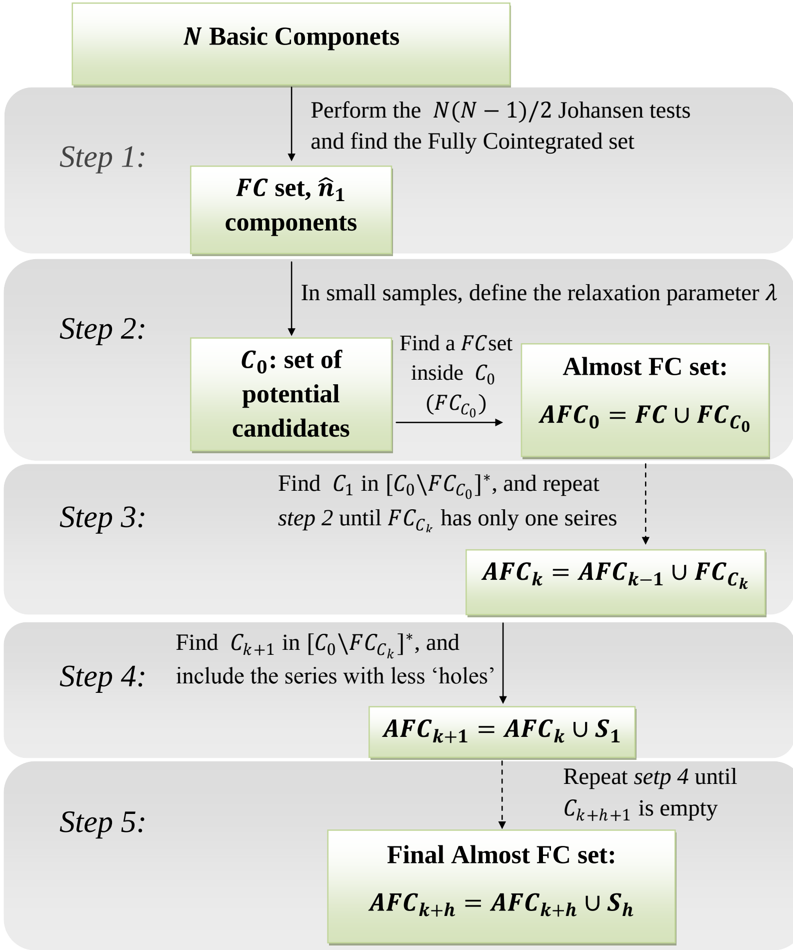


Figure 1: Sketch diagram for the algorithm to construct the set of components with a common trend

Notes:
- * The symbol '\' represents the set difference operator, so that $A \setminus B = \{x \in A : x \notin B\}$.

VII Empirical application: US CPI

In this section we apply the pairwise procedure with outliers correction analyzed in previous sections to the US CPI.

As mentioned in the introduction, our main objective is to model and forecast all the components of an aggregate and we do this with a single-equation approach that incorporates long run restrictions discovered with the pairwise procedure. By means of this procedure we formed what we called 'fully cointegrated' subsets, which have the property that all the series inside them share a single common trend, or equivalently $n_j - 1$ cointegration relationships, with n_j being the number of series inside the j^{th} subset.

The absence of economic theory linking disaggregated prices in the long run could make the concept of *cointegration* to sound inadequate for this application. However, this observation does not preclude the existence of linear combinations between CPI components which cancel unit roots and are useful to obtain better forecasting results. The absence of theory only implies that these relationships may not be expected to be *permanent* as, for example, is the relationship between income and consumption. For this reason in this section the concept of cointegration should be interpreted as *common unit roots restrictions*[3].

The results of the application are summarized in a single and easily readable table that, as we discuss later, constitutes a powerful tool for decision makers, both when interested in the aggregated picture and/or in specific sectors.

VII.1 Data

The CPI break down used in this analysis correspond to the maximum disaggregation level available to the public in the *Bureau of Labor Statistics* (seasonally un-adjusted CPI-U for all urban consumers) for the period 1999.1 − 2016.12 (216 observations). The total number of components is 174. Not all the series have data for the whole sample period, after dropping those with less than 150 observations we keep 169 components. From these series we exclude nine that evolve by steps (regulated prices) so that we end up with 160 series which, considering 2016 weights, represent 92% of the CPI[4]. Among the remaining series, *Owners' equivalent rent of primary residence* weights approximately 24% of the CPI, hence, in order to avoid the global results to be driven by our ability to forecast a single series, we also exclude it form the analysis. Thus, all in all, we will work with 159 series, the remaining ones are neither considered for the construction of the fully cointegrated subsets, nor for the forecasting exercises.

VII.2 Outliers' analysis

As argued in §V.1 the presence of outlying observations can generate devastating effects on parameter estimates and inferential conclusions if not adequately treated, especially in cointegration tests. In that section we designed a strategy for dealing with this issue in the context of the pairwise approach, this strategy requires an individual analysis of outliers for each of the 159 components.

As described in §V.1 the outliers search for the components is carried out in individual models for the differenced components using *Autometrics* with Impulse Indicator Saturation (IIS). Autometrics is a model selection algorithm which, starting from a General Unrestricted Model (GUM) that includes all potentially relevant regressors, and using a multiple path search, reduces the GUM to a simpler model that encompasses it and passes a battery of diagnostic tests (see Doornik (2009)). When applied jointly with IIS, *Autometrics* includes one impulse for each observation and keeps the relevant impulses (see Santos et al. (2008)). The GUM in which we perform the outliers search of each component is:

[3] We are grateful to David Hendry for this observation.

[4] The nine excluded series are: Tuition other school fees and childcare, College tuition and fees, Elementary and high school tuition and fees, Child care and nursery school, Technical and business school tuition and fees, Postage, Delivery services, Limited service meals and snacks, Other lodging away from home including hotels and motels

$$\Delta x_{it} = c + \sum_{j=1}^{5} \phi_j \Delta x_{it-j} + \phi_{12} \Delta x_{it-12} + \sum_{j=1}^{11} S_{jt} + \epsilon_{it}, \tag{12}$$

where x_{it} is the log of the price index of component i, and S_j are centered seasonal dummies.

We select the impulses in two steps. First, we use a target size[5] of 0.25% to select lags, seasonal dummies and impulses, and store the retained impulses. In a second step we consider the same GUM augmented with the retained impulses and a target size of 5% with no IIS.

To make tables legible, components are grouped into six broad categories: non-energy industrial durable goods (MAN Dur), processed food (PF), services (SERV), non-processed food (NPF), non-energy industrial non-durable goods (MAN No Dur) and energy (ENE)[6].

Table 3 summarizes the results. Four main observations emerge from it: *(i)* the average number of outliers in the components is 3.6, which represent 2.0% of the observations (last two columns of row 7), *(ii)* energy and services prices are the most contaminated with a mean proportion of 2.5% and 3.2% of outlying observations per component, respectively (last column of rows 6 and 2), *(iii)* 49% of the outliers are large (larger than 4σ in absolute value), *(iv)* large outliers are more typical in services and energy prices representing 60% and 55% of the total number of outliers, respectively.

Lists of 'highly contaminated' series (5% or more outlying observations) and 'clean' series (no outlying observations) are included in appendix B. While the highly contaminated series are 14 out of the 159 series we are dealing with and represent 16.6% of the weight of those series, the clean ones are are 29 series and weight 23.1%.

Table 3: Mean number of outliers by size and category

		L+	S+	S-	L-	**Mean**	**Mean (% of T)**
(1)	NPF (11 comp.)	1.1	1.8	0.7	0.7	**4.4**	**2.0%**
(2)	ENE (7 comp)	1.3	1.3	0.7	1.1	**4.4**	**2.5%**
(3)	PF (48 comp.)	0.9	1.0	0.7	0.4	**3.0**	**1.7%**
(4)	MAN_dur (50 comp.)	0.5	0.8	0.7	0.5	**2.4**	**1.3%**
(5)	MAN_NoDur (6 comp.)	0.0	0.3	0.7	0.3	**1.3**	**0.7%**
(6)	Serv (37 comp)	2.2	1.5	0.8	1.3	**5.8**	**3.2%**
(7)	TOTAL(159 comp)	5.9	6.8	4.3	4.3	**3.6**	**2.0%**
(8)	**PROP.**	**30%**	**31%**	**20%**	**19%**	**100%**	

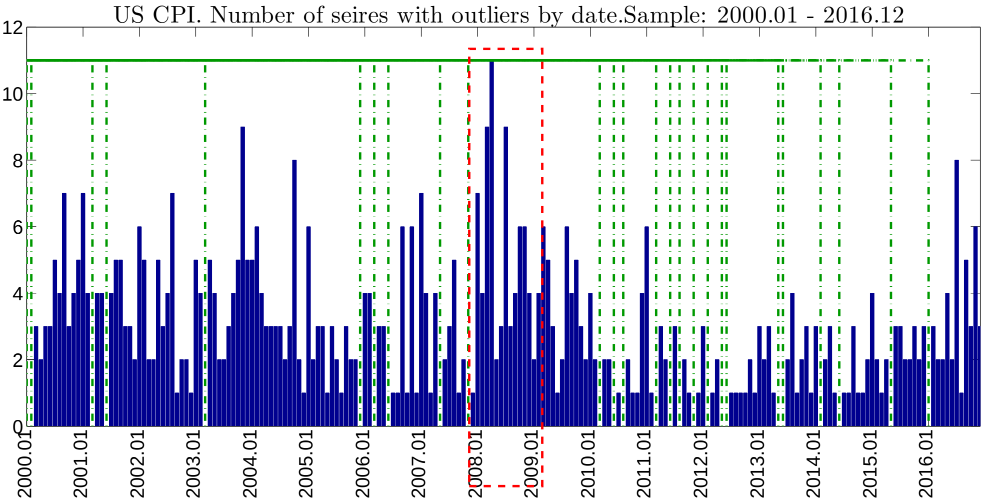
Numbers in parenthesis after the category name indicate the number of series in the category.
L+: Large (larger than 4σ) and positive outliers.
S+: Small (smaller than or equal to 4σ) and positive outliers.
L-: Large and negative outliers.
S-: Small and negative outliers.

