposed in the paper is especially suitable for monthly variables and can be easily extended to include exogenous variables, but does not fit well to estimate causal econometric models, because the economic theory usually available for building them applies to the aggregate more than to the components. In any case proving forecasts for the aggregate and its components one has detailed information which in occasions could provide hints about the factors causing the forecast behaviour of the aggregate. Nonetheless, the paper shows how to make a joint use of causal forecasts of an aggregate and of the forecasts based on our method. Usually the latter will be more accurate, but the former will provide an explanation of the forecasts in terms of determining variables which should not be ignored. In many occasions these forecasts could be linked by a simple regression offering an economic explanation for the forecasts derived in this paper.

References

Beaulieu, J.J. and J.A. Miron, (1990), Seasonal unit roots and deterministic seasonals in aggregate U.S. data, Journal of Econometrics Conference on Seasonality and Econometric Models, Montreal, May.

Capistrán, C, (2006), On comparing multi-horizon, Economics Letters 93, 176–181.

Clements M.P. and D.F. Hendry, (1995), Forecasting in cointegrated system, Journal of applied econometrics, 10, 127-146.

Dickey D.A. and Fuller, W.A., (1979), “Distribution of the estimators for autoregressive time series with a unit root”, Journal of the American Statistical Association, 74, 427-431.

Duarte, C. and Rua, Antonio (2007), Forecasting inflation through a bottom-up approach: how bottom is bottom?, Economic Modelling, 24, 941-953.

European Forecasting Network (2003), “EFN Report on the Euro Atrea Outlook”, Autum.

Engle R and C.W.J. Granger (1987), Cointegration and error-correction representation, estimation and testing, Econometrica, 56, 143-159.

Engle R.F. and S. Kozicki (1993), Testing for common features, Journal of Business and Economic Statistic, 11, 369-395

Espasa, A., M. L. Matea, M. C. Manzano and V. Catasus (1987), "La inflación subyacente en la economía española: estimación y metodología", Boletín Económico del Banco de España, marzo, 32-51.

Espasa, A. (2002), Forecasting Inflation in the European Monetary Union: A disaggregated approach by countries and by sectors, (con E. Senra y R. Albacete), European Journal of Finance, vol 8.

Espasa, A. (2004), A time series disaggregated model to forecast GDP in the euro zone, Proceedings of the colloquium on modern tools for business cycle analisis, Monograph of official statistics, Eurostat.

Engle R.F. and B.S. Yoo (1987), Forecasting and testing in co-integrated system, Journal of Econometrics, 35, 143-159.

Franses P.H. (1991), Seasonality, non-stationary and the forecasting of monthly time series, International Journal of Forecasting, 7, 199-208

Ghose D. (1995), Linear aggregation in cointegrated systems, Journal of Economic Dynamic and Control, 19, 1011-1032.

Hendry, D. F. and K. Hubrich (2006). Forecasting aggregates by disaggregates, European Central Bank Working Paper 589.

Hubrich K. (2003), Forecasting euro area inflation: does aggregating forecast by HICP component improve forecast accuracy?, ECB, Working Paper Nº. 247.

Hylleberg S., R.F. Engle, C.W.J. Granger and B.S. Yoo, (1990), Seasonal integration and cointegration, Journal of Econometrics, 44, 215-238.

Kohn, R. (1982). When is an aggregate of a time series efficiently forecast by its past?, Journal of Econometrics (18): 337–349.

Lütkepohl H. (1984), Linear transformations vector ARMA processes, Journal of Econometrics, 26, 283-293.

Lütkepohl, H. (1987). Forecasting Aggregated Vector ARMA Processes, Springer-Verlag.

Peña, D. and Poncela, P. (2004), Forecasting with non-stationary dynamic factor models, Journal of Econometrics, 119, 291,321.

Reisel, G.C. and S.K. Ahn (1992), Vector autoregressive models with unit roots and reduced ranks structure: estimation, likelihood ratio test and forecasting, Journal of Time Series Analysis, 13, 353-375.

Rose, D. E.(1977). Forecasting aggregates of independent ARIMA processes, Journal of Econometrics (5): 323–345.

Stock J and M.W. Watson (1988), Testing for common trends, Journal of American statistical association, 83, 1097-1107.

Tiao G.C. and I. Guttman (1980), Forecasting contemporal aggregates of multiple time series, Journal of Econometrics, 12, 219-230.

Tiao G.C. and R.S. Tsay (1989), Model specification in multivariate time series (with comments), Journal of Royal Statistical Society, Series B,51, 157-213.

Vahid F. and R.F. Engle (1993), Common trends and common cycles, Journal of Applied Econometrics, 8, 341-360

Vahid F. and R.F. Engle (1997), Codependent cycles, Journal of Econometrics, 80, 199-221

Vahid F. and J.V. Issler (2001), The importance of common cyclical features in VAR analysis: a monte carlo satudy, Journal of Econometrics.

Wei, W. W. S. and Abraham, B. (1981). Forecasting contemporal time series aggregates, Communications in Statistics - Theory and Methods A10: 1335–1344.

Appendix 1, the advantage of disaggregating when the components have a common trend.

In the presence of a common trend in the components of an aggregate variable, the EDF condition is very specific. Let us illustrate it for the case in which the data generation process (DGP) is known and the components are fully cointegrated. The result shows that the EDF condition can not be satisfied for an arbitrary W weight vector when there exits full cointegration among n the components.

The aggregate variable $X_t = (x_{1t}, ..., x_{nt})'$ is the weighted average of components, $X_t = W' x_t = \sum_{i=1}^{n} w_i x_{it}$ with $W' = (w_1, ..., w_n)$.

Assumption 1: We have an aggregate variable composed by a ***n*** dimensional non-stationary vector of basic components, contains n pure random walks generated from a single unit root (a common stochastic trend.)

.

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).

