Working paper

2020-10

Statistics and Econometrics

ISSN 2387-0303

What do international energy prices have in common after taking into account the key drivers?

Máximo Camacho, Ángela Caro and Daniel Peña

uc3m | Universidad Carlos III de Madrid

What do international energy prices have in common after taking into account the key drivers?

Maximo Camacho*
Department of Quantitative Methods for the Economy, University of Murcia
and
Angela Caro†
Department of Statistics, University Carlos III of Madrid
and
Daniel Peña‡
Department of Statistics, Institute UC3M-BS of Big Data
University Carlos III of Madrid

December 17, 2020

Abstract

Differences across international energy prices are driven by many factors, but what do energy prices have in common? We analyze global and group-specific co-movements in the energy market by means of a Dynamic Factor Model with Cluster Structure. A new extension of the model is included which allows to evaluate the effect of macroeconomic variables which are country-specific over energy prices. A Monte Carlo experiment is carried out in order to test the estimation performance of the proposed extension. In a Big Data scenario of 30 countries and 12 industrial sectors we find that the co-movements between energy prices are related to groups of countries and may be classified within four groups. The connections within groups may be explained by high prices of a specific fuel type.

Keywords: Energy prices, Dynamic Factor Model, Clustering, Penalized regression

*Supported by the program Groups of Excellence of the region of Murcia, the Fundación Séneca, Science and Technology Agency of the region of Murcia project 19884/GERM/15.

†Supported by the Spanish Ministerio de Ciencia, Innovación y Universidades under grant FPU15/03983 and the program Groups of Excellence of the region of Murcia, the Fundación Séneca, Science and Technology Agency of the region of Murcia project 19884/GERM/15.

‡Partially supported by the Spanish Agencia Estatal de Investigación under grant PID2019-109196GB-I00.

1 Introduction

Energy economics and energy policy are two important concerns for the development of countries. Governments and investors have to pay close attention to international energy prices given that they affect not only the competitiveness of industry yet households. Recently, the intervention of governments in the regulation of energy is gaining even more importance, since the environment is suffering the consequences of the improper use of energy fuels.

We already know that there exist differences in the energy market structures across countries due to production and transportation costs, trade restrictions, and contractual terms, between others factors. Nevertheless, after taking into account these differences, we are interested in the existence of common factors that describe a global behavior in the international energy market, together with group-specific factors explaining energy prices related to regions, countries or industrial sectors.

After three decades since the pioneering work of Griffin (1980) about energy economics and policy, research in this topic has increased considerably according to the economic events. The lack of data about international energy prices has been an obstacle for researchers, nevertheless, the establishment of statistical agencies and developments in data storage and the internet have provided new tools which are capable to analyze big data sets about energy. Previous works have analyzed energy prices paying attention to individual fuel types, for example Brown et al. (2008) studies what affects natural gas and crude oil prices in USA, and Nick and Thoenes (2014) pays attention to natural gas prices in Germany. International oil prices have been considered by Van Benthem and Romani (2009) and Aastveit et al. (2015) together with domestic end-use energy and demand, respectively. How energy prices relate to energy comsumption and investment was examined in Mahadevan and Asafu-Adjaye (2007) and Bretschger (2015) using the consumer price index and energy use as proxies for energy price. Only a few of these articles consider different fuel types together in their analysis of energy prices, and in such cases they are used as control variables.

Up to our knowledge, we are pioneers in analyzing the co-movements of international energy prices in a bigdata scenario of 30 countries and 12 industrial sectors. The data set is from Sato et al. (2019). We consider for the analysis the Dynamic Factor Model with Cluster Structure (DFMCS) which allows us to investigate if there exists a group structure between international energy prices, to characterize the heterogeneity of the global energy market based on industry, country or region, to quantify the extent to which "crisis" affected the global energy prices, and to identify the sources that explain the cross-section variations in energy prices through control variables which are country-specific.

We extend the methodology proposed in Alonso et al. (2020) in order to study the effect of control variables, which are country specific, over energy prices. We also run a Monte Carlo simulation to evaluate the performance, in finite samples, of Alonso et al. (2020) clustering procedure when we take into account control variables. Results provide useful interpretations about the existence of trading groups of countries in the global energy market.

The rest of the article is organized as follows. Section 2 introduces the model. The estimation method and the Monte Carlo experiment are described in Section 3. Section 4 presents the data and Section 5 contains the empirical results. Finally, some concluding remarks and future extensions are given in Section 6.

2 Theoretical framework

In this section we introduce an extension of the DFMCS framework to analyze what drives energy prices for 12 industrial sectors and over 30 OECD and non-OECD countries. It consists in including in the DFMCS an 'observed' component of macroeconomic variables which are country-dependent.

Let $t = 1, ..., T$ represents the time index and $i = 1, ..., N$ the cross-section index, the unknown and fixed number of groups is S and $G = g_1, ..., g_N$ is the group membership with $g_i \in 1, ..., S$. The number of countries is Q and $C = c_1, ..., c_N$ represent the country membership with $c_i \in 1, ..., Q$. There are N_j units within group j $(j = 1, ..., S)$ such that $N = \sum_{j=1}^{S} N_j$. The response variable of the ith unit, observed at time t, y_{it}, is defined as

$$y_{it} = x_{it}'\beta_{c_i} + f_{0,t}'\lambda_{0,i} + f_{g_i,t}'\lambda_{g_i,i} + \varepsilon_{it}. \quad (1)$$

where x_{it} is a $p \times 1$ vector of observable variables, $f_{0,t}'$ is a $r_0 \times 1$ vector of unobserved global factors affecting all the N series in the sample, $f_{g_i,t}'$ is a $r_j \times 1$ vector of unobserved group-specific factors that affect only the units in group g_i, β_{c_i} is a $p \times 1$ vector of unknown regression coefficients for country q $(q = 1, ..., Q)$, $\lambda_{0,i}$ and $\lambda_{g_i,i}$ are the corresponding factor loadings, and $\varepsilon_{i,t}$ is the unit specific error. Here β_{c_i} is common for all i units belonging to country q.

In line with Wang (2008) identifying conditions for large dimensional factor models, we consider the following assumptions to identify the model and make possible the estimation. Let Λ_0 and Λ_j be the corresponding matrix of factor loadings, then we assume that $\Lambda_0'\Lambda_0 = \mathbf{I}_{r_0}$, $\Lambda_j'\Lambda_j = \mathbf{I}_{r_j}$ for

$j = 1, ..., S$, the $r = \sum_{j=0}^{S} r_j$ covariance matrices of the factors are diagonal, $\Lambda_0'\Lambda_j = \mathbf{0}_{r_0 \times r_j}$ and $\Lambda_j'\Lambda_i = \mathbf{0}_{r_j \times r_i}$ for $j \neq i$.

3 Estimation method

The model is estimated using the procedure introduced in Alonso et al. (2020). This method works well and seems to be better than the one in (Ando and Bai, 2017) in terms of estimation of factors and loadings, as shown by their Monte Carlo simulation results.

Given the DFMCS in (1), we just have information about the left hand side and the control variables included in matrix X. The unknown parameters to be estimated are the number of groups, S, the number of common and group-specific factors, r_0 and r_j, the corresponding factors and its loadings, $\widehat{F}_0, \widehat{F}_1, ..., \widehat{F}_S, \widehat{\Lambda}_0, \widehat{\Lambda}_1, ..., \widehat{\Lambda}_S$, the membership of each variable to a given group, g_i, and the sensitivity to observable factors, β_{c_i}. We need to modify Alonso et al. (2020) procedure, AGP in what follows, because they do not consider exogenous variables and we want to take into account the effect of macroeconomic variables, which are country-specific, over energy prices.