[5]The the significance level for retaining a regressor of the GUM

[6]Note that this grouping is not perfect for a component could include prices belonging to two broad categories

As a by product of the components' outliers analysis we end up with a cross-sectional distribution of outliers at each point in time, i.e; at each month of the sample we have the estimated outliers of the 159 components of the CPI. Although a deep study of these distributions exceeds the objective of this paper, we present some initial results.

Figure 2 shows the number of series with outliers at each of the 216 months of the sample. As it shows, the distribution by dates if far form uniform, with some months having 11 series with outliers and some others months (26) with none. Interestingly, there seems to be a concentration around years 2008-2009, the sub-prime crisis period (red box of the figure).



Note: Green doted lines indicate dates at which there are no contaminated series.

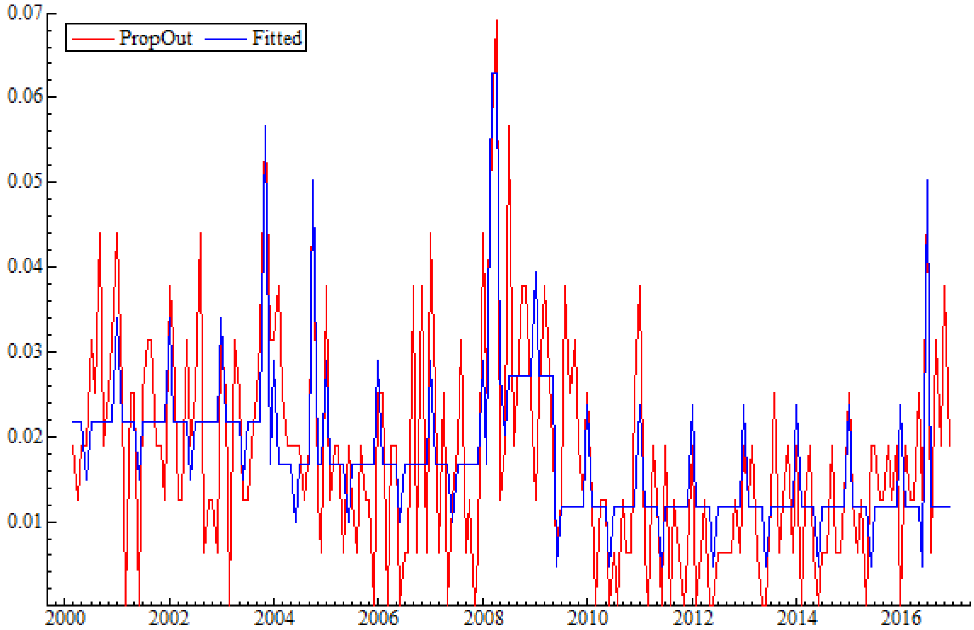
Figure 2: Distribution of series with outliers by date

To confirm that there is a mean shift in the number of contaminated components during the sub-prime crisis, we estimate a model for the proportion of series with outliers including as potential regressors lags 1 to 5, seasonal dummies and choosing the *Autometrics* option IIS+SIS for outliers and breaks detection[7]. Results are summarized in figure 3, from where four important conclusions can be drawn: (i) in January there are, on average, more series with outliers, (ii) in June the proportion of series with outliers is reduced, (iii) there is a significant and positive step from 2008.2 to 2009.5, and (iv) after 2009.5 the mean proportion of series with outliers is lower than before the crisis.

The conclusion is clear; the pattern by which the US CPI is congested of series perturbed by outliers along the time is not random. It has some seasonality and experiences changes in mean related with the general economic conditions.

Finally it is noteworthy that the exhaustive outliers' search we made is also relevant for modeling and forecasting the CPI itself. This is so for outliers in the components are also

[7]The option IIS+SIS, apart from including an impulse in each observation, it includes a Step. See Doornik et al. (2013)



Estimaiton Sample is: 2000.3 - 2016.12					
	Coefficient	Std.Error	t-value	t-prob	Part.R^2
Constant	0.012	0.001	11.600	0.000	0.413
CSeasonal	0.012	0.003	4.680	0.000	0.103
CSeasonal_5	-0.007	0.003	-2.760	0.006	0.038
I:2003(11)	0.035	0.010	3.480	0.001	0.059
I:2004(10)	0.033	0.010	3.350	0.001	0.055
I:2016(7)	0.039	0.010	3.880	0.000	0.073
S1:2003(11)	0.005	0.002	2.440	0.016	0.030
S1:2008(2)	-0.046	0.007	-6.450	0.000	0.178
S1:2008(4)	0.036	0.008	4.750	0.000	0.105
S1:2009(5)	0.015	0.003	5.280	0.000	0.127

Notes:

- Steps ($S1:$) take the value one from the first observation until the date indicated in the name of the step, and zero form then on.
- CSeasonal is the centered seasonal variable corresponding to January.

Figure 3: Changes in the mean proportion of series with outliers along the sample

outliers in the aggregate but, as we argue below, in many cases they can be estimated only in the components. In order to use the individual outliers in a model for the CPI we construct the *aggregated outlier* series as the weighted sum of all individual outliers multiplied by their coefficients and include this series in a model for the CPI. Since the individual outliers will enter the CPI weighted by the corresponding component's weight, we expect the coefficient of the *aggregated outlier* not to differ significantly from one. The reason for expecting a unitary coefficient for the *aggregated outlier* can be easily seen by writing:

$$CPI_t = \sum_{i=1}^{N} w_{it} C_{i,t}, \tag{13}$$

where C_{it} represents component i at period t and w_{it} its weight. Now, components can be expressed as the sum of their *core* plus their outliers:

$$C_{i,t} = C_{i,t}^{*} + \sum_{j=1}^{Q_i} \gamma_{ij} O_{ij,t} \equiv C_{i,t}^{*} + \sum_{j=1}^{Q_i} O_{ij,t}^{*}, \tag{14}$$

where Q_i is the number of outliers in component i, $O_{ij,t}$ is the variable representing the j-*th* outlier of component i, γ_{ij} its coefficient, and $O_{ij,t}^{*} = \gamma_{ij} O_{ij,t}$.

The *aggregated outlier* series is defined as:

$$AggOut_t = \sum_{i=1}^{N} w_{it} \sum_{j=1}^{Q_i} O_{ij,t}^{*} \tag{15}$$

Plugging (14) and in (13) and using (15) we get:

$$CPI_t = \sum_{i=1}^{N} w_{it}(C_{i,t}^{*} + \sum_{j=1}^{Q_i} O_{ij,t}^{*}) = \sum_{i=1}^{N} w_{it} C_{i,t}^{*} + AggOut_t \tag{16}$$

Expression 16 implies that $AggOut_t$ will have an unitary coefficient in a model for the CPI.

For assessing the usefulness this variable to model the CPI we compare three simple models. Starting from the GUM;

$$\Delta CPI_t = c + \sum_{i=1}^{4} \phi_i \Delta CPI_{t-i} + \phi_{12} \Delta CPI_{t-12} + \sum_{s=1}^{11} \rho_i S_{it} + \epsilon_t,$$

where S_{it} is a centered seasonal dummy, we consider three possibilities to be estimated with $Autometrics$:

i. *Only IIS:* IIS is applied in previous GUM.

ii. *Only AggOut:* The GUM is augmented with the series of $AggOut_t$ (IIS is not used).

iii. *AggOut:* IIS is applied in a the augmented GUM.

Table 4 includes model selection criteria for the three possibilities. As it shows, the two models including $AggOut_t$ outperform model (i). Interestingly, model (iii) seems to be the best option. This last result suggest two conclusions: *(a)* some components' outliers

-which are also outliers of the CPI- are not identifiable in a model for the aggregate, and *(b)* some CPI's outliers -which must be present in some component- are not identifiable in component's models, probably because these observations correspond to small outliers of the same sign in more than one component.

Table 4: Comparison of different models for the CPI

	AIC	SIC	$\boldsymbol{Adj.R^2}$
Only IIS	-9.14	-8.90	0.65
Only AggOutl	-9.22	-9.02	0.67
AggOutl + IIS	-9.33	-9.06	0.71

Basic GUM: $\Delta CPI_t = c + \sum_{i=1}^{4} \phi_i \Delta CPI_{t-i} + \phi_{12}\Delta CPI_{t-12} + \sum_{s=1}^{11} \rho_i S_{it} + \epsilon_t$.
Only IIS: IIS is applied in previous GUM.
Only AggOutl: The GUM is augmented with the series of $AggOut_t$ (IIS is not used).
AggOutl: IIS is applied in a the augmented GUM.

The *p-value* for the null that $AggOut's$ coefficient is equal to 1 is 0.16 thus, as expected, it is not rejected.

VII.3 Pairwise tests' results

Since the pairwise approach does not deal with seasonal unit roots, we performed previous seasonal unit root test tests as proposed by Osborn et al. (1988) to all the components. The results indicate that they do not show seasonal unit roots in general and that the assumption of only one unit root, linear growth and deterministic seasonality seems sensible (details are available upon request).