Proof:

Let an aggregate process, X_t. To simplify the proof, we consider two components case for the disaggregate DGP, $X_t = (x_{1t}, x_{2t})'$. The disaggregate DGP is $\Delta X_t = \Pi X_{t-1} + e_t = \alpha' \beta X_{t-1} + e_t$. Now, we can write the disaggregate model on the form:

$$\Delta x_t = \begin{pmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{pmatrix} = \begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}$$

First Part:

First, we show the aggregate DGP of two CI(1,1) components is a I(1). For simplicity, the aggregate variable, X_t is the contemporaneous sum of the components, $X_t = x_{1t} + x_{2t}$

Aggregation theory tell us that it is represented by a ARMA(2,1) process, see Lütkepohl (1984), $(1-\phi_1 L - \phi_2 L^2)Y_t = (1-\theta L)a_t$.
(1)

On the other hand, the DGP is given by,

$$x_t = \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} 1+\pi_{11} & \pi_{12} \\ \pi_{21} & 1+\pi_{22} \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}$$
(2)

If the previous disaggregate system is solved; it can be expressed by two independent univariante processes represent by two differential equations of second order.

$$((1-\pi_{11}L)(1-\pi_{22}L) - (\pi_{21}\pi_{12})L^2)x_{1t} = (1-\pi_{22}L)e_{1t} + \pi_{12}Le_{2t}$$
$$((1-\pi_{11}L)(1-\pi_{22}L) - (\pi_{21}\pi_{12})L^2)x_{2t} = (1-\pi_{11}L)e_{2t} + \pi_{21}Le_{1t} \quad (3)$$

The sum of equations (3) produces an aggregate representation as:

$$((1-\pi_{11}L)(1-\pi_{22}L) - (\pi_{21}\pi_{12})L^2)X_t = (1-\pi_{11}L)e_{2t} + \pi_{21}Le_{1t} + (1-\pi_{22}L)e_{1t} + \pi_{12}Le_{2t}$$
(4)

That aggregate representation of the DGP shares the unit root with the both previous equations since they share the same characteristic equation.

$$\lambda^2 - (\pi_{11} + \pi_{22})\lambda + ((\pi_{11}\pi_{22}) - (\pi_{21}\pi_{12})) = 0$$
(5)

Then, the aggregate process is a I(1) and it shares the former unit root.

Second Part:

From the error correction representation, it can be shown that the aggregate model has a greater mean square error than the disaggregate one.

Due to the existence of a common trend, the DGP is written by,

$$\Delta x_t = \begin{pmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{pmatrix} = \begin{pmatrix} \alpha_1 & -\beta\alpha_1 \\ \alpha_2 & -\beta\alpha_2 \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}$$

$$\alpha_1 = -a_{12}a_{21}/(1-a_{22})$$
$$\beta = (1-a_{22})/a_{21}$$
$$\alpha_2 = a_{21}$$

(6)

Writing the DGP in levels and aggregating,

$$x_t = \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} 1+\alpha_1 & -\beta\alpha_1 \\ \alpha_2 & 1-\beta\alpha_2 \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}$$

(7)

Then, the aggregate representation would be,

$$\begin{aligned} &(1-L)(1-(1+\alpha_1-\beta\alpha_2)L)X_t = \\ &= e_{1t}+e_{2t}-(1-\beta\alpha_2)Le_{1t}-\beta\alpha_1 Le_{1t}-(1+\alpha_1)Le_{2t}+\alpha_2 Le_{2t} = \\ &= e_{1t}+e_{2t}-e_{1t-1}(1+\beta(\alpha_1-\alpha_2))-e_{2t-1}(1+(\alpha_1-\alpha_2)) \end{aligned} \quad (8)$$

Operating, the aggregate representation can be written as,

$$(1-L)(1-(1+\alpha_1-\beta\alpha_2)L)X_t = e_{1t}+e_{2t}-e_{1t-1}(1+\beta(\alpha_1-\alpha_2))-e_{2t-1}(1+(\alpha_1-\alpha_2)) \quad (9)$$

Now, we present the aggregate and disaggregate models:

Aggregate Model: $(1-L)X_t = a_t \dfrac{(1+\theta L)}{(1-\phi L)}$

And disaggregate model: $\Delta C_t = \begin{pmatrix} \Delta c_{1t} \\ \Delta c_{2t} \end{pmatrix} = \begin{pmatrix} \alpha_1 & -\beta\alpha_1 \\ \alpha_2 & -\beta\alpha_2 \end{pmatrix} \begin{pmatrix} c_{1t-1} \\ c_{2t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}$

Thus, the one period ahead forecast in the disaggregate model case is, $\Delta X_{t+1/t} = W'\Delta x_{t+1/t} = W'\alpha\beta\Delta x_t$, and in the aggregate one, $\Delta X_{t+1/t} = \phi\Delta X_t + \theta a_t$.

The forecasting error of period h=1 is for the disaggregate case, $\Delta X_{t+1} - W'\Delta x_{t+1/t} = e_{1t+1} + e_{2t+1}$ with $Var(\Delta X_{t+1} - W'\Delta x_{t+1/t}) = W'\Omega W > 0$.
In the aggregate case, $\Delta X_{t+1} - \Delta X_{t+1/t} = (e_{1t+1} + e_{2t+1}) + \left[\theta a_t - e_{1t}(1+\beta(\alpha_1-\alpha_2)) - e_{2t}(1+(\alpha_1-\alpha_2))\right]$ with $Var(\Delta X_{t+1} - W'\Delta x_{t+1/t}) = W'\Omega W + K > 0$ and $K > 0$, since cov(e_{it+1}, e_{jt} ,)=0 $\forall t$ and $\forall i, j$.

The necessary condition for both errors are equal is,

$$\left[\theta a_t - e_{1t}(1+\beta(\alpha_1-\alpha_2)) - e_{2t}(1+(\alpha_1-\alpha_2))\right] = 0$$
(10)

Now, we show that this equality is not satisfying for CI(1,1), except for some special cases.

