Within Country Sector Sensitivity

Marko Bogoevski, Viktor Domazetoski

June 2020

1 Introduction

Consider a world with $J \geq 1$ countries (economies): $country_1, ..., country_J$ and $S \geq 1$ sectors (industries): $sector_1, ..., sector_S$ as a network $G = (V, E)$ of $n = JS$ nodes in which each node represents a country-industry pair, where $V = 1, ..., n$ is the set of nodes and E is the set of edges to be defined shortly. Country-industry pairs $(\hat{i}, r)$ are mapped to the nodes in V with

$$(\hat{i}, r) \rightarrow (\hat{i} - 1)S + r \tag{1}$$

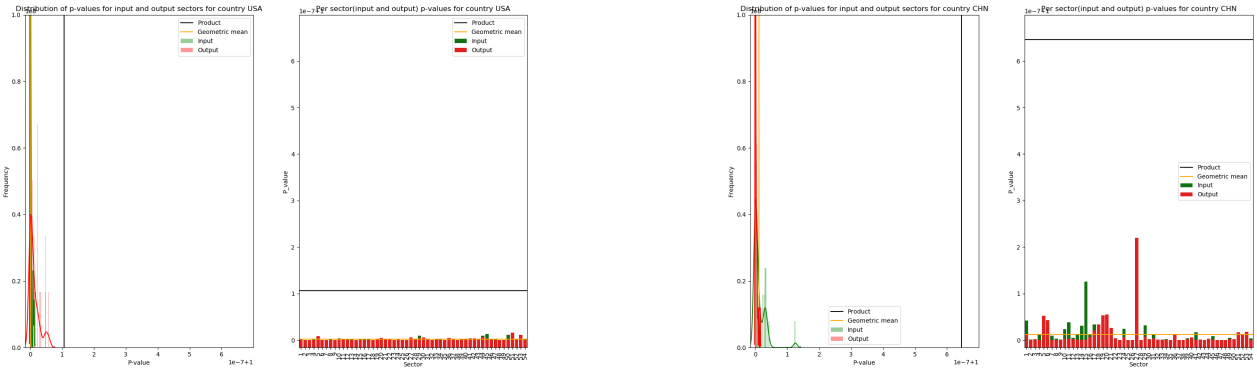
, for $\hat{i} = 1, ..., J$ and $r = 1, ..., S$. Note that the nodes $1, ..., S$ correspond to the $country_1$, the nodes $S + 1, ..., 2S$ are related to the $country_2$, and so on. The WIOD includes detailed data for J countries (indexed by $\hat{i}$ or $\hat{j}$) and S sectors (indexed by r or s) organized as two tensors: 4-order tensor $Z \in R^{J \times J \times S \times S}$ with entries $z^{rs}_{\hat{i}\hat{j}}$ describing the intermediate purchases (input flows) by industry s in country $\hat{j}$ from sector r in country $\hat{i}$; and 3-order tensor $F \in R^{J \times J \times S}$ with entries $f^{r}_{\hat{i}\hat{j}}$ denoting the final use in each country $\hat{j}$ of output originating from sector r in country $\hat{i}$.

Now consider a constriction on the mapping given by (1), that links between any two sets of different country nodes are prohibited, meaning node number i referring to country $\hat{k} \in 1, ..., J$ can only be linked to nodes in the interval $[(\hat{k}-1)S+1, \hat{k}S]$. Thus, a link between two nodes exist iff. their inverse mapping corresponds to the same country index $\hat{i}$. Informally, a link between two nodes exists if both nodes in G refer to the same country. This splits the previously fully connected network G in J fully connected components each of size S. The analysis here focuses on analyzing each separate fully connected network, namely extracting information of the form of within country sector sensitivity. The analysis utilizes within country data from the WIOD for years 2000-2014 and data for Macedonia for years 2010 and 2015. A subset from the tensor Z is used to produce the sensitivity values of each sector seen as input and output for each country with the subset entries being in the form $z^{rs}_{\hat{i}\hat{j}}$ where $\hat{i} = \hat{j}$ must hold true. Iterating for $r \in 1...S$, $s \in 1...S$ and $\hat{i} \in 1...J$ results in utilizing only diagonal squares of the WIOD data, where diagonal squares would be submatrices of Z of shape $S \times S$, each fulfilling the requirement that its main diagonal is a subarray of the main diagonal of Z and all main diagonals of the submatrices are mutually exclusive. In this way, each submatrix describes the intermediate purchases (input flows - alternatively output flows) by industry s from sector r in each country.

The Cobb Douglas model referenced in (**??**) has been run on each forementioned submatrix for all J countries alongside MKD for every mentioned year resulting in a total of 14*J+2 data points. Each data point represents a vector $v \in R^d$ where $d = 3 * S$. The first S values represent each sector's p-value if seen as input, the second S values represent each sector's p-value if seen as output and the last S values represent each sector's c-values. The assumptions of the Cobb Douglas model propose that p is a multiplier of the gross world product - GWP, where $p = 1$ depicts a stable sector (no effect), $p > 1$ depicts an increase and $p < 1$ depicts a decrease. An input sector with $p = 1$ would mean that the shock induced in that sector had a stable effect(none) on the overall GWP where that sector is only seen as an input sector in the intermediate trade, $p > 1$ would mean that the shock in that sector induced an increase in the GWP and $p < 1$ would mean that the shock induced a decrease in the GWP. The same analogous interpretation can be made for the output sectors. It is worth noting that the results from the data analysis are only a product of a simulated 1% shock increase in the links throughout the whole economy of a country. Namely the shock matrix is generated by analyzing the rate of change of A given a static rate of change in L of 0.01. In reality, this would not be the case. Also, the GWP in this sense stands for the GDP of the country since the analysis is based on within country sector communications. The c-values correspond to the level of affection each sector is viable to. These values sum to 100 and are percents which give a heuristic weight for each sector on how likely is that sector to be the most affected one if shock was induced throughout each sector of the country in question.The updated GWP equation in (??) can be easily narrowed to fit the within country analysis by constricting the indices $i = (1, r)$ and $j = (1, s)$ rather than having $i = (\hat{i}, r)$ and $j = (\hat{j}, s)$. Exploratory data analysis was performed on this dataset for the year 2010 and the results are represented below.