In order to obtain economically and statistically sensible cointegration relationships between the components of the CPI we consider only those which satisfy the following four conditions: *(i)* the cointegration relationship does not require a deterministic trend, *(ii)* coefficients of both prices are statistically significant, *(iii)* the bivariate VAR characteristic polynomial's second largest root is not close to one, and *(iv)* the cointegration relationship is stable over time.

For the outlier's corrected series (see §V.1), cointegration tests are performed at the 1% of significance and the number of lags for each pair is determined with the AIC in a model with one cointegration restriction and without trend in the cointegration relationship. Centered seasonal dummies are included in all models. We do not use the small samples correction designed in §V.2 because we our time series are long enough.

For grouping the components by subsets we consider the strategy summarized in §VI, which basically consists of using the results of the pairwise cointegration tests for discovering subsets of series in which all possible pairs are cointegrated. The series in these subsets share a unique common trend. Subsets with less than five series were disregarded. We found 7 subsets that jointly include 41 series, which represent 25.8% of the components and 23.5% of the total weight we are considering.

Table 5 details of the outcome of the procedure. To make the table legible, we use the same broad categories as above: non-energy industrial durable goods (MAN Dur), processed food (PF), services (SERV), non-processed food (NPF), non-energy industrial non-durable goods (MAN No Dur) and energy (ENE).

One conclusion of the table is that subsets of series sharing a common trend cannot, in general, be assigned to a single broad category. However, in almost all the cases, more than 85% of the subset weight is explained by two broad categories. The exceptions are blocks 2 and 3 for which the two most important categories explain around 78.% of the block's weights.

This observation has two relevant implications: first the ad-hoc method proposed by Boivin and Ng (2006) for extracting non-pervasive common factors would not work for the US CPI; second, although a 'labeling' strategy that matches blocks with single broad categories is not possible, in many cases, this could be done using just two categories.

Another relevant conclusion from table 5 concerns the comparison of the distribution of the 6 categories of prices in the CPI (first row of the table) and in the fully cointegrated subsets (last row of the table). This comparison indicates that non-energy industrial durable goods are under represented in the subsets, and non-processed food is over represented. The rest of the categories have a similar participation in the CPI and in the subsets. The specific components inside each subset are detailed in appendix A.

Table 5: Detailed results of the Pairwise procedure: number of series and proportion of weight by broad categories and blocks

	Serv		ManD		ENE		PF		ManND		NPF		**TotQ**	**TotW**
	Q	W	Q	W	Q	W	Q	W	Q	W	Q	W		
%Tot	23.3	51.0	31.4	20.3	4.4	11.6	30.2	11.4	3.8	3.3	6.9	2.5	**159**	**100**
Set1	1	70.0	2	5.4	0	0.0	3	15.5	0	0.0	3	9.1	**9**	**3.6**
Set2	4	52.5	0	0.0	0	0.0	2	21.5	1	26.0	0	0.0	**7**	**1.8**
Set3	2	45.0	2	32.7	0	0.0	0	0.0	0	0.0	1	22.3	**5**	**1.7**
Set4	2	81.2	1	2.1	0	0.0	2	16.7	0	0.0	0	0.0	**5**	**4.7**
Set5	1	88.1	2	7.2	0	0.0	1	2.1	0	0.0	1	2.7	**5**	**4.7**
Set6	0	0.0	0	0.0	2	27.3	3	72.7	0	0.0	0	0.0	**5**	**0.8**
Set7	2	24.6	0	0.0	1	64.7	1	4.9	0	0.0	1	5.8	**5**	**6.3**
%	**29.3**	**58.1**	**17.1**	**5.0**	**7.3**	**18.3**	**29.3**	**11.6**	**2.4**	**2.0**	**14.6**	**5.1**	**41.0**	**23.5**

Columns Q indicate the percentage of series in each category (first row) and the number of series in each Subset.
Columns W indicate the total weight of each category in the CPI (first row) and the proportion of the weight of each category in each Subset.
Last column contains the total weight of the blocks.

VII.4 Forecasting the US CPI and all its components

In this section we forecast all of the 159 components of the CPI in single-equation models that include cointegration relationships discovered by means of the pairwise approach.

For building the single-equation models we use the automatic model selection algorithm *Autometrics* with IIS (see the beginning of §VII.2 for a brief description of *Autometrics*).

Since for each component, the process of building the econometric model is subject to a set of diagnostic tests included in *Autometrics*, we can conclude that they are reasonable for empirical applications. Additionally, as the components aggregate to the CPI, we can apply another test to the models for the disaggregates. It consists of comparing the forecast of the aggregate obtained indirectly by aggregating the forecasts of the components, with a direct forecast from a scalar model of the aggregate. We denote the indirect approach by I-PW (the 'I' stands for indirect and 'PW' for pairwise) and the direct one by D. The latter is our baseline model.

The pairwise strategy I-PW would not only provide models to analyze all the components, but it could also be an instrument to obtain more accurate forecasts of the aggregate. This could be so because it incorporates more information than the corresponding direct forecast and could palliate the curse of dimensionality in the number of parameters by considering the restrictions implied by cointegration. Therefore, our approach to forecast the aggregate is an intermediate one between the direct approach and the vector-model approach (a full information method, that in our case of interest is not feasible).

Given our interest in forecasting all the components, the comparison with direct approaches should not be used as a definitive criterion for assessing the forecasting performance of our procedure, we should use some disaggregated baseline. Therefore, we also compare the forecasting performance of I-PW with the disaggregated forecasts using univariate models for each component, denoted as I-B (indirect basic).

For these approaches (D, I-PW and I-B) we consider three broad possibilities depending on the regressors to be included in the General Unrestricted Model (GUM). Apart from own lags, seasonal dummies, and cointegration relationships (when it is the case), we may include: a) No other regressor, b) Lags of the aggregated CPI (only for the indirect procedures), c) Lags of twenty nine broad categories which add up to the CPI. We denote this option as *Dissaggregated information* (DI)

For each of the three possibilities, in the I-PW procedure series not belonging to any fully cointegrated subset can be modeled individually or all together in a scalar model for the sub-aggregate that adds up all these series. Abusing notation we label this last possibility as *I-PW-GP*, for Guerrero and Peña (2003).

Thus, we have six different I-PW possibilities, three I-B, and two direct. For the D and I-B alternatives, we add an additional possibility consisting of including dynamic factor models estimated from all the disaggregates (D-DFM and I-B-DFM). Therefore, we end up with 12 alternatives.

Table 6 includes a summary of the equations for the different forecasting procedures. From options a to c above, only option a is included in the table, the other two options are simple extensions.

Equations in table 6 represent the initial GUMs from where models are selected using *Autometrics* with Impulse Indicator Saturation. The selection is carried out in two steps. First we use a target size of 0.25% to select variables, lags and impulses. Retained impulses are stored. In a second step we consider the same GUM augmented with the retained impulses and a target size of 5% with no IIS.

As explained in §VII.1 the 159 components we are dealing with do not represent 100% of the CPI, we call the subaggregate formed by these components as CPI^*. To forecast the CPI we consider the following model:

$$\Delta logCPI_t = c + \lambda_0 \Delta logCPI_t^* + \sum_{i=1}^{11} \gamma_i S_{i,t} + \epsilon_t, \tag{17}$$

where the h steps ahead forecast of CPI_t^*, required to compute h steps ahead forecasts of the CPI, is computed by aggregating the components' forecasts.

VII.4.1 Components' forecasts from the I-PW-CPI approach

Figure 4, figure 5, table 7 and table 8 summarize the detailed forecasts of the CPI components obtained with the I-PW-CPI approach described above.

Figure 4 includes all the components' forecasts in a single plot, and figure 5 contains box plots for each of the 24 months of years 2015 and 2016. Each box describes the cross-sectional distribution of the components at each point in time.

An important conclusion that emerges from these two figures is that the dispersion of the forecasts for the different components seems to be smaller than the observed values of 2015 (compare the sizes of the boxes and the length of the whiskers of figure 5 in 2015 and 2016).

Table 7 gives 12 step ahead forecasts for the average annual rates of growth of US CPI and its components for 2016 (made with information up to December 2015). The table uses green shadows to indicate that the point forecasts of the components are below the lower bound of the confidence interval for the CPI, and red for the components' forecasts above the corresponding upper bound. Italics and bold letters are used to indicate that the weight of a particular component is relatively high (see the *Notes to table 7* below for a detailed description).

