Master Thesis: Study on the Differences between Central Bank and Non Central Bank Affiliated Research in Banking Competition Publications

Wiktor Owczarz
(I6052598)

August 2018

Supervised by: J.W.B. Bos & R.J Almeida

Master in Economics and Financial Research
School of Business and Economics,
Maastricht University

Abstract

This work presents a novel examination of banking competition articles, analysing if authors affiliated with central banks could be more prone to report their findings with a notion of uncertainty, which results from the possible impact such research findings could have on the market. A set of methods that can be used to analyse articles is introduced and applied to a sample of 69 Panzar-Rosse banking competition studies, including 15 papers whose authors are associated with central bank institution. The results do not support the hypothesis that the central bank researchers are more prone to vagueness while reporting their results. Still, the evidence suggests that central bank research reports significant results more frequently.

Keywords: Panzar-Rosse, Banking competition, Natural Language Processing

---

1. Introduction

The role of the independent central bank in modern economy is frequently discussed in scientific literature (Bernanke (2004), Muchlinski (2011)). The recent findings especially underline the complexity of the interaction between central banks and financial markets and the crucial role this relation plays in shaping the expectations about monetary policies and interest rates. The empirical findings published by Blinder et al. (2008) and Issing (2005) confirm that the communication between central banks and financial markets can influence forming of the expectations by agents participating in the market. These findings constitute evidence that the language interactions between central banks and the financial markets is an indispensable part of creating effective monetary policy (Muchlinski, 2011). What is more, shaping monetary policy is not the sole responsibility of the central bank. This institution also has authority over financial stability and bank supervision (Bhattacharya and Boot, 1998). Consequently, next to monetary policy, financial stability and bank supervision might be another area where policy is strongly related to the language it is communicated with.

As pointed out by Greenspan (2003), what adds even more complexity to the problem is the fact that the central bank's communicates cannot be considered as pure disclosure of information. The communication between central bank and financial markets have traces of uncertainty and ambiguity. The author argues that this uncertainty of the monetary policy landscape is not just one of its characteristics, but it is rather what defines that monetary landscape.

Analysis of the uncertainty involved in communication can lead to better understanding of the direction of the monetary policy, that central bank might want to pursue, what in turn results in forming of more precise monetary policy expectations. Due to this potential advantage, this subject is the focal point of many researchers' work. Therefore, the relation between central banks and the way they influence monetary policies through the communicated information is already thoroughly studied (Boukus and Rosenberg (2006), Ehrmann (2005), Kahveci and Odaba (2016), Masciandaro and Romelli (2016), Stajner et al. (2017)). However, the influence this communication and the involved uncertainty has on banking environment remains understudied, hence it will be the subject of this work.

This paper focuses on a set of banking competition papers, written by researchers associated with academia or central bank, and investigates if there are any characteristics signalling that central banks' researchers are more prone to being vague and cautious about reported results. It is hypothesised that the researches working at central bank, while conducting a scientific study, might behave in this way as they are influenced by reputation of their job and possible impact their findings might have on the market. This cautiousness might not only be observed on the linguistic level but also be reflected in the empirical studies. The reason behind the carefulness in communicating the results is the high cost in case of making an inaccurate or false statement about situation on a banking market. As a consequence, when conducting empirical studies, central bank researchers might be more inclined to make Type II Error, maintaining status quo (overcautiously restraining from any signals of market power abuse), since the cost of making Type I Error (false signalling market power abuse) is high. On the contrary, it is hypothesised that the group of university researches is under different level of public scrutiny and is therefore less conservative while publishing their results.

The purpose of this paper is to introduce a set of tools that allows to analyse and measure differences between banking competition studies. The aim is to quantify if there are significant, measurable dissimilarities between banking competition papers published by the two groups of researchers.

The remainder of the paper is structured as follows. First, concepts of Panzar-Rosse H-statistic and hedging are introduced. Second, data collection process is described. Third, methods to compare reported test statistics, measure uncertainty within a text and similarity between texts are discussed. Each section that describes a new method is followed by short empirical application. Finally, the papers is concluded with summary of the results.

2. Related concepts

2.1. Panzar-Rosse Model

This paper focuses on a single competition model, the Panzar-Rosse (P-R) model (Panzar and Rosse, 1987). Despite the fact that P-R H-statistic was claimed to be an unreliable measure of the market power (Bikker et al., 2012), it is still widely used in empirical literature (see article list in appendix A). The reason for that is model's input data, which is easily accessible and reliable. Aim of this study is to present data analysis methods, which will be based on P-R model, but could also be applied to analyse data from other competition models. Other common ways to evaluate market power, which could be used instead of P-R model and which are not discussed in this study, are shown in Table 1 (Bikker and Spierdijk, 2017).