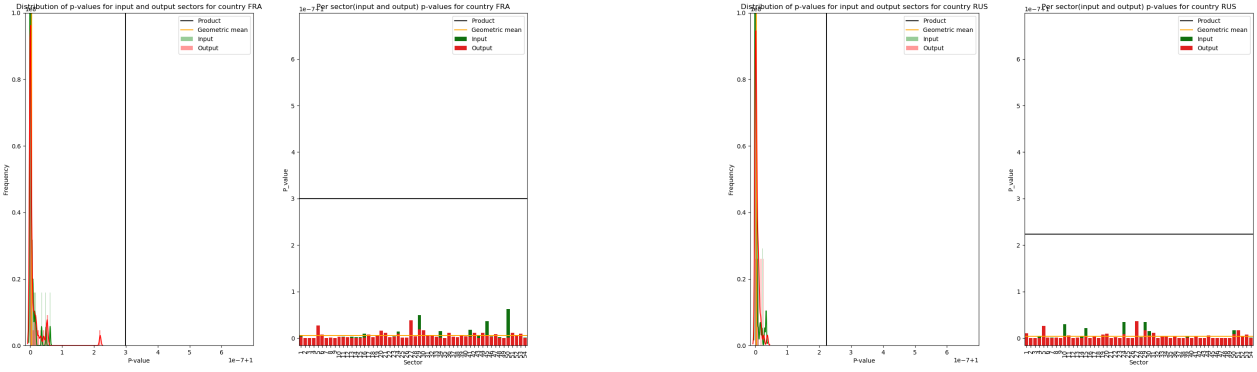
2 Analysis of p-values

Separate groups of countries are identified based on their current GDP values. These groups act as referent bodies when analyzing the p and c values for every country. USA, CHN, JPN, DEU and IND are labeled as high GDP representatives; FRA, CAN, RUS, ESP and TUR are labeled as medium GDP representatives; MK, ROW, CYP, EST, LTU are labeled as low GDP representatives. Also a merged group of representatives is formed consisting of the countries USA, CHN, FRA, RUS, MK and EST to compare MK results more easily to central representatives of each forementioned GDP group.



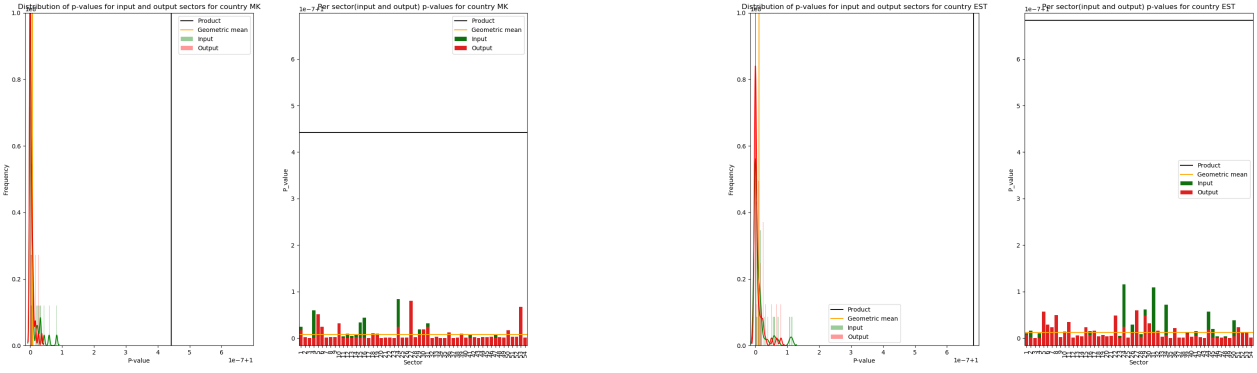
(a) USA

(b) CHN



(c) FRA

(d) RUS



(e) MK

(f) EST

Figure 1: P distribution for input and output sectors for central representatives

Figure 1 showcases the distribution of the p-values for the each of countries from the central representatives for each sector given on the left subplot and the values itself for each sector on the right subplot. The right skewed form of the distribution holds for most of the countries regardless of their position relative to their GDP value. We also see that most of the values are concentrated around 1 and differ in scales of $1e^{-}7$ to $1e^{-}8$. The minor scale of the variations could be possibly explained by the small(0.01) shock which is induced upon the input-net matrix A, but the form of the distribution should be invariant