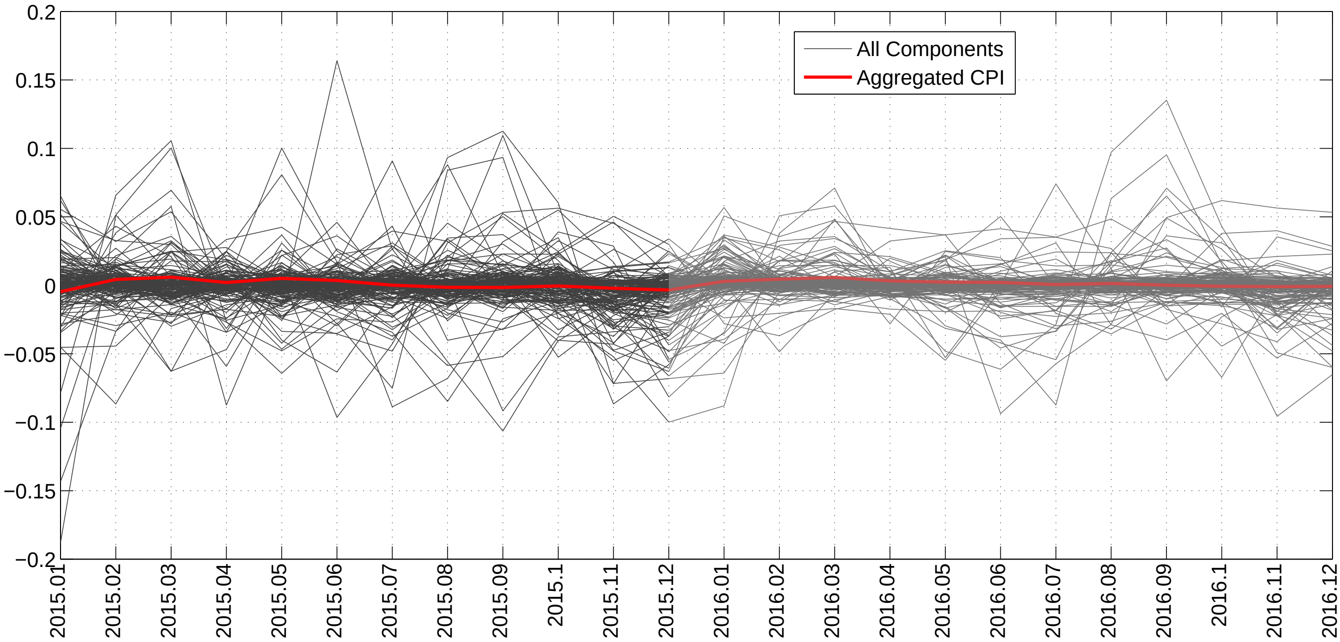
This table constitutes a powerful instrument for analyzing inflation. One the one hand, if the user is interested just in a global analysis, the table gives information of recent evolution and expected dynamics for next year. Furthermore, the color map constitutes a fast and deep informative description about what explains the recent evolution and the forecasts: Are all the components growing at similar rates? Are there some highly inflationary and

Table 6: Summary of the forecasting exercises

Model	Description
D	$\Delta cpi_t = c + \sum_{k=1}^{K} \phi_k \Delta cpi_{t-k} + \phi_{12}\Delta cpi_{t-12} + \phi_{24}\Delta cpi_{t-24} + \sum_{i=1}^{11} \gamma_i S_{i,t} + \epsilon_t$
D-DFM	$\Delta cpi_t = c + \sum_{k=1}^{K} \phi_k \Delta cpi_{t-k} + \phi_{12}\Delta cpi_{t-12} + \phi_{24}\Delta cpi_{t-24} +$ $\sum_{k=1}^{K} \delta_k F_{t-k} + \sum_{i=1}^{11} \gamma_i S_{i,t} + \epsilon_t$
I-PW	*case i)* Series inside some subset $\Delta x_{i,t} = c + \sum_{r=1}^{R_i} \alpha_{i,r} CR_{r,t-1} + \sum_{k=1}^{K_i} \Delta\phi_k x_{i,t-k} + \phi_{12}\Delta x_{i,t-12} + \phi_{24}\Delta x_{i,t-24} +$ $\sum_{j=1}^{J} \theta_j \Delta SubAggCT_{i,t-j} + \sum_{i=1}^{11} \gamma_i S_{i,t} + \epsilon_{i,t}$ *case ii)* Series not in any subset $\Delta x_{i,t} = c + \sum_{k=1}^{K} \phi_k \Delta x_{i,t-k} + \phi_{12}\Delta x_{i,t-12} + \phi_{24}\Delta x_{i,t-24} + \sum_{i=1}^{11} \gamma_i S_{i,t} + \epsilon_{i,t}$
I-PW-GP	For series inside some subset, same as I-PW case i. For the others, only its sub-aggregate is forecast in a model with the same structure as I-PW case ii
I-B	$\Delta x_{i,t} = c + \sum_{k=1}^{K} \phi_k \Delta x_{i,t-k} + \phi_{12}\Delta x_{i,t-12} + \phi_{24}\Delta x_{i,t-24} + \sum_{i=1}^{11} \gamma_i S_{i,t} + \epsilon_{i,t}$
I -B-DFM	$\Delta x_{i,t} = c + \sum_{k=1}^{K} \phi_k \Delta x_{i,t-k_1} + \phi_{12}\Delta x_{i,t-12} + \phi_{24}\Delta x_{i,t-24} +$ $\sum_{k=1}^{K} \delta_k F_{t-k} + \sum_{i=1}^{11} \gamma_i S_{i,t} + \epsilon_{i,t}$

- Lower case letters denote logarithms.
- All the equations represent the initial GUMs form where models are selected using *Autometrics* with Impulse indicator saturation. The selection is carried out in two steps. First we use a target size of 0.25% to select variables, lags and impulses. Retained impulses are stored. In a second step we consider the same GUM augmented with the retained impulses and a target size of 5% with no IIS.
- $K = J = 5$.
- *SubAggCT* stands for the subaggregate formed by the series inside the corresponding fully cointegrated subset.
- $S_{i,t}$ are centered seasonal dummies.
- In models D-DFM and I-B-DFM the q-dimensional factors (F) are computed from the difference of all the components. The optimal number of factor is chosen with the information criteria of Bai and Ng (2002). The factors are forecast in a VAR model, where lags are selected with *Autometrics* with IIS. The same two step procedure applies in this case.

some others highly deflationary? Can inflationary and deflationary components be grouped in some broad category? Have the dynamics of the components been stable in the recent past? Which are the main drivers of the aggregate's forecast? All these important questions



Dark lines are observed values and light ones are forecasts.

Figure 4: Components' forecasts at 2015.12 ($\Delta logCPI_t$)

can be answered quite rapidly just by analyzing the color map.

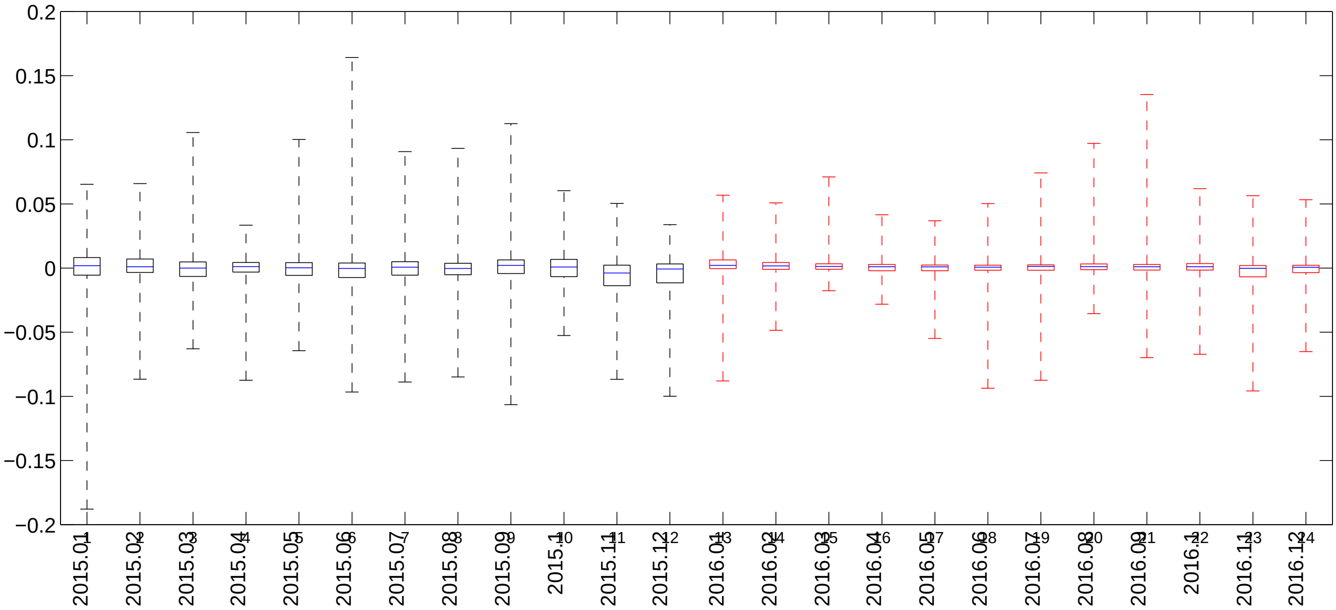
On the other hand, if the user is interested in a particular component, she will have not only her component's forecast, but also a fast comparison with the rest of the prices in the same category and with all the components of the CPI.

Table 7 shows that the components' forecasts below the mentioned lower bound belong mainly to the category of non-energy industrial goods. Components' forecasts above the upper bound correspond mainly to non-processed food and energy goods, having special impact in the headline inflation the prices in the energy group.

An important number of components in the group of non-energy industrial goods show expected negative inflation, being especially negative the expectations in *Televisions* (-18.2%), *Personal computers and peripheral equipment* (-11.9%), and *Men's shirts and sweaters* (-10.2%).

Table 8 classifies all the components according to our ability to forecast them. Red colors are used for components' which are relatively hard to forecast (large Root Mean Squared Forecast Errors) and green for relatively easy to forecast ones. Additionally, highly inflationary components (those with red color in column *2014* of table 7) are marked with a dotted shadow (see the *Notes to table 8* below for a detailed description).

Three main conclusions can be extracted form table 8: (i) almost all energy and non-processed food components are relatively hard to forecast, (ii) almost all services are relatively easy to forecast, (iii) in non-energy industrial goods and processed food there is not a clear pattern.



- The edges of the box are the 25^{th} and 75^{th} percentiles of the 159 components' forecasts.
- The whiskers extend from the minimum to the maximum.
- Black color indicates observed values, red, forecasts.