AGP includes a prior step in which the observed time series, y_{it}, are cleaned from additive outliers and level shifts, see Alonso et al. (2020) for more details about the cleaning procedure. We extend AGP to consider the estimation of the regression coefficients, β_{c_i} and divide the estimation of model (1) in five steps:

1. Estimate the country-dependent regression coefficients, β_{c_i}, by minimizing

$$L(\beta_{c_i}) = \sum_{i=1}^{N} \parallel y_i - X_i\beta_{c_i} \parallel^2 + T \sum_{i=1}^{N} p_i(\beta_{c_i}). \tag{2}$$

where

$$p_i(\beta_{c_i}) \equiv p_{\kappa_{c_i},\gamma}(\beta_{c_i}) = \sum_{c_i=1}^{p_{c_i}} p_{\kappa_{c_i},\gamma}(\mid \beta_{c_i} \mid),$$

is the SCAD (*smooth clipped absolute deviation*) penalty of (Fan and Li, 2001) as suggested in Ando and Bai (2017), with $\kappa_{c_i} > 0$ and $\gamma = 3.7$, which minimize a Bayesian risk criteria for the regression coefficients. The size of the penalty which is control by κ_{c_i}, for $c_i \in 1, ..., Q$, is the same for those cross-sectional units related to country q with $q = 1, ..., Q$ in the sample. Given $\widehat{\beta}_{c_i}$, we subtract the corresponding 'observed' component from each time series, y_i, obtaining

$$y_i^* = y_i - X_i\widehat{\beta}_{c_i}.$$

2. Estimate an initial set of global factors F_0 and their corresponding loadings Λ_0. The number of global factors r_0 is obtained using the test proposed in Caro and Peña (2020). The factors are estimated by $\widehat{f}_{0t} = \widehat{\lambda}_0 y_t^*$, and the common component by $c_t = \widehat{\Lambda}_0 \widehat{\Lambda}_0' y_t^*$, where $\widehat{\Lambda}_0$ is the estimated matrix of factor loadings which columns are the eigenvectors of the *combined dynamic correlation matrix* of the observed data, $\mathbf{R}_{k_0}$, introduced in Caro and Peña (2020), associated to the r_0 largest eigenvalues.

3. Apply the clustering algorithm proposed in Alonso and Peña (2019), based on similar linear dependency measures between time series, to the estimated common component $\widehat{\Lambda}_0 \widehat{f}_{0,t}$. Once the optimal number of groups S is calculated using a modification of the Silhouette algorithm proposed by Rousseeuw (1987), the memberships g_i for $i = 1, ..., N$ are obtained.

4. Obtain the number of group-specific factors r_j for $j = 1, .., S$ using the CP test taking into acount the time series $y_{i,t}$ belonging to each group. The corresponding factors $F_1, ..., F_S$ and their loadings $\Lambda_1, ..., \Lambda_S$ are estimated as describe in step (2) for the global factors. In this step, all the global and group-specific factors from steps 2 and 4 are compared and classified according to the decision rules in Alonso et al. (2020) based on empirical canonical correlations.

5. As suggested in Alonso et al. (2020), it must be verified that the groups are a consequence of specific factors and not due to different loadings corresponding to global factors. Therefore, group-specific residuals $v_t = y_t^* - \widehat{\Lambda}_0 \widehat{f}_{0,t}$ are obtained and used to re-estimate the group-specific factors. Finally, each group must be analyzed to check the existence of at least one specific factor.

Our objective in the next sections is, by means of a Monte Carlo simulation, to evaluate the clustering performance of AGP under different data generation processes, and to applied the proposed extention to the analysis about international energy prices.

3.1 Monte Carlo experiment

In this section we simulate a data structure similar to the one in our data set about energy prices and evaluate the clustering performance of AGP procedure. We consider three data-generating processes (DGP) and we set the number of countries in each DGP to be three, such that $c_i \in 1, 2, 3$. Each variable y_i for $i = 1, ..., N$ is generated as

$$y_i = X_i \beta_{c_i} + F_0 \lambda_{0,i} + F_{g_i} \lambda_{g_i,i} + \varepsilon_i, \tag{3}$$

where the r-dimensional global common factor $f_{0,t}$ is a vector of $U(0,1)$ variables and the corresponding elements of the loading matrix Λ_0 follow $U(-2,2)$ distribution, the r_j-dimensional group-specific factor $f_{g_i,t}$ ($j = 1, ..., S$), is a vector of $N(0,1)$ variables, and each element of the factor loading matrix Λ_j is generated from the $N(0.5j, 1)$ distribution.

The number of columns of X_i is set to $p = 30$, while the true number of predictors is $q = 3$. Each of the elements of X_i is generated from $N(0,1)$ distribution. The non-zero true parameter values of β_{c_i} are set to be $(4, 3.5, 3)$ for country 1, $(-2.5, -2, -2.5)$ for country 2, and $(1, 0.5, 1.5)$ for country 3. These non-zero elements are put into the first three elements of β_{c_i}, for example, the true parameter vector is $\beta_1 = (4, 3.5, 3, 0, 0, ..., 0)'$ for country 1. We set the number of groups $S = 3$. We assume that each country has 99 series, the first 33 series of each country belong to group 1, the second 33 series, from 34 to 66, of each country belong to group 2, and the last 33 series, from 67 to 99, of each country belong to group 3. We consider the sample sizes $N = \{297, 594\}$ and the number of time observations $T = \{100, 200\}$.

First DGP assumes that the N-dimensional vector ε_t has a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\sigma_e^2 \mathbf{I}_N$. Second DGP has non-homoscedastic errors with cross-sectional dependence such that $\varepsilon_{it} = 0.2e_{it}^1 + \delta_t e_{it}^2$, where $\delta_t = 1$ if t is odd and zero if t is even, and the N-dimensional vectors $e_t^1 = (e_{1t}^1, ..., e_{Nt}^1)'$ and $e_t^2 = (e_{1t}^2, ..., e_{Nt}^2)'$ follow multivariate normal distributions with mean $\mathbf{0}$ and covariance matrix $S = (s_{ij})$ with $s_{ij} = 0.3^{|i-j|}\sigma_e^2$ and e_t^1 and e_t^2 are independent. The third DGP presents errors which are serial and cross-sectional correlated such that, $\varepsilon_{it} = 0.2\varepsilon_{i,t-1} + e_{it}$, where $t = 1, ..., T$, the N-dimensional vector $e_t = (e_{1t}, ..., e_{NT})'$ follows multivariate normal distributions with mean $\mathbf{0}$ and covariance matrix $S = (s_{ij})$ with $s_{ij} = 0.3^{|i-j|}\sigma_e^2$. We consider the noise variances $\sigma_e^2 = 1, 2$.

We generate 100 replications using each of the three data-generating process. In each replication the proposed procedure is applied to the simulated data in order to select, simultaneously, the number of groups, the number of global common factors, the number of group-specific pervasive factors and the size of the regularization parameter. We set the possible numbers of group-specific and global factors to range from zero to eight. The number of groups ranges from two to twelve. Possible candidates for the regularization parameter κ_i are $\kappa_i = 10^{-3+0.25k}$ for $k = 0, ..., 12$.

Table 1 shows the mean of the selected number of clusters for the 100 iterations, and below each mean the total number of iterations out of 100 where the true number was selected.
We see how AGP tends to underestimate the number of groups, S. The method suffers with the increase of variance under DGP_1, over all when $T = 100$. Under DGP_2 and DGP_3, where errors are allowed to present serial and limited cross-section correlation, AGP provides very accurate

Table 1: Mean of the selected number of clusters (first row) and number of iterations out of 100 where the true number of clusters was selected (second row).

	DGP_1		DGP_2		DGP_3	
(T, N)	$\sigma^2 = 1$	$\sigma^2 = 2$	$\sigma^2 = 1$	$\sigma^2 = 2$	$\sigma^2 = 1$	$\sigma^2 = 2$
(100, 297)	2.96	2.08	2.98	2.89	2.92	2.51
	97	29	98	89	93	62
(100, 594)	3.00	2.17	3.00	2.90	3.00	2.72
	100	25	100	91	100	72
(200, 297)	3.00	2.58	3.01	2.99	3.00	2.95
	100	60	99	99	100	95
(200, 594)	3.00	2.88	3.00	3.00	3.00	3.00
	100	88	100	100	100	100

results.

The similarity between the original data clustering and the estimated one is measured by the Adjusted Rand Index in Hubert and Arabie (1985) using the Permutation Model,

$$ARI(C, C') = \frac{\sum_{i=1}^{S} \sum_{j=1}^{S'} \binom{\#(C_i \cap C'_j)}{2} - \sum_{i=1}^{S} \binom{\#(C_i)}{2} \sum_{j=1}^{S'} \binom{\#(C'_j)}{2} / \binom{n}{2}}{\left(\sum_{i=1}^{S} \binom{\#(C_i)}{2} + \sum_{j=1}^{S'} \binom{\#(C'_j)}{2} \right) / 2 - \sum_{i=1}^{S} \binom{\#(C_i)}{2} \sum_{j=1}^{S'} \binom{\#(C'_j)}{2} / \binom{n}{2}},$$

where C is the original partition with S groups, and C' the estimated one with S' groups, and it represents the probability that C and C' will agree on a randomly chosen pair. Table 2 reports the mean of the Adjusted Rand Index (ARI) for the 100 iterations. The closer the index is to 1 the better the agreement between the real partition and the estimated one.