Measure	**Definition**	**Reference**
K-firm concentration ratio	Sum of top k market shares	Savings (1970)
Herfindal-Hirschman index	Sum of squared market shares	Cowling and Waterson (1976)
Lerner Index	1 - (marginal cost/price)	Lerner (1934)
Rothschild conduct index	Related to the product of market demand elasticity and Learner index	Rothschild (1942)
Boone index	Relative profit differences	Boone (2008)

Table 1: Empirical measures of market power

The Panzar-Rosse competition model used in this work does not only require data that is relatively easy to gather, but is also straight-forward to estimate. Panzar-Rosse revenue test relates gross revenue to input prices and other firm specific control variables. Assuming single-output production function with n-input prices the P-R empirical equation is the following:

$$logTR = \alpha + \sum_{i=1}^{n} \beta_i logw_i + \sum_{j=1}^{J} \gamma_j logCF_j + \varepsilon \tag{1}$$

where TR is total revenue, w_i denotes the price of the i-th input factor and CF_j the firm specific control. The H-statistic is defined as sum of input elasticities:

$$H = \sum_{i=1}^{n} \beta_i \tag{2}$$

H-statistics can take different value ranges, which are reported in Table 2.

Values	**Competitive environment**
H $\leq$ 0	Monopoly equilibrium: banks operate independently as under monopoly profit maximisation conditions
0 $<$ H $<$ 1	Monopolistic competition free entry equilibrium
H = 1	Perfect competition, full efficient capacity utilisation

Table 2: Source: Bikker (2004)

2.2. Hedging

Since the notion of uncertainty is essential for this analysis, it is important to define what is considered as linguistic uncertainty and how it can be defined in the context of this study.

In linguistics, uncertainty is related to concept of hedging (Hyland, 1998). Hedging is defined as any linguistic mean, which could indicate that a writer either does not wish to express complete commitment to the truth or that this commitment is not expressed categorically. This dual lack of commitment is what proposed methodology tries to capture both on linguistic and on data level.

In the literature, hedging is associated with occurrence of hedging cues, phrases, also referred to as *weasel words* (Hyland, 1998). In this paper a list of 420 hedge cues will be used to validate if a sentence introduces uncertainty. The list of cues created for the purpose of this study, is a combination of two dictionaries. One of the sources is Harvard General Inquirer (GI) word list[1], mainly used in Psychology and Sociology, capturing documents' attributes and classifying them as positive, negative, strong or weak statements. The *IF* category from the GI list is used in this work, as it includes the words that denote feelings of uncertainty, doubt and vagueness e.g. *appear, barley, likely.* Since GI list might not cover phrases that indicate uncertainty in financial field specifically, a dictionary created by Loughran and Mcdonald (2016) was added to the GI list, extending it by words such as, among others *volatility, fluctuation, likelihood.* The exact usage of the word lists will be described in more detail in later sections. A sentence of an article can be classified as uncertain in the context of this study, if it includes at least one of the hedge cues.

3. Data

Data used for the purpose of this study includes a set of banking competition articles, from which text and reported statistics were extracted. In addition, the list of words indicating uncertainty is used for hedging classification.

3.1. Articles collection and data extraction

Articles collection was done over three month period and concluded in March 2018. The set of banking competition papers was initially chosen by selecting articles tagged with keyword *Panzar-Rosse*, using search engines of Business Source Complete (EBSCO) and Research Papers in Economics (REPEC) databases[2]. An additional source of selected articles was Bikker et al. (2012) work, where a sample of 31 papers using Panzar-Rosse model, was listed.
Further selection of the articles was done using the following eligibility criteria, what limited the final sample to 71 articles.

Eligibility criteria:

1. Paper is written in English.
2. Panzar-Rosse is a single model used (included in keywords).
3. Article is an empirical study of a banking markets not a theory paper. Theoretical papers with empirical applications are exceptions included in the sample.
4. Paper can be both published or unpublished.

Inclusion rules are based on Stanley et al. (2013) guidelines for economic meta-study. The English language criterion is set in order to reduce the complexity of text analysis. Single model papers were collected to simplify the comparison of reported statistics. For obvious reasons, lack of reported statistics, theoretical papers are

[1]Source: http://www.wjh.harvard.edu/ inquirer/homecat.htm
[2]Source: http://repec.org

excluded. Finally, collection of both published and unpublished papers increases the diversity in the sample and reduces risk of the publication bias. Including the unpublished articles in the sample could however increase the cost of comparing the articles, since the authors do not have to comply to the standards set by the journals and as a consequence, the style used in working papers might be of a lower quality. The difference concerning the methodology used should not be an issue in this study, since all papers use Panzar-Rosse model, which is well established.

During the collection process of the articles, not only the text files were gathered, but also the qualitative information about the publication and the authors. Information about papers includes: title, authors, publication date, analysed countries and time frame of the data used. The authors' information includes: name, surname and institution that the author is affiliated with[3]. The affiliation of the authors is available in REPEC database, which has a significant advantage over other data sources, since it allows to directly access authors affiliation data. Affiliation is represented as a number between 0 and 1 and can be defined as proportion of papers written while affiliated with one institution over total research output of a given author. One author can have multiple affiliations and they all sum up to the total of 1. In this work in case a researcher is associated with both university and central bank, all work of this person is considered to be connected to central bank. In case REPEC affiliation score was not available, most current workplace was filled using results from the popular search engine.