Figure 5: Box plots of the observed and forecast components at 2015.12 ($\Delta logCPI_t$)

Notes to table 7

- Columns *10-15* include the mean of annual growth of each price ($\Delta_{12} logP_t$) in December for years 2010 to 2015.

- Columns *2016* include the forecast of $\Delta_{12} logP_{2016.12}$ with information up to December 2015.

- Names Reference: *Man Dur:* Non energy industrial durable goods, *PF:* Processed food, *SERV:* Non energy services, *NPF:* Non-processed food, *ENE:* Energetic goods and services, *ManND:* Industrial non-durable goods.

- Color references for columns *10-15*

  - Green: the 12 step ahead forecast for $\Delta_{12} logP_{2016.12}$ is smaller than the lower bound of the 80% confidence interval for the mean of ($\Delta_{12} logP_t$) in December of years 2010 to 2015.

  - Red: the 12 step ahead forecast for $\Delta_{12} logP_{2016.12}$ is larger than the upper bound of the 80% confidence interval for the mean of ($\Delta_{12} logP_t$) in December of years 2010 to 2015.

  - White: the 12 step ahead forecast for $\Delta_{12} logP_{2016.12}$ is inside the 80% confidence interval for the mean of ($\Delta_{12} logP_t$) in December of years 2010 to 2015.

  - Standard deviation for the mean of ($\Delta_{12} logP_t$) at December of years 2010 to 2015 is computed as $\frac{\sigma}{\sqrt{6}}$ where σ is the sample standard deviation $\Delta_{12} logP_t$. Autocovariances at lags 12 and larger are ignored, and the confidence interval is constructed assuming normality.

- Color references for columns *2016*
  - Green: the 12 step ahead forecast at December 2015 for $\Delta_{12}logP_{2016.12}$ is smaller than the lower bound of the 80% confidence interval for the forecast of $\Delta_{12}logCPI_{2016.12}$.
  - Red: the 12 step ahead forecast at December 2015 for $\Delta_{12}logP_{2016.12}$ is larger than the upper bound of the 80% confidence interval for the forecast of $\Delta_{12}logCPI_{2016.12}$.
  - White: the 12 step ahead forecast at December 2015 for $\Delta_{12}logP_{2016.12}$ is inside the 80% confidence interval for the forecast of $\Delta_{12}logCPI_{2016.12}$.
  - Blue italics: indicate component's weights larger than the average weight $(1/N)$
  - Blue bold: indicate component's weights larger than the 3%.
  - Standard deviation for the 12 step ahead forecast error of $\Delta_{12}logCPI_t$ is computed as the historical out of sample Root Mean Squared Forecast Error (see table 9).
  - Grey shadows in components' names are just to distinguish between categories.

Table 7: Components' forecasts ($\Delta_{12} log P_t$)

MAN Dur	10-15	2016
Men's suits,	-1.2	7.7
Men's furnish	2.3	1.5
Men's shirts	-0.4	-10.2
Men's pants a	1.8	0.6
Boys' apparel	1.9	-0.3
Women's outer	3.1	3.0
Women's dress	1.8	-8.2
Women's suits	-1.3	*-0.5*
Women's under	2.1	0.3
Girls' appare	-0.8	-2.7
Men's footwea	1.4	0.4
Boys' and gir	2.6	1.8
Women's footw	0.3	0.9
Infants' and	0.9	-0.1
Watches	0.9	0.2
Jewelry	1.0	0.3
Personal comp	-9.3	-11.9
Computer soft	-5.7	-4.9
Telephone har	-6.0	-8.3
Miscellaneous	-1.2	-0.9
Floor coverin	-1.7	0.5
Window coveri	-2.0	-4.3
Other linens	-5.0	-5.2
Bedroom furni	-1.2	0.3
Living room,	-0.9	-1.0
Other furnitu	-3.1	-3.3
Major applian	-2.2	-0.5
Other applian	-1.8	-2.2
Clocks, lamps	-6.2	-5.5
Dishes and fl	-5.7	-2.6
Nonelectric c	-0.9	-0.4
Tools, hardwa	-0.4	*-0.4*
Eyeglasses an	1.0	1.5
Televisions	-18.5	-18.2
Other video e	-9.4	-8.0
Video discs a	-1.6	-0.9
Audio equipme	-5.7	-5.3
Recorded musi	-1.7	-1.0
Sports vehicl	1.1	0.6
Sports equipm	-2.7	-1.4
Photographic	-5.0	-6.8
Photographers	1.9	1.1
Toys	-5.3	-6.0
Sewing machin	0.9	0.0
Music instrum	0.4	-0.1
New vehicles	1.1	*0.2*
Used cars and	0.7	*-1.5*
Leased cars a	-3.0	*-1.2*
Tires	0.9	1.6
Vehicle acces	2.4	2.8
MAN Dur		**-1.2**

PF	10-15	2016
Flour and pre	1.5	3.0
Breakfast cer	1.1	1.6
Rice, pasta,	0.5	0.9
Bread	2.2	2.4
Fresh biscuit	2.4	3.2
Cakes, cupcak	1.5	1.1
Other bakery	1.5	0.5
Uncooked grou	8.4	6.2
Uncooked beef	8.4	5.6
Uncooked beef	7.1	0.2
Uncooked othe	9.3	5.7
Bacon, breakf	5.4	2.4
Ham	5.7	4.7
Pork chops	4.9	0.3
Other pork in	5.9	1.9
Other meats	3.3	3.1
Chicken	3.1	2.0
Other poultry	3.5	2.8
Processed fis	3.0	2.2
Eggs	5.8	-4.4
Ice cream and	2.8	1.1
Other dairy a	1.9	2.2
Canned fruits	1.3	2.5
Frozen fruits	1.4	1.8
Other process	1.4	2.2
Carbonated dr	0.8	1.5
Frozen noncar	3.2	2.1
Nonfrozen non	0.2	1.5
Coffee	2.3	-0.2
Other beverag	0.6	0.8
Sugar and art	0.0	1.6
Candy and che	1.4	2.9
Other sweets	1.4	1.8
Butter and ma	5.9	2.1
Salad dressin	0.7	0.6
Other fats an	2.3	1.1
Soups	0.2	1.6
Frozen and fr	0.6	0.8
Snacks	2.8	0.6
Spices, seaso	1.8	0.9
Baby food	2.2	1.7
Other miscell	1.5	*0.6*
Beer, ale, an	1.4	1.9
Distilled spi	0.5	0.9
Wine at home	-0.1	0.8
Alcoholic bev	2.5	2.6
Cigarettes	3.2	*4.0*
Tobacco produ	2.9	3.8
PF		**1.8**

SERV	10-15	2016
Educational b	4.9	5.5
Wireless tele	-2.6	*0.2*
Internet serv	0.7	*0.4*
Full service	2.3	*2.3*
Limited servi	2.3	*2.5*
Food from ven	1.8	2.4
Other food aw	2.2	3.2
Haircuts and	1.4	*2.1*
Legal service	2.4	3.1
Funeral expen	2.1	3.0
Laundry and d	1.8	2.5
Apparel servi	2.8	-0.8
Financial ser	3.2	2.6
Rent of prima	2.4	*3.0*
Other lodging	2.2	*1.7*
Tenants' and	3.2	2.2
Garbage and t	2.4	2.8
Household ope	1.7	*2.8*
Physicians' s	2.3	*1.5*
Dental servic	2.5	*3.0*
Services by o	1.6	2.0
Hospital serv	5.3	*5.5*
Nursing homes	3.1	3.4
Cable and sat	2.5	*2.9*
Pets and pet	0.7	*0.2*
Pet services	3.4	3.8
Other recreat	1.2	*2.2*
Newspapers an	3.8	2.5
Recreational	-1.4	-0.4
Car and truck	0.1	0.9
Motor vehicle	1.8	*2.4*
Motor vehicle	4.0	*4.7*
State motor v	0.4	0.5
Parking and o	3.0	2.8
Airline fares	1.2	*-2.2*
Other interci	0.8	-0.1
Intracity tra	3.1	2.4
SERV		**2.6**

NPF/ENE/ManND	10-15	2016
Fresh fish an	4.7	2.5
Milk	4.0	0.4
Cheese and re	3.7	4.1
Apples	2.9	10.7
Bananas	0.7	-1.3
Citrus fruits	3.2	5.9
Other fresh f	1.4	1.7
Potatoes	2.0	1.1
Lettuce	-0.9	-5.9
Tomatoes	2.0	2.8
Other fresh v	2.1	1.3
Fuel oil	2.5	-4.6
Propane, kero	2.0	7.2
Electricity	1.7	*-1.2*
Utility (pipe	-0.8	*0.3*
Water and sew	5.5	*4.7*
Gasoline (all	-0.1	*4.5*
Other motor f	3.8	9.7
Personal care	0.1	*0.1*
Hair, dental,	-0.4	0.0
Cosmetics, pe	0.5	0.9
Household cle	-0.5	0.5
Household pap	1.6	-3.7
Miscellaneous	0.5	1.0
NPF		**2.2**
ENE		**2.0**
ManNonDur		**0.0**

	CPI	Fore(h12)
2015	**0.73**	**2.05**
2016	**2.04**	**2.04**

confidence 80%

2014

(green)	Fore < CPI - 1.28
	Fore = CPI +- 1.28
(red)	Fore > CPI + 1.28
italics	Weight > 100/159
bold	Weight larger than 3

confidence 80%

2010-2015

(green)	Fore 16 < Mean(10-15)
	Fore 16 = Mean(10-15)
(red)	Fore 16 > Mean(10-15)

Notes to table 8

- The table classifies all the components according to our ability to forecast them.
- Columns *h1*, *h6* and *h12* refer to forecasts horizons 1, 6 and 12 respectively
- Using all the components' RMSFE for each forecast horizon we compute the quintiles and classify the components according to the quintile to which they belong.
- Names Reference: *Man Dur:* Non energy industrial durable goods, *PF:* Processed food, *SERV:* Non energy services, *NPF:* Non-processed food, *ENE:* Energetic goods and services, *ManND:* Industrial non-durable goods.
- Color reference:
  - Dark Red: $Q4 \leq RMSFE_i < Q5$. (relatively 'hard' to forecast).
  - Light Red: $Q3 \leq RMSFE_i < Q4$.
  - White: $Q2 \leq RMSFE_i < Q3$.
  - Dark Green: $Q1 \leq RMSFE_i < Q2$.
  - Light Green: $RMSFE_i < Q1$ (relatively 'easy' to forecast).
  - Grey shadows in components' names are just to distinguish between categories.