Table 2: Clustering performance evaluation using the Adjusted Rand Index.

	DGP_1		DGP_2		DGP_3	
(T, N)	$\sigma^2 = 1$	$\sigma^2 = 2$	$\sigma^2 = 1$	$\sigma^2 = 2$	$\sigma^2 = 1$	$\sigma^2 = 2$
(100, 297)	0.8392	0.3195	0.8690	0.7511	0.8269	0.6399
(100, 594)	0.9248	0.3140	0.9172	0.8898	0.9372	0.7387
(200, 297)	0.9251	0.5508	0.9144	0.9027	0.9074	0.8658
(200, 594)	0.9792	0.8887	0.9738	0.9204	0.9393	0.9727

In general, AGP procedure shows pretty good allocation performance. Nevertheless, as we mention before for Table 1, AGP is very sensitive to the increase of variance under homoscedastic errors and time dimension $T = 100$.

4 Data

We analyze the Energy Price Index with fixed weights (FEPI) constructed in Sato et al. (2019). The data set includes 30 OECD and non-OECD countries and 12 sectors for the time period 1995-2015. Countries and sectors included in the sample are listed in Table 3. The cross-section dimension is $N = 30 \times 12 = 360$ and the time-series dimension is $T = 21$ years. Previous to the analysis the data are first differenced in order to achieve stationarity. Using subscripts $i = 1, ..., 30$ for the country, $s = 1, ..., 12$ for the sector and $t = 1, ..., 21$ for time, the $FEPI_{ist}$ is defined as

$$FEPI_{ist} = \sum_j w_{is}^j \cdot log(P_{it}^j) \tag{4}$$

where $w_{is}^j = \sum_j \frac{F_{is}^j}{\sum_j F_{is}^j}$ is the time invariant weight with F_{is}^j being the input quantity of fuel type j in tons of oil equivalent for sector s in country i, and P_{it}^j denotes the real price of fuel type j per toe of aggregate industry in country i at time t in constant 2010 USD. In our case, we choose the weights to be the average of the weights from 1995-2015. The FEPI captures only energy price changes that come from changes in fuel prices, and not through changes in the mix fuel inputs. The fuel types considered are electricity, natural gas, coal and oil.

From a general point of view, there are many factors that influence directly or indirectly energy prices, for example Dahl (2015) enumerates the following: population growth, demographic shifts and elongation of human life, income growth, environmental concerns, technology (investment capital available), renewable energies, waste storage and proliferation, government intervention, transportation/travel (moving freight, commuting, recreation and tourism, socializing, shopping, other services, industry travel), household heating, cooling, transport, and nuclear energy. We consider the effect of some of these factors over energy prices in our analysis. Having into account the data scarcity, we just include those control variables which are available for all the countries in the sample. Table 4 describes the control variables together with their transformations.

Figure 1 plots 12 time series, each one corresponding to the FEP of an industrial sector of Australia, the first country in the sample. We observe that Textile & leather, Transport equipment and Wood & wood products sectors are the ones with less variability, whereas Construction, Iron & stell and Mining & quarrying are the ones with largest variability. It is clear that FEP for

Table 3: Countries and sector coverage

Countries		**Sectors**
Australia	Japan	Chemical & petrochemical
Austria	Korea, Republic of	Construction
Belgium	Mexico	Food & tobacco
Brazil	Netherlands	Iron & steel
Canada	New Zealand	Machinery
Croatia	Norway	Mining & quarrying
Cyprus	Poland	Non-Ferrous metals
Czech Republic	Portugal	Non-metallic minerals
Denmark	Romania	Paper, pulp & print
Finland	Slovakia	Textile & leather
France	Sweden	Transport equipment
Germany	Switzerland	Wood & wood products
Greece	Turkey	
Hungary	United Kingdom	
Italy	United States of America	

Table 4: Control variables

	Transformation
Energy imports	1
Energy intensity level of primary energy	1
GDP per unit of energy use	1
Renewable energy consumption	1
Inflation rate	0
Population growth	0
Electricity production from renewable sources	1
Combustible renewables and waste	1

NOTES: Transformation = 1 denotes that the series is in growth rates.
Transformation = 0 denotes no transformation is needed.

Construction sector was the most affected by the financial crisis in 2008, followed by FEP in Mining and Quarrying sector, both of them have not yet recover since then. This analysis gives the intuition that it may exists a cluster structure in the data given by group-specific factors related to industrial sectors. Figure 2 represents the construction sector for each of the 30 countries in the sample. The most different country is Turkey, which FEP experiment large decreases around 1998 and 2004, and the one having positive values at the end of the sample period is Brazil. We observe different patterns and magnitudes between countries which could be interpreted as the existence of group-specific factors related to countries.

We implement the DFMCS to the FEP data set in order to analyze whether there exits a cluster structure describing FEP based on countries/industrial sectors, or just global factors. Empirical results are given in next section.

Figure 1 here.

Figure 2 here.

5 Empirical results

The Monte Carlo simulations clearly show that AGP procedure performs well when estimating the number of groups S and the allocations within groups. Previous to the analysis, the observed time series y_i, are cleaned from additive outliers and level shifts as proposed in Alonso et al. (2020).

First, we run the SCAD-penalized regression, in order to evaluate the effect of control variables in international FEP. The penalization parameter κ_{c_i} is optimized using a BIC criterion and the function `penalized`, in the MATLAB toolbox: 'penalized', see McIlhagga (2016). The estimation results show one relevant coefficient for each country. Given the close relation between prices and inflation, it is the control variable which affects the FEPs in a larger number of countries: Japan, Switzerland, USA, Australia, Austria, Belgium, Cyprus, Czech Republic and Denmark. Energy imports affect energy prices in Greece, Sweden and United Kingdom; Energy intensity level of primary energy is significant in Korea and Norway; GDP per unit of energy use in Romania; Renewable energy consumption in Italy, the Netherlands and Brazil; Population growth in Finland and Germany; Electricity production from renewable sources in France, Mexico, New Zealand and Slovakia; and Combustible renewables and waste in Croatia, Hungary, Poland and Turkey.

Once we estimate $\widehat{\beta}_{c_i}$ we subtract the observed component from the observed data, y_i, obtaining $y_i^* = y_i - X_i\widehat{\beta}_{c_i}$.

Second, we estimate the number of common factors in y_i^* using the CP eigenvalue ratio test based on correlation matrices. As suggested by Lam and Yao (2012) we applied the test twice for the possible existence of weak factors. The estimated number of global factors is $r_0 = 2$, see the CP ratio of eigenvalues in Figure 3. Each factor explains 72.7% and 7.3% of the total variability, respectively. The time series plots of CP estimated factors and its corresponding loadings are plotted in Figures 4 and 5, respectively. Loadings corresponding to the first global factor take negative values for all the series in the sample with the exception of Brazil FEP. This factor is able to capture the global dynamic of the differenced series, for example taking negative values during the 2000s energy crisis and the global recession in 2008. The loadings of the second factor takes positive and negative values across countries with differences in magnitude. The largest positive ones correspond to Slovakia, New Zealand and Poland FEP, and the largest negative ones to Greece and Cyprus.

Figure 3 here.

Figure 4 here.

Third, we apply the algorithm proposed in Alonso and Peña (2019), based on the Generalized Cross Correlation (GCC) measure of linear dependency between time series, to the common component. After applying a hierarchical clustering with average linkage to the dissimilarity matrix obtained from $\widehat{GCC}$ measure, we consider a modification of the Silhouette algorithm proposed by Rousseeuw (1987), which give us the number of groups $S = 6$ together with the allocation of each series to one of the groups, $G = g_1, ..., g_N$.

