FORECASTING DISAGGREGATES BY SECTORS AND REGIONS. THE CASE OF INFLATION IN THE EURO AREA AND IN SPAIN.

25th February 2013

Gabriel Pino[#], J.D. Tena[+] and Antoni Espasa[*]

Abstract

We study the performance of different strategies to forecast 969 and 600 monthly price indexes disaggregated by sectors and geographical areas in Spain, regions, and in the EA12, countries, in order to get a detailed picture of inflation and relative prices in both economies. We take the curse of dimensionality problem into consideration by dealing with ARIMA models and spatial bi-dimensional vector equilibrium correction (SVeqCM) models, where the price indexes for each sector are allowed to be cointegrated with prices in neighbouring areas using different definitions of neighbourhood. We find that geographical disaggregation forecasts are reliable at the regional level in Spain since they improve the forecast accuracy of the headline national inflation. These highly disaggregated forecasts can be used for competitive and other type of macro and regional analysis. The above result does not hold for country disaggregation in the EA12. It seems that the level of economic integration is an important factor to make the geographical analysis useful and that in the Euro Area the regional analysis within countries seems very appropriate.

Key Words: spatial cointegration, contiguity matrix, curse of dimensionality, dynamic factors models.

[#] Gabriel Pino, Universidad de Concepción, Departamento de Economía, Victoria 471, Chile, e-mail: gapino@udec.cl and Southern Illinois University, Economic Department, 1000 Faner Drive, Carbondale, IL 62901, USA, e-mail: gabriel.pino@siu.edu.

[+] Juan de Dios Tena, Universidad Carlos III, Departamento de Estadística, Università di Sassari, Dipartamento di Economia e Impresa, CRENoS and Instituto Flores de Lemus, C/Madrid 126. 28903 Getafe (Madrid), Spain, e-mail: jtena@est-econ.uc3m.es.

[*] Antoni Espasa, Universidad Carlos III, Departamento de Estadística and Instituto Flores de Lemus, C/Madrid 126. 28903 Getafe (Madrid), Spain, e-mail: antoni.espasa@uc3m.es.

Antoni Espasa and Juan de Dios Tena acknowledge financial support from Ministerio de Educación y Ciencia, ECO2009-08100 and ECO2012-32401.

1. Introduction:

Forecasting a headline rate of inflation by considering the information of the disaggregates have received recently a lot of attention; see among others Hendry and Hubrich (2011).Also, Espasa and Mayo (2012) argue that in a variable like inflation the aggregate and all its disaggregates –in their case sector disaggregates- matter for policy and investment decisions and focus their attention in forecasting both aggregate and disaggregates taking into account some of the common features in the latter. There exists a large amount of information on a consumer price index of any developed economy, since the Statistical Offices provide breakdowns of the corresponding CPI by sectors and regions. This vast information is very useful to really understand the behaviour of the inflation in a given country or economic area, particularly by knowing the relative performance of the prices indexes through sectors around the different regions or members' states. This interest becomes evident in two recent contributions by Beck et al. (2009) and Beck et al. (2011) that highlight the importance of considering regional factors and a combination of regional and sectoral components respectively to explain the heterogeneity of disaggregated inflation rates in the EMU. The former identifies the presence of area wide and national components that drive the dynamics of prices in 70 European regions. However, a national factor extracted from series that ignore disaggregation by sectors could represent a mixture of national-specific component and other non national specific factors. This issue is solved in Beck et al. (2011) who extract aggregate, sector, country specific and regional orthogonal components of 730 inflation series. They find indeed that region-specific idiosyncratic components explain most of the variation of prices. The importance of considering all the potential sources of variation of price series at the national, regional and by sector level constitutes a motivation for this paper. However, unlike the previous authors, our interest is mainly the forecast evaluation of the disaggregated price series. Two important additional differences with this literature relate to the fact that 1) we consider the possible presence of cointegration between neighbour prices instead of based the analysis on differentiated series; and 2) to the consideration of 57, instead of only 11, different sectors. The last point is important to avoid the problem of aggregation bias that could result from grouping together very heterogeneous sectors.

Providing forecasts of disaggregated price indexes constitutes a valuable task in itself as it allows central bankers and entrepreneurs to identify how different sectors along regions are affected by different types of economic shocks in order to design an efficient monetary policy, to undertake investment decisions or to receive valuable signals about a possible lack of competitiveness. But those economic agents do not only need to know the past of this detailed information but also its forecasts.

In this paper we are interested in formulating a forecasting procedure for all the disaggregates of a macro-variable like inflation at the highest level of its breakdown by sectors and regions. The procedure

in itself is important because we could easily have around one thousand disaggregates to forecasts.
The interest of adding the geographical dimension in the breakdown by sector in forecasting the headline inflation rate in the Euro Area, was arouse by Espasa and Albacete (2007) who show that in an indirect forecast of this rate by aggregating the forecasts of the components, the breakdown by the double criteria provides more accurate forecasts than breakdowns by a single one. These authors work with VEqCM models using a simple breakdown in just two sectors and five country groups. Espasa and Mayo (2012) show the importance of considering common trends and common cycles in forecasting inflation by a full breakdown by sectors that in the case of the US inflation implies 160 sectors. These authors forecast the disaggregates by single-equation models which include restrictions derived from the existence of common features between the components. They also show that the forecasts of the components by ARIMA models are not very accurate. The above previous literature implies that in trying to forecast a great number of components the consideration of restrictions between them is crucial. The main characteristic in our approach over those previous papers consists in using double criteria in the breakdown with the maximum number of components in each case. Our work represents a first attempt to forecasts components at this highest disaggregated level considering some restrictions between the components. In this framework a general approach for considering restrictions is too complex to start with it. Since Espasa and Mayo (2012) show the interest in restrictions as common trends from the sector breakdown, in this paper we restrict ourselves to the study of restrictions from the regional breakdown and particular spatial cointegration restrictions. The results obtained could show which could be relevant or unimportant aspects to be considered for further research in formulating a more complex forecasting procedure which incorporates a more general way to include restrictions between the great number of components which are present in the type of problem in which we are interested.
The forecasts of a high number of components will be useful if they are reliable. This could be tested for each component, but it seems not enough. Since the components add up to an aggregate, for the mentioned reliability one needs to test that the forecast of the aggregate by aggregating the forecasts of the components is, at least, not significantly worse than the alternative forecasts of the aggregate using the same information set. Thus one aim of the paper is also to study different procedures to forecast the disaggregated price indexes by sectors and regions and to show that those forecasts are reliable because they add up to a forecast of the aggregate which is not worse than alternative forecasts.

As mentioned above, in this study we want to focus the attention on the cointegration restrictions in a given sector price through regions. In fact, in this paper we do not pay attention to the national sectors cointegration restrictions studied in Espasa and Mayo (2012) and concentrate our efforts in studying sector cointegration through regions. Therefore we have a question of spatial cointegration, in which only one level of neighbours exists but we do not know the appropriate definition of neighbour and we need to try with different ones and chose the one which performs best.

The existence of spatial cointegration could be very different in breaking down the CPI of an economic area, like the Euro Area, in sectors through country members than applying such break to the sectors of the regions in a state economy. For this reason we apply our analysis to the 12 states of the Euro Area 12 and to the 17 regions of the Spanish economy.

It is widely recognized in the literature, that the question of which is the best procedure, direct or indirect, for forecasting an aggregate is mainly empirical. In our case, the question is if because of the curse of dimensionality the double breaking by sectors and regions performs worse than the breaks using just one criterion, sectors or regions. In the case of Spain we pass from 57 price indexes for the sectors available at the national level to 969 price indexes when considering these sectors within each one of the 17 Spanish regions.

This paper analyzes different strategies to forecast 600 and 960 monthly price indexes disaggregated by sectors and geographical areas in the Euro Area 12 and in Spain respectively. We deal with the curse of dimensionality problem by specifying and estimating ARIMA models as well as alternative spatial bi-dimensional vector equilibrium correction (SVeqCM) models where the price indices for each sector in a particular geographical area is allowed to be cointegrated with the corresponding price in neighbouring areas using different definitions of neighbourhood based on geographical, economic and sociological considerations.

Independently of the empirical answer to the question in the previous paragraph, when dealing with several hundreds of time series the presentation of results and forecasts in a way which could be simple to capture the main traits is quite crucial, otherwise many people is going to ignore the valuable outcome of this type of exercises. Using blanks, stars and colors we show, for instance, that results about the 969 Spanish price indexes can be presented in a friendly way.

The structure of this paper is as follows. Next section describes and analyzes the main features of the time series used in the paper. Section 3 presents and discusses the different methodologies considered to forecast inflation in Spain and the Euro Area 12 and a discussion of the forecasting results under these methodologies can be found in Section 4. Some concluding remarks follow in Section 5.

2. Data Description

We use both aggregate price indices as well as information related to different sectors and geographical areas. More specifically, we consider the following series: 1) the aggregate EA12 and Spanish Consumer Price Index; 2) price indexes for 50 sectors in the EA12 and 57 sectors in the Spanish economy; 3) aggregate price indexes for each of the 12 EA countries and the 17 Spanish regions; and 4) disaggregated price indexes the 12 EA countries and for 57 different sectors in the 17 available regions in Spanish [1]. Price series for the different Spanish regions (aggregated and disaggregated by sectors) are available from the Spanish Statistical Office (http://www.ine.es). At the European level, disaggregated price series by sectors and countries were obtained from the European Commission (http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home). Spanish series cover the 1993:01-2009:12 period while EA12 series are available from 1996:01 to 2009:12.We use data up to 2005 to estimate the models and the remaining four years (2006:01-2009:12) to compare the forecasts obtained under different strategies.