The collection of the data was done manually, in spite of the API access offered by the REPEC database, which should allow to automatise this process[4].

In empirical applications of presented methods, author's affiliation is used to divide collected articles into two groups: papers with author affiliated with the central bank and others, university or private institution researches. After collection of articles and the related information, each paper was assigned a unique ID, starting with letter *p* followed by three digits e.g. *p001* in order to simplify reporting of the results.

Second part of the data gathering process is the quantitative data extraction. There are two types of data extracted from each paper in the sample. First, it is the linguistic data, namely a set of words and phrases used in the article. This extraction was followed by conversion of text into text vectors, which was successful in case of 69 of the papers in the sample[5]. Out of the 69 papers left, 15 are affiliated with central bank.

Second type of quantitative data extracted from papers were the reported statistics: p-values, t-statistics and Panzar-Rosse H-statistics. Not all papers reported all three values, however most of the studies reported either p-value, t-statistics or standard errors. In such cases, standard errors, were transformed into p- and t-values. T-statistics were transformed into p-values[6]. This transformation was not implemented the other way around, specifically no p-value was transformed into t-value, as many of the reported p-values were equal to zero. The remaining studies that did not report any of the three statistics, used star system to report significance. These

[3]Complete data can be accessed at: https://osf.io/nybg3/

[4]Although access to the API was granted by the Federal Reserve Bank of St. Louis for the purpose of this study, this API required static IP to connect to, which was not available to us. API access is an opportunity to explore in the future research.

[5]The two excluded papers were photocopies, not the original PDF files.

[6]Both transformations were made only when the sample size was provided.

studies were excluded from the sample. All of the statistics were manually extracted from the documents, since the documents use different table formats, what makes automating such process very difficult. The final set of articles and reported statistics consists of:

- t-statistics from 51 articles, incl. 14 central bank papers,
- p-values from 50 papers, incl. 14 central bank papers,
- h-statistics from 65 articles, incl. 15 central bank papers.

3.2. Additional data sources

Next to the lists of words discussed in Section 2.2, Wikipedia's sentences with weasel annotations are used to train the hedging classifier.

This set of Wikipedia's sentences was created in 2010 at the University of Szeged as part of collaborative task - CoNLL-2010. The purpose was to detect sentences containing uncertainty [7]. Each of the sentences included was manually annotated as uncertain or not, what was indicated by for instance the usage of complex verb phrases or using statements, which could evoke questions such as *Who says that?*, *Whose opinion is this?* or *How many people think so?*[8].

This paper uses the same data set of manually annotated sentences to train the hedging classifier that will later on be used to classify the sentences from our sample of banking competition papers.

4. Methodology

This section describes methods that can be used to measure the level of uncertainty of an individual paper. These methods can be divided into one group of methods related to data analysis and the other group related to text analysis. The data analysis section describes ways to analyse reported test statistics. Text analysis part focuses on the following three areas: text cleaning, measuring uncertainty and measuring similarity. Each method is followed by an empirical application, presented using the sample of 69 collected Panzar-Rosse papers.

4.1. Reported Data Analysis

In the annual presidential address, Harvey (2017) raises concern that majority of the results presented by the financial research community are significant. By doing so, author attempts to raise awareness of the actual meaning of the p-value and the publication bias. The problem of significant results is also relevant in the context of this paper, as the aim of this study is to measure the "opposite" of the publication bias. The hypothesis proposed here suggests that central bank researchers could be more prone to reporting insignificant, inconclusive findings.

Being more prone to reporting inconclusive findings, and therefore probability of making Type II error cannot be calculated directly from the reported statistics. As it is difficult to measure and quantify authors' willingness to report more false negative findings, this paper proposes an alternative approach to measuring Type II error propensity, by investigating if reported statistics show any abnormal patterns. Such

[7]Source: http://rgai.inf.u-szeged.hu/conll2010st/
[8]Source: http://www.aclweb.org/anthology/W10-3001

approach is valid in this domain specific case, since we focus on analysing single, well-established model with marginal possibility for deviation in methodology. Furthermore, most of the banking competition studies rely on public data sources. These two factors minimise the possibility to falsify results during data collection or analysis process. Consequently, the only part where researchers could engage in p-hacking is when reporting their results. (Harvey, 2017).

In order to look for any abnormal patterns in the reported statistics, the analysis focuses on comparing the distributions between the two groups of researchers, those associated with central bank and those affiliated with other institutions. It is assumed that on average the two groups should get comparable results. For the purpose of comparing the distributions, two sample Kolmogorov-Smirnov (KS) test can be used. KS test is a non-parametric test used to compare distance between empirical distribution functions, with null hypothesis stating that the samples are drawn from the same distribution.

$$D_{n,m} = \sup_x |F_{1,n}(x) - F_{2,m}(x)| \tag{3}$$

$D_{n,m}$ is the KS statistics, where $F_{1,n}(x)$ is the distribution function of the first sample with sample size n. However, simply testing for similarity between samples is not sufficient to grasp the whole scope of the compared distributions. In order to understand characteristics of the distributions more extensively, one should look at their features combined with a visual examination. The proposed approach is advantageous in a way that it is relatively simple to implement and it provides the initial understanding of the between group differences. On the contrary, this approach only answers the question if the analysed groups report significant results more or less frequently, and not if the willingness to accept Type II error is higher in one of the groups. Furthermore, since social sciences studies are not always comparable and homogeneous, one has to control for countries analysed, analysis period and data source.