Proof: If it exits an aggregate representation for the disaggregate process, that representation is function of the aggregate components.

$$(1-L)X_t = a_t \frac{(1+\theta L)}{(1-\phi L)}$$

Known the process, for the aggregate model, $(1-L)X_t = a_t \dfrac{(1+\theta L)}{(1-\phi L)}$, the one-period-ahead forecasting error in the aggregate case,

$$\Delta X_{t+1} - \Delta X_{t+1/t} = a_{t+1}$$

The error in the disaggregate case is,

$$\Delta X_{t+1} - W'\Delta x_{t+1/t} = e_{t+1}$$

According that necessary condition for existing an aggregate process equivalence to the sum of the disaggregate process, $a_{t+1} = W'e_{t+1} \quad \forall t$

Then, the previous necessary condition (10) is re-write as:

$$\left[\theta(e_{1t} + e_{2t}) - e_{1t}(1+\beta(\alpha_1 - \alpha_2)) - e_{2t}(1+(\alpha_1 - \alpha_2))\right] = 0$$
(10)

Operating from (10),

$$e_{1t}\left[\theta - (1+\beta(\alpha_1 - \alpha_2))\right] + e_{2t}\left[\theta - (1+(\alpha_1 - \alpha_2))\right] = 0$$
$$e_{1t}\left[\theta - (1+\beta(\alpha_1 - \alpha_2))\right] = -e_{2t}\left[\theta - (1+(\alpha_1 - \alpha_2))\right]$$
(11)

equations (10) and (11) are true at the sometime if, $\theta - (1+\beta(\alpha_1 - \alpha_2)) = 0$ and $\theta - (1+(\alpha_1 - \alpha_2)) = 0$. 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, 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.

Appendix 2, the advantage of disaggregating when the components have common serial correlation.

In the presence of a CSCF in the components of an aggregate variable, the EDF condition is very specific. Let us illustrate it for the case in which the data generation process (DGP) is known and the components share a CSCF in their first differences. The result shows that the EDF condition can not be satisfied for an arbitrary W weight vector when there exits CSCF among n the components.

The aggregate variable $X_t = (x_{1t},..., x_{nt})'$ is the weighted average of components, $X_t = W' x_t = \sum_{i=1}^{n} w_i x_{it}$ with $W' = (w_1,..., w_n)$.

Assumption 1: We have an aggregate variable composed by a ***n*** dimensional non-stationary vector of basic components, whose first differences contains n stationary process with a proportional lag structure.

.

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).

Proof:

Let an aggregate process, $X_t = (x_{1t}, x_{2t})'$. To simplify the proof, we consider two components case for the disaggregate DGP. The disaggregate DGP is

$\Delta X_t = \Pi \Delta X_{t-1} + e_{tt}$ .Now, we can write the disaggregate model on the form:

$$\Delta X_t = \begin{pmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{pmatrix} = \begin{pmatrix} \pi_1 & \pi_2 \\ \gamma\pi_1 & \gamma\pi_2 \end{pmatrix} \begin{pmatrix} \Delta x_{1t-1} \\ \Delta x_{2t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}$$

First Part:

First, we have shown the aggregate DGP of two non-stationary processes is another non-stationary process. Then, the first differences of the aggregate is stationary

Second Part:

From the disaggregate representation, it can be shown that the aggregate model has a greater mean square error than the disaggregate one. Due to the existence of CSC, the DGP is written by;

$$\Delta X_t = \begin{pmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{pmatrix} = \begin{pmatrix} \pi_1 & \pi_2 \\ \gamma\pi_1 & \gamma\pi_2 \end{pmatrix} \begin{pmatrix} \Delta x_{1t-1} \\ \Delta x_{2t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}$$

Then, the aggregate representation would be,

$$(1-L)(1-(1+\pi_1+\gamma\pi_2)L)\Delta X_t = e_{1t} + e_{2t} - e_{1t-1}(1+\pi_2(1-\gamma)) - e_{2t-1}(1+\pi_1(1-\gamma))$$

Now, we present the aggregate and disaggregate models:

Aggregate Model: $\Delta X_t = a_t \dfrac{(1+\theta L)}{(1-\phi L)}$

And disaggregate model: $\Delta X_t = \begin{pmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{pmatrix} = \begin{pmatrix} \pi_1 & \pi_2 \\ \gamma\pi_1 & \gamma\pi_2 \end{pmatrix} \begin{pmatrix} \Delta x_{1t-1} \\ \Delta x_{2t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}$

Thus, the one period ahead forecast in the disaggregate model case is, $\Delta X_{t+1/t} = W'\Delta x_{t+1/t}$, and in the aggregate one, $\Delta X_{t+1/t} = \phi\Delta X_t + \theta a_t$.

The forecasting error of period h=1 is for the disaggregate case, $\Delta X_{t+1} - W'\Delta x_{t+1/t} = e_{1t+1} + e_{2t+1}$ with $Var(\Delta X_{t+1} - W'\Delta x_{t+1/t}) = W'\Omega W > 0$.

In the aggregate case, $\Delta X_{t+1} - \Delta X_{t+1/t} = (e_{1t+1} + e_{2t+1}) + \left[\theta a_t - e_{1t}(1+\pi_2(1-\gamma)) - e_{2t}(1+\pi_1(1-\gamma))\right]$ with $Var(\Delta X_{t+1} - W'\Delta x_{t+1/t}) = W'\Omega W + K > 0$ and $K > 0$, since cov(e_{it+1}, e_{jt})=0 $\forall t$ and $\forall i, j$.

The necessary condition for both errors are equal is,

$$\left[\theta a_t - e_{1t}(1+\pi_2(1-\gamma)) - e_{2t}(1+\pi_1(1-\gamma))\right] = 0$$

Now, we show that this equality is not satisfying for components with CSC, except for some special cases.

Proof: If it exits an aggregate representation for the disaggregate process, that representation is function of the aggregate components.

$$(1-L)X_t = a_t \frac{(1+\theta L)}{(1-\phi L)}$$

Known the process, for the aggregate model, $(1-L)X_t = a_t \frac{(1+\theta L)}{(1-\phi L)}$, the one-period-ahead forecasting error in the aggregate case,

$$\Delta X_{t+1} - \Delta X_{t+1/t} = a_{t+1}$$

The error in the disaggregate case is,

$$\Delta X_{t+1} - W'\Delta X_{t+1/t} = W' e_{t+1}$$

According that necessary condition for existing an aggregate process equivalence to the sum of the disaggregate process, $a_{t+1} = W' e_{t+1} \quad \forall t$

Then, the previous necessary condition is re-write as:

$$\left[\theta(e_{1t} + e_{2t}) - e_{1t}(1+\pi_2(1-\gamma)) - e_{2t}(1+\pi_1(1-\gamma))\right] = 0$$

Operating from (4),

$$e_{1t}\left[\theta-(1+\pi_2(1-\gamma))\right]-e_{2t}\left[\theta-(1+\pi_1(1-\gamma))\right]=0$$

$$e_{1t}\left[\theta-(1+\pi_2(1-\gamma))\right]=e_{2t}\left[\theta-(1+\pi_1(1-\gamma))\right]$$

equations (4) and (6) are true at the sometime if $\pi_1=\pi_2$ or $\gamma=1$. Then, we get the same forecast error for both models, aggregate and disaggregate.