Results from the clustering are available in the Empirical Section in the Appendix. First of all, we observe that groups are classified according to countries instead to industrial sectors. For example, group 1 includes over all FEP of Cyprus, Mexico, USA, Switzerland, Greece, Romania and the Netherlands. Given that USA is the largest trading partner of Mexico it was expected that both of them were in the same cluster. We also observe that series of Construction from different countries are clustered in this group. Group 2 contains principally FEP of United Kingdom, Czech Republic, Poland, Austria, Germany, Slovakia and Denmark. This cluster is representative of countries with high industrial electricity prices according to the study of International industrial energy prices from the Department for Business, Energy & Industrial Strategy at the government of the UK (https://www.gov.uk). From this study we also conclude that group 3 represents mainly FEP of countries with high industrial gas prices, for example Finland, Japan, Korea, France, Italy, Sweden, and Portugal between others. Group 4 includes over all FEP from Hungary, Norway, Belgium and New Zealand. Finally, group 5 represents FEP from Austria and Brazil, and group 6 contains three series which may be consider atypical.

Figure 5 here.

Fourth, once we have the membership of each series to the corresponding group, we estimate the group-specific factors and their corresponding loadings using the time series in each group. The test estimates 3, 2, 3, and 10 factors in groups 1, 2, 3 and 4 respectively. Groups 5 and 6 do not have common factors. These specific-group factors may contain some of the global factors, for these reason all the 'global' and 'group-specific' factors are classified following the rules proposed by Alonso et al. (2020) based on the empirical canonical correlation between each global factor from step (2) and the group-specific factors from step (4). The first factor from the common component is classified as global, given that it is highly correlated with two or more groups, and the second factor is classified as specific because it is highly correlated with group 1.

Finally, we subtract the global common component from y_i^* and re-estimate the group-specific factors. The new estimates are 1, 2, 3, and 5 factors in groups 1, 2, 3 and 4, respectively. Group 4 is the group with largest number of factors, this is because each factor is representative of the FEPs in a specific country. For example, see upper plot in 6, the first group-specific factor represents the dynamics in Hungary, whereas the second group-specific factor, bottom plot, is representative of FPE in Norway. The rest of the bar plots representing the group-specific factor loadings are available the Appendix.

Figure 6 here.

In summary, our DFM with cluster structure has 5 groups, 4 of them with group-specific factors and one group only only affected by the global factor.

6 Concluding remarks

This article has three main contributions: first, we have presented an extension of the methodology proposed by Alonso et al. (2020) for DFMCS. The goodness of fit from this extension has been evaluated in a Monte Carlo experiment and has allowed us to evaluate the effect of macroeconomic variables which are country-specific over a large sample of international energy prices; second, the number of global and group-specific factors are estimated using the test proposed in Caro and Peña (2020); third, the factors and their corresponding loadings are estimated following the approach Caro Navarro and Peña (2018) based on correlation matrices.

Results from the application of international energy prices have provided useful interpretations about the existence of co-movements between energy prices related to group of countries instead of groups related to industrial sectors. Country connections within groups may be also explained by the high price of a specific fuel type. This analysis gives new insights for public policy decision making, to formulate and implement environmental policies, and for energy market investors to diversify their portfolios.

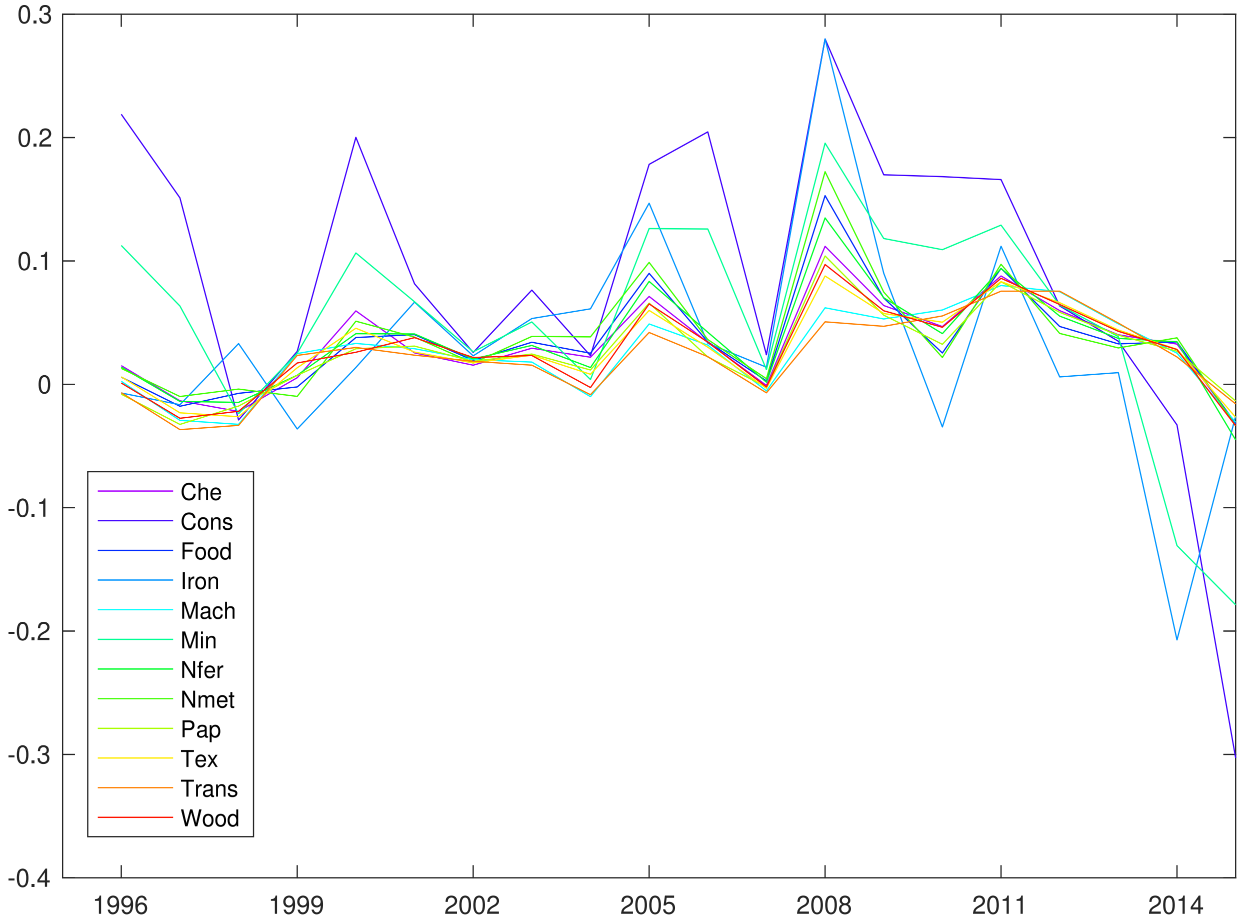

Figure 1: Fixed Energy Price for 12 industrial sectors in Australia from 1995 to 2015