4.2. Application: reported data analysis

Tables 3, 4 and Figures 1, 2 reported below represent descriptive statistics and distribution plots for t-statistics and p-values reported in the sample of banking competition papers used in this study. The results show that although both groups report high t-statistics and low p-values, there is a visible difference between them. Articles associated with central banks report significant results more frequently. The fat right tail of the central banks' reported t-statistics and its median value being more than 4 times higher than the value of the non central bank papers show clear skewness in the central bank's sample. The Kolmogorov-Smirnov test also confirms that both distributions of the samples are not the same as indicated by the p-value equal to 0. Discussed results are opposite to what was expected. Data shows that central bank's P-R articles present findings, which are more frequently significant rather than insignificant as initially hypothesised.

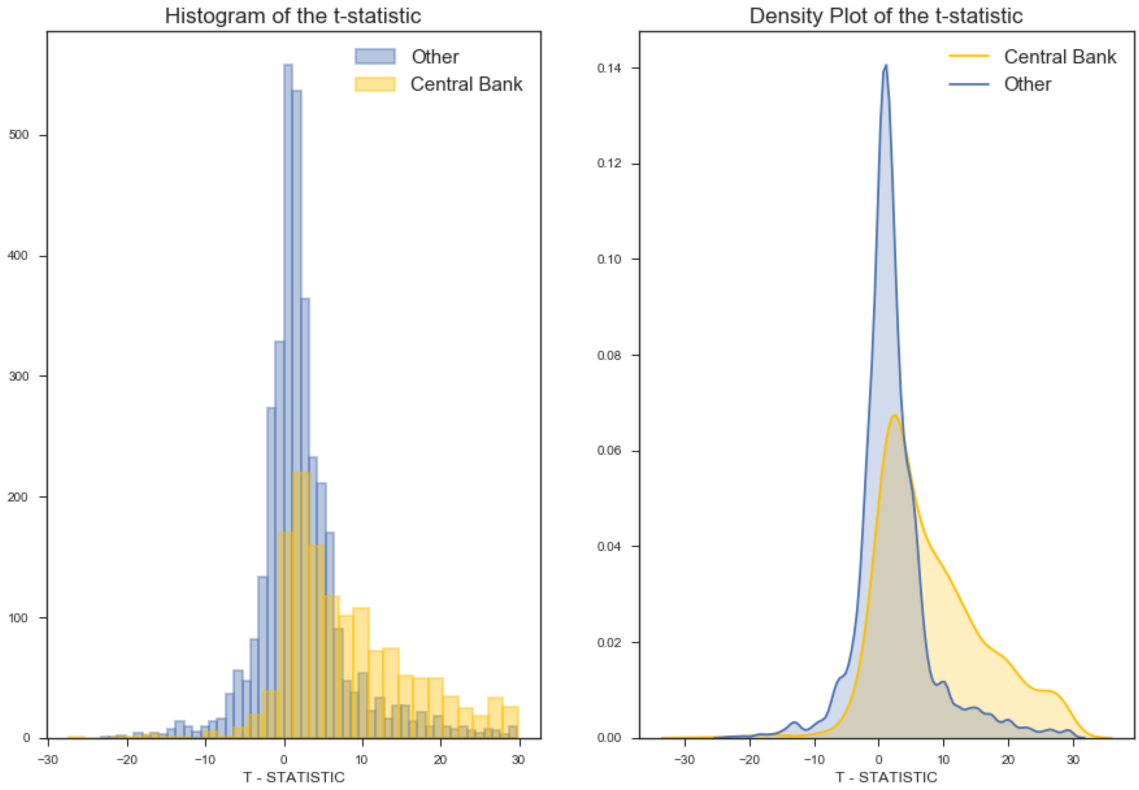

Figure 1: Distribution of t-statistic across articles

	Central Bank	**Other**
mean	8.11	2.23
standard deviation	8.18	5.75
75%	13.15	4.1
median	6.15	1.42
25%	2.17	-0.32
Observations	1360	3627
Kolmogorov-Smirnov (p-value)	0.00	