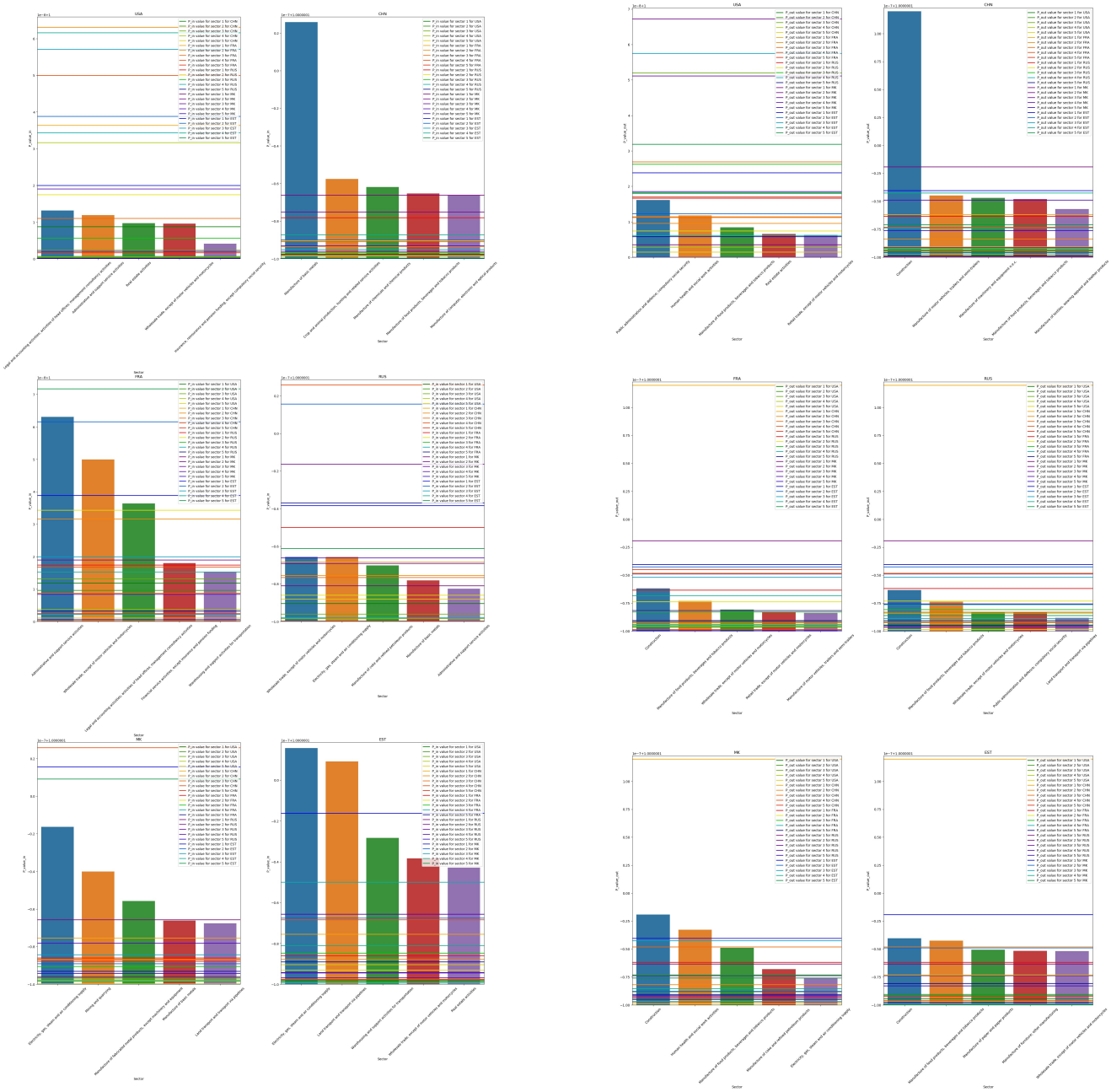
to the scale of the shock.The orange vertical line depicts the geometric mean p-value for each country. Other measures, such as median value were analyzed separately but it should be noted that the median value is not certain to represent a valid choice if the goal is to find correlation with the GDP values of the countries since sectors belonging in the top 50 percentiles could have very well large comparative p-values and be much more important towards contributing to the GDP rather than sectors in the bottom 50 percentiles. The geometric mean is invariant to this since it represents a weighted multiplier of each p-value of each sector.

First, to clarify, the meaning of a certain p_j value in a sector when seen as input in a country represents a geometric mean by itself of the p^{uv} values of the shock P matrix where $u = j$ and v is iterated from 1...S. Each of these separate values in turn represents the GDP multiplier of that certain link if positive shock was induced only in it. In this way, we estimate an overall GDP multiplier for that sector when seen as input. The p_j value for an output sector is analogous with the only change being that the geometric mean is now calculated from the p^{uv} values of the shock P matrix where $v = j$ and u is iterated from 1...S. This discussion enforces the idea that a geometric mean of all the p-values for input and output sectors would represent a valid positive correlated variable of the GDP multiplier of the country as a whole given the static 0.01 shock in each internal link. Another measure is also plotted as a black vertical line which represents the product of the p-values. These 2 values should have almost full positive correlation and the ordering of the countries based on them should be preserved. They both in a way represent the GDP multiplier when an equivalent shock is induced in every link of the country's economy, the difference being in that fact the the product is consistent with the assumptions of the model and depicts the effects of a non-normalized 1% shock. It can be also seen that only 1 value for the product and the geometric mean is plotted. This is because the values for input and output for both the product and geometric mean are the same. This should be further analyzed mathematically.

Results show that rescaling every geometric mean value and product value for every country and comparing the values implies that the current GDP value of a country doesn't necessarily have a positive correlation with them. EST, which is a relatively low GDP representative has bigger geo-mean value and product value in comparison to USA. EST also showcases one of the biggest geo-mean and product value closely followed by CHN, which would, according to the model, indicate that EST although being a relatively low GDP valued country, has a growth potential bigger than most of the other countries if the same percentual shock of %1 was induced in every link in every country.

Although each p_j value alongisde the geo-mean and the product value is regarded as a multiplier of the GDP, in reality the current GDP of the country dictates the relative static growth in comparison to other countries, so even though EST internally possibly has the biggest percentual growth, externally it is minor compared to other high GDP countries. Theoretically though, if conditions don't change, EST should be able to reach high GDP values in a distant future and precede other countries, but unless EST can boost its GDP momentarily it should be practically impossible since the ratio between the high GDP country values and EST GDP is much greater than the ratio of the p-values. Consequentially, let GDP_{1,t_1} be USA's GDP at time t_1 and GDP_{2,t_1} be EST's GDP at time t_1. Also, let p_1 be USA's product p-value at time t_1 and p_2 be EST's product p-value at time t_1. At time t_2, we can heuristically expand equation (22) of the Cobb Douglas model and derive the relationship between the product p-value for the country and the GDP value of that country with $GDP_{1,t_2} = GDP_{1,t_1} * p_1$, analogous for GDP_{2,t_2}. Let's also presume $GDP_{1,t_1}/GDP_{2,t_1} = x$ and x tends to a high number(690.75 at current times). In order for this ratio to converge to 1 or lower in a certain future, the ratio p_1/p_2 should converge to $1/x$. This would mean that EST would need to find a way to boost its internal GDP growth x(its p multiplier-product) times greater than that of USA's at a current time which seems impossible considering the current low p-values of both countries. MK shows similar high geo-mean and product p-values when compared to EST and is ranked 3rd highest from the six presented representatives.