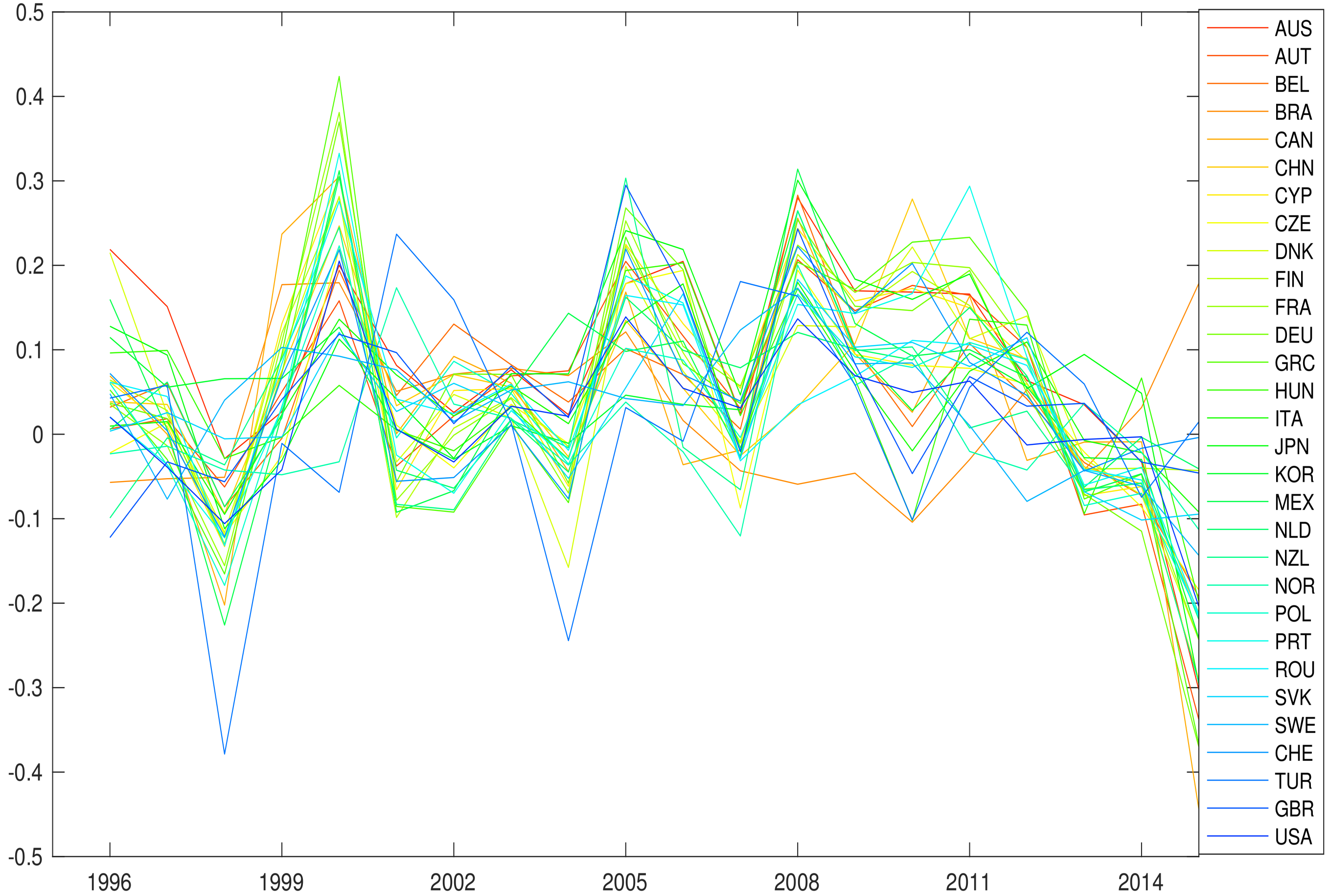

Figure 2: Construction sector Fixed Energy Prices for each country in the sample.

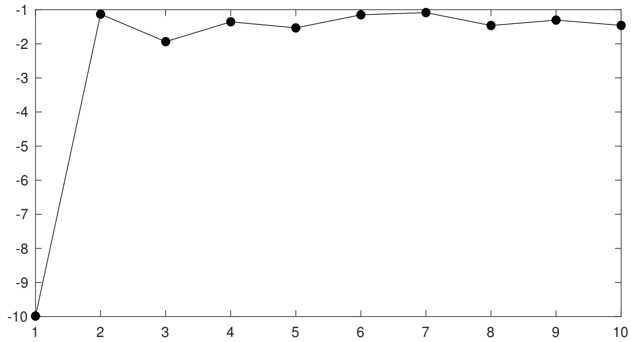

Figure 3: CP ratio of eigenvalues for the estimation of the initial factors.

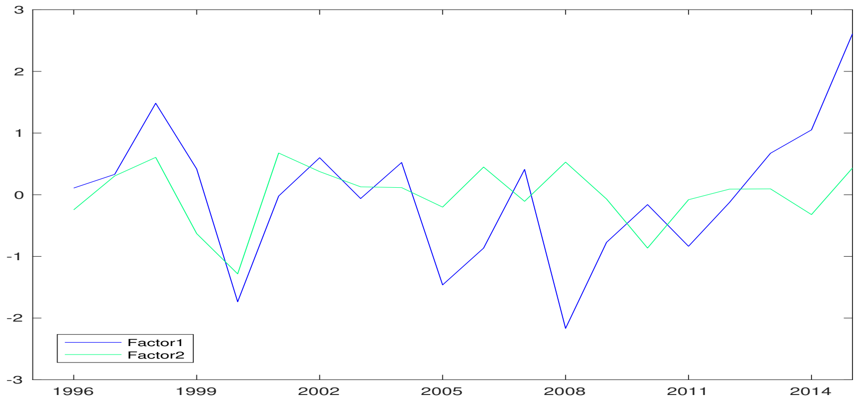

Figure 4: Estimated initial factors.

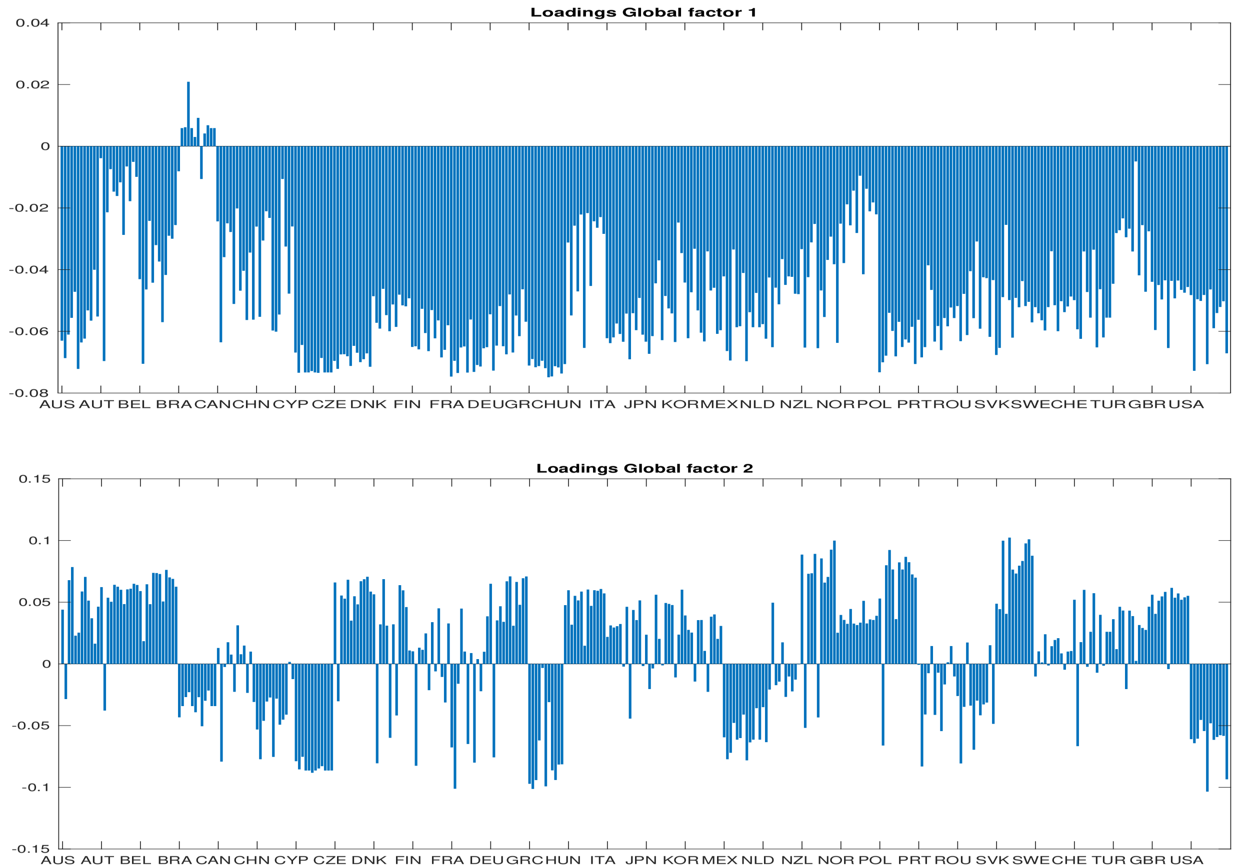

Figure 5: Estimated loadings of the two initial factors.