Conclusion:

If the process presents CSC and disaggregate DGP is known, 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 all coefficients are identical in each equation

(2) If the disaggregate processes are identical

Appendix 3, listing of Basic components for the euro area, UK and UD in sets of common features, S1, S2 and S1(2).

List 1 EA basic components with the common trend (S1)

Weights	Class	Description - Price Index -
1.55%	NPF	Vegetables
5.21%	MAN	Garments
0.79%	MAN	Materials for the maintenance and repair of the dwelling
0.74%	MAN	Water supply
0.34%	MAN	Carpets and other floor coverings
0.63%	MAN	Household textiles
0.65%	MAN	Glassware, tableware and household utensils
0.34%	MAN	Motor cycles, bicycles and animal drawn vehicles
0.35%	MAN	Major durables for indoor and outdoor recreation including musical instruments
0.58%	MAN	Gardens, plants and flowers
0.94%	MAN	Newspapers and periodicals
0.35%	MAN	Miscellaneous printed matter; stationery and drawing materials
0.45%	SERV	Sewerage collection
0.09%	SERV	Repair of furniture, furnishings and floor coverings
0.12%	SERV	Repair of household appliances
2.43%	SERV	Maintenance and repair of personal transport equipment
0.89%	SERV	Other services in respect of personal transport equipment
0.75%	SERV	Passenger transport by road
0.44%	SERV	Combined passenger transport
0.06%	SERV	Other purchased transport services
1.59%	SERV	Cultural services
1.76%	SERV	Package holidays
7.31%	SERV	Restaurants, cafés and the like
0.92%	SERV	Canteens
1.69%	SERV	Accommodation services
8.54%	SERV	Miscellaneous goods and services

List 2 EA basic components with the common serial correlation (S2)

Weights	Class	Description - Price Index -
3.6%	NPF	Meat
1.1%	NPF	Fruit
0.5%	ENE	Liquid fuels
0.2%	ENE	Solid fuels
0.6%	ENE	Heat energy
3.6%	ENE	Fuels and lubricants for personal transport equipment
2.44%	PF	Bread and cereals
2.2%	PF	Milk, cheese and eggs
0.5%	PF	Oils and fats
1.1%	PF	Sugar, jam, honey, chocolate and confectionery
0.4%	PF	Food products n,e,c,
0.4%	PF	Coffee, tea and cocoa
0.9%	PF	Mineral waters, soft drinks, fruit and vegetable juices

0.8%	PF	Wine
1.4%	MAN	Footwear including repair
1.0%	MAN	Non-durable household goods
0.6%	MAN	Equipment for the reception, recording and reproduction of sound and pictures
0.2%	MAN	Photographic and cinematographic equipment and optical instruments
0.4%	MAN	Information processing equipment
0.6%	MAN	Games, toys and hobbies
0.3%	MAN	Equipment for sport, camping and open-air recreation
0.3%	MAN	Miscellaneous printed matter; stationery and drawing materials
0.9%	SERV	Services for the maintenance and repair of the dwelling
0.4%	SERV	Refuse collection
0.1%	SERV	Passenger transport by sea and inland waterway
0.1%	SERV	Other purchased transport services
0.2%	SERV	Postal services
1.8%	SERV	Package holidays

List 3 EA basic components with the second common trend (S1(2))

Weights	Class	Description - Price Index -
1.11%	NPF	Fruit
0.51%	ENE	Liquid fuels
0.17%	ENE	Solid fuels
1.00%	MAN	Major household appliances whether electric or not and small electric household appliances
0.38%	MAN	Information processing equipment
0.45%	SERV	Passenger transport by railway
0.49%	SERV	Passenger transport by air

List 4 UK basic components with the common trend (S1)

Weights	Class	Description - Price Index -
1.60%	NPF	Vegetables
1.30%	PF	Sugar, jam, honey, chocolate and confectionery
4.80%	MAN	Clothing
0.90%	MAN	Footwear including repair
5.10%	SERV	Actual rentals for housing
4.60%	ENE	Electricity, gas and other fuels
0.80%	MAN	Major household appliances whether electric or not and small electric household appliances
0.50%	MAN	Glassware, tableware and household utensils
0.40%	MAN	Other medical products; therapeutic appliances and equipment
0.20%	SERV	Dental services
0.80%	SERV	Hospital services
4.40%	MAN	Motor cars
0.30%	MAN	Motor cycles, bicycles and animal drawn vehicles
2.30%	SERV	Maintenance and repair of personal transport equipment
2.30%	MAN	Communications
0.60%	MAN	Equipment for the reception, recording and reproduction of sound and pictures
0.70%	MAN	Recording media
0.10%	SERV	Repair of audio-visual, photographic and information processing equipment

2.00%	MAN	Games, toys and hobbies
0.80%	MAN	Pets and related products; veterinary and other services for pets
0.50%	MAN	Books
0.60%	MAN	Newspapers and periodicals
0.60%	MAN	Miscellaneous printed matter; stationery and drawing materials
2.70%	SERV	Package holidays
0.20%	SERV	Insurance connected with the dwelling
0.20%	SERV	Insurance connected with health

List 5 UK basic components with the common serial correlation (S2)