Table 3: t-statistic

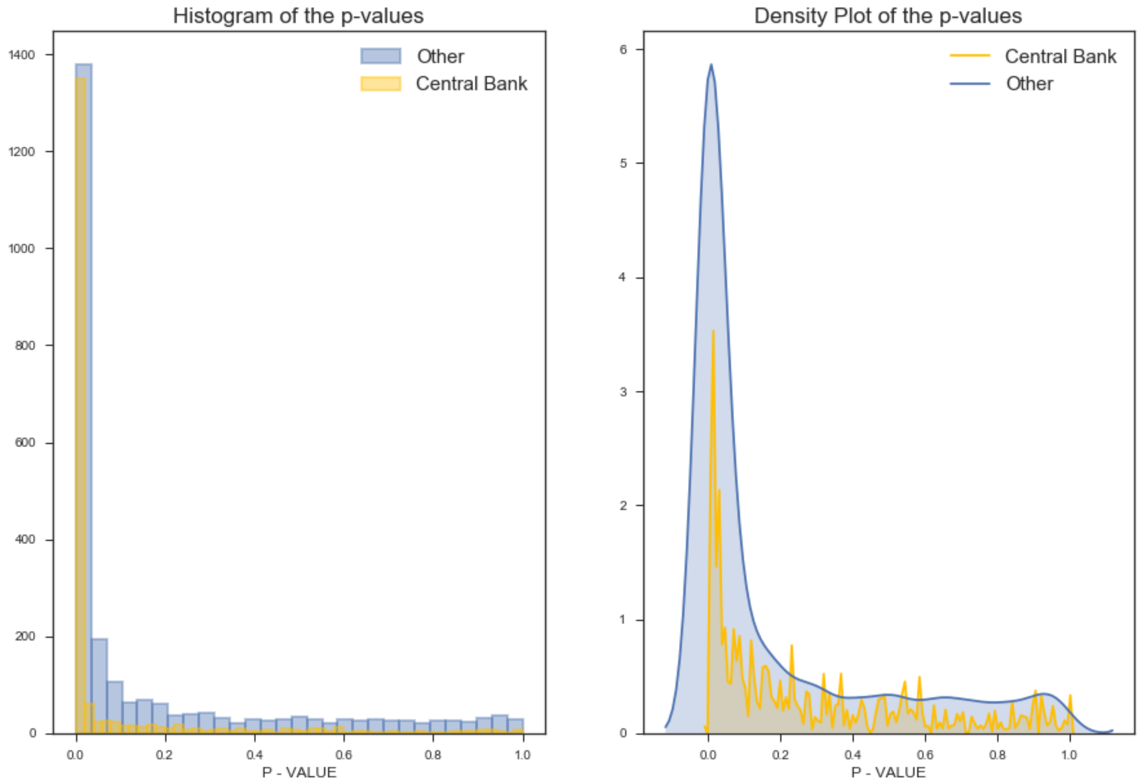

Figure 2: Distribution of p-value across articles

	Central Bank	**Other**
mean	0.081	0.173
standard deviation	0.199	0.277
75%	0.018	0.024
median	0	0.02
25%	0	0
Observations	1456	2879
Kolmogorov-Smirnov (p-value)	0.00	

Table 4: p-values

Table 5 and Figure 3 below report distribution characteristics of the h-statistics of papers analysed. The results of h-statistics analysis presented below do not signal a significant difference as compared to the results of t- and p-statistics results presented above. H-statistics show a mean difference significant at 5% level, however this result is mainly driven by Bikker et al. (2012) paper.

The work of Bikker et al. (2012) analyses a sample of over 100 countries, which is much higher than the number of countries in other papers. The resulting high number of h-statistics reported in this paper skews the counts. Bikker et al. (2012) paper being included in the group of central bank papers causes that the number of observations from only 15 papers is much higher as compared to the group of 50 non central bank papers.

Despite the fact that the mean difference is significant at 5% level, it could be argued that the h-statistics distributions are still similar to each other when accounting for their mean, median and standard deviation.

In addition, the bimodal distribution of h-statistic in the sample of university affiliated papers is to be mentioned, as the same feature is not present in the central bank papers' sample. This also has an impact on the KS test, which concludes, as expected, that the distributions are dissimilar. All things considered, main characteristic of the distributions are comparable although not exactly the same. Our hypothesis suggests that central banks prefer to report values for the h-statistics that are within range $0 < H \leq 1$. Such range indicates monopolistic competition equilibrium, suggesting no market power abuse in order not to raise concern of the financial market participants (Table 2). Despite the fact that the h-statistics are within the expected range, it is difficult to argue that the reported results differ with respect to the affiliation of the author with the central bank or the lack of it.

	Central Bank	**Other**
mean	0.528	0.492
standard deviation	0.36	0.366
75%	0.776	0.76
median	0.613	0.54
25%	0.33	0.22
Observations	1366	1036
Kolmogorov-Smirnov (p-value)	0.0015	