For the EA12 Eurostat offers weights of the different countries in order to map the aggregate inflation rate in the EA12 with the national inflation series. However, for Spanish regions this information is not available from the Spanish Office for National Statistics (INE). The problem is solved using as weights each region's share of expenditure in Spanish expenditure. Indeed, the inflation series obtained by this aggregation is almost identical to the official Spanish inflation rate. Weights at the sector level, on the other hand, are freely available from the Spanish Office for National Statistics (INE) and the Eurostat. These institutions formulate the aggregate price index for each region based on a chain Laspeyres price index in both cases. During the forecasting exercise, we aggregate inflation projections by using weights computed with information up to the last available period.[2,3]

In tables 1 and 2 we report descriptive statistics, similar to those in Beck et al. (2011),for Spain and the EA12 respectively. . In general,and consistently with Beck et al. (2011), Imbs et al. (2005) and Pesaran and Smith (1995), disaggregated inflation shows low levels of persistency that indicate that persistence of aggregated inflation comes as a result of aggregation bias that is generated by aggregating heterogeneous price series. A second fact we observe for both Spain and the EA12 is more heterogeneity across sectors than across geographical areas. Also, the last column of the tables indicate a relatively higher degree of comovement within sectors than among different sectors for a given region or country. This enhances the importance of taking into account links by sectors in different geographical areas in order to capture the dynamics of disaggregated series in an accurate way. Moreover, fairly heterogeneous values for mean and volatility for each of disaggregate series suggest the convenience of using disaggregated models by sectors and regions to have a complete picture of the Spanish and European inflation.

Figures of price series in levels are not shown to save space; however, the inspection reveals that most of them grow smoothly during the period under consideration. Series in first differences, on the

[1] A description of the sectors, regions, and countries is found in the Appendix.

[2] In the Spanish case, the Ceuta and Melilla region was broken down in two regions since 2007 and therefore it is not possible to have the complete series. Hence, given the low weight of these two regions that only represent 0.2% of the total national expenditure, in this particular case we restrict our analysis to the aggregated price index for Ceuta y Melilla in all the cases by aggregating both series since 2007 according to the share in the total Spanish expenditure.

[3] In the case of the EA12 Consumer Price Index, the only irregularities are for the series of education in Belgium and other major durables for recreation and culture in Austria that are only available from 1999:12. Therefore, in these two cases, models were specified and estimated using the information available from that date. Also, other major durables for recreation and culture in the case of Spain was only available from 2006:01 and it was dropped from the analysis and weights were rescaled for this fact.

other hand, show regular crossing points and no obvious trend. Additionally, some series such as lamb, fish, and potatoes in the case of the Spanish communities and vegetables, package holidays, and accommodation services in case of EA12 countries exhibit a clear seasonal behaviour.

For a formal test on the number of unit roots in the series we employed the methodology proposed by Osborn et al. (1988) (OCSB henceforth) who extended the procedure of Hasza and Fuller (1982) to seasonal time series for monthly data. Although we are aware of other more sophisticated procedures to investigate the presence of seasonal unit roots such as the tests proposed by Franses (1991) and Beaulieu and Miron (1993), we choose the OCSB test because of simplicity enables us to determine whether or not to take seasonal differences instead of testing for unit roots one by one at each of the harmonic frequencies of the seasonal cycle.

Results of the test for the disaggregate commodities indicates that at the 5% confidence level the majority of the price series requires only one regular difference (and no seasonal differences) to become stationary. For example, at the 5% significant level, results of the tests indicate that for the five biggest Spanish communities, Andalusia, Catalonia, Madrid, Basque Country, and Valencia, the 77%, 77%, 75%, 72%, and 74% of their sectors can be considered integrated of order one respectively (the average of this proportion for the 17 Spanish communities is 77%). Moreover, at the same confidence level, in the EA12 countries these percentages are 88%, 80%, 84%, and 80% for Germany, Spain, France, and Italy which represent about the 80% of the weighting in the inflation of the EA12 (the average of this proportion for the 12 countries is 85%). Also, in the OCSB equation the null of not significant seasonal dummies is rejected at the 5% in 47%, 49%, 53%, 53%, and 42% of the series in Andalusia, Catalonia, Madrid, Basque Country, and Valencia respectively (the average for the 17 communities is 42%) whereas in the EA12 countries this hypothesis can be rejected in the 42%, 58%, 52%, and 46% of the cases for Germany, Spain, France, and Italy (the average for the 12 countries is 52%). As a robustness exercise, for the annual rate of inflations in each of the sectors in the different Spanish regions and countries in Europe we run the Pesaran (2006) panel unit root test that allows for cross sectional (spatial) dependence. Results of the test indicate that the null hypothesis is rejected at the 5% in all the cases for Spain and also for practically all the EA12 series with the only exception of actual rental for houses.

Consistently with this analysis, we specify econometric models in the following sections by assuming that the different price series are generated by unit root processes and allowing for deterministic seasonality in the cases seasonal dummies are jointly significant. However, for robustness we also consider projections obtained under ARIMA models based on alternative hypothesis about the number of unit roots in the models.

3. Strategies to forecast regional inflation by sectors in Spain and the Euro Area

In this Section we present the strategies to forecast annual inflation rates disaggregated by sectors and geographical areas in Spain and the EA12 for the period 2006:01-2009:12. We evaluate this forecast based on models applied to different degrees of disaggregation. More specifically, for both Spain and the EA12, we compare results obtained from a benchmark strategy, denoted by B, based on a simple ARIMA model specified for the aggregate inflation in Spain and the EA12, with those obtained from a number of alternative strategies that consider different econometric specifications. These strategies can be split in two main groups. The first one refers to the use of ARIMA models applied to disaggregated series by sectors and geographical areas in Spain and the EA12. The second approach is based on the specification and estimation of alternative spatial vector equilibrium correction (SVeqCM) models in where the price indexes for each sector is allowed to be cointegrated with prices in neighbouring geographical areas using different definitions of neighbourhood based on geographical, economic and sociological considerations as well as alternative definitions of neighbourhood based on cointegration tests.

The different approaches correspond to different ways to deal with the curse of dimensionality. In fact, under the first strategy each of the individual time series is restricted to depend only on its own past values whereas in the second strategy we allow for the presence of a long-run equilibrium between prices in the same sector for two different geographical areas but each of the individual series are not directly affected by the evolution of similar series in not neighborhood areas.

In all the cases, we forecast inflation by following a recursive scheme; see for example Faust et al. (2005) and West (2006). Under this approach, the size of the sample used to estimate the parameters of the different model grows as one makes predictions for subsequent observations allows the parameter of the econometric models to change according to the new information of the sample.

In the remaining of this section we explain the main features of the two big groups of methodologies used in this paper to forecast inflation in Spain and The EA12.

3.1. Disaggregated ARIMA models by sectors and geographical areas.

The first alternative strategy (A1 henceforth), obtains inflation forecasts in the EA12 and Spain from aggregating projections of ARIMA models for each of the 12 European countries and 18 Spanish regions respectively. Under the second strategy, denoted by A2, we specify ARIMA models for price indices in 57 Spanish sectors and 50 sectors in the EA12. The third strategy (A3), considers both sector and geographic disaggregation. Thus inflation forecasts in each Spanish region and each European country can be obtained from the aggregation of projections in the different sectors of that specific geographical area and they can be aggregated again to obtain the overall inflation forecast in Spain and the EA12.
In all cases, our ARIMA models are specified using the TRAMO/SEATS automatic procedure; see Gomez and Maravall (1996).

3.2. Vector Equilibrium Correction (VeqCM) Models with Spatial Cointegration

We also consider VeqCM models to characterize the dynamic pattern for each of the disaggregated (by sectors and geographical areas) price series. The baseline model takes the form

$$\Delta p_{i,j,t} = \gamma_{i,j} + \alpha_{i,j}[\beta_{i,j}{}' \quad \delta_{i,j}]\begin{bmatrix} p_{i,j,t-1} \\ 1 \end{bmatrix} + \Phi_{i,j}\Delta p_{i,j,t-1} + \Gamma_{i,j}D_{1t} + \Psi_{i,j}D_{2t} + \varepsilon_{i,j,t} \quad (1)$$

where $p_{i,j,t}$ is a (2x1) vector containing (logs of) price levels in sector i for a region or country and its neighbour to be defined; $\alpha_{i,j}$ and β_{ij} are respectively the (2x1) adjustment and cointegration vectors; $\delta_{i,j}$ is a scalar which allows for a constant in the cointegration relationship; ; $\gamma_{i,j}$ is a (2x1) vector of intercept parameters; D_{1t} includes seasonal dummies and Γ_{ij} is the matrix of parameters associated to these interventions; D_{2t} are centered seasonal dummies that only takes nonzero values from 2002:01 to take into account the structural break in the seasonal pattern of many disaggregated series in Spain and the EA12 and $\Psi_{i,j}$ is the matrix of parameters associated to this second group of interventions; and ε_{ijt} is a (2x1) vector of serially uncorrelated errors.