Weights	Class	Description - Price Index -
1.00%	NPF	Fruit
3.40%	ENE	Fuels and lubricants for personal transport equipment
1.50%	PF	Milk, cheese and eggs
0.40%	PF	Coffee, tea and cocoa
1.00%	PF	Mineral waters, soft drinks, fruit and vegetable juices
0.50%	PF	Beer
2.30%	PF	Tobacco
2.80%	MAN	Furniture and furnishings, carpets and other floor coverings
0.70%	MAN	Household textiles
0.80%	MAN	Major household appliances whether electric or not and small electric household appliances
0.40%	MAN	Other medical products; therapeutic appliances and equipment
0.30%	MAN	Motor cycles, bicycles and animal drawn vehicles
2.30%	MAN	Communications
0.50%	MAN	Information processing equipment
0.80%	MAN	Jewellery, clocks and watches
0.30%	MAN	Other personal effects
0.60%	MAN	Pharmaceutical products
0.10%	SERV	Repair of household appliances
0.50%	SERV	Domestic services and household services
2.30%	SERV	Maintenance and repair of personal transport equipment
2.10%	SERV	Education
10.00%	SERV	Restaurants, cafés and the like
0.80%	SERV	Hairdressing salons and personal grooming establishments
1.10%	SERV	Social protection

List 6 UK basic components with the second common trend (S1(2))

Weights	Class	Description - Price Index -
1.00%	NPF	Fruit
3.40%	ENE	Fuels and lubricants for personal transport equipment
0.30%	PF	Food products n.e.c.
0.50%	PF	Beer
1.10%	MAN	Water supply and miscellaneous services relating to the dwelling
0.40%	MAN	Photographic and cinematographic equipment and optical instruments
0.50%	MAN	Information processing equipment
0.90%	MAN	Other major durables for recreation and culture
0.40%	MAN	Equipment for sport, camping and open-air recreation

0.10%	SERV	Repair of household appliances
3.50%	SERV	Transport services
1.10%	SERV	Recreational and sporting services
10.00%	SERV	Restaurants, cafés and the like
1.10%	SERV	Social protection

List 8 US basic components with the common trend (S1)

Weights	Class	Description - Price Index -
0.06%	NPF	Other poultry including turkey
0.11%	NPF	Eggs
0.22%	NPF	Other fresh fruits
0.06%	NPF	Lettuce
0.08%	NPF	Tomatoes
0.19%	ENE	Fuel oil
1.16%	ENE	Utility (piped) gas service
0.31%	PF	Milk
0.21%	PF	Other beverage materials including tea
0.11%	PF	Other fats and oils including peanut butter
0.10%	PF	Soups
0.11%	MAN	Window coverings
0.18%	MAN	Other linens
0.11%	MAN	Indoor plants and flowers
0.14%	MAN	Men's suits, sport coats, and outerwear
0.19%	MAN	Men's furnishings
0.22%	MAN	Men's shirts and sweaters
0.12%	MAN	Women's outerwear
0.10%	MAN	Women's dresses
0.70%	MAN	Women's suits and separates
0.26%	MAN	Girls' apparel
0.15%	MAN	Boys' and girls' footwear
0.31%	MAN	Women's footwear
0.05%	MAN	Watches
0.20%	MAN	Other motor fuels
0.06%	MAN	Sewing machines, fabric and supplies
0.08%	SERV	Distilled spirits at home
0.73%	SERV	Airline fare
1.05%	SERV	Wireless telephone services
0.29%	SERV	Internet services and electronic information

List 9 US basic components with the common serial correlation (S2)

Weights	Class	Description - Price Index -
0.19%	NPF	Uncooked beef steaks
0.25%	NPF	Other meats
0.27%	NPF	Chicken
0.08%	NPF	Apples

0.08%	NPF	Citrus fruits
0.22%	NPF	Other fresh fruits
0.08%	NPF	Potatoes
0.08%	NPF	Tomatoes
0.19%	ENE	Fuel oil
0.11%	ENE	Propane, kerosene, and firewood
3.00%	ENE	Electricity
1.16%	ENE	Utility (piped) gas service
2.96%	ENE	Gasoline (all types)
0.20%	PF	Breakfast cereal
0.21%	PF	Cakes, cupcakes, and cookies
0.31%	PF	Milk
0.02%	PF	Frozen noncarbonated juices and drinks
0.32%	PF	Nonfrozen noncarbonated juices and drinks
0.19%	PF	Candy and chewing gum
0.06%	PF	Other sweets
0.06%	PF	Salad dressing
0.31%	PF	Snacks
0.07%	PF	Baby food
0.43%	PF	Other miscellaneous foods
0.73%	PF	Cigarettes
0.14%	MAN	Men's suits, sport coats, and outerwear
0.22%	MAN	Men's shirts and sweaters
0.17%	MAN	Men's pants and shorts
0.20%	MAN	Boys' apparel
0.12%	MAN	Women's outerwear
0.10%	MAN	Women's dresses
0.70%	MAN	Women's suits and separates
0.35%	MAN	Women's underwear, nightwear, sportswear and
0.26%	MAN	Girls' apparel
0.31%	MAN	Women's footwear
0.18%	MAN	Infants' and toddlers' apparel
6.93%	MAN	New and used motor vehicles
0.20%	MAN	Other motor fuels
0.15%	MAN	Vehicle accessories other than tires
0.07%	MAN	Motor vehicle body work
0.14%	MAN	Televisions
0.48%	MAN	Pets and pet products
0.15%	MAN	Newspapers and magazines
0.12%	MAN	Recreational books
0.21%	MAN	Personal computers and peripheral equipment
2.61%	MAN	Personal care
0.30%	SERV	Food at employee sites and schools
0.14%	SERV	Food from vending machines and mobile vendors
5.96%	SERV	Rent of primary residence
0.16%	SERV	Housing at school, excluding board
24.43%	SERV	Owners' equivalent rent of primary residence
2.04%	SERV	Motor vehicle insurance
0.18%	SERV	Parking and other fees

0.73%	SERV	Airline fare
0.17%	SERV	Other intercity transportation
4.77%	SERV	Medical care services
1.21%	SERV	Cable and satellite television and radio service
1.45%	SERV	College tuition and fees
0.43%	SERV	Elementary and high school tuition and fees
1.05%	SERV	Wireless telephone services
0.29%	SERV	Internet services and electronic information

List 10 US basic components with the second common trend (S1(2))