Table 5: H-statistic

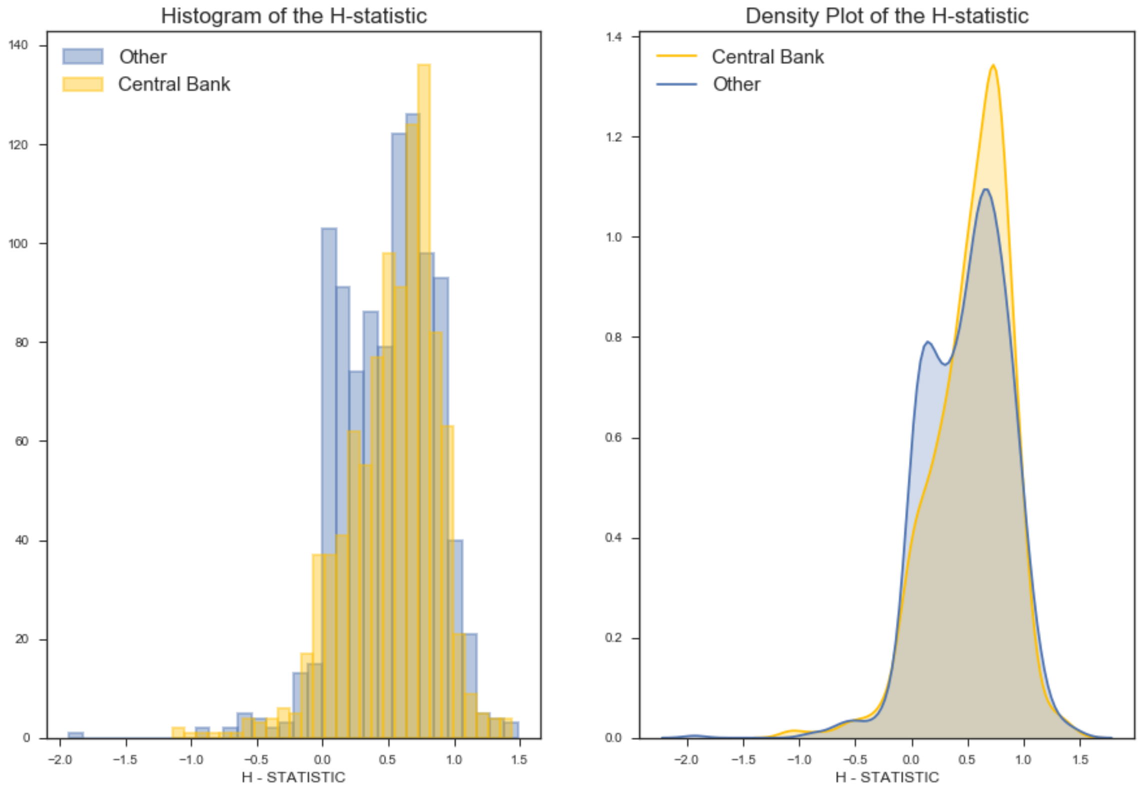

Figure 3: Distribution of H-statistics across articles

The following Figure 4 depicts another representation of h-statistics, reporting boxplots of h-statistics per article. The aim of such visualisation is to look at the heterogeneity between the two groups of researchers in more detail and to mimic the methodology applied in traditional meta study research. Although Figure 4 is not a typical forest plot, which is usually used for the purpose of comparing treatment-control differences in meta study research, it still allows to analyse cross paper dissimilarities and h-statistic distributions on a paper level. Figure 4 confirms previous results by showing no clear heterogeneity between observed groups.

The main disadvantages of the data analysis presented is the fact that it does not take into consideration the noise in the data. It is a severe fault as one of the goals of the robust meta analysis is to account for it (Stanley et al., 2013). The inability to exclude noise from data is caused by not being able to standardise the measured effect (h-statistic), due to various sets of countries and different time frames analysed in papers.