The intuition of this VeqCM model is very similar in nature to the Space-Time AR models proposed by Giacomini and Granger (2004). They propose a model that assumes that spatial effects take one period to become manifest and ignores dependence beyond the first temporal and spatial lag. Two important differences between that paper and our approach are: (1) we allow for a cointegration relationship with the neighbour price; and (2) VeqCM systems with two equations are considered, one for the regional price and another for the neighbour price (instead of imposing neighbouring series to be exogenous as in Giacomini and Granger, (2004).

The number of lags in equation (1) is chosen to be equal to 1 as this is the specification that minimizes the Schwarz and Akaike criterion in almost all cases in both Spain and the Euro Area. In fact, this specification is the most parmonious way to describe both the short and long run dynamics of the prices series. Besides, note that the proposed model allows for a constant (but not a deterministic linear trend) in the cointegration relationship. This is because, a deterministic linear trend in the cointegration relationship amounts to imposing the assumption that prices in the different geographical areas diverge as the forecasting period increases. This specification is not useful to forecast as the linear deterministic trend in the cointegration equation can be interpreted as a proxy for other variables not included in the model and it is reasonable to think that they could be subject to structural shocks during the forecasting period.

In many Spanish and European series there is a structural break from the period 2002:01 that can be explained by a methodological change in the way that series were collected. This is the case, for example, of different prices for shoes and clothes in both Spain and the EA12. We account for this change in the seasonal pattern by allowing the set of seasonal dummy variables in (1) to have a different impact before and after the break period. Then, we test for each new observation using an F-statistics, whether seasonality can be captured with or without a structural break or if there is seasonality at all. In the initial estimation and during the forecasting exercise, at each period $T+h$, a F-test for deterministic seasonality is run.

We consider a specification similar to (1) for the following alternative definition of neigbours : 1) Price indexes in the whole area (Spain or EA12) for that sector (C1); 2) Geographical areas with similar economic growth (C2); 3) Similar per-capita income (C3); 4) Similar macroeconomic conditions (C4); 5) Similar density of population (C5); 6) geographical contiguity (C6).[4]

Besides, we used two other definitions of neighbourhood. The first one (C7) is based on the cointegration test proposed by Johansen (1995) and consider that the neighbours for a price index in sector i and region j is the average of all the price indices for that sector in all the other regions for which the null of no cointegration is rejected at the 5% level. The second strategy (C8) defines neighbourhood from an ADF tests applied to relative price indices for all the possible pairs of geographical areas in a given sector. Then we consider as the set of neighbours the average of all the prices for which the null of nonstationarity is rejected at the 5%.

In the case of the last two strategies, the econometric tests for cointegration and unit roots are repeated for each period during the forecasting exercise. This allows for a flexible definition of neighbours that could be different at different time periods. In the few cases where we did not find long run relationships either under C7 or C8, an unrestricted VAR model for variables in first differences with the region in which the model shows the lowest Schwarz criterium for the whole system of equations.

Note that each these definitions impose a single concept of neighbourhood for all the price indexes across sectors. However, it seems natural to assume that different concept of neighbourhoods could be applied to different sectors. In order to account for this fact, an the strategy C9) is considered that select the model with the lowest Schwarz criterion in the relevant equation for each of the strategies already presented (A3 and C1 to C8).

It is also possible that there is some combination of spatial VeqCM and ARIMA models that are not considered in any of the previous strategies and could improve the forecast of overall inflation. In order to explore this issue, two *ex-post* additional strategies are defined. The first one (C10) select for each individual price index the strategy (A3 and C1 to C8) that provides the best individual inflation forecast according to RMSFE and then aggregate all of them to obtain the overall inflation forecast for Spain and EA12. The second one (C11) consists of estimating inflation in a given sector for Spain or the EA12 using the best strategy from comparing the RMSFE obtained from an aggregated ARIMA model and that obtained from the best strategy according to all the alternative definitions of neighbourhoods. Then, inflation forecasts in the different sectors are aggregated to estimate the overall rate of inflation in Spain or the EA12. Note that RMSFE under strategies C10 and C11 can only be obtained after inflation data is known. Therefore they cannot be consider as competing strategies by as a way to observe the best forecast that could be obtained if the best model was used in each case.

[4] A description of the series contained in the different groups of neighbours for each strategy can be obtained from the authors upon request.

4. Results

4.1. Cointegration analysis and forecasting inflation in Spanish regions.

One important problem in order to evaluate the forecast of the Spanish inflation is the high degree of erraticity in the inflation rates after the economic crisis at the end of 2008. Hence, for robustness we evaluate the performance of the different forecasting strategies for both the periods 2006:1-2009:12 and 2006:01-2008:12. Table 3 shows the root mean square forecast error (RMSFE) of the benchmark strategy and the RMSFE of each of the alternative strategies relative to the benchmark. A relative RMSFE smaller than one indicates an improvement of this forecast with respect to the benchmark one. The table also indicates whether the forecasts are significantly different using the modified Diebold and Mariano (1995) test proposed by Harvey et al. (1997).

[INSERT TABLE 3 AROUND HERE]

As expected, the economic crisis has influenced negatively the accuracy of forecasts under all the strategies. However, main conclusions about the relative efficiency of the different methodologies are unaffected by this consideration. In fact, we find in both cases that geographical consideration alone is not relevant or even disruptive while the use of sector disaggregation only implies a significant improvement. Moreover, using double disaggregation by sectors and geographical areas marginally improve inflation forecast compared to the case that only considers disaggregated models by sectors. In fact, the modified DM test to compare strategies A2 and A3 takes values of 1.62 and 1.51 for 1 and 4 periods ahead that although are not significant at the conventional levels, they are close to the rejection areas at the 5% and they are indeed rejected at the 10% significant level. For longer horizons, the values of these statistics surpass in many cases the critical values at the 5% significant level. For example, the values of the statistics for the 9, 10, 11 and 12 periods ahead forecast are 3.82, 3.61, 3.68 and 2.17 and the null hypothesis is rejected at the 5% in all these cases.

[INSERT FIGURE 1 AROUND HERE]

A relevant question is whether to use an univariate ARIMA model or a spatial bi-dimensional vector equilibrium correction (VeqCM) model to forecast inflation for each of the 969 disaggregated series. For simplicity we focus the analysis in the comparison of strategies A3 and C1 with respect to A2 (in case of sector inflation forecast) and A1 (in case of geographical inflation forecast). We do this because from Table 3, strategy C1 provides a slightly better inflation forecast than the other approaches (C2 to C6). Also, although we have shown in Table 3 that including 2009 in the analysis has an important effect in the accuracy of the forecast, conclusions about the relative efficiency of the different strategies uphold when 2009 is considered. For this reason, in the remaining of this section is based on analysis that includes 2009 in all cases.

Tables 4 and 5 show the best strategy to forecast inflation at horizons one and twelve months respectively for each of the series under analysis. Table 6 summarizes the main results by sectors and regions. For the one-step ahead forecast, it can be observed that the best forecast for the aggregate of a region is always obtained under strategy C1. Also, the majority of the 969 inflation series are better forecasted by considering cointegration relationships with Spain instead of using ARIMA models. For longer horizons, see table 6, the opposite is true. This is consistent with Christoffersen and Diebold (1998) who find that vector equilibrium correction (VeqCM) models are particularly useful to forecast in the short run as it identifies situations of disequilibrium and indicates the dynamic of the variables in the model to return to equilibrium in the subsequent periods. On the other hand, if the purpose of the analysis is to predict inflation in each of the 57 sectors, a substantial number of them are better forecasted by specifying ARIMA models to the aggregate series.

[INSERT TABLE 4 AROUND HERE]
[INSERT TABLE 5 AROUND HERE]
[INSERT TABLE 6 AROUND HERE]

A series of robustness exercises were also run but not explicitly reported here for the sake of brevity[5]. First, following and important branch of the economic literature on macroeconomic modelling[6], we have also estimate dynamic common factors by principal components for all the regional inflation series in each of the sectors and then we have used the approaches described by Boivin and Ng (2006) and Schumacher and Breitung (2008) to forecast each of the 969 inflation series. However, these strategies do not improve the forecast of the inflation series is most cases. Second, we considered model transformations based on additional differencing the SVeqCM to reduce forecast-error biases, see Hendry (2006), with no important differences in the results of the analysis. In a final set of experiments, for each of the 969 disaggregated prices we specify single equations where we allow the dependent variable to react to different cointegration relationships following a spirit similar to Aron and Muelbauer (2012). In this case we include in each of the equations reactions to price differences between the region in question and each of the other regions. However, this specification did not improve inflation forecast in most cases.

The analysis developed in this section is not only useful to forecast inflation but also to get a better insight about cointegration links. This is particularly useful to identify the sectors and areas prone to be affected by problems of competitiveness. Table 7 classifies Spanish regions according to its number of cointegrated sectors, but not clear pattern emerge fromthis table. Table 8 identifies the sectors that have the most and the least number of regions cointegrated with Spain for the groups of food, industry and energy and serviceThe presence of long-run relationships in food is more common in fish, cereales and alcoholic drinks; in other goods in clothes and footwear; and in services medical services, some education items, publications, some repairing services, rental apartments and recreational objects.
[INSERT TABLE 7 AROUND HERE]
[INSERT TABLE 8 AROUND HERE]

4.2. Cointegration analysis and forecasting inflation in the EA12.

[5] Results of these experiments are available from the authors upon request.