Weights	Class	Description - Price Index -
0.07%	NPF	Pork chops
0.16%	NPF	Fresh fish and seafood
0.08%	NPF	Apples
0.08%	NPF	Citrus fruits
0.08%	NPF	Potatoes
0.15%	NPF	Canned fruits and vegetables
0.06%	NPF	Other processed fruits and vegetables including
0.12%	PF	Rice, pasta, cornmeal
0.23%	PF	Other bakery products
0.15%	PF	Ice cream and related products
0.16%	PF	Other dairy and related products
0.02%	PF	Frozen noncarbonated juices and drinks
0.32%	PF	Nonfrozen noncarbonated juices and drinks
0.12%	PF	Coffee
0.06%	PF	Other sweets
0.31%	PF	Snacks
0.05%	PF	Tobacco products other than cigarettes
0.05%	MAN	Floor coverings
0.35%	MAN	Outdoor equipment and supplies
0.30%	MAN	Miscellaneous household products
0.19%	MAN	Men's furnishings
0.22%	MAN	Men's footwear
0.18%	MAN	Infants' and toddlers' apparel
0.17%	MAN	Video discs and other media, including rental of
0.09%	MAN	Photographers and film processing
0.04%	MAN	Music instruments and accessories
2.61%	MAN	Personal care
0.30%	SERV	Food at employee sites and schools
0.32%	SERV	Beer, ale, and other malt beverages at home
5.96%	SERV	Rent of primary residence
0.18%	SERV	Parking and other fees
0.81%	SERV	Land-line telephone services, local charges

List 10 Summary of all EA Basic components

Harmonized indices of consumer prices - Euro area

Weights	Description - Price Index -		S1F	S1(2)F	S2F
3.6%	Meat	NPF			S2F
1.0%	Fish and seafood	NPF			
1.1%	Fruit	NPF		S1(2)F	S2F
1.6%	Vegetables	NPF	S1F		
2.1%	Electricity	ENE			
1.3%	Gas	ENE			
0.5%	Liquid fuels	ENE		S1(2)F	S2F
0.2%	Solid fuels	ENE		S1(2)F	S2F
0.6%	Heat energy	ENE			S2F
3.6%	Fuels and lubricants for personal transport equipment	ENE			S2F
2.4%	Bread and cereals	PF			S2F
2.2%	Milk, cheese and eggs	PF			S2F
0.5%	Oils and fats	PF			S2F
1.1%	Sugar, jam, honey, chocolate and confectionery	PF			S2F
0.4%	Food products n,e,c,	PF			S2F
0.4%	Coffee, tea and cocoa	PF			S2F
0.9%	Mineral waters, soft drinks, fruit and vegetable juices	PF			S2F
0.5%	Spirits	PF			
0.8%	Wine	PF			S2F
0.7%	Beer	PF			
2.5%	Tobacco	PF			
0.0%	Clothing materials	MAN			
5.2%	Garments	MAN	S1F		
0.2%	Other articles of clothing and clothing accessories	MAN			
1.4%	Footwear including repair	MAN			S2F
0.8%	Materials for the maintenance and repair of the dwelling	MAN	S1F		
0.7%	Water supply	MAN	S1F		
2.5%	Furniture and furnishings	MAN			
0.3%	Carpets and other floor coverings	MAN	S1F		
0.6%	Household textiles	MAN	S1F		
1.0%	Major household appliances whether electric or not and small ele	MAN		S1(2)F	
0.6%	Glassware, tableware and household utensils	MAN	S1F		
0.4%	Tools and equipment for house and garden	MAN			
1.0%	Non-durable household goods	MAN			S2F
0.3%	Motor cycles, bicycles and animal drawn vehicles	MAN	S1F		
4.4%	Motor cars	MAN			
0.9%	Spares parts and accessories for personal transport equipment	MAN			
0.6%	Equipment for the reception, recording and reproduction of sound	MAN			S2F
0.2%	Photographic and cinematographic equipment and optical instrur	MAN			S2F
0.4%	Information processing equipment	MAN		S1(2)F	S2F

Harmonized indices of consumer prices - Euro area

Weights	Description - Price Index -		S1F	S1(2)F	S2F
0.5%	Recording media	MAN			
0.4%	Major durables for indoor and outdoor recreation including music	MAN	S1F		
0.6%	Games, toys and hobbies	MAN			S2F
0.3%	Equipment for sport, camping and open-air recreation	MAN			S2F
0.6%	Gardens, plants and flowers	MAN	S1F		
0.5%	Pets and related products; veterinary and other services for pets	MAN			
0.7%	Books	MAN			
0.9%	Newspapers and periodicals	MAN	S1F		
0.3%	Miscellaneous printed matter; stationery and drawing materials	MAN	S1F		S2F
0.2%	Cleaning, repair and hire of clothing	SERV			
5.9%	Actual rentals for housing	SERV			
0.9%	Services for the maintenance and repair of the dwelling	SERV			S2F
0.4%	Refuse collection	SERV			S2F
0.4%	Sewerage collection	SERV	S1F		
0.6%	Other services relating to the dwelling n,e,c,	SERV			
0.1%	Repair of furniture, furnishings and floor coverings	SERV	S1F		
0.1%	Repair of household appliances	SERV	S1F		
0.8%	Domestic services and household services	SERV			
3.7%	Health	SERV			
2.4%	Maintenance and repair of personal transport equipment	SERV	S1F		
0.9%	Other services in respect of personal transport equipment	SERV	S1F		
0.5%	Passenger transport by railway	SERV		S1(2)F	
0.7%	Passenger transport by road	SERV	S1F		
0.5%	Passenger transport by air	SERV		S1(2)F	
0.1%	Passenger transport by sea and inland waterway	SERV			S2F
0.4%	Combined passenger transport	SERV	S1F		
0.1%	Other purchased transport services	SERV	S1F		S2F
0.2%	Postal services	SERV			S2F
2.7%	Telephone and telefax equipment and services	SERV			
0.1%	Repair of audio-visual, photographic and information processing	SERV			
0.0%	Maintenance and repair of other major durables for recreation an	SERV			
1.0%	Recreational and sporting services	SERV			
1.6%	Cultural services	SERV	S1F		
1.8%	Package holidays	SERV	S1F		S2F
1.1%	Education	SERV			
7.3%	Restaurants, cafés and the like	SERV	S1F		
0.9%	Canteens	SERV	S1F		
1.7%	Accommodation services	SERV	S1F		
8.5%	Miscellaneous goods and services	SERV	S1F		