Figure 2 shows the sectors with the strongest affective powers(seen both as input and output) of the central representative countries. These sectors are those which identify with the highest input/output p-values. Referring to the previous clarification of the p-values, we can define the affection of the sector in this case as the level of positive impact on the GDP of the country if shocks are only regarded in every link where they act as input(if the p-value for the sector is p_in) or every link where they act as an output sector(if the p-value is p_out). For each country, the p-values of the other representatives for the same top 5 sectors are also shown to provide referential view. We can see that the **Construction** sector is depicted as the most affectious sector for every country except USA when treated as an output sector, closely followed by **Manufacture of food products, beverages and tobacco products** and other types of manufacturing sectors. USA's most affectious output sectors on the other hand seem to represent more low-valued GDP contributive sectors such as the **Public administration and defence, compulsory social security** sector and **Human health and social work activities**, the latter also being the 2nd most affectious output sector in MK. This would imply that USA tends to have more stable high contributive GDP sectors when compared to lower GDP valued countries internally. When regarding input sectors, the **Electricity, gas, steam and air conditioning** supply takes the place analogous to its output parallel Construction in RUS,EST and MK and the **Legal and accounting activities, management consultancy activities** sector alongside **Administrative and support services** and **Real estate activities** for USA and FRA. We can see that the most affectious input sector for CHN is **Manufacture of basic metals** which makes a lot of sense since all its output sectors are heavily reliant on this one and CHN overall can be seen as a heavily focused manufacturing country. We can see that CHN dominates most of the manufacturing sectors since its p-values are way higher than any of the other central representatives. We can also see that FRA dominates the **Administrative and support service activities** sector over USA. This plot makes a lot of sense since it represents in a way the sectors which contribute towards most of the success(GDP) of the country and thus gives



(a) Input sectors

(b) Output sectors

Figure 2: Merge analysis of p-values for sectors for provided countries

an image of the country's distribution of goods and focus. Each subplot is an image of the strategy each of these countries employed in 2010 to distribute their labor.

Having seen the difficulty of interpreting the geo-mean p-values for countries at the current moment, the next analysis will focus on interpreting the distributions of the p-values for certain sectors throughout each of the representative countries. Figure 3 represents the distribution of the p-values for the top 6 most affectious sectors from the previous analysis for each of the central representative countries. We can see here that the sector Electricity, gas and steam distribution seen as both input and output is heavily dominated by the lower GDP countries(EST,MK) and slightly RUS. The Construction sector is highly affectious for CHN and for the lower GDP countries if seen as output, but not so much if seen as input. PA and defence is more affectious if seen as output sector and we can that the same applies in USA. On the other hand, the Legal and accounting activities sector is more affectious if regarded as input for both USA, EST and the most for FRA. A long reached conclusion from this would be that, low GDP contributing sectors such as Legal and accounting activities, tend to make an overall better magnification on the GDP of the country if used as input sectors, meaning these sectors should profit and contribute more positively if they are buying goods from other sectors during a positive shock effect, and this is seen in both low GDP countries such as EST as in high GDP countries such as USA. Electricity, gas, steam and air conditioning

supply seems to follow the same trend but for lower valued GDP countries, not so much for higher ones. Construction follows the trend of Eletricity, gas, steam and air conditioning supply for low valued GDP countries only if treated as an output sector, meaning it should profit and contribute more positively if the sector is selling goods to other sectors during a positive shock effect. PA and defence follows the trend of Legal and accounting activities, if used as an output sector.