[6] See, for example, Marcelino et al. (2006) and the references therein.

In the same vein that the Spanish inflation case, the Table 9 shows the RMSFE for the benchmark strategy and the relative mean square forecast errors (RMSFE) obtained under different strategies. For simplicity, we omit from this table results with strategies C2 to C11 given that they are very similar to the ones obtained under C1. Table 10 reports a comparative evaluation of strategies A2, A3 and C1 following Harvey et al. (1997).

Unlike the Spanish case, our results do not support disaggregation by regions and sectors in order to forecast EA12 headline inflation. In fact, strategy A2 provides a lower, although not significantly different, RMSFE for all forecast horizons than the alternative disaggregated methodologies. The different importance of the geographical dimension in the analysis of Spanish and AE12 inflation data could reflect that spatial links are stronger between the regions of a country than in the countries of an economic union.

[INSERT TABLE 9]

[INSERT TABLE 10]

However, note that forecasting aggregate inflation in the EA12 is not the only scope of the 600 models considered by each of the disaggregated strategies A3 and C1. Indeed, the analyst could be interested in forecasting inflation by countries, by sectors or both.

When the purpose is to forecast the inflation for a particular country, strategies A3 and C1 provides a better performance than A1 in all cases for the one-step ahead forecast. This superiority of disaggregate models to forecast inflation by countries is also evident at longer horizons. For example, for the 12 step-ahead forecasts, strategies A3 and C1 outperform A1 in 8 out of 12 countries.

Table 11 shows the relative performance of strategies A2, A3 and C1 in order to forecast inflation in each of the EA12 sectors. At horizon one, A2 is clearly the best strategy in 44 out of 50 sectors from which they are significant at the 0.05 in 32 cases. Inflation forecast could only be significantly improved by considering either A3 or C1 in 3 sectors.

[INSERT TABLE 11]

It is also of interest to compare strategies A3 and C1 to forecast inflation in each of the 50 sectors of the 12 countries. Focusing on the 1 step-ahead forecast, cointegration improves inflation forecast compared to simple extrapolative devices in at least 8 out of 12 countries for sectors: s1: bread and cereals, s3: fish and seafood, s5: oils and fat, s14: beer, s38: major appliances and s48: recreational objects. Note that all of them correspond to tradable goods whose prices, due to the possibility of arbitrage, are not expected to diverge through countries.

Cointegration results for countries and sectors are shown in tables 12 and 13 respectively. We do not observe important differences by countries. Regarding cointegration by sectors, the main features are broadly consistent with those obtained at the regional level for Spain. Specifically, perishable food products and local services do not cointegrate much while long run relationships are stronger in homogeneous and durable tradable goods.

[INSERT TABLE 12]

[INSERT TABLE 13]

5. Concluding remarks

In this paper we have studied the performance of different strategies to forecast 969 and 600 monthly price indexes disaggregated by sectors and geographical areas in Spain and the EA12 respectively. We have dealt with the curse of dimensionality problem avoiding modelling with vectors of dimension higher than two. Thus we specify and estimate ARIMA models as well as alternative spatial bi-dimensional vector equilibrium correction (SVeqCM) models where the price indexes for each sector is allowed to be cointegrated with prices in neighbouring geographical areas using different definitions of neighbourhood based on geographical, economic and sociological considerations. In the case of Spain we show that the forecasts of the disaggregated price indexes by sectors and regions provides a huge information about inflation and relative prices is trustworthy since by aggregating them we obtain a more accurate forecast of the aggregated inflation than the ones proved by other much simpler methods.

The results for Spain show that disaggregating by just one criterion, sector or region, the former is more efficient than the latter in forecasting the Spanish headline inflation (this is more determining for the EA12). These results confirm the ones given by Espasa and Albacete (2007), who use a much more reduced disaggregation levels. On the other hand, disaggregating using both criteria, the results improve over the ones using only sector disaggregation. The is relevant because it points out that passing from dealing with 57 aggregated-sector series to work with 969 sectors through all the regions, we do not get worse results and we can provide a much wider forecasting information through sectors within regions. This turns to be really important when putting the results for the highly disaggregated series in relative terms with respect to Spain or other regions. Including spatial cointegration restrictions does not help in improving the aggregate forecast, but for sectors within the regions, cointegration helps in about the 50% of the 969 cases. In this sense the paper provides evidence, see table 9, of which are the sectors that can be considered with higher cointegration levels.

For the EA12, geographical disaggregation is not very important even when taken in consideration jointly with the sector disaggregation. This suggests that the degree of economic integration is very important to make the geographical disaggregation analysis useful *and that in the Euro Area the regional analysis within countries seems very appropriate.*

Much work remains to be done to explore, for example, how these approaches can be used to forecast inflation in the US and to check whether a better forecast of inflation in the Euro Area could be obtained by modelling spatial links across regional prices in each country. Another question for the future is to consider cointegration through sectors as in Espasa and Mayo (2012) and spatial cointegration together.

References

Aron, J. and Muellbauer, J., (2012), "Improving forecasting in an emerging economy, South Africa: Changing trends, long run restrictions *and disaggregation", International Journal of Forecasting, 28, 456-476.*

Beaulieu, J.J. and Miron, J.A., (1993), "Seasonal unit roots in aggregate U.S data", Journal of Econometrics, 55, 305-328.

Beck, G., Hubrich , K., and Marcellino, M. (2009), "Regional inflation dynamics within and across euro area countries and a comparison with the United States", Economic Policy, 24, 141-184.

Beck, G., Hubrich , K., and Marcellino, M. (2011), "On the importance of sectoral and regional shocks for price-setting", CEPR Discussion Papers 8357, C.E.P.R. Discussion Papers.

Boivin, J. and Ng, S., (2006), "Are more data always better for factor analysis?", Journal of Econometrics, 132, 169-194.

Christoffersen, P.F. and Diebold, F.X., (1998), "Cointegration and Long-Horizon Forecasting", Journal of Business and Economic Statistics, 16, 450-458.

Diebold, F. and Mariano, R., (1995), "Comparing predictive accuracy", Journal of Business and Economic Statistics, 13, 253-265. Published again in Journal of Business and Economic Statistics, 20(1), 135-145, January 2002.

Espasa, A. and Albacete, R., (2007), "Econometric Modelling for Short-Term Inflation Forecasting in the EMU", Journal of Forecasting, 26, 303-316.

Espasa, A. and Mayo, I., (2012), "Forecasting aggregates and disaggregates with common features", International Journal of Forecasting, forthcoming.

Faust, J., Rogers, J., and Wright, J., (2005), "News and Noise in G-7 GDP Announcements", Journal of Money, Credit and Banking, 37, 403-19.

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

Giacomini, R. and Granger, C.W.J., (2004), "Aggregation of space-time processes", Journal of Econometrics, 118, 7-26.

Gomez, V. and Maravall, A., (1996), "Programs TRAMO and SEATS, Instruction for Users (Beta version: September 1996)", Banco de España, WP9628, *Harvey, D., Leybourne, S., and Newbold, P., (1997), "Testing the Equality of Prediction Mean Square Errors", International Journal of Forecasting, 13, 281-291.*

Hasza, D.P. and W.A. Fuller, (1982), "Testing for Nonstationary Parameter Specifications in Seasonal Time Series Models", The Annals of Statistics, 10, 1209-1216.

Hendry, D.F., (2006), "Robustifying forecast for equilibrium-correction models", Journal of Econometrics, 135, 399-426.

Hendry, D.F. and Hubrich, K., (2011), "Combining disaggregate forecasts or combining disaggregate information to forecast an aggregate", Journal of Business and Economic Statistics, 29, 216-227.

Imbs, J., Mumtaz, H., Ravn, M.O and Rey, H., (2005), "PPP strikes back: Aggregation and the real exchange rate.", Quarterly Journal of Economics, 120:1-44.

Johansen, S., (1995), Likelihood-based inference in cointegrated vector autoregressive models, Oxford University Press.

Osborn, D.R., Chui, A.P.L., Smith, J.P. and Birchenhall, C.R., (1988), "Seasonality and the order of integration for consumption", Oxford Bulletin of Economics and Statistics, 50, 361-377.

Pesaran M., (2006), "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure", Econometrica, 74, 967-1012.

Pesaran, M.H. and Smith, R., (1995), "Estimating long-run relationships from dynamic heterogeneous panels", Journal of Econometrics, 68:79-113.

Schumacher, C. and Breitung, J., (2008), "Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data", International Journal of Forecasting, 24, 386-398.

West, K.D. (2006), "Forecast Evaluation", Handbook of Economic Forecasting, Elliott G., Granger C.W.J., and Timmermann A. (eds.), North Holland.