List 11 Summary of all UK Basic components

Harmonized indices of consumer prices - Monthly Data - UK -

Weights	Description - Price Index -		S1F	S1(2)F	S2F
1.6%	Vegetables	NPF	S1F		
1.0%	Fruit	NPF		S1(2)F	S2F
0.5%	Fish and seafood	NPF			
2.3%	Meat	NPF			
3.4%	Fuels and lubricants for personal transport equipment	ENE		S1(2)F	S2F
4.6%	Electricity, gas and other fuels	ENE	S1F		
2.3%	Tobacco	PF			S2F
0.5%	Beer	PF		S1(2)F	S2F
1.0%	Wine	PF			
0.6%	Spirits	PF			
1.0%	Mineral waters, soft drinks, fruit and vegetable juices	PF			S2F
0.4%	Coffee, tea and cocoa	PF			S2F
0.3%	Food products n.e.c.	PF		S1(2)F	
1.3%	Sugar, jam, honey, chocolate and confectionery	PF	S1F		
0.2%	Oils and fats	PF			
1.5%	Milk, cheese and eggs	PF			S2F
1.7%	Bread and cereals	PF			
0.3%	Other personal effects	MAN			S2F
0.8%	Jewellery, clocks and watches	MAN			S2F
2.3%	Electrical appliances for personal care; other appliances,	MAN			
0.6%	Miscellaneous printed matter; stationery and drawing mate	MAN	S1F		
0.6%	Newspapers and periodicals	MAN	S1F		
0.5%	Books	MAN	S1F		
0.8%	Pets and related products; veterinary and other services f	MAN	S1F		
0.5%	Gardens, plants and flowers	MAN			
0.4%	Equipment for sport, camping and open-air recreation	MAN		S1(2)F	
2.0%	Games, toys and hobbies	MAN	S1F		
0.9%	Other major durables for recreation and culture	MAN		S1(2)F	
0.7%	Recording media	MAN	S1F		
0.5%	Information processing equipment	MAN		S1(2)F	S2F
0.4%	Photographic and cinematographic equipment and optical	MAN		S1(2)F	
0.6%	Equipment for the reception, recording and reproduction o	MAN	S1F		
2.3%	Communications	MAN	S1F		S2F
0.5%	Spares parts and accessories for personal transport equip	MAN			
0.3%	Motor cycles, bicycles and animal drawn vehicles	MAN	S1F		S2F

Harmonized indices of consumer prices - Monthly Data - UK -

Code	Number	Weights	Description - Price Index -		S1F	S1(2)F	S2F
cp044	b21	4.4%	Motor cars	MAN	S1F		
cp045	b22	0.4%	Other medical products; therapeutic appliances and equip	MAN	S1F		S2F
cp051	b23	0.6%	Pharmaceutical products	MAN			S2F
cp052	b24	0.6%	Non-durable household goods	MAN			
cp0531_532	b25	0.6%	Tools and equipment for house and garden	MAN			
cp0533	b26	0.5%	Glassware, tableware and household utensils	MAN	S1F		
cp054	b27	0.8%	Major household appliances whether electric or not and sn	MAN	S1F		S2F
cp055	b28	0.7%	Household textiles	MAN			S2F
cp0561	b29	2.8%	Furniture and furnishings, carpets and other floor covering	MAN			S2F
cp0562	b30	1.1%	Water supply and miscellaneous services relating to the dw	MAN		S1(2)F	
cp0611	b31	1.0%	Materials for the maintenance and repair of the dwelling	MAN			
cp0612_613	b32	0.9%	Footwear including repair	MAN	S1F		
cp0621_623	b33	4.8%	Clothing	MAN	S1F		
cp092	b49	0.2%	Insurance connected with health	SERV	S1F		
cp0931	b50	0.2%	Insurance connected with the dwelling	SERV	S1F		
cp0932	b51	1.1%	Social protection	SERV		S1(2)F	S2F
cp0933	b52	0.8%	Hairdressing salons and personal grooming establishment	SERV			S2F
cp0934_935	b53	1.7%	Accommodation services	SERV			
cp0941	b54	1.1%	Canteens	SERV			
cp0942	b55	10.0%	Restaurants, cafés and the like	SERV		S1(2)F	S2F
cp0951	b56	2.1%	Education	SERV			S2F
cp0952	b57	2.7%	Package holidays	SERV	S1F		
cp0953_954	b58	2.1%	Cultural services	SERV			
cp096	b59	1.1%	Recreational and sporting services	SERV		S1(2)F	
cp10	b60	0.1%	Repair of audio-visual, photographic and information proce	SERV	S1F		
cp1111	b61	3.5%	Transport services	SERV		S1(2)F	
cp1112	b62	0.7%	Other services in respect of personal transport equipment	SERV			
cp112	b63	2.3%	Maintenance and repair of personal transport equipment	SERV	S1F		S2F
cp1211	b64	0.8%	Hospital services	SERV	S1F		
cp1212_121	b65	0.2%	Dental services	SERV	S1F		
cp1231	b66	0.2%	Medical services; paramedical services	SERV			
cp1232	b67	0.5%	Domestic services and household services	SERV			S2F
cp124	b68	0.1%	Repair of household appliances	SERV		S1(2)F	S2F
cp1252	b69	0.8%	Services for the maintenance and repair of the dwelling	SERV			
cp1253	b70	5.1%	Actual rentals for housing	SERV	S1F		