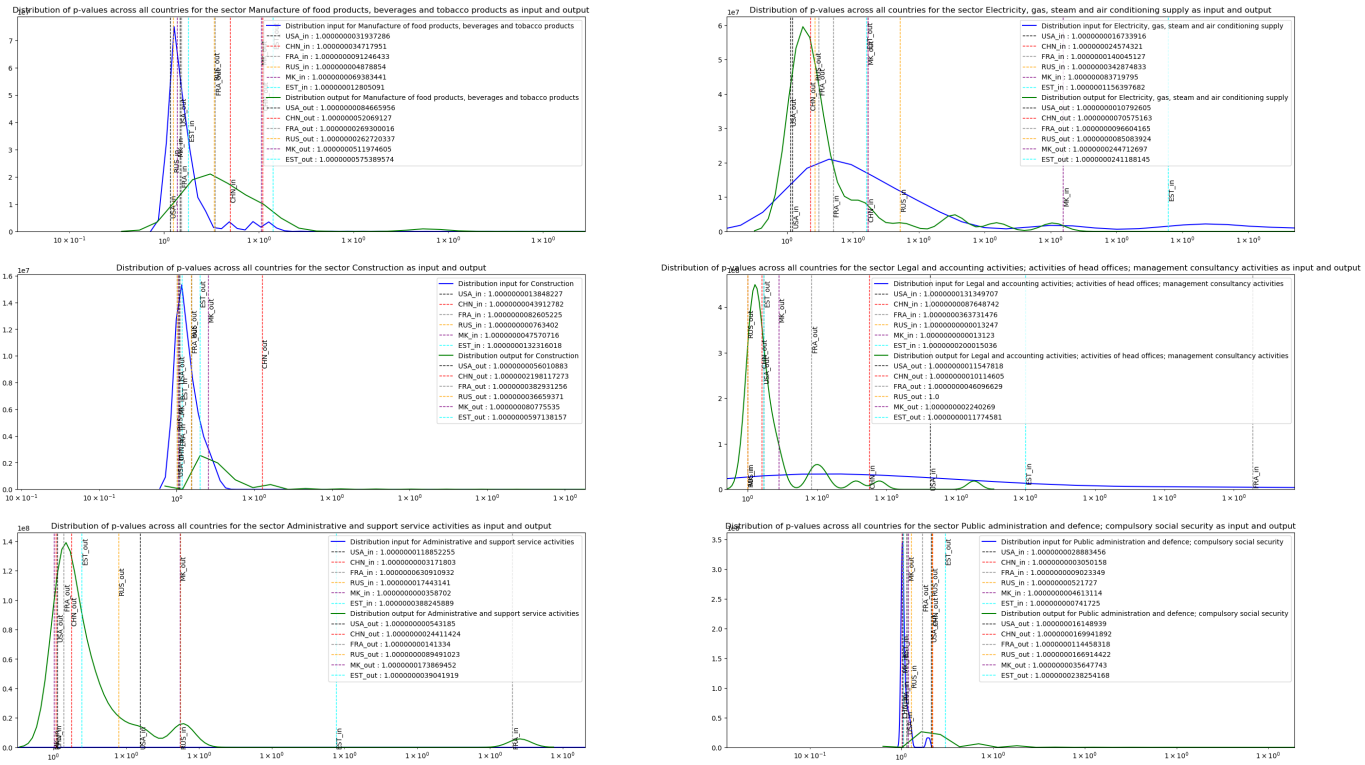


Figure 3: P distribution for the sector seen as input/output for only the central representative countries

Theoretically, if the Cobb-Douglas model shows consistent, one could scheme an economical distribution of goods throughout each of the sectors according to these p-values to maximize each sector's productivity if the shock is known. This would mean refocusing certain sectors to invest more as input sectors rather than output or vice versa to maximize the GDP growth. Of course this remark goes both ways since the p-values themselves are heavily dependent on the current distribution of shares in the representative household of each country for each sector(beta values from the Cobb-Douglas model), so the only thing preventing this is the choice of initialization of the current economy in the country.

3 Analysis of c-values

First, if we were to interpret the c_j value of the dataset we could roughly state that it represents a scalar which corresponds to how likely(and how much) that sector is affected by every other sector if every other sector is seen as input for it. It is important to note here that these values are invariant to the logarithm of the Hicks productivity shock and the shocks induced in the A matrix. In this sense, the c-values can be regarded as being closer to the 'sensitivity' of a sector and the p-values as being closer to the level of affection the sectors can induce. Figure 4 depicts the distribution of these values. The sectors **Construction**, **Real estate activities**, **Wholesale trade** and **Eletricity, gas, steam and air conditioning supply** make up most of the overall sensitivity. This means that these sectors turned out as being the most affected ones overall by every other input sector. It is worth noting that this affection doesn't correlate necessarily with positive p-values(¿1), but also negative ones(¡1). The values in the C matrix from which these values are derived are heavily reliant on the beta values of each sector. This makes values in different rows of the C matrix non comparable since they are normalized on different scales. This is the main reason why we are interested only in the orderings of the sectors and their frequencies in being in the top 3 affected sectors by a certain input sector. From this plot we can conclude that the Construction sector is easily affected by a change in the economy of the input sectors it sells to, but also makes a big impact on them since it acts as an output sector to them and its p_{out} value is high. The same can be inferred for Real estate activities and Wholesale trade, except of motor vehicles and motorcycles. Sectors near the end of the barplot theoretically are sectors which don't necessarily depend(or are affected by) most of the other sectors they buy from comparatively to the other sectors in the higher spots. Next, we will see the distribution of the c-values for each of the central representatives. This will show us which are the most sensitive sectors and by what scale in each of these countries.

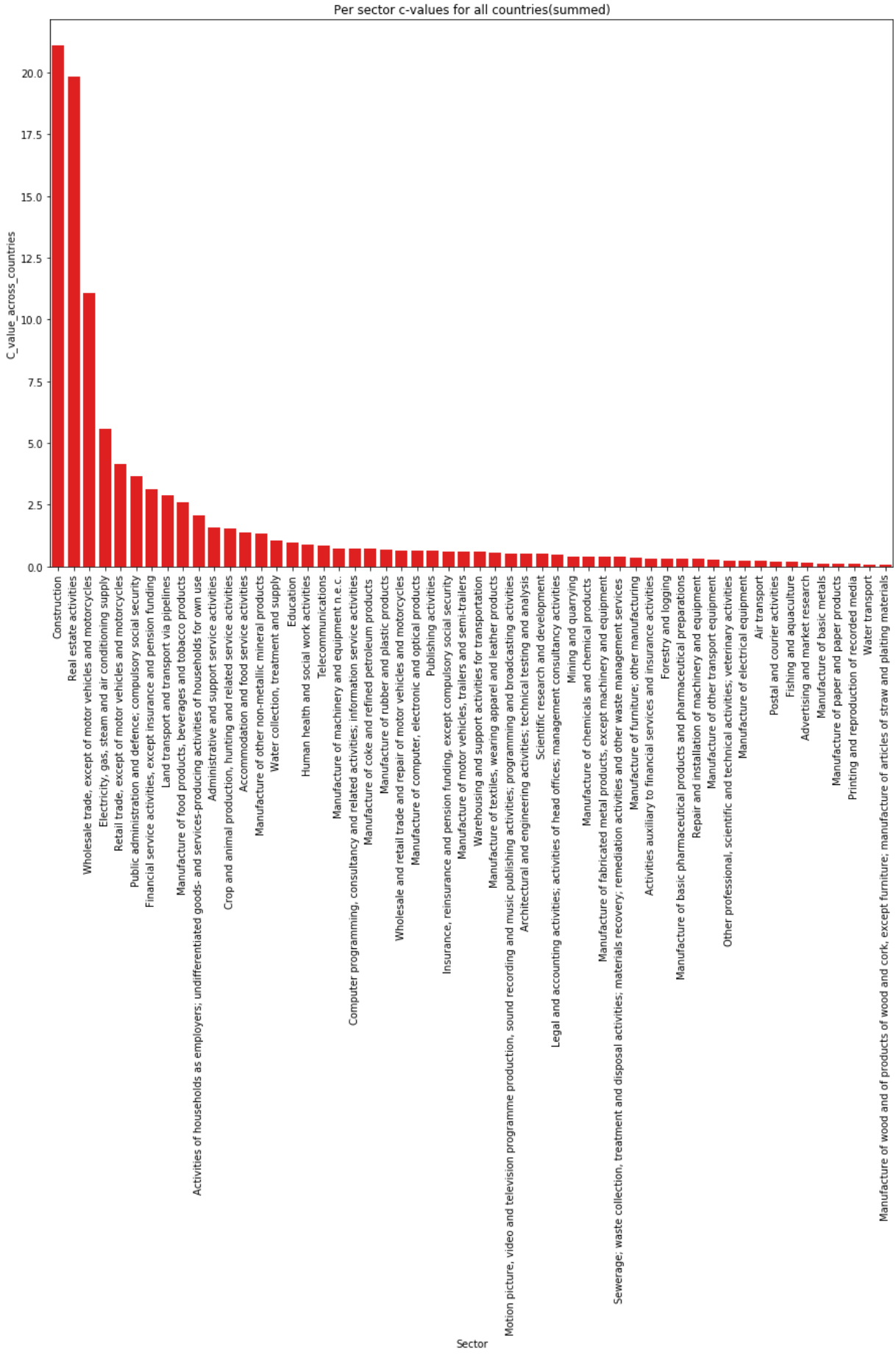
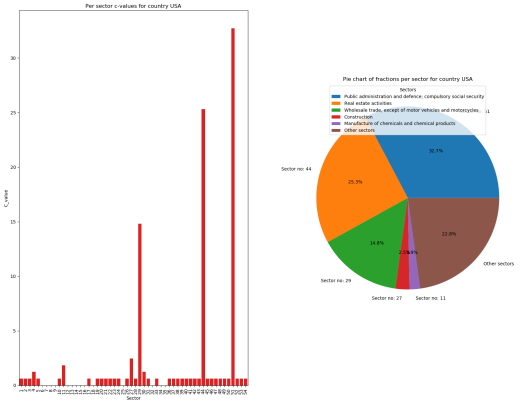
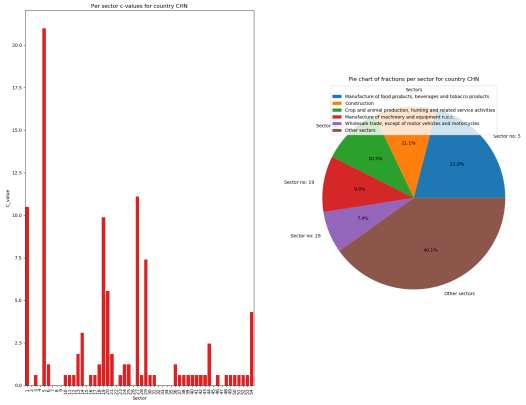