Table 1. *Descriptive statistics: Spanish inflation, disaggregation by sectors and regions*

	Level		Volatility		Persistence		Disp	Corr(xi,x)
	Mean	Std	Mean	Std	Mean	Std		
Overall Inflation	2.93	0.15	14.77	1.89	0.20	0.03	17.44	0.20
				Autonomous Communities				
Andalusia	2.81	1.53	14.43	16.39	0.23	0.18	17.35	0.20
Aragon	2.92	1.57	16.37	15.44	0.18	0.19	19.34	0.20
Asturias	2.90	1.57	16.75	18.05	0.17	0.17	20.13	0.16
Balearic Island	3.04	1.59	14.47	10.61	0.14	0.21	16.01	0.23
Canary Island	2.69	1.57	13.10	11.05	0.19	0.17	14.65	0.25
Cantabria	2.84	1.46	16.05	14.00	0.16	0.19	18.95	0.21
Castilla y Leon	2.87	1.54	14.79	15.23	0.25	0.17	18.18	0.20
Castilla la Mancha	2.88	1.48	15.84	16.74	0.22	0.17	19.10	0.21
Catalunya	3.17	1.45	14.56	14.92	0.20	0.19	17.11	0.19
C. Valenciana	2.87	1.50	14.57	15.38	0.19	0.18	17.25	0.21
Extremadura	2.72	1.60	16.33	16.19	0.15	0.19	19.05	0.20
Galicia	2.92	1.41	15.58	16.11	0.23	0.17	17.82	0.21
Madrid	2.83	1.58	13.49	12.59	0.17	0.19	15.89	0.20
Murcia	3.14	1.49	17.70	18.67	0.15	0.20	20.86	0.21
Navarra	3.15	1.47	17.79	15.50	0.15	0.21	20.24	0.18
Basque Country	3.08	1.57	14.92	15.43	0.18	0.17	17.87	0.18
Rioja	3.22	1.44	19.71	22.30	0.16	0.17	23.02	0.14
				Sectors				
s1	2.16	0.32	6.54	1.40	0.07	0.13	5.04	0.64
s2	4.15	0.66	10.49	2.81	0.23	0.15	7.22	0.64
s3	3.29	0.54	12.43	3.18	0.27	0.13	9.08	0.68
s4	3.80	0.67	55.78	9.49	0.46	0.07	23.51	0.89
s5	1.85	0.44	28.97	6.18	0.34	0.06	11.86	0.90
s6	2.40	0.36	49.93	11.45	0.10	0.06	18.42	0.91
s7	1.95	0.37	6.59	1.56	0.18	0.14	5.00	0.67
s8	2.21	0.37	42.32	11.36	-0.08	0.06	18.63	0.88
s9	3.26	0.52	12.75	3.16	0.03	0.09	10.06	0.61
s10	2.76	0.65	20.27	5.62	0.27	0.13	13.46	0.76
s11	2.51	0.28	14.97	1.35	0.53	0.09	5.92	0.90
s12	2.10	0.26	9.34	1.68	0.13	0.10	6.05	0.77
s13	3.13	0.20	30.31	2.30	0.57	0.06	10.33	0.92
s14	4.41	0.49	10.62	1.96	0.69	0.04	3.98	0.93
s15	3.70	0.43	10.49	2.67	0.29	0.17	7.96	0.66

s16	4.63	0.35	13.09	1.72	0.57	0.04	4.93	0.92
s17	2.20	0.32	8.62	1.88	0.12	0.15	6.83	0.62
s18	5.10	0.83	86.71	19.70	0.43	0.07	41.26	0.85
s19	2.83	0.37	17.36	2.65	0.49	0.13	9.37	0.82
s20	0.73	0.53	9.37	2.57	0.16	0.18	6.80	0.66
s21	2.39	0.22	4.90	1.26	0.11	0.09	4.08	0.60
s22	1.53	0.51	10.43	2.37	-0.03	0.12	7.87	0.61
s23	2.82	0.36	7.65	2.84	0.20	0.14	5.54	0.66
s24	6.79	0.13	22.04	1.13	-0.01	0.08	2.85	0.96
s25	1.91	0.42	47.06	6.33	0.28	0.04	7.67	0.98
s26	1.82	0.58	59.67	6.70	0.23	0.03	8.74	0.98
s27	1.81	0.80	69.31	11.60	0.16	0.04	10.87	0.98
s28	2.66	0.35	33.58	7.45	0.20	0.06	9.57	0.94
s29	2.56	0.53	34.35	7.12	0.24	0.03	9.32	0.95
s30	2.73	0.70	48.03	5.86	0.21	0.04	10.17	0.96
s31	2.36	0.67	45.85	10.46	0.22	0.04	12.58	0.95
s32	4.30	0.31	6.68	1.79	0.10	0.08	5.17	0.57
s33	4.50	0.36	3.71	1.10	0.33	0.13	2.74	0.62
s34	2.86	0.41	11.96	2.31	0.15	0.08	4.60	0.90
s35	4.12	0.23	5.52	1.71	0.14	0.08	4.39	0.52
s36	2.95	0.45	6.72	2.72	0.17	0.06	3.80	0.76
s37	2.22	0.40	16.78	3.72	0.12	0.05	5.82	0.93
s38	0.21	0.31	3.63	0.92	0.10	0.07	2.86	0.54
s39	2.87	0.26	5.38	1.22	0.11	0.06	4.16	0.60
s40	1.54	0.27	6.70	1.50	0.09	0.10	5.62	0.55
s41	4.36	0.38	6.23	1.21	0.13	0.08	4.30	0.65
s42	3.96	0.39	6.48	1.49	0.12	0.06	3.12	0.82
s43	0.18	0.23	9.42	0.68	-0.04	0.05	2.83	0.95
s44	3.01	0.09	10.71	0.64	0.43	0.01	1.45	0.99
s45	4.83	0.45	12.36	2.91	0.03	0.12	6.91	0.62
s46	4.41	0.46	9.76	2.85	0.20	0.07	4.23	0.87
s47	0.37	0.14	14.08	0.64	0.01	0.00	0.35	1.00
s48	-2.31	0.59	5.69	0.94	0.20	0.13	3.91	0.72
s49	2.75	0.18	5.37	0.41	0.18	0.05	1.71	0.88
s50	2.89	0.48	10.93	4.79	-0.18	0.15	9.60	0.39
s51	4.71	0.98	12.76	1.95	0.03	0.09	6.12	0.64
s52	4.55	0.64	13.69	4.70	0.06	0.08	4.42	0.81
s53	4.80	0.02	17.48	0.25	-0.07	0.00	0.13	1.00

s54	3.12	0.41	4.73	1.39	0.07	0.10	3.76	0.53
s55	2.85	0.32	4.24	0.76	0.16	0.09	3.31	0.62
s56	4.13	0.21	10.24	2.09	0.11	0.05	3.43	0.95
s57	3.55	0.25	5.36	0.78	0.10	0.07	2.90	0.78

This table reports descriptive statistics for monthly inflation rates disaggregated by sectors and regions. The reported statistics include the weighted mean and the standard deviation (std) of the time-series means of all inflation series included in a given group (level), the weighted mean and the standard deviation (std) of the time series standard deviation of all inflation series included in a given group (volatility), the weighted mean and the standard deviation (std) of the persistence measures of all inflation series included in a given group, the average over time of the cross-sectional dispersion of all inflation series included in a given group and the weighted mean of the correlation of all inflation series included in a given group with the group aggregate inflation rate. The measure for persistence is based on the weighted sum of the first order autocorrelation for all the series.

Table 2. Descriptive statistics: EA12 inflation, disaggregation by sectors and countries

	Level		Volatility		Persistence		Disp	Corr(xi,x)
	Mean	Std	Mean	Std	Mean	Std		
Overall Inflation	1.99	0.48	14.09	9.17	0.12	0.07	20.64	0.15
				Countries				
Germany	1.62	1.83	14.47	27.11	0.08	0.23	24.85	0.06
Austria	1.70	1.49	14.19	15.94	0.08	0.19	19.54	0.12
Belgium	1.89	1.52	21.56	38.79	0.11	0.29	34.00	0.14
Spain	2.69	1.81	14.27	19.03	0.29	0.20	20.17	0.16
Finland	1.70	1.80	15.22	15.83	-0.02	0.16	19.84	0.20
France	1.54	1.86	10.88	14.18	0.12	0.25	16.14	0.18
Greece	3.31	1.94	28.26	30.85	0.06	0.21	34.72	0.26
Netherland	1.97	2.18	18.13	23.71	0.05	0.15	26.13	0.17
Ireland	2.45	2.71	15.62	17.81	0.13	0.23	19.13	0.24
Italy	2.26	1.31	11.63	14.97	0.11	0.26	14.19	0.19
Luxembourg	2.32	1.65	14.92	14.28	-0.10	0.24	16.57	0.26
Portugal	2.52	1.65	13.81	15.88	0.14	0.19	18.06	0.13
				Sectors				
s1	2.11	0.67	3.76	1.33	0.57	0.16	2.47	0.75
s2	1.87	0.56	5.77	2.78	0.42	0.19	4.82	0.65
s3	2.61	0.47	12.56	5.14	-0.11	0.19	10.43	0.56
s4	1.54	0.61	7.18	2.74	0.55	0.11	3.93	0.72
s5	1.42	0.69	13.55	8.10	0.50	0.19	10.05	0.46
s6	2.26	0.81	37.62	25.83	0.36	0.21	33.17	0.59
s7	1.52	1.42	44.06	22.12	0.29	0.21	32.44	0.75
s8	1.64	0.52	3.98	1.37	0.31	0.22	3.28	0.54