Table 8: Components RMSFE

MAN dur	h1	h6	h12
Men's suits,			
Men's furnish			
Men's shirts			
Men's pants a			
Boys' apparel			
Women's outer			
Women's dress			
Women's suits			
Women's under			
Girls' appare			
Men's footwea			
Boys' and gir			
Women's footw			
Infants' and			
Watches			
Jewelry			
Personal comp			
Computer soft			
Telephone har			
Miscellaneous			
Floor coverin			
Window coveri			
Other linens			
Bedroom furni			
Living room,			
Other furnitu			
Major applian			
Other applian			
Clocks, lamps			
Dishes and fl			
Nonelectric c			
Tools, hardwa			
Eyeglasses an			
Televisions			
Other video e			
Video discs a			
Audio equipme			
Recorded musi			
Sports vehicl			
Sports equipm			
Photographic			
Photographers			
Toys			
Sewing machin			
Music instrum			
New vehicles			
Used cars and			
Leased cars a			
Tires			
Vehicle acces			

PF	h1	h6	h12
Flour and pre			
Breakfast cer			
Rice, pasta,			
Bread			
Fresh biscuit			
Cakes, cupcak			
Other bakery			
Uncooked grou			
Uncooked beef			
Uncooked beef			
Uncooked othe			
Bacon, breakf			
Ham			
Pork chops			
Other pork in			
Other meats			
Chicken			
Other poultry			
Processed fis			
Eggs			
Ice cream and			
Other dairy a			
Canned fruits			
Frozen fruits			
Other process			
Carbonated dr			
Frozen noncar			
Nonfrozen non			
Coffee			
Other beverag			
Sugar and art			
Candy and che			
Other sweets			
Butter and ma			
Salad dressin			
Other fats an			
Soups			
Frozen and fr			
Snacks			
Spices, seaso			
Baby food			
Other miscell			
Beer, ale, an			
Distilled spi			
Wine at home			
Alcoholic bev			
Cigarettes			
Tobacco produ			

SERV	h1	h6	h12
Educational b			
Wireless tele			
Internet serv			
Full service			
Limited servi			
Food from ven			
Other food aw			
Haircuts and			
Legal service			
Funeral expen			
Laundry and d			
Apparel servi			
Financial ser			
Rent of prima			
Other lodging			
Tenants' and			
Garbage and t			
Household ope			
Physicians' s			
Dental servic			
Services by o			
Hospital serv			
Nursing homes			
Cable and sat			
Pets and pet			
Pet services			
Other recreat			
Newspapers an			
Recreational			
Car and truck			
Motor vehicle			
Motor vehicle			
State motor v			
Parking and o			
Airline fares			
Other interci			
Intracity tra			

NPF/ENE/ManND	h1	h6	h12
Fresh fish an			
Milk			
Cheese and re			
Apples			
Bananas			
Citrus fruits			
Other fresh f			
Potatoes			
Lettuce			
Tomatoes			
Other fresh v			
Fuel oil			
Propane, kero			
Electricity			
Utility (pipe			
Water and sew			
Gasoline (all			
Other motor f			
Personal care			
Hair, dental,			
Cosmetics, pe			
Household cle			
Household pap			
Miscellaneous			

RMSE < Q1
Q1 <= RMSFE < Q2
Q2 <= RMSFE < Q3
Q3 <= RMSFE < Q4
Q4 <= RMSFE < Q5

VII.5 Forecasting comparison

Table 9 includes the results of an out of sample[8] forecasting exercise for the evaluation period 2010.1 − 2016.12. At each month of this period the 12 forecasting models described above are estimated using information up to the previous month, and multi-step ahead forecasts are produced for horizons $H = 1$ to $H = 12$. The computation of the fully cointegrated subsets, and the corresponding cointegration relationships, is carried out only each December. Hence, in PW approaches we are using less information than the truly available, except for January.

First row of table 9 includes the root mean squared forecast error (RMSFE) of $\Delta_{12}log(CPI)$ from horizons $H = 1$ to $H = 12$ of the direct procedure. All the other values in the table are ratios with respect to the first row. The best procedures are highlighted in green and the worst in red. Table 10 includes pvalues of the Diebold-Mariano tests for comparing forecasting accuracy of selected methods.

Direct approaches

The use of disaggregated information in a scalar model for the aggregate, as proposed by Hendry and Hubrich (2011), deteriorate the results in short horizons. In longer horizons (9 to 12) the procedure D-DI produce similar results as those of the baseline. The inclusion of dynamic factors extracted form the disaggregates deteriorates the baseline in short horizons and produce some forecasting gains in $H = 11$ and $H = 12$.

Indirect approaches

Table 9 shows that mixed methods that combine the forecasts of certain components with the forecast for the sub-aggregate of the remaining ones (methods labeled a GP at the beginning of this section) are the clear winners of the forecasting competition, both in short and long horizons. The forecasting of gains these procedures are much more remarkable for long horizons. Focus, for example, in line 9 of table 9: While the relative RMSFE in horizons 1 to 5 is around 0.95, in horizon 12 this figure reduces to 0.83. As we comment below, these reductions are statistically significant.

The worst performers are, in general, the disaggregated methods that do not incorporate any restriction (those labeled as I-B). As can be seen in table 9, red shadows appear mainly in lines corresponding to I-B procedures.

When including dynamic factors in the models for the disaggregates, results show some improvements with respect to I-B procedures (line 7 of the table).

In the I-PW approach, components that do not belong to any fully cointegrated subset are forecast with univariate models as in I-B. As table 9 shows, this approach performs similarly to I-DFM and cannot beat the baseline. However, as argued above, when components outside fully cointegrated subsets are forecast altogether in a single model for its

[8]Since we use the currently available series, which contains data revisions that were not available when the data was published for the first time, our exercise is not strictly out of sample but a *pseudo* out of sample exercise.

subaggregate, forecasting gains are remarkable, in long horizons the RMSFE is reduced in almost 20 percentage points. This reduction is economically very important and statistically significant, as table 10 shows.

This result highlights the great importance that disaggregating may have, but also suggests that the disaggregation cannot be done indiscriminately as in the I-B approach. The forecasting gains appear when the only components that are forecast individually are those belonging to some fully cointegrated subset. This is a clear indication that pairwise procedure proposed in this paper shows a way for selecting the components for which computing individual forecasts worth it.

Table 9: Relative RMSFE $\Delta_{12}log(CPI)$. (First row: RMSFE for the baseline. All the others are ratios with respect to the first)

		H=1	H=2	H=3	H=4	H=5	H=6	H=7	H=8	H=9	H=10	H=11	H=12
(1)	**D (baseline)**	**0.23**	**0.40**	**0.52**	**0.60**	**0.68**	**0.77**	**0.88**	**0.98**	**1.06**	**1.13**	**1.20**	**1.30**
(2)	D-DI-2	1.09	1.16	1.15	1.15	1.14	1.12	1.10	1.08	1.06	1.05	1.04	1.04
(3)	D-DFM	1.03	1.05	1.10	1.13	1.14	1.11	1.09	1.07	1.05	1.01	0.97	0.96
(4)	I-B	1.08	1.10	1.16	1.17	1.15	1.13	1.11	1.10	1.11	1.10	1.10	1.10
(5)	I-B-CPI	1.11	1.14	1.18	1.19	1.19	1.17	1.13	1.11	1.11	1.11	1.10	1.09
(6)	I-B-DI	1.04	1.07	1.13	1.14	1.14	1.11	1.09	1.11	1.11	1.10	1.10	1.09
(7)	I-DFM	1.07	1.08	1.11	1.12	1.12	1.10	1.07	1.07	1.07	1.07	1.09	1.09
(8)	I-PW	1.05	1.10	1.16	1.18	1.17	1.16	1.11	1.09	1.09	1.08	1.08	1.07
(9)	I-PW-CPI	1.06	1.08	1.12	1.15	1.14	1.11	1.07	1.07	1.08	1.09	1.08	1.06
(10)	I-PW-CPI-GP	0.93	0.93	0.96	0.96	0.94	0.90	0.88	0.87	0.86	0.84	0.84	0.83
(11)	I-PW-DI	1.05	1.11	1.18	1.21	1.19	1.16	1.11	1.09	1.08	1.08	1.08	1.06
(12)	I-PW-DI-GP	1.05	1.02	1.01	0.96	0.91	0.86	0.85	0.85	0.85	0.84	0.84	0.82

- See table 6 for a description of each exercise.
- Dark red entrances highlight the loser procedure.
- Light red indicates procedures j for which $RMSFE_{loser} - RMSFE_j \leq 0.01$.
- Dark green indicates the best procedure.
- Light green indicates procedures j for which $RMSFE_j - RMSFE_{winner} \leq 0.01$.