Figure 4: Summed C distribution for the sectors for each country

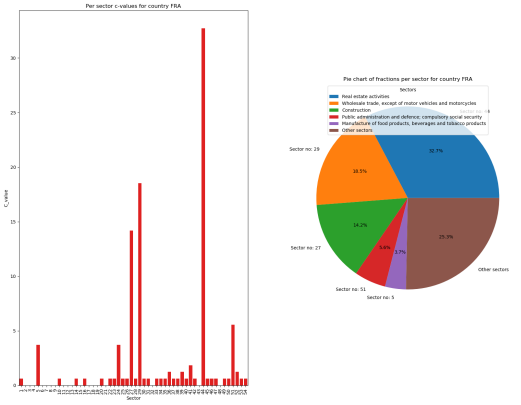
Figure 5 represents the results. The top 5 most sensitive sectors of each of the countries are shown with their corresponding c values. All other c-values are summed into the 'Other sectors' pie fraction. Besides the pie plot there is also a histogram in order to better visualize the distribution of the c values in the Other sectors compartment. USA's most sensitive sector is the Public administration and defence sector which also has the biggest p_{out} value for USA.



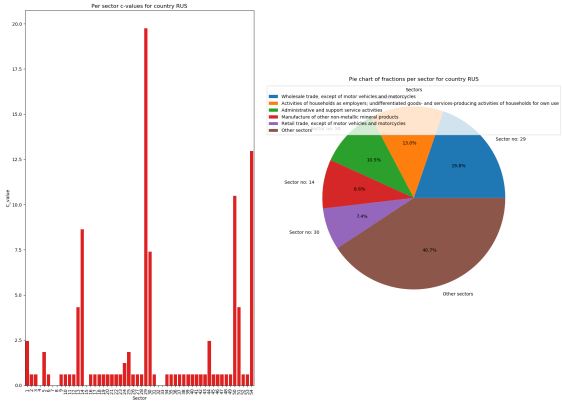
(a) USA



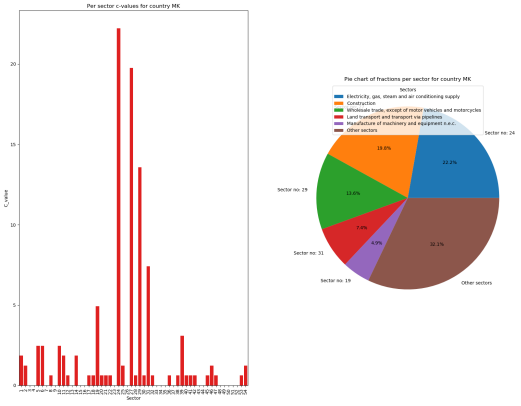
(b) CHN



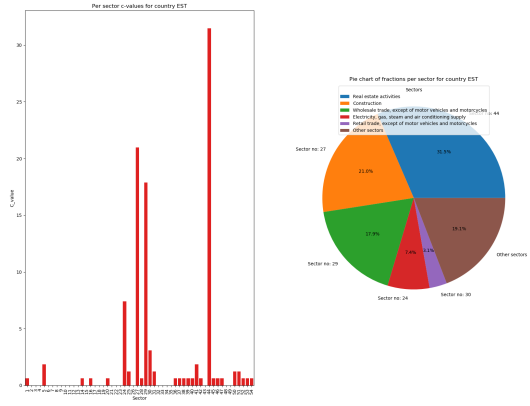
(c) FRA



(d) RUS



(e) MK



(f) EST

Figure 5: C distribution for sectors for central representatives

This again, means that this sector is easily affected by a change in the economy of the input sectors it sells to but also affects them the most out of all the other sectors when they buy from it. Real estate activities and the Construction sectors

are present in every country's top 5 most sensitive sectors as expected since they both topped places in the summed c-values distribution. It is also interesting to notice that most of the countries have a uniform sensitivity distribution but this could be explained by the fact that the c-values are proportional to the number of times each of the sectors was included in the top 3 affected sectors by each input sector. This nearly uniform distribution of the sectors which aren't top sectors proposes the idea that a sector is either highly sensitive to a lot of input sectors or its sensitivity to any one of them is low. MKD's most sensitive sector is the Electricity, gas, steam and air conditioning supply which also proved to be the one with the highest p_{in} value for MKD. We can see here the most sensitive sectors in each country are usually those which correlate with high p_{in} or p_{out} values. This is normal to expect since the c-values from the C matrix are in a way specific normalization of these values.