List 12 Summary of all US Basic components

Consumer Price Index - CPI Monthly Data - US

Weights	Description - Price Index -		SF1	S1(2)F	S2F
0.23%	Uncooked ground beef	NPF			
0.09%	Uncooked beef roasts	NPF			
0.19%	Uncooked beef steaks	NPF			S2F
0.05%	Uncooked other beef and veal	NPF			
0.12%	Bacon, breakfast sausage, and related	NPF			
0.07%	Ham	NPF			
0.07%	Pork chops	NPF		S1(2)F	
0.09%	Other pork including roasts and picnics	NPF			
0.25%	Other meats	NPF			S2F
0.27%	Chicken	NPF			S2F
0.06%	Other poultry including turkey	NPF	SF1		
0.16%	Fresh fish and seafood	NPF		S1(2)F	
0.14%	Processed fish and seafood	NPF			
0.11%	Eggs	NPF	SF1		
0.08%	Apples	NPF		S1(2)F	
0.07%	Bananas	NPF			
0.08%	Citrus fruits	NPF		S1(2)F	S2F
0.22%	Other fresh fruits	NPF	SF1		S2F
0.08%	Potatoes	NPF		S1(2)F	S2F
0.06%	Lettuce	NPF	SF1		
0.08%	Tomatoes	NPF	SF1		S2F
0.23%	Other fresh vegetables	NPF			
0.15%	Canned fruits and vegetables	NPF		S1(2)F	
0.08%	Frozen fruits and vegetables	NPF			
0.06%	Other processed fruits and vegetables including	NPF		S1(2)F	
0.19%	Fuel oil	ENE	SF1		S2F
0.11%	Propane, kerosene, and firewood	ENE			S2F
3.00%	Electricity	ENE			S2F
1.16%	Utility (piped) gas service	ENE	SF1		S2F
2.96%	Gasoline (all types)	ENE			S2F
0.04%	Flour and prepared flour mixes	PF			
0.20%	Breakfast cereal	PF			S2F
0.12%	Rice, pasta, cornmeal	PF		S1(2)F	
0.24%	Bread	PF			
0.11%	Fresh biscuits, rolls, muffins	PF			
0.21%	Cakes, cupcakes, and cookies	PF			S2F
0.23%	Other bakery products	PF		S1(2)F	
0.31%	Milk	PF	SF1		S2F
0.29%	Cheese and related products	PF			
0.15%	Ice cream and related products	PF		S1(2)F	
0.16%	Other dairy and related products	PF		S1(2)F	
0.32%	Carbonated drinks	PF			
0.02%	Frozen noncarbonated juices and drinks	PF		S1(2)F	S2F
0.32%	Nonfrozen noncarbonated juices and drinks	PF		S1(2)F	S2F
0.12%	Coffee	PF		S1(2)F	
0.21%	Other beverage materials including tea	PF	SF1		
0.05%	Sugar and artificial sweeteners	PF			
0.19%	Candy and chewing gum	PF			S2F
0.06%	Other sweets	PF		S1(2)F	S2F
0.07%	Butter and margarine	PF			
0.06%	Salad dressing	PF			S2F
0.11%	Other fats and oils including peanut butter	PF	SF1		
0.10%	Soups	PF	SF1		
0.31%	Frozen and freeze dried prepared foods	PF			
0.31%	Snacks	PF		S1(2)F	S2F
0.25%	Spices, seasonings, condiments, sauces	PF			
0.07%	Baby food	PF			S2F
0.43%	Other miscellaneous foods	PF			S2F
0.49%	Alcoholic beverages away from home	PF			
0.73%	Cigarettes	PF			S2F
0.05%	Tobacco products other than cigarettes	PF		S1(2)F	
0.05%	Floor coverings	MAN		S1(2)F	
0.11%	Window coverings	MAN	SF1		
0.18%	Other linens	MAN	SF1		
1.03%	Furniture and bedding	MAN			
0.22%	Major appliances	MAN			

Consumer Price Index - CPI Monthly Data - EEUU

Weights	Description - Price Index -		S1F	S1(2)F	S2F
0.13%	Other appliances	MAN			
0.34%	Clocks, lamps, and decorator items	MAN			
0.11%	Indoor plants and flowers	MAN	SF1		
0.07%	Dishes and flatware	MAN			
0.10%	Nonelectric cookware and tableware	MAN			
0.21%	Tools, hardware and supplies	MAN			
0.35%	Outdoor equipment and supplies	MAN		S1(2)F	
0.38%	Household cleaning products	MAN			
0.25%	Household paper products	MAN			
0.30%	Miscellaneous household products	MAN		S1(2)F	
0.78%	Household operations	MAN			
0.14%	Men's suits, sport coats, and outerwear	MAN	SF1		S2F
0.19%	Men's furnishings	MAN	SF1	S1(2)F	
0.22%	Men's shirts and sweaters	MAN	SF1		S2F
0.17%	Men's pants and shorts	MAN			S2F
0.20%	Boys' apparel	MAN			S2F
0.12%	Women's outerwear	MAN	SF1		S2F
0.10%	Women's dresses	MAN	SF1		S2F
0.70%	Women's suits and separates	MAN	SF1		S2F
0.35%	Women's underwear, nightwear, sportswear and	MAN			S2F
0.26%	Girls' apparel	MAN	SF1		S2F
0.22%	Men's footwear	MAN		S1(2)F	
0.15%	Boys' and girls' footwear	MAN	SF1		
0.31%	Women's footwear	MAN	SF1		S2F
0.18%	Infants' and toddlers' apparel	MAN		S1(2)F	S2F
0.05%	Watches	MAN	SF1		
0.31%	Jewelry	MAN			
6.93%	New and used motor vehicles	MAN			S2F
0.20%	Other motor fuels	MAN	SF1		S2F
0.23%	Tires	MAN			
0.15%	Vehicle accessories other than tires	MAN			S2F
0.07%	Motor vehicle body work	MAN			S2F
0.26%	Internal and respiratory over-the-counter drugs	MAN			
0.11%	Nonprescription medical equipment and supplies	MAN			
0.14%	Televisions	MAN			S2F
0.03%	Other video equipment	MAN			
0.17%	Video discs and other media, including rental of	MAN		S1(2)F	
0.10%	Audio equipment	MAN			
0.08%	Audio discs, tapes and other media	MAN			
0.48%	Pets and pet products	MAN			S2F
0.32%	Pet services including veterinary	MAN			
0.33%	Sports vehicles including bicycles	MAN			
0.27%	Sports equipment	MAN			
0.07%	Photographic equipment and supplies	MAN			
0.09%	Photographers and film processing	MAN		S1(2)F	
0.23%	Toys	MAN			
0.06%	Sewing machines, fabric and supplies	MAN	SF1		
0.04%	Music instruments and accessories	MAN		S1(2)F	
0.15%	Newspapers and magazines	MAN			S2F
0.12%	Recreational books	MAN			S2F