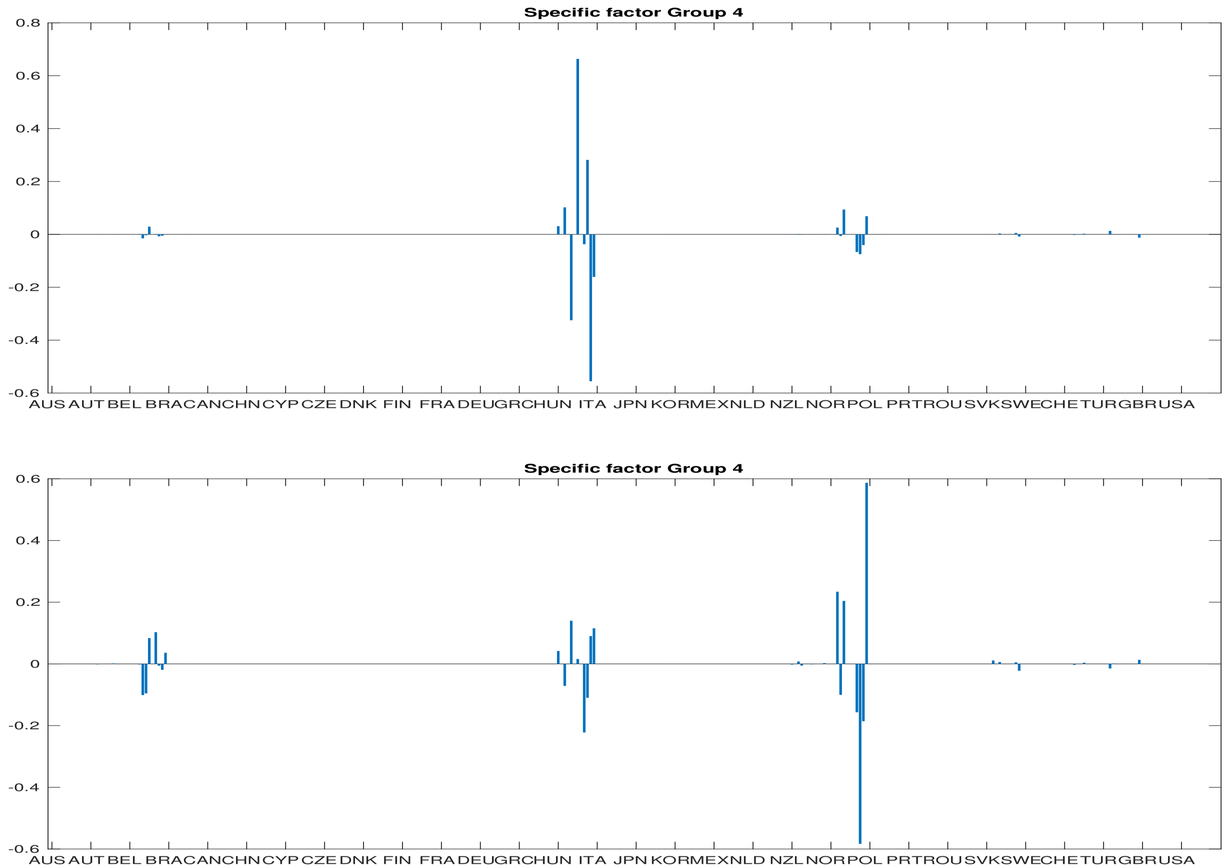

Figure 6: Estimated loadings of the first and second specific-factors in Cluster 4.

References

Aastveit, K. A., Bjørnland, H. C., and Thorsrud, L. A. (2015). What drives oil prices? Emerging versus developed economies. *Journal of Applied Econometrics*, 30(7):1013–1028.

Alonso, A. M., Galeano, P., and Peña, D. (2020). A robust procedure to build dynamic factor models with cluster structure. *Journal of Econometrics*, 216(1):35 – 52.

Alonso, A. M. and Peña, D. (2019). Clustering time series by linear dependency. *Statistics and Computing*, 29(4):655–676.

Ando, T. and Bai, J. (2017). Clustering huge number of financial time series: A panel data approach with high-dimensional predictors and factor structures. *Journal of the American Statistical Association*, 112(519):1182–1198.

Bretschger, L. (2015). Energy prices, growth, and the channels in between: Theory and evidence. *Resource and Energy Economics*, 39:29–52.

Brown, S. P., Yucel, M. K., et al. (2008). What drives natural gas prices? *Energy Journal*, 29(2):45.

Caro, A. and Peña, D. (2020). A test for the number of factors in dynamic factor models. *Submitted*.

Caro Navarro, Á. and Peña, D. (2018). Estimation of the common component in dynamic factor models. *Working Paper of the Statistics Department at Carlos III University of Madrid*.

Dahl, C. (2015). *International energy markets: understanding pricing, policies, & profits*. PennWell Books.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, 96(456):1348–1360.

Griffin, J. M. (1980). Henry b. steele. *Energy Economics and Policy (New York: Acadmic Press, 1980)*.

Hubert, L. and Arabie, P. (1985). Comparing partitions. *Journal of classification*, 2(1):193–218.

Lam, C. and Yao, Q. (2012). Factor modeling for high-dimensional time series: inference for the number of factors. *The Annals of Statistics*, 40(2):694–726.

Mahadevan, R. and Asafu-Adjaye, J. (2007). Energy consumption, economic growth and prices: A reassessment using panel vecm for developed and developing countries. *Energy policy*, 35(4):2481–2490.

McIlhagga, W. H. (2016). penalized: A matlab toolbox for fitting generalized linear models with penalties.

Nick, S. and Thoenes, S. (2014). What drives natural gas prices? A structural var approach. *Energy Economics*, 45:517–527.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65.

Sato, M., Singer, G., Dussaux, D., and Lovo, S. (2019). International and sectoral variation in industrial energy prices 1995–2015. *Energy Economics*, 78:235–258.

Van Benthem, A. and Romani, M. (2009). Fuelling growth: what drives energy demand in developing countries? *The Energy Journal*, pages 91–114.

Wang, P. (2008). Large dimensional factor models with a multi-level factor structure: identification, estimation and inference. *Unpublished manuscript, New York University.*

Appendix

Table 5: Fixed energy prices series included in Cluster 1.

Cluster 1
"CYP.Che" "CYP.Cons" "CYP.Food" "CYP.Iron" "CYP.Mach" "CYP.Min" "CYP.Nfer"
"CYP.Nmet" "CYP.Pap" "CYP.Tex" "CYP.Trans" "CYP.Wood"
"MEX.Che" "MEX.Cons" "MEX.Food" "MEX.Iron" "MEX.Mach" "MEX.Min" "MEX.Nfer"
"MEX.Nmet" "MEX.Pap" "MEX.Tex" "MEX.Trans" "MEX.Wood"
"USA.Che" "USA.Cons" "USA.Food" "USA.Iron" "USA.Mach" "USA.Min" "USA.Nfer"
"USA.Nmet" "USA.Pap" "USA.Tex" "USA.Trans" "USA.Wood"
"CHN.Che" "CHN.Cons" "CHN.Food" "CHN.Iron" "CHN.Mach" "CHN.Min" "CHN.Nfer"
"CHN.Nmet" "CHN.Tex" "CHN.Wood"
"GRC.Che" "GRC.Cons" "GRC.Food" "GRC.Iron" "GRC.Min" "GRC.Nfer" "GRC.Nmet"
"GRC.Pap" "GRC.Tex" "GRC.Trans"
"ROU.Che" "ROU.Cons" "ROU.Food" "ROU.Mach" "ROU.Min" "ROU.Nfer" "ROU.Nmet"
"ROU.Pap" "ROU.Tex" "ROU.Wood"
"NLD.Che" "NLD.Cons" "NLD.Food" "NLD.Mach" "NLD.Nmet" "NLD.Tex"
"FRA.Che" "FRA.Cons" "FRA.Min" "FRA.Nmet"
"CAN.Cons" "CAN.Min" "CAN.Tex" "CAN.Wood"
"PRT.Cons" "PRT.Food" "PRT.Min" "PRT.Nmet"
"DNK.Cons" "DNK.Min" "DNK.Nmet"
"FIN.Cons" "FIN.Trans"
"KOR.Iron" "KOR.Nmet"
"NZL.Cons" "NZL.Min"
"AUS.Cons"
"AUT.Cons"
"CZE.Cons"
"DEU.Cons"
"ITA.Nmet"
"POL.Cons"
"CHE.Cons"
"TUR.Mach"

Table 6: Fixed energy prices series included in Cluster 2.