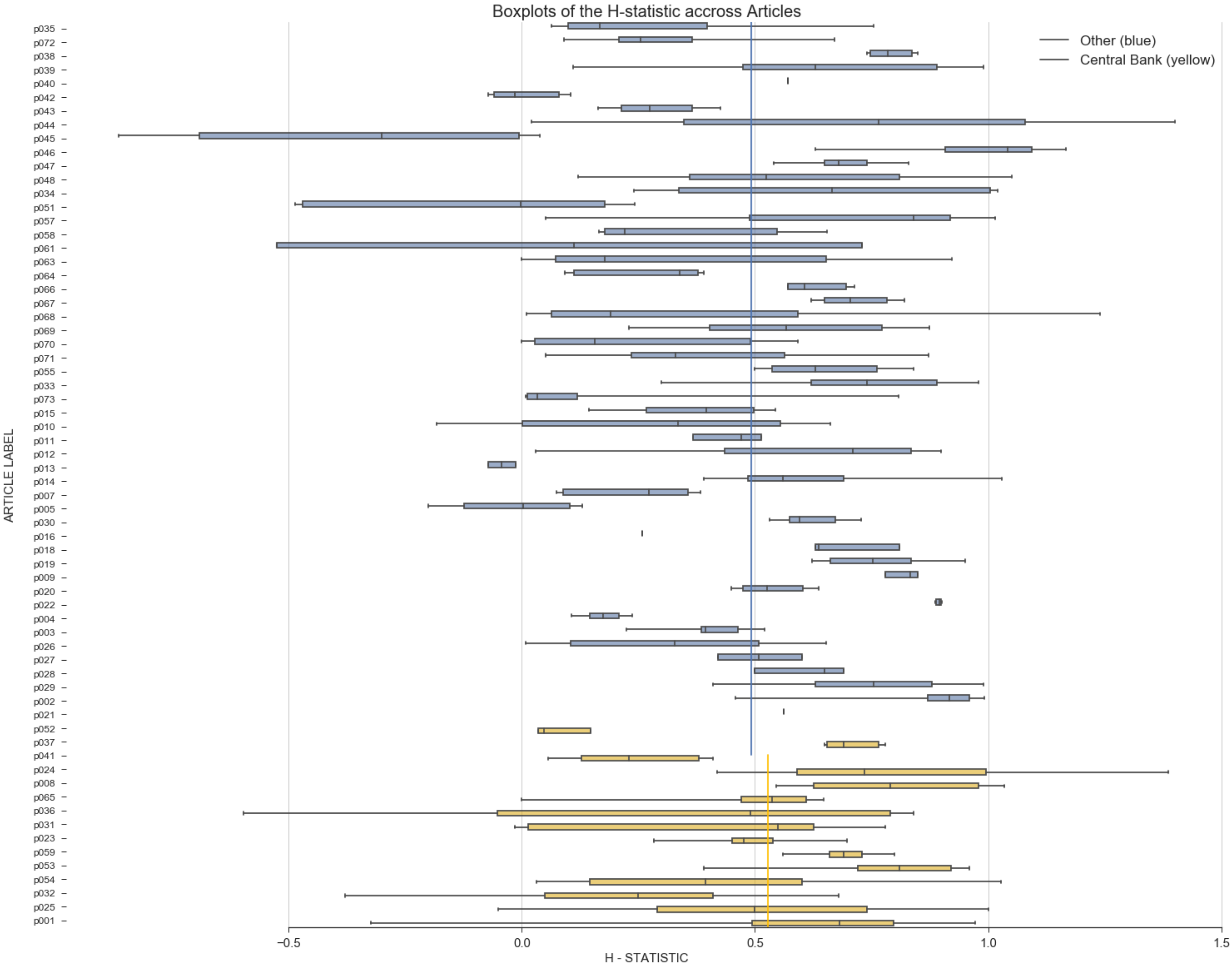

Figure 4: Boxplots of h-statistics per article

4.3. Text Analysis

The main purpose of the text analysis methods is to extract not only the direct message from the document itself but also to discover the hidden unintended information (Loughran and Mcdonald, 2016). In order to achieve that, cleaning and normalising of the text data is an essential step, which helps to remove the unnecessary noise before analysing the actual linguistic data.

4.3.1. Text cleaning

Natural language data cleaning process consists of the following steps that should be taken in order to achieve consistency of the results.

The first step concentrates on the reference lists of articles in the sample. Since this analysis focuses on specific domain, with many reoccurring sources, the reference list can be a misleading indicator of documents' similarity. Therefore, to avoid any bias, all references are removed from the documents. Next step is to unify the text data. First, all words are transformed to lowercase, as including capital letters does not bring any additional value to the analysis. Further, all non-unicode characters are excluded, only alphabetical and numeric characters remain. Afterwards, English stop words are also removed. Stop words are words, which are most commonly used in English language such as: *a, the, to, on*, which bring no conceptual value in text. In this paper stop words included in Python package SpaCy were used[9]. In addition, these stop word list was adjusted for the purpose of this study, since we are interested in understanding vagueness and uncertainty. Words such as *about, few, more, most, some* can signal hedging and were therefore not excluded. Final unification step was lemmatization, which is finding a base of the word. Lemmatization allows to group the inflected words together so they can be analysed as one. For example, word base for *am, are, is* is *be*, and for *car, cars, car's* is *car*. Another possible technique for finding a common base of words is stemming, however since lemmatization is considered to be superior to this approach (Loughran and Mcdonald, 2016), it will not be used in this study.

The next step of the cleaning process is transformation of the document into vectors of word counts, the so-called bag-of-words technique (BoW). The underlying assumption of BoW approach is that order of the words and thus their direct context is not important (Loughran and Mcdonald, 2016). This is an obvious simplification but it does allow to reduce dimensionality of the document so that it can be further analysed using linguistic techniques. Most common technique of constructing BoW, is to collapse a document down to a term-document matrix consisting of rows of words and columns of word counts (Loughran and Mcdonald, 2016).

The final step of the text cleaning process it to normalise vectors of word counts. It is debated if and how word counts should be normalised (Zobel and Moffat, 1998). A common agreement is that in most cases using word counts is not desirable, since it is related to the document length. One of advised weighting techniques to use instead, is term frequency inverse-document frequency (TF-IDF). TF-IDF allows to measure not only how important a term is within a single document but also its importance within whole corpus of the documents.

The inverse document frequency is calculated as:

$$idf_t = log\frac{N}{df_t} \tag{4}$$

where df_t is the number of documents in a collection that contain a term t. N represents the total number of documents.

TF-IDF is represented as:

$$tf - idf_{t,d} = \begin{cases} \frac{1+log(tf_{t,d})}{1+log(a_d)}\ log\frac{N}{df_t} & \text{if } tf_{t,d} \geq 1 \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where $tf_{t,d}$ is a raw frequency count of term t in document d and a_d is the average number of words in document d.