The next analysis will focus on clustering the countries based on the distribution of their c-values. The goal is to identify groups of similar countries where similarity is measured by the similarity of the distributions of the sensitivities of the sectors in each country. This will allow to label certain groups as being more susceptible to certain changes since certain targeted sectors are affected much more and also see where MKD stands in comparison.

4 Clustering based on c, within cluster analysis

After running k-means++ and GMM with a tied covariance matrix on the c-values and analyzing the corresponding BCV(between cluster variance) vs WCV(within cluster variance) ideal elbow position on the first algorithm and the BIC score from the second, the potential number of clusters ranged in the interval from 4 to 10. Any value from 4-7 was empirically expected as a number sufficiently rational to analyze given the current problem so six was chosen as the number of clusters. The clustering using both algorithms had a slight overlap and the most intuitive groups (based on GDP metrics and known wealth distributions) were chosen as follows:

- Cluster 0: CAN, DEU, FRA, GRC, HUN, IRL, KOR, NOR, USA
- Cluster 1: LTU, MLT, ROU, SVK, ROW, MK
- Cluster 2: BRA, CHE, CHN, IDN, JPN, MEX, RUS, TWN
- Cluster 3: AUS, AUT, BEL, CZE, DNK, ESP, EST, GBR, ITA, NLD, POL, SWE
- Cluster 4: BGR, CYP, FIN, HRV, LUX, LVA, PRT, SVN
- Cluster 5: IND, TUR

What's interesting to notice is that cluster 0 contains 2 high and 2 medium GDP representatives, cluster 2 contains 2 high and 1 medium GDP representatives, cluster 5 contains 1 high and 1 medium representatives, cluster 3 contains 1 medium and 1 low GDP representatives, cluster 4 contains 1 low GDP representative and cluster 1 contains 3 low GDP representatives from the aforementioned three representative groups. We can also see that most of the generalization holds throughout the whole cluster for every cluster, so the clustering in a way hierarchicaly organized the countries as a weighted function of their GDP values and c-distributions introducing potential correlation between the two.

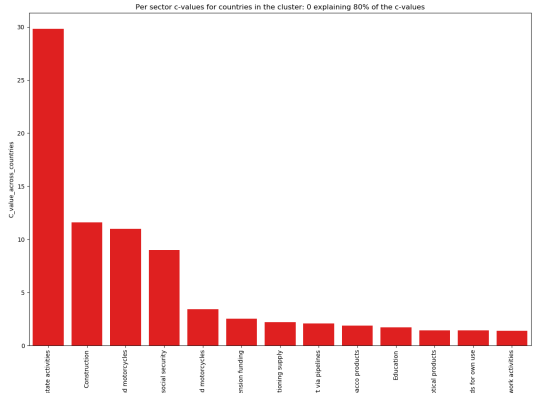
Within cluster c-analysis was performed to further investigate this correlation and see each group's preferential c-distribution and most sensitive sectors.

Figure 6 shows each cluster's most sensitive sectors which sensitivity cumulates to 80%. The results are depicted by stating the distribution of the c-values for the top affected sectors and the contributions towards that sensitivity of each of the countries in the corresponding GDP groups in each cluster.

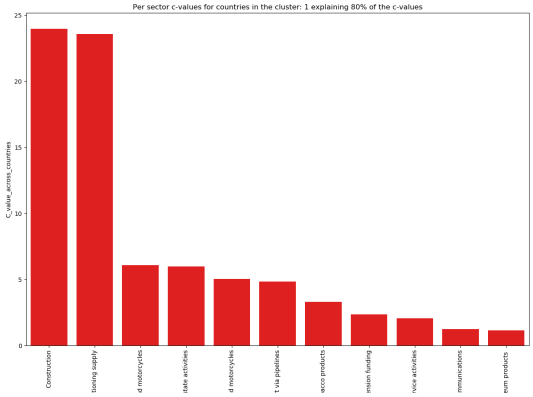
In cluster 0: CAN, DEU, FRA, GRC, HUN, IRL, KOR, NOR, USA we find the most sensitive sectors with their according ranking percentages as: Real estate(30%), Construction(11.5%), wholesale trade(11%) and Public administration and defence(9%). After these ones come Public administration and defence, Retail trade, Financial service activities and Electricity... Since most of the sensitivity ranking is taken up by Real estate, we can state that this is the prevalent sensitive sector in this cluster. Most of its sensitivity comes from sensitivity regarded in DEU(12.184%), FRA(12.184%), CAN(9.885%) and USA(9.425%).

Cluster 2's most sensitive sectors consisting of: BRA, CHE, CHN, IDN, JPN, MEX, RUS, TWN are: Wholesale trade(17.9%), Construction(12.2%), Retail trade(9.028%) and Food products, beverages and tobacco(6.1%), followed by Administrative and support service activities, Activities of households, Financial services and Real estate. We can see couple of similiraties with cluster 0's distribution but Real estate here is not regarded as a very sensitive sector. It is rather replaced by Wholesale trade and Construction. In Wholesale trade, contributors towards the sensitivity are JPN(14.22%), RUS(13.793%) and CHN(5.172%). On the other hand in Construction, contributors are JPN(19.62%), CHN(11.4%) and RUS(0.633%). Retail trade is characterized by its top contributors which are RUS(10.256%), JPN(3.42%) and CHN(0.855%). The same analysis is done on the remaining clusters resulting in the following distributions.