s9	1.61	0.49	3.52	1.84	0.32	0.15	3.45	0.45
s10	0.69	0.53	8.70	5.26	0.44	0.21	7.37	0.67
s11	1.32	0.57	4.68	2.62	0.17	0.11	4.50	0.50
s12	1.41	0.93	5.11	5.89	0.10	0.21	5.17	0.37
s13	1.86	0.78	4.31	2.21	0.18	0.20	3.94	0.51
s14	1.90	0.76	5.50	3.24	0.12	0.18	5.02	0.36
s15	5.23	0.84	17.25	3.00	0.05	0.14	10.33	0.46
s16	0.96	1.24	49.65	33.15	0.06	0.19	36.98	0.83
s17	1.47	1.20	42.81	25.00	0.10	0.17	29.88	0.83
s18	2.54	1.07	3.18	2.65	0.12	0.31	3.42	0.33
s19	2.51	0.77	4.40	1.34	0.10	0.27	3.41	0.48
s20	3.21	0.94	7.26	17.32	0.09	0.16	6.42	0.56
s21	3.15	1.04	15.35	6.07	0.15	0.07	10.33	0.71
s22	1.58	0.73	5.68	4.17	-0.08	0.20	5.40	0.53
s23	1.05	0.99	16.98	14.61	0.02	0.15	16.05	0.62
s24	-0.63	0.74	5.15	8.15	-0.04	0.16	6.64	0.43
s25	2.70	0.50	5.55	2.34	0.06	0.22	4.17	0.40
s26	1.78	0.66	6.67	13.52	0.03	0.13	10.97	0.52
s27	1.23	0.79	3.08	1.94	-0.12	0.18	3.08	0.40
s28	1.05	0.78	3.28	2.16	0.33	0.17	3.42	0.47
s29	2.81	1.16	6.88	3.46	-0.02	0.15	5.44	0.30
s30	2.24	0.68	8.72	3.74	-0.01	0.06	4.55	0.48
s31	0.82	0.62	5.56	2.91	-0.04	0.18	4.58	0.43
s32	0.74	0.56	5.72	3.39	0.03	0.11	5.14	0.42
s33	1.14	0.45	3.87	1.74	0.10	0.12	3.33	0.41
s34	3.31	0.56	31.51	4.55	0.29	0.14	12.08	0.91
s35	3.29	0.90	4.34	1.81	0.10	0.29	3.65	0.56
s36	2.45	1.15	8.26	7.35	-0.13	0.15	6.54	0.46
s37	2.94	1.06	17.57	7.96	-0.16	0.13	13.22	0.60
s38	2.56	1.66	19.21	19.47	-0.03	0.11	13.81	0.38
s39	-2.60	0.90	10.86	2.29	0.05	0.07	7.59	0.44
s40	-6.13	1.55	7.06	3.98	0.21	0.29	6.41	0.53
s41	1.43	0.71	6.45	4.44	-0.02	0.10	4.51	0.34
s42	0.88	0.52	8.56	4.03	0.09	0.23	7.02	0.64
s43	2.40	0.54	11.03	21.33	-0.09	0.13	7.83	0.36
s44	2.19	0.55	4.40	2.89	0.02	0.10	4.25	0.41
s45	3.58	1.43	86.36	39.08	-0.23	0.26	58.44	0.58
s46	3.14	0.94	12.31	7.09	0.06	0.12	5.42	0.50

s47	2.56	0.84	2.95	3.13	0.18	0.15	3.07	0.58
s48	2.49	1.03	5.44	2.90	-0.01	0.16	4.07	0.50
s49	3.10	0.63	56.82	32.61	0.00	0.25	41.22	0.65
s50	2.23	0.61	3.31	1.40	0.07	0.13	2.81	0.55

This table reports descriptive statistics for monthly inflation rates disaggregated by sectors and countries. The reported statistics include the weighted mean and the standard deviation (std) of the time-series weighted means of all inflation series included in a given group (level), the weighted mean and the standard deviation (std) of the time series standard deviation of all inflation series included in a given group (volatility), the weighted mean and the standard deviation (std) of the persistence measures of all inflation series included in a given group, the average over time of the cross-sectional dispersion of all inflation series included in a given group and the weighted mean of the correlation of all inflation series included in a given group with the group aggregate inflation rate. The measure for persistence is based on the weighted sum of the first order autocorrelation for all the series.. The measure for persistence is the first order autocorrelation.

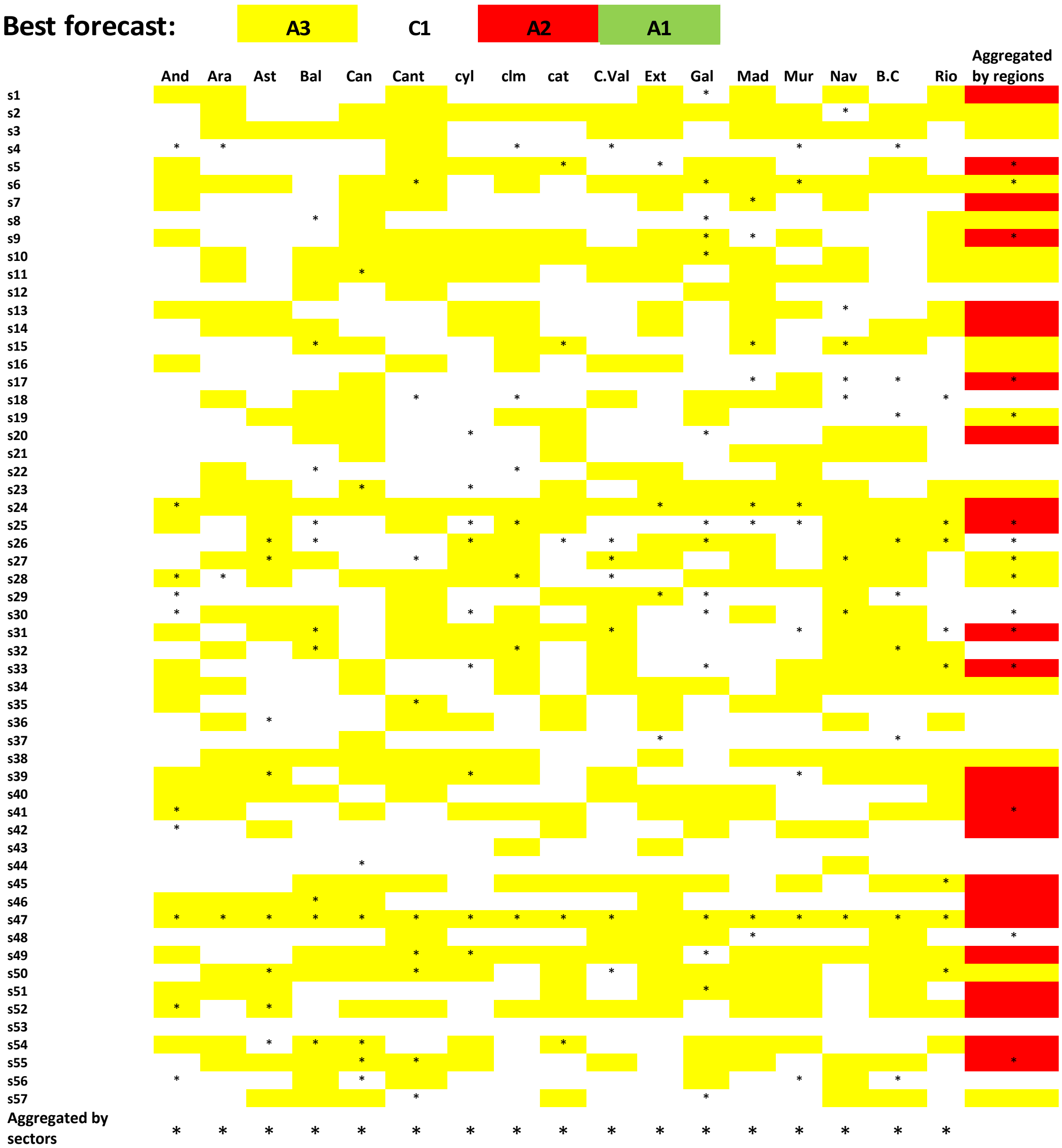
Table 3. *Spanish RMSFE of the Benchmark strategy and relative RMSFE with respect to Benchmark under alternative strategies.*