Table 10 includes p-values for the Diebold-Mariano tests for comparing forecasting accuracy of some selected methods. The table shows that all the arguments made above are still valid when instead of just looking at relative RMSFE we analyze the statistical significance of the differences.

Table 10: P-values of Deibold-Mariano tests for selected comparisons

	H=1	H=2	H=3	H=4	H=5	H=6	H=7	H=8	H=9	H=10	H=11	H=12
9 vs. 7	0.42	0.44	0.43	0.22	0.28	0.39	0.50	0.41	0.37	0.28	0.46	0.14
9 vs. 1	0.14	0.10	0.02	0.01	0.01	0.02	0.05	0.09	0.09	0.06	0.06	0.11
9 vs. 4	0.37	0.29	0.19	0.36	0.37	0.24	0.15	0.15	0.20	0.30	0.22	0.02
10 vs.1	0.03	0.07	0.20	0.19	0.06	0.01	0.01	0.01	0.01	0.00	0.00	0.00
10 vs.7	0.01	0.02	0.02	0.02	0.01	0.01	0.00	0.01	0.00	0.00	0.00	0.00

- Row names refer to the lines of table 9
- The null hypothesis is that the RMSFE are equal.

VIII Concluding Remarks

In this paper we studied the properties of a pairwise procedure for testing cointegration between all possible pairs between the components of an aggregate at the maximum level of disaggregation and applied it to the US CPI. This procedure allows to discover subsets of series that share a unique common trend (*fully cointegrated subsets*).

As our methodology does not rely on any cross-sectional averaging procedure, we need neither to assume pervasiveness of the common trends, nor to impose special restrictions on serial or cross-correlation of idiosyncratic components. Furthermore, we do not need the cross-sectional dimension to go to infinity.

For making the procedure relevant in empirical applications we extended it for making it robust to data irregularities and short samples issues. The application to all the components of the US CPI offer several important conclusions. First, the outliers' analysis at the components level is very convenient both for studying the components themselves and for modeling the aggregate. We found that the pattern by which the US CPI is congested of series perturbed by outliers along the time is not random; it has some seasonality and experiences changes in mean related with the general economic conditions. We also found that an indicator for the aggregate outliers (AO_t) constructed by aggregating all the outliers of the components, is a relevant variable for modeling the aggregate.

In the construction of subsets of components that share a unique common trend (fully cointegrated subsets), we concluded that the usual practice of extracting common factors from components that belong to the same broad category of the CPI (energy, services, durable manufactures, non-durable manufactures, processed food, and non-processed food) is not well founded because the components in the fully cointegrated subsets do not, in general, belong to the same broad category.

Our proposal for forecasting the aggregate is to do it indirectly, by constructing single-equation models for each component including the restrictions derived form the fully cointegrated subsets and, then, adding up the components' forecast. In a forecasting competition exercise we compared the ability of our procedure for forecasting the aggregate with other direct and indirect alternatives. The results show that disaggregation could be greatly relevant for forecasting, but it cannot be done indiscriminately. Indiscriminate disaggregation can derive in worse forecasting accuracy than direct methods. Our results suggest that the pairwise approach shows a promising way for choosing useful disaggregations. When doing individual forecasts just for the components that belong to the fully cointegrated subsets, the root mean squared forecast error of the aggregate is reduced in almost 20 percentage points in long horizons.

The main theoretical result of the paper is that pairwise cointegration tests inside fully cointegrated subsets are asymptotically equivalent, in the sense that the probability that all tests deliver the same conclusion tends to 1 as T goes to infinity, independently of the number of series. Thus, multiple testing is not an issue for pairs of components inside a fully cointegrated subset. This result is valid both when N is fixed and when it goes to infinity. Additionally, we showed that the risk of including wrong components in the estimated fully cointegrated subsets, as well as the risk of wrongly discovering subsets composed by outsiders, can be easily controlled.

Monte Carlo experiments confirm the asymptotic results and show a good small samples behavior for alternative data structures.

We also showed that the pairwise approach can be extended for sets of macro variables (not necessarily components of a single one) with *general* and *sectorial* trends. Theoretical results still apply to this case, which requires testing cointegration in all pairs and in some triplets of series. Monte Carlo experiments also showed a good performance in discovering the sectors.

References

Altissimo, F., B. Mojon, and P. Zaffaroni (2007). Fast micro and slow macro: Can aggregation explain the persistence of inflation? *Federal Reserve Bank of Chicago WP 2007-02*.

Ando, T. and J. Bai (2015). Panel data models with grouped factor structure under unknown group membership. *Journal of Applied Econometrics*.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica 71*(1), 135–171.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*(1), 191–221.

Bailey, N., S. Holly, and M. H. Pesaran (2015). A two-stage approach to spatio-temporal analysis with strong and weak cross-sectional dependence. *Journal of Applied Econometrics*.

Bailey, N., G. Kapetanios, and M. H. Pesaran (2015). Exponent of cross-sectional dependence: Estimation and inference. *Journal of Applied Econometrics*.

Beck, G., K. Hubrich, and M. Marcellino (2015). On the importance of sectoral and regional shocks for price-setting. *Journal of Applied Econometrics , forthcomming.*

Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological) 57*(1), 289–300.

Bils, M. and P. J. Klenow (2004). Some evidence on the importance of sticky prices. *Journal of Political Economy 112*(5), 947–985.

Boivin, J., M. P. Giannoni, and I. Mihov (2009). Sticky prices and monetary policy: Evidence from disaggregated us data. *The American Economic Review 99*(1), 350–384.

Boivin, J. and S. Ng (2006). Are more data always better for factor analysis? *Journal of Econometrics 132*(1), 169–194.

Bollobás, B. and P. Erdös (1976). Cliques in random graphs. In *Mathematical Proceedings of the Cambridge Philosophical Society*, Volume 80, pp. 419–427. Cambridge Univ Press.

Breitung, J. and S. Eickmeier (2015). Analyzing business cycle asymmetries in a multi-level factor model. *Economics Letters 127*, 31–34.

Bron, C. and J. Kerbosch (1973). Algorithm 457: finding all cliques of an undirected graph. *Communications of the ACM 16*(9), 575–577.

Castle, J., J. Doornik, and D. Hendry (2011). Evaluating automatic model selection. *Journal of Time Series Econometrics 3*(169), 239–246.

Castle, J. L., J. A. Doornik, and D. F. Hendry (2012). Model selection when there are multiple breaks. *Journal of Econometrics 169*(2), 239–246.

Castle, J. L. and D. F. Hendry (2010). Nowcasting from disaggregates in the face of location shifts. *Journal of Forecasting 29*(1-2), 200–214.

Cheung, Y.-W. and K. S. Lai (1993). Finite-sample sizes of johansen's likelihood ratio tests for cointegration. *Oxford Bulletin of Economics and Statistics 55*(3), 313–328.

Clark, T. E. (2006). Disaggregate evidence on the persistence of consumer price inflation. *Journal of Applied Econometrics 21*(5), 563–587.

Clements, M. P. and D. F. Hendry (1995). Forecasting in cointegrated systems. *Journal of Applied Econometrics 10*(2), 127–146.

Cubadda, G. and E. Scambelloni (2015). Index-augmented autoregressive models: Representation, estimation, and forecasting. *Working Paper*.

Doornik, J. (2009). Autometrics. In *Castle and Shephard, The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry.*, pp. 88–121. Oxford University Press.

Doornik, J., D. Hendry, and B. Nielsen (1998). Inference in cointegrating models: Uk m1 revisited. *Journal of economic surveys 12*(5), 533–572.

Doornik, J., D. Hendry, and F. Pretis (2013). Step-indicator saturation. *University of Oxford. Working Paper* (658).

Doz, C., D. Giannone, and L. Reichlin (2012). A quasi–maximum likelihood approach for large, approximate dynamic factor models. *Review of Economics and Statistics 94*(4), 1014–1024.

Engle, R. and S. Kozicki (1993). Testing for common features. *Journal of Business & Economic Statistics 11*(4), 369–380.

Espasa, A. and I. Mayo-Burgos (2013). Forecasting aggregates and disaggregates with common features. *International Journal of Forecasting 29*(4), 718–732.

Espasa, A., E. Senra, and R. Albacete (2002). Forecasting inflation in the european monetary union: A disaggregated approach by countries and by sectors. *The European Journal of Finance 8*(4), 402–421.

Giacomini, R. and C. Granger (2004). Aggregation of space-time processes. *Journal of Econometrics 118*(1), 7–26.

Guerrero, V. M. and D. Peña (2003). Combining multiple time series predictors: a useful inferential procedure. *Journal of Statistical Planning and Inference 116*(1), 249–276.

Harbo, I., S. Johansen, B. Nielsen, and A. Rahbek (1998). Asymptotic inference on cointegrating rank in partial systems. *Journal of Business & Economic Statistics 16*(4), 388–399.

Hendry, D. and K. Hubrich (2011). Combining disaggregate forecasts or combining disaggregate information to forecast an aggregate. *Journal of business & economic statistics 29*(2), 216–227.

Hendry, D. F. (1995). On the interactions of unit roots and exogeneity. *Econometric reviews 14*(4), 383–419.

Imbs, J., H. Mumtaz, M. O. Ravn, and H. Rey (2005). Ppp strikes back: Aggregation and the real exchange rate. *The Quarterly Journal of Economics 120*(1), 1–43.

Johansen, S. (1992). Cointegration in partial systems and the efficiency of single-equation analysis. *Journal of Econometrics 52*(3), 389–402.

Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive models*, Volume 9. Cambridge University Press.