Cluster 2
"UK.Che" "UK.Cons" "UK.Food" "UK.Iron" "UK.Mach" "UK.Nfer" "UK.Nmet" "UK.Pap"
"UK.Tex" "UK.Trans" "UK.Wood"
"CZE.Che" "CZE.Food" "CZE.Iron" "CZE.Mach" "CZE.Nfer" "CZE.Nmet" "CZE.Pap"
"CZE.Tex" "CZE.Trans" "CZE.Wood"
"POL.Che" "POL.Food" "POL.Mach" "POL.Nfer" "POL.Nmet" "POL.Pap" "POL.Tex"
"POL.Trans" "POL.Wood"
"AUS.Che" "AUS.Food" "AUS.Iron" "AUS.Nfer" "AUS.Nmet" "AUS.Pap" "AUS.Tex"
"AUS.Wood"
"DEU.Che" "DEU.Iron" "DEU.Min" "DEU.Nfer" "DEU.Pap" "DEU.Tex" "DEU.Trans"
"DEU.Wood"
"SVK.Che" "SVK.Cons" "SVK.Iron" "SVK.Min" "SVK.Nfer" "SVK.Nmet" "SVK.Wood"
"DNK.Che" "DNK.Iron" "DNK.Nfer" "DNK.Pap" "DNK.Tex" "DNK.Trans"
"JPN.Iron" "JPN.Nfer" "JPN.Nmet" "JPN.Pap" "JPN.Trans"
"TUR.Che" "TUR.Min" "TUR.Nfer" "TUR.Pap" "TUR.Tex"
"KOR.Che" "KOR.Mach" "KOR.Pap" "KOR.Tex"
"NOR.Che" "NOR.Cons" "NOR.Min" "NOR.Nmet"
"BEL.Che" "BEL.Food" "BEL.Nmet"
"ITA.Nfer" "ITA.Pap" "ITA.Trans"
"CAN.Iron" "CAN.Nfer"
"FIN.Nfer" "FIN.Pap"
"HUN.Iron" "HUN.Nmet"
"CHE.Che" "CHE.Pap"
"BRA.Iron"
"FRA.Iron"
"GRC.Wood"
"NLD.Iron"
"NZL.Nmet"

Table 7: Fixed energy prices series included in Cluster 3.

Cluster 3
"SWE.Che" "SWE.Cons" "SWE.Food" "SWE.Iron" "SWE.Mach" "SWE.Min" "SWE.Nfer"
"SWE.Nmet" "SWE.Pap" "SWE.Tex" "SWE.Trans" "SWE.Wood"
"FIN.Che" "FIN.Food" "FIN.Iron" "FIN.Mach" "FIN.Min" "FIN.Nmet" "FIN.Tex"
"FIN.Wood"
"ITA.Che" "ITA.Cons" "ITA.Food" "ITA.Iron" "ITA.Mach" "ITA.Min" "ITA.Tex"
"ITA.Wood"
"PRT.Che" "PRT.Iron" "PRT.Mach" "PRT.Nfer" "PRT.Pap" "PRT.Tex" "PRT.Trans"
"PRT.Wood"
"FRA.Food" "FRA.Mach" "FRA.Nfer" "FRA.Pap" "FRA.Tex" "FRA.Trans" "FRA.Wood"
"CHE.Food" "CHE.Mach" "CHE.Min" "CHE.Nmet" "CHE.Tex" "CHE.Trans" "CHE.Wood"
"CAN.Che" "CAN.Food" "CAN.Mach" "CAN.Nmet" "CAN.Pap" "CAN.Trans"
"KOR.Cons" "KOR.Food" "KOR.Min" "KOR.Nfer" "KOR.Trans" "KOR.Wood"
"JPN.Che" "JPN.Cons" "JPN.Food" "JPN.Mach" "JPN.Min" "JPN.Tex"
"NLD.Min" "NLD.Nfer" "NLD.Pap" "NLD.Trans" "NLD.Wood"
"AUS.Mach" "AUS.Min" "AUS.Trans"
"DNK.Food" "DNK.Mach" "DNK.Wood"
"DEU.Food" "DEU.Mach" "DEU.Nmet"
"TUR.Cons" "TUR.Nmet" "TUR.Trans"
"HUN.Cons" "HUN.Min"
"ROU.Iron" "ROU.Trans"
"BEL.Cons"
"CHN.Trans"
"CZE.Min"
"GRC.Mach"
"NZL.Wood"
"POL.Min"
"UK.Min"

Table 8: Fixed energy prices series included in Clusters 4, 5 and 6.

Cluster 4
"HUN.Che" "HUN.Food" "HUN.Mach" "HUN.Nfer" "HUN.Pap""HUN.Tex" "HUN.Trans"
"HUN.Wood"
"NOR.Food" "NOR.Iron" "NOR.Mach" "NOR.Pap" "NOR.Tex" "NOR.Trans" "NOR.Wood"
"BEL.Iron" "BEL.Mach" "BEL.Min" "BEL.Nfer" "BEL.Pap" "BEL.Tex"
"NZL.Che" "NZL.Food" "NZL.Iron" "NZL.Nfer" "NZL.Pap" "NZL.Trans"
"SVK.Food" "SVK.Mach" "SVK.Pap" "SVK.Tex" "SVK.Trans"
"TUR.Food" "TUR.Iron" "TUR.Wood"
"CHE.Iron" "CHE.Nfer"
"AUT.Food" "AUT.Nmet"
"BEL.Trans" "BEL.Wood"
"BRA.Nfer"
"JPN.Wood"
"POL.Iron"
Cluster 5
"AUT.Che" "AUT.Iron" "AUT.Mach" "AUT.Min" "AUT.Nfer" "AUT.Pap" "AUT.Tex"
"AUT.Trans" "AUT.Wood"
"BRA.Cons" "BRA.Food" "BRA.Mach" "BRA.Min" "BRA.Pap" "BRA.Tex" "BRA.Trans"
"BRA.Wood"
"NZL.Mach" "NZL.Tex"
"NOR.Nfer"
Cluster 6
"BRA.Che" "BRA.Nmet"
"CHN.Pap"

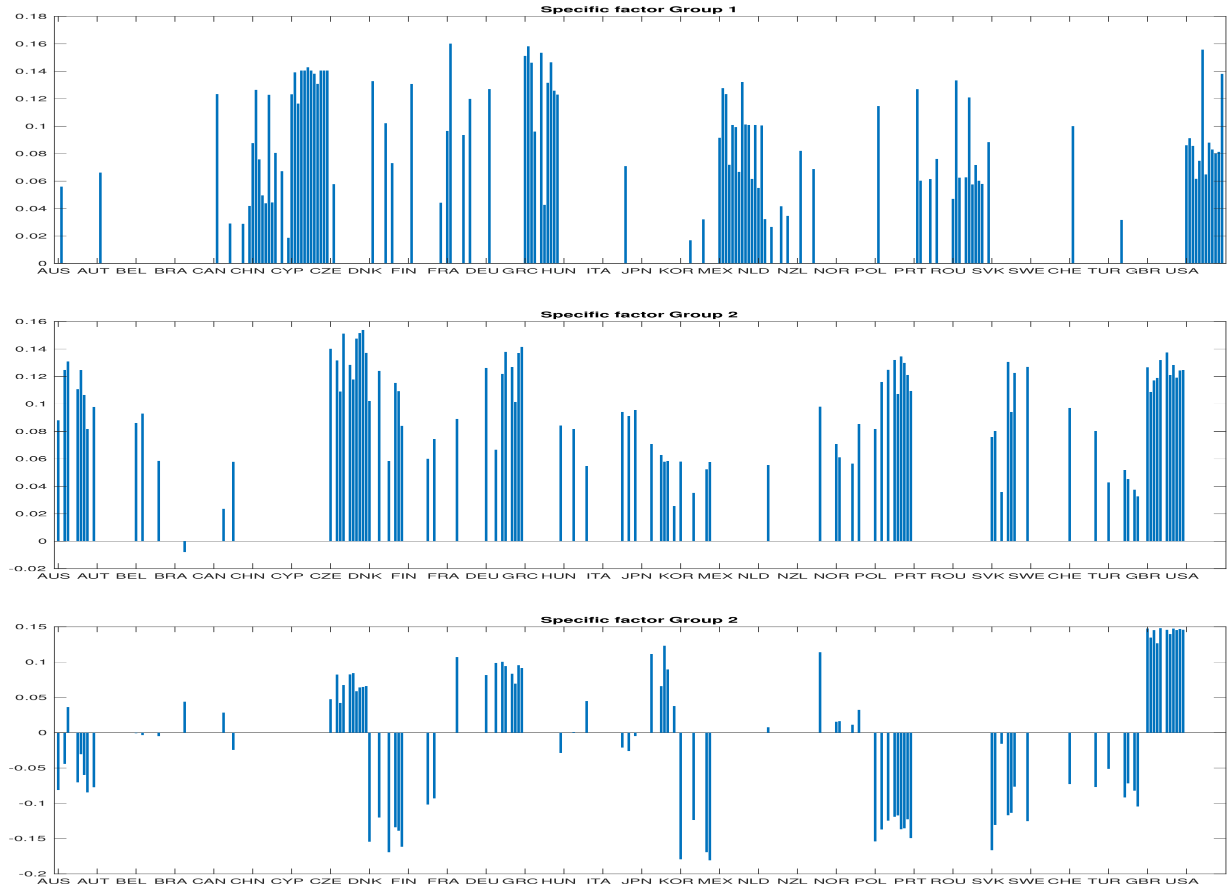

Figure 7: Estimated loadings of the specific-factor in Cluster 1, first row, and the first and second specific-factors in Cluster 2, second and third row, respectively.