	Period 2006:01-2009:12				**Period 2006:01-2008:12**			
	1P	**4P**	**8P**	**12P**	**1P**	**4P**	**8P**	**12P**
B	0.66	1.97	2.93	3.65	0.56	1.51	1.85	2.14
A1	0.91(*)	0.99	1.03	1.07	0.89 (*)	1.04	1.04	1.10
A2	0.63(**)	0.71(*)	0.80	0.84	0.66 (**)	0.77	0.75	0.83
A3	0.62(**)	0.69(*)	0.77(*)	0.82	0.64 (**)	0.75	0.71	0.79
C1	0.60(**)	0.70(*)	0.82(*)	0.87	0.64 (**)	0.77 (*)	0.74 (*)	0.82
C2	0.60(**)	0.70(*)	0.82(*)	0.89	0.64 (**)	0.77 (*)	0.76 (*)	0.85
C3	0.60(**)	0.70(**)	0.83(*)	0.90	0.64 (**)	0.77 (*)	0.76 (*)	0.86
C4	0.60(**)	0.71(*)	0.82(*)	0.88	0.64 (**)	0.77 (*)	0.75 (*)	0.83
C5	0.60(**)	0.70(*)	0.81(*)	0.88	0.64 (**)	0.77 (*)	0.75 (*)	0.85
C6	0.63(**)	0.72(*)	0.83(*)	0.89	0.68 (**)	0.79 (*)	0.77 (*)	0.86
C7	0.60(**)	0.70(**)	0.82(*)	0.89	0.64 (**)	0.77 (*)	0.74 (*)	0.84
C8	0.60(**)	0.71(*)	0.82(*)	0.88	0.64 (**)	0.77 (*)	0.73 (*)	0.83
C9	0.62(**)	0.70(*)	0.79(*)	0.85	0.64 (**)	0.76	0.72	0.80
C10	0.59(**)	0.68(*)	0.77(*)	0.82	0.61 (**)	0.75 (*)	0.70 (*)	0.80
C11	0.57(**)	0.68(*)	0.76(*)	0.82	0.61 (**)	0.74 (*)	0.70 (*)	0.78

B: ARIMA model for the aggregate series; A1: ARIMA models applied to regions; A2: ARIMA models applied to sectors; A3: ARIMA models applied to sectors and regions; C1: SVeqCM with the whole area; C2: SVeqCM with similar economic growth; C3: SVeqCM with similar per-capita income; C4:SVeqCM wiht similar macroeconomic conditions; C5: VeqCM with similar density of population; C6: VeqCM with geographical contiguity; C7: SVeqCM with cointegrated regions using Johansen (1995); C8: SVeqCM with regions with stationary relative prices; C9: SVeqCM with neighbours selected according to the Schwarz criterium; C10: SVeqCM with the smallest RMSFE for each individual series; C11: SVeqCM with the smallest RMSFE for a given sector.

** (*) denotes rejection at the 0.01 (0.05) significance level relative to the benchmark strategy by using the modified Diebold and Mariano (1995) test as proposed by Harvey et al. (1997).

Table 4. *Best forecasting strategy according to RMSFE. One period ahead.*



C1 better than A3 #	And	Ara	Ast	Bal	Can	Cant	cyl	clm	cat	C.Val	Ext	Gal	Mad	Mur	Nav	B.C	Rio
	34(5)	30(2)	33(2)	32(4)	25(2)	26(3)	35(5)	28(3)	32(1)	33(4)	25(2)	33(9)	26(4)	32(5)	26(4)	27 (6)	30(2)

*Note: ** denotes rejection at the 0.05 significance level by using the modified Diebold and Mariano (1995) test as proposed by Harvey et al. (1997). This test has been run by comparing strategies A3 and C1 for each individual series, and A2 (A1) with respect to the best of A3 and C1 for aggregated regions (aggregated sectors).

≠ Between brackets the number of cases significant at the 5% level.

Table 5. *Best forecasting strategy according to RMSFE. Twelve period ahead.*

Best forecast: A3 C1 A2 A1

	And	Ara	Ast	Bal	Can	Cant	cyl	clm	cat	C.Val	Ext	Gal	Mad	Mur	Nav	B.C	Rio	Aggregated by regions
s1		*															*	
s2																		
s3																		
s4				*	*													*
s5																		
s6		*	*			*												
s7																	*	*
s8								*				*	*					
s9												*				*	*	
s10										*						*		
s11	*		*						*			*				*		
s12												*						
s13																		
s14	*													*			*	
s15			*	*	*				*	*								
s16			*															
s17		*						*										
s18																		
s19			*										*					*
s20		*					*					*						
s21	*								*			*		*	*			
s22			*				*		*					*				*
s23							*											
s24			*					*										
s25	*	*															*	*
s26		*	*			*								*	*		*	
s27					*				*				*					
s28										*								
s29	*						*									*		
s30				*														*
s31				*					*			*						
s32						*		*										*
s33							*					*				*	*	*
s34						*										*		
s35				*		*										*		*
s36																		
s37		*						*					*	*				*
s38								*				*						*
s39																		
s40				*							*				*	*	*	
s41									*					*				
s42											*				*		*	
s43	*			*	*		*		*	*	*	*	*	*	*	*	*	*
s44							*	*		*			*			*	*	
s45															*			
s46																		
s47		*				*										*		
s48			*	*	*			*					*		*		*	
s49											*							*
s50	*						*						*					
s51						*			*			*		*				*
s52		*							*									
s53																		*
s54								*	*									
s55														*				
s56			*															*
s57						*						*				*		
Aggregated by sectors																		
C1 better than A3 ≠	27(3)	28(4)	27(6)	23(6)	27(2)	26(2)	26(5)	27(6)	24(6)	26(4)	29(1)	28(8)	26(3)	28(4)	28(4)	29(7)	32(6)	

Note: * denotes rejection at the 0.05 significance level by using the modified Diebold and Mariano (1995) test as proposed by Harvey et al. (1997). This test has been run by comparing strategies A3 and C1 for each individual series, and A2 (A1) with respect to the best of A3 and C1 for aggregated regions (aggregated sectors).

≠ Between brackets the number of cases significant at the 5% level.

Table 6*. Best forecast under different strategies: A3, C1, A2. Spain and Euro Area 12.*

	Spain			Euro Area		
	Sectors	***Regions***	***Disaggregated***	***Sectors***	***Countries***	***Disaggregated***
			1 step-ahead forecast			
A2	24(8)			42(32)		
A3	16(4)	0(0)	462(78)	2(1)	2(2)	336(56)
C1	17(3)	17(17)	507(57)	6(2)	10(6)	264(42)
			4 step-ahead forecast			
A2	13(5)			34(20)		
A3	21(5)	9(9)	507(99)	10(1)	3(1)	324(69)
C1	23(0)	8(8)	462(67)	6(0)	9(3)	276(31)
			8 step-ahead forecast			
A2	17(0)			28(9)		
A3	19(1)	15(6)	522(79)	16(2)	3(0)	363(80)
C1	21(2)	2(2)	447(65)	6(0)	8(0)	237(41)
			12 step-ahead forecast			
A2	17(1)			22(4)		
A3	19(2)	15(0)	508(65)	18(3)	3(1)	368(44)
C1	21(4)	2(0)	461(74)	10(1)	5(0)	232(30)

The number of cases in which the strategy is significantly better at the 5% than the second best strategy is shown between brackets.

Table 7. *Cointegration by Spanish regions*

Region	Cointegration with Spain (I)	Cointegration with other regions (II)
Val	50	30
PV	48	32
Ast	48	32
Ara	48	27
Nav	47	35
Rio	47	30
Gal	46	27
Can	45	31
CYL	45	25
Cat	45	25
I.Can	44	39
And	43	32
Bal	43	30
Ext	43	28
CLM	42	29
Mad	41	25
Mur	35	25

All the estimation use the sample 1993:01-2009:12 and use a confident level of 5% to decide about cointegration.

(I) Number of sectors in each region that are cointegrated with the corresponding sector in Spain. (II) Number of sectors which are cointegrated wit bthe corresponding sector in more than 13 regions.

Table 8. Spain. *Cointegration by Sectors*

Highly cointegrated sectors	Low cointegrated sectors
Food	
Highly cointegrated sectors	**Low cointegrated sectors**
Potatoes, Lamb, Preserved fruits, Fish, Crustaceans, molluscs and processed fish, Cereals, Bread, Alcoholic drinks, sugar.	Tobacco, Fresh fruit, Coffee, cacao and infusions, Vegetables, Milk, Pork and Eggs.
Industry and energy	
Highly cointegrated sectors	**Low cointegrated sectors**
Men’s clothes, Women’s clothes, Clothes for babies and children, Men’s footwear.	Textile and home accessories, Major appliances, Non durable household items.
Service	
Highly cointegrated sectors	**Low cointegrated sectors**
Rented apartments, Recreational objects, Publications, Repair of footwear, Primary School, Complements and repair, medical services, secondary school, other expenses in education	Personal transportation, mail and communication.

Highly cointegrated sectors are those for which the corresponding regional series cointegrate with Spain in at least 15 regions and cointegrate with other regions in at least 14 cases.

Low cointegrated sectors are those which cointegrate with Spain in 10 to 14 regions and with other regions in 9 to 13 cases.

Table 9. *RMSFE of the Benchmark strategy and relative RMSFE with respect to Benchmark under alternative strategies (Euro Area 12).*

	Period 2006:01-2009:12				**Period 2006:01-2008:12**			
	1P	**4P**	**8P**	**12P**	**1P**	**4P**	**8P**	**12P**
B	0.37	1.05	1.73	2.30	0.33	0.83	0.99	1.25
A1	1.12	1.13	1.10	1.08	1.06 (*)	1.07	1.10	1.06
A2	0.77(**)	0.81(*)	0.84	0.95	0.70 (**)	0.88	0.89	1.00
A3	0.81(**)	0.93	0.98	0.99	0.79 (**)	0.93	0.97	0.94
C1	0.78(**)	0.87(*)	0.90	0.95	0.73 (**)	0.89	0.90	0.93

B: ARIMA model for the aggregate series; A1: ARIMA models applied to regions; A2: ARIMA models applied to sectors; A3: ARIMA models applied to sectors and regions; C1: SVeqCM with the whole area.