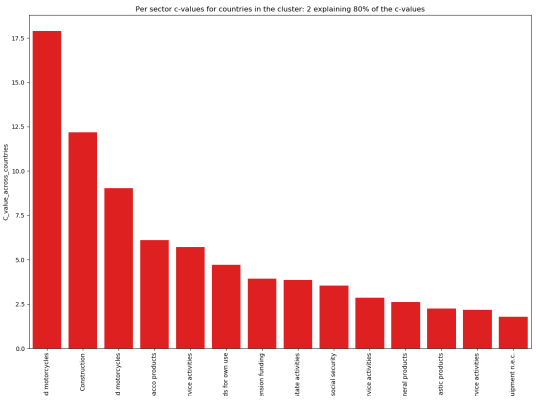
Cluster 5: IND, TUR. We find dominant sensitive sectors: Land transport(24.7%), Real estate(11.42%), Construction(9.877%) After this come Crop and animal production, Activities of households, Retail trade and Non-metalic minerals.



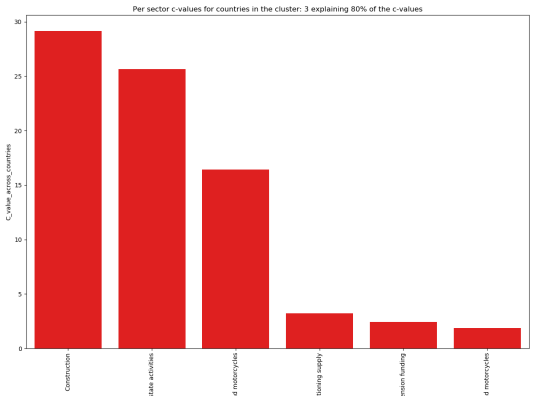
(a) Cluster 0



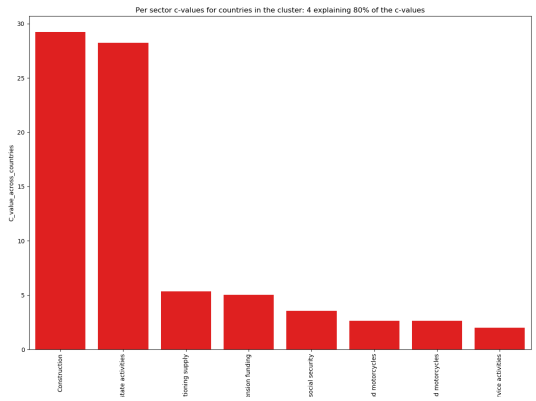
(b) Cluster 1



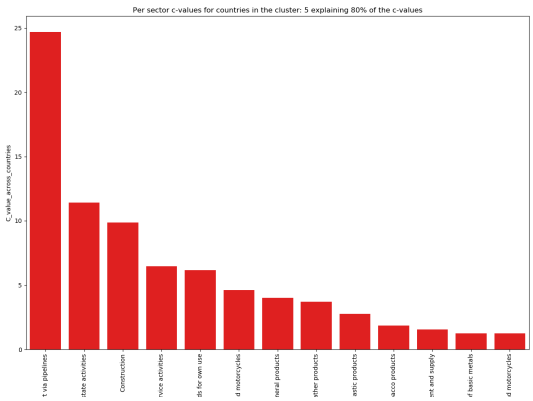
(c) Cluster 2



(d) Cluster 3



(e) Cluster 4



(f) Cluster 5

Figure 6: C distribution for sectors for central representatives

In Land transport, contributors are TUR(55%) and IND(45%) In Real estate, contributors are TUR(94.5%) and IND(5.5%). In Construction, contributors are IND(87.5%) and TUR(12.5%).

Cluster 3: AUS, AUT, BEL, CZE, DNK, ESP, EST, GBR, ITA, NLD, POL, SWE. We find dominant sectors: Construction(29.167%), Real estate(25.67%), Wholesale trade(16.41%), followed by Electricity, gas, Financial service activities and Retail trade. We can see that this cluster is similar to cluster 0 and thus slightly with cluster 1 with the exception being that most of the sensitivity here comes from Construction followed by Real estate rather than the inverse order. In Construction, contributors are ESP(9%) and EST(6%). In Real estate, contributors are EST(10.22%) and ESP(6.212%). In Wholesale trade, contributors are EST(9.1%) and ESP(8.464%).

Cluster 4: BGR, CYP, FIN, HRV, LUX, LVA, PRT, SVN. We find dominant sectors: Construction(29.244%), Real estate(28.241%) After these follow Electricity, Financial service activities, Public administration and Wholesale trade. This cluster finds many similarities regarding the most sensitive sectors with cluster 3, but what differs in it are its secondary most

sensitive sectors which seem to be sectors regarding Public administration and Wholesale trade rather than Retail trade. In Construction, contributors are CYP(11.61%). In Real estate, contributors are CYP(12.842%).

Cluster 1: LTU, MLT, ROU, SVK, ROW, MK. We find dominant sectors: Construction(24%), Electricity,gas,steam(23.56%), followed by Wholesale trade, Real estate, Retail trade and Land transport. In Construction, contributors(from representatives) are ROW(19.3%), MK(13.3%) and LTU(6.438%). In Electricity,gas and steam contributors(from representatives) are ROW(21%), LTU(17.467%) and MK(15.721%). Construction yet again, as in the previous hierarchical 3 GDP clusters is regarded as being the most sensitive. Electricity, gas and steam though is an interesting, almost identically sensitive sector here defining the cluster characteristic.

From this analysis we can conclude that the clusters regarding the higher GDP countries tend to characterize themselves with having Real estate, Wholesale trade and Land transport as the most sensitive sectors while the clusters representing the lower GDP countries have more sensitive Construction, Electricity, gas and steam sectors followed then by Real estate and Wholesale trade.

References

.