Johansen, S. and K. Juselius (2014). An asymptotic invariance property of the common trends under linear transformations of the data. *Journal of Econometrics 178*, 310–315.

Johansen, S., R. Mosconi, and B. Nielsen (2000). Cointegration analysis in the presence of structural breaks in the deterministic trend. *The Econometrics Journal 3*(2), 216–249.

Johansen, S. and B. Nielsen (2009). An analysis of the indicator saturation estimator as a robust regression estimator. In *Castle, and Shephard (2009), The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry.*, pp. 1–36. Oxford University Press.

Juselius, K. (2006). *The cointegrated VAR model: methodology and applications.* Oxford University Press, USA.

Juselius, K. (2015). Haavelmo's probability approach and the cointegrated var. *Econometric theory 31*(02), 213–232.

Karadimitropoulou, A. and M. León-Ledesma (2013). World, country, and sector factors in international business cycles. *Journal of economic dynamics and control 37*(12), 2913–2927.

Kiviet, J. F. and G. D. Phillips (1992). Exact similar tests for unit roots and cointegration. *Oxford Bulletin of Economics and Statistics 54*(3), 349–367.

Lütkepohl, H. (1984). Linear transformations of vector arma processes. *Journal of Econometrics 26*(3), 283–293.

Lütkepohl, H., P. Saikkonen, and C. Trenkler (2003). Comparison of tests for the cointegrating rank of a var process with a structural shift. *Journal of Econometrics 113*(2), 201–229.

Lütkepohl, H., P. Saikkonen, and C. Trenkler (2004). Testing for the cointegrating rank of a var process with level shift at unknown time. *Econometrica 72*(2), 647–662.

Lunnemann, P. and T. Mathä (2004). How persistent is disaggregate inflation? an analysis across eu 15 countries and hicp sub-indices. *Working paper, European Central Bank* (415).

Matula, D. W. (1976). The largest clique size in a random graph. *Department of Computer Science, Southern Methodist University, Technical Report CS 7608*.

Mayo, I. and A. Espasa (2009). Forecasting aggregates and disaggregates with common features. Technical report, Universidad Carlos III de Madrid. Departamento de Estadística, Working Paper (later published as Espasa and Mayo-Burgos (2013)).

Moench, E., S. Ng, and S. Potter (2013). Dynamic hierarchical factor models. *Review of Economics and Statistics 95*(5), 1811–1817.

Newman, M. (2009). *Networks: an introduction*. OUP Oxford.

Nielsen, B. (2004). On the distribution of likelihood ratio test statistics for cointegration rank. *Econometric reviews 23*(1), 1–23.

Nielsen, B. and A. Rahbek (2000). Similarity issues in cointegration analysis. *Oxford Bulletin of Economics and Statistics 62*(1), 5–22.

Osborn, D. R., A. P. Chui, J. P. Smith, and C. R. Birchenhall (1988). Seasonality and the order of integration for consumption*. *Oxford Bulletin of Economics and Statistics 50*(4), 361–377.

Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica: Journal of the Econometric Society 57*, 1361–1401.

Perron, P. (2006). Dealing with structural breaks. *Palgrave handbook of econometrics 1*, 278–352.

Perron, P. and G. Rodríguez (2003). Searching for additive outliers in nonstationary time series*. *Journal of Time Series Analysis 24*(2), 193–220.

Perron, P. and T. J. Vogelsang (1992). Nonstationarity and level shifts with an application to purchasing power parity. *Journal of Business & Economic Statistics 10*(3), 301–320.

Pesaran, H. (2007). A pair-wise approach to testing for output and growth convergence. *Journal of Econometrics 138*(1), 312–355.

Pesaran, M. H., T. Schuermann, and S. M. Weiner (2004). Modeling regional interdependencies using a global error-correcting macroeconometric model. *Journal of Business & Economic Statistics 22*(2), 129–162.

Saikkonen, P. and H. Lütkepohl (2000). Testing for the cointegrating rank of a var process with structural shifts. *Journal of Business & Economic Statistics 18*(4), 451–464.

Saikkonen, P. and H. Lutkepohl (2002). Trend adjustment prior to testing for the cointegrating rank of a vector autoregressive process. *Journal of Time Series Analysis 21*(4), 435–456.

Santos, C., D. Hendry, and S. Johansen (2008). Automatic selection of indicators in a fully saturated regression. *Computational Statistics 23*(2), 317–335.

Stock, J. and M. Watson (2007). Why has us inflation become harder to forecast? *Journal of Money, Credit and Banking 39*, 3–33.

Trenkler, C., P. Saikkonen, and H. Lütkepohl (2007). Testing for the cointegrating rank of a var process with level shift and trend break. *Journal of Time Series Analysis 29*(2), 331–358.

Zivot, E. and J. Wang (2007). *Modeling financial time series with S-Plus*, Volume 191. Springer Science & Business Media.

Appendix A Detail of the fully cointegrated subsets

Table A.1: Detail of the fully cointegrated subsets. Estimation sample: 1999.1-2016.12

	Weight in %	Category
Subset 1		
Potatoes	0.11	NPF
Lettuce	0.09	NPF
Tomatoes	0.12	NPF
Frozen noncarbonated juices and drinks	0.02	PF
Other beverage materials including tea	0.15	PF
Frozen and freeze dried prepared foods	0.38	PF
Men's suits, sport coats, and outerwear	0.13	ManDur
Photographic equipment and supplies	0.06	ManDur
Physicians' services	2.49	Serv
Subset 2		
Cakes, cupcakes, and cookies	0.25	PF
Soups	0.13	PF
Cosmetics, perfume, bath, nail preparations and implements	0.47	ManNoDur
Apparel services other than laundry and dry cleaning	0.04	Serv
Garbage and trash collection	0.42	Serv
Car and truck rental	0.15	Serv
Parking and other fees	0.34	Serv
Subset 3		
Other fresh vegetables	0.37	NPF
Men's shirts and sweaters	0.25	ManDur
Girls' apparel	0.30	ManDur
Financial services	0.34	Serv
State motor vehicle registration and license fees	0.41	Serv
Subset 4		
Other meats	0.38	PF
Spices, seasonings, condiments, sauces	0.41	PF
Women's outerwear	0.10	ManDur
Educational books and supplies	0.24	Serv
Limited service meals and snacks	3.59	Serv
Subset 5		
Apples	0.12	NPF
Butter and margarine	0.10	PF
Women's dresses	0.21	ManDur
Computer software and accessories	0.12	ManDur
Full service meals and snacks	4.10	Serv
Subset 6		
Fuel oil	0.15	ENE
Other motor fuels	0.07	ENE
Uncooked ground beef	0.27	PF
Uncooked beef steaks	0.24	PF
Pork chops	0.08	PF
Subset 7		
Other fresh fruits	0.37	NPF
Electricity	4.08	ENE
Bread	0.31	PF
Other food away from home	0.36	Serv
Dental services	1.19	Serv

Appendix B Additional results of the outliers' analysis

Table B.1: Series with no Outliers. Estimation sample: 1999.1-2016.12

	Series with no outliers			
		W (%)	**Category**	**W of Cat (%)**
1	Citrus fruits	0.21	NPF	2.5
2	Utility (piped) gas service	1.16	ENE	11.6
3	Cakes, cupcakes, and cookies	0.25	PF	
4	Bacon, breakfast sausage, and related products	0.19	PF	
5	Other pork including roasts and picnics	0.11	PF	
6	Chicken	0.40	PF	
7	Processed fish and seafood	0.18	PF	
8	Frozen fruits and vegetables	0.12	PF	
9	Sugar and artificial sweeteners	0.07	PF	
10	Soups	0.13	PF	11.4
11	Beer, ale, and other malt beverages at home	0.40	PF	
12	Wine at home	0.35	PF	
13	Men's furnishings	0.28	ManDur	
14	Boys' apparel	0.22	ManDur	
15	Women's suits and separates	0.67	ManDur	
16	Jewelry	0.20	ManDur	
17	Dishes and flatware	0.07	ManDur	
18	Nonelectric cookware and tableware	0.10	ManDur	20.3
19	Toys	0.43	ManDur	
20	Tires	0.33	ManDur	
21	Personal care products	1.02	ManNoDur	
22	Hair, dental, shaving, and miscellaneous personal care products	0.54	ManNoDur	3.3
23	Household cleaning products	0.50	ManNoDur	
24	Rent of primary residence	11.49	Serv	
25	Other lodging away from home including hotels and motels	1.06	Serv	
26	Tenants' and household insurance	0.50	Serv	
27	Car and truck rental	0.15	Serv	
28	Motor vehicle maintenance and repair	1.70	Serv	51.0
29	Other intercity transportation	0.26	Serv	

Table B.2: Highly contaminated series (5% or more observations are outliers). Estimation sample: 1999.1-2016.12

		W (%)	**Category**	**W of Cat (%)**
1	Bananas	0.13	NPF	2.5
2	Electricity	4.08	ENE	11.6
3	Tobacco products other than cigarettes	0.07	PF	11.4
4	Cigarettes	0.88	PF	
5	Used cars and trucks	2.90	ManDur	20.3
6	Haircuts and other personal care services	0.89	Serv	
7	Intracity transportation	0.41	Serv	
8	Financial services	0.34	Serv	
9	Physicians' services	2.49	Serv	
10	Parking and other fees	0.34	Serv	51.0
11	Food from vending machines and mobile vendors	0.12	Serv	
12	Internet services and electronic information providers	1.04	Serv	
13	State motor vehicle registration and license fees	0.41	Serv	
14	Wireless telephone services	2.54	Serv	