** (*) denotes rejection at the 0.01 (0.05) significance level relative to the benchmark strategy by using the modified Diebold and Mariano (1995) test as proposed by Harvey et al. (1997).

Table 10 *Comparison of strategies A2, A3 and C1.*

	One period ahead			*Twelve periods ahead*	
	C1	*A3*		*C1*	*C3*
A2	*-0.24*	*-1.51*	*A2*	*0.08*	*-0.40*
C1		*-0.84*	*C1*		*-0.93*

** (*) denotes rejection at the 0.01 (0.05) significance level relative to the benchmark strategy by using the modified Diebold and Mariano (1995) test as proposed by Harvey et al. (1997).

Table 11 *Best forecasting strategy according to the RMSFE, one and twelve periods ahead.*

Best forecast: A3 C1 A2

	One period-ahead	Twelve periods-ahead
s1	A2 *	A3
s2	A2 *	A2
s3	A2 *	A2
s4	A2	A3
s5	A2 *	A2
s6	A2	A3
s7	C1 *	A3
s8	A2	A3
s9	A2 *	A3 *
s10	A2	A3
s11	C1	A3
s12	A2 *	A3
s13	A2 *	A3
s14	A2 *	A3
s15	A2	C1
s16	A2 *	A2
s17	A3 *	A3 *
s18	A2 *	A2 *
s19	A2	A3
s20	A2 *	A2
s21	A3	A2
s22	A2 *	A3
s23	A2	C1
s24	A2 *	A2
s25	A2 *	A2 *
s26	A2 *	A2
s27	A2 *	A3 *
s28	A2 *	A2 *
s29	A2 *	A2
s30	A2	C1
s31	A2 *	A3
s32	A2 *	A2
s33	A2 *	A2
s34	C1	C1
s35	A2 *	C1
s36	A2 *	A3
s37	A2	A2
s38	A2 *	A2
s39	A2	C1 *
s40	A2 *	C1
s41	A2 *	A2 *
s42	C1 *	A3
s43	A2 *	A2
s44	A2	A2
s45	A2 *	A2
s46	C1	C1
s47	A2 *	C1
s48	A2 *	C1
s49	A2 *	A2
s50	A2 *	A2

Table 12 *Cointegration by EA12 countries*

Country	Cointegration with the EA12 (I)	Cointegration with other EA12 countries (II)
Spa	44	24
Por	44	21
Bel	43	25
Ger	43	24
Fra	42	31
Ita	42	31
Gre	41	25
Irl	41	24
Aus	41	21
Lux	40	23
Hol	40	22
Fin	40	19

All the estimation use the sample 1993:01-2009:12 and use a confident level of 5% to decide about cointegration.

(I) Number of sectors in each region that are cointegrated with the corresponding sector in the EA12 using a sample 1993:01-2009:12. (II) Number of sectors which are cointegrated with a similar sector in more than 6 countries.

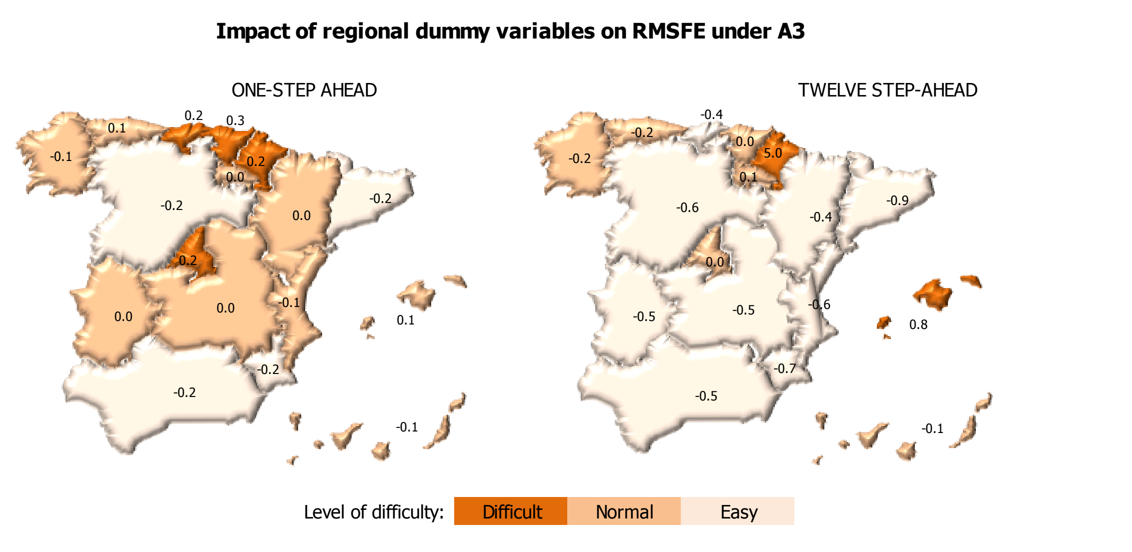
Table 13. ***EA12*** *Cointegration by Sectors*

Food	
Highly cointegrated sectors	*Low cointegrated sectors*
Vegetables, Wine, Mineral waters, soft drinks, fruit and vegetable juices.	*Meat, Milk, cheese and eggs, Bread and cereals, Coffee, tea and cocoa.*
Industry and energy	
Highly cointegrated sectors	*Low cointegrated sectors*
Clothing, Audio-visual, photographic and information processing equipment, Glassware, tableware and household utensils, Motor cars, Non-durable household goods, Tools and equipment for house and garden	*Tools and equipment for house and garden, Other recreational items and equipment, Fuels and lubricants for personal transport equipment*
Service	
Highly cointegrated sectors	*Low cointegrated sectors*
Accommodation services, Package holidays, Miscellaneous goods and services,	*Transport services, Miscellaneous services relating to the dwelling*

Highly cointegrated sectors are those who cointegrate with EA12 in at least 12 cases and cointegrate with other regions in at least 8 cases.

Low cointegrated sectors are those who cointegrate with EA12 11 in at least 8 cases and with other regions in at least 8 cases.

Figure 1. *Classification of Spanish regions according to how difficult they can be forecasted.*



Appendix

Time Series

We use time series for the following disaggregate products in the case of Spain: S1: Cereals; S2: Bread; S3: Beef; S4: Lamb; S5: Pork; S6: Bird; S7: Other meat; S8: Fish; S9: Crustaceans, molluscs and processed fish; S10: Eggs; S11: Milk; S12: Milk products; S13: Oil and fats; S14: Fresh fruit; S15: Preserved fruit; S16: Vegetables; S17: Preserved vegetables; S18: Potatoes; S19: Coffee, cacao and infusions; S20: Sugar; S21: Other food products; S22: Non-alcoholic drinks; S23: Alcoholic drinks; S24: Tobacco; S25: Men's clothes; S26: Women's clothes; S27: Clothes for babies and children; S28: Complements and Repairs; S29: Men's footwear; S30: Women's footwear; S31: Footwear for babies and children; S32: Repair of footwear; S33: Rented apartments; S34: Heating, lighting and water distribution; S35: Own apartments; S36: Furniture and floor coverings; S37: Textile and home accessories; S38: Major appliances; S39: Household items; S40: Non durable household items; S41: Home services; S42: Medical services; S43: Medicines and other chemical products; S44: Personal transportation; S45: Public urban transportation; S46: Public intercity transportation; S47: Mail and communications; S48: Recreational objects; S49: Publications; S50: Recreation; S51: Primary school; S52: Secondary school; S53: University; S54: Other expenditures in education; S55: Personal items; S56: Tourism and hotels; and S57: Other goods and services.

We use time series for the following disaggregate products in the case of EA12: S1: Bread and cereals; S2: Meat; S3: Fish and seafood; S4: Milk, cheese and eggs; S5: Oils and fats; S6: Fruit; S7: Vegetables; S8: Sugar, jam, honey, chocolate and confectionery; S9: Food products n.e.c; S10: Coffee, tea and cocoa; S11: Mineral waters, soft drinks, fruit and vegetable juices; S12: Spirits; S13: Wine; S14: Beer; S15: Tobacco; S16: Clothing; S17: Footwear including repair; S18: Actual rentals for housing; S19: Maintenance and repair of the dwelling; S20: Water supply and miscellaneous services relating to the dwelling; S21: Electricity, gas and other fuels; S22: Furniture and furnishings, carpets and other floor coverings; S23: Household textiles; S24: Major household appliances whether electric or not and small electric *household appliances; S25: Repair of household appliances; S26: Glassware, tableware and household utensils; S27: Tools and equipment for house and garden; S28: Non-durable household goods; S29: Domestic services and household services; S30: Health; S31: Motor cars; S32: Motor cycles, bicycles and animal drawn vehicles; S33: Spares parts and accessories for personal transport equipment; S34: Fuels and lubricants for personal transport equipment; S35: Maintenance and repair of personal transport equipment; S36: Other services in respect of personal transport equipment; S37: Transport services; S38: Postal services; S39: Telephone and telefax equipment and services; S40: Audio-visual, photographic and information processing equipment; S41: Other major durables for recreation and culture; S42: Other recreational items and equipment, gardens and pets; S43: Recreational and cultural services; S44: Newspapers, books and stationery; S45: Package holidays; S46: Education; S47: Restaurants, cafés and the like; S48: Canteens; S49: Accommodation services; S50: Miscellaneous goods and services.*