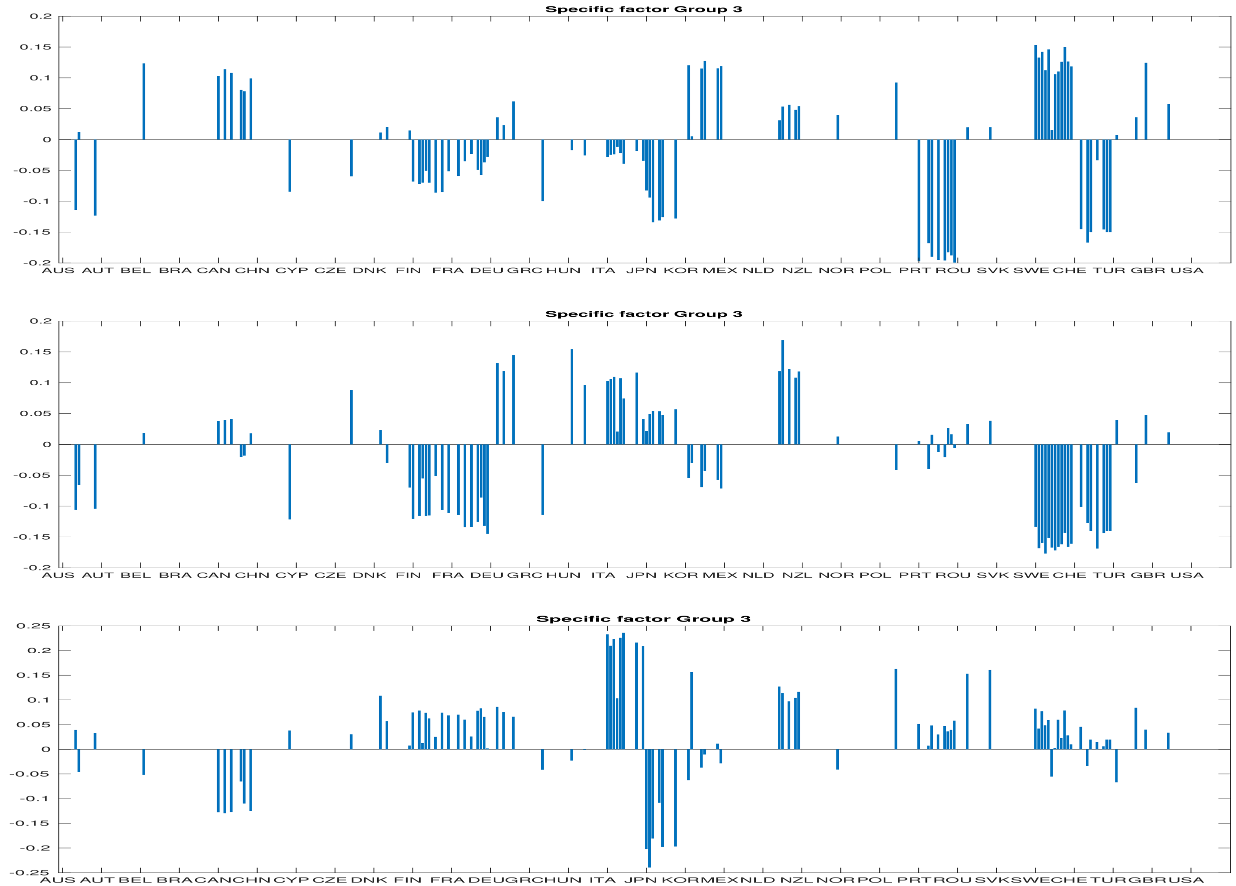

Figure 8: Estimated loadings of the first, second and third specific-factors in Cluster 3.

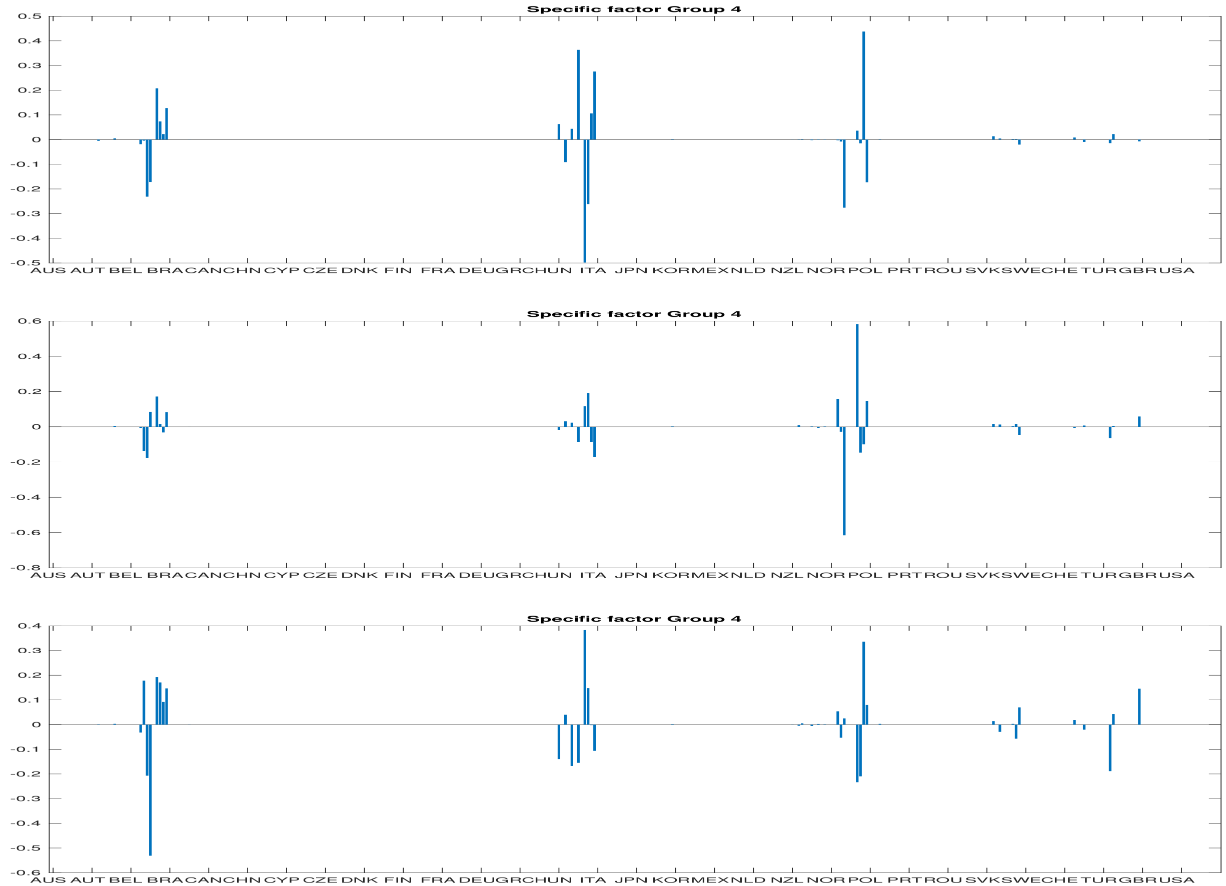

Figure 9: Estimated loadings of the first, second and third specific-factors in Cluster 4.