Next section will use raw word counts approach for words associated with sentiment (uncertainty), as well as TF-IDF approach for sentence classification in order to provide comparative measure of the uncertainty.

[9]Source: https://spacy.io

4.3.2. Measuring Uncertainty: Uncertainty Index

In this section we will focus on two approaches that can be used to identify and measure uncertainty. First one is based on raw frequency counts of the hedge cues. Second approach aims to identify uncertain sentences and calculate their share within a text (TF-IDF).

Our hypothesis suggests that central banks have higher propensity to use hedging cues to introduce vagueness to text. In order to capture uncertainty within a single document we propose simple uncertainty index, which measures proportion of hedge words in text (d).

$$\text{uncertainty index}_d = \frac{\sum \text{hedge words}_d}{\sum \text{words}_d} \tag{6}$$

The index will allow to analyse if and how frequently terms listed as hedge cues are used. The comparison will be done using collected articles sample to see if any documents use vague structures more frequently.

4.4. Application: uncertainty index

In appendix B, Tables 11 and 12 show ranking of each document based on their uncertainty index score. We can observe that 6 out of 15 central banking papers are in the top 20 most uncertain papers. Philip Davis and Karim (2013) (central bank affiliation) paper was evaluated to use most uncertain cues, as 8.9% of the words are categorised as being uncertain. Second central bank affiliated study that uses most hedge words is Bikker (1998) with score of 7.4%. Two most vague documents written by university researchers are Nathan and Neave (1989) and Carbó et al. (2009) with scores of 8.2% and 7.75% respectively. The average uncertainty index for both groups is 5.1%, which suggest that there is no major difference between them. Furthermore, when analysing which hedge cues are most frequent in the samples (Table 6), the usage of similar words further supports no major difference between the two groups.

Group	**Hedge cues**
All articles	differ, depend, risk, indicate, some, suggest, most, consider, many
Central Bank	risk, differ, depend, some, most, consider, may, indicate, suggest
Others	differ, depend, indicate, risk, some, suggest, most, consider, may

Table 6: Most frequently used hedge cues by group (in descending order)

Additional to the hedge cues, statistical cues were investigated. Although not many examples of the latter were found, it is interesting to list the most commonly used ones in the order of priority.

Group	Statistical cues
All articles	reject, hypothesis, significant, error, alternative, null hypothesis, significance level
Central Bank	**significant**, reject, hypothesis, error, null hypothesis, alternative, reject null
Others	reject, hypothesis, significant, p-value, error, alternative, null hypothesis, significance level

Table 7: Most frequently used statistical cues by group (in descending order)

Significant is the most frequently used statistical cue, seen in the central bank papers, which is in line with the analysis of reported statistics (Section 4.2).

4.5. Measuring uncertainty: sentence classification

The approach proposed above focuses on raw frequencies of the hedge cues within text, hence it assumes that every time a hedge cue occurs, it increases the level of overall vagueness in a paper. However, language is much more complex and occurrence of a single cue might not make the full sentence vague. Therefore, in order to account for this language complexity another approach for detecting vagueness will be applied, using sentence classifiers similar to the those described by Medlock and Briscoe (2007) and Ganter and Strube (2009). The goal of sentence classification is to train an algorithm to detect if the statement is factual or not, in most objective manner. There are three algorithms that can be considered for such classification task: Naïve Bayes, Support Vector Machine and Random Forest.

Naïve Bayes applies Bayes rule to the classification task by calculating probability of belonging to certain class (c) being a document (d), $P(c|d)$. One important assumption of the Naïve Bayes is the conditional independence, which is aligned with BoW techniques. Bayes classifier calculates probability of belonging to a certain class by evaluating individual features (x_i) of the document (d).

The advantages of using the Naïve Bayes in text analysis are the efficiency of the algorithm and its robustness to irrelevant alternatives. Furthermore, it behaves well with many equally important features, which is a common characteristic in language processing (Medlock and Briscoe, 2007). On the other hand, the main disadvantage of this method is its inability to classify new words, since zero probabilities cannot be conditioned away, and prediction becomes zero.

Second widely used method is Support Vector Machine (SVM) (Light et al. (2004), Medlock and Briscoe (2007), Velldal (2011)). SVM has two objectives in the optimisation problem, firstly maximising the distance between two decision boundaries (hyperplanes) and secondly correctly classifying all observations. The major advantage of the SVM is that it is a universal learner, it works well with both linear and nonlinear classification rules. Moreover, SVM is independent of the dimensions of the feature space. The independence is an inherited feature, since SVM optimises boundaries margin and not the data itself (Joachims, 1998).

The last classification algorithm used is random forest, which is based on decision trees. The main advantage of the this approach is that in comparison to Naïve Bayes, it works better with new words, since they are excluded from the branching decision process. Furthermore, it uses implicit feature selection, as the decision trees are trained using bagging (Breiman et al., 1984). However, random forest is not frequently used in text analysis for several reasons. First, text analysis uses very sparse data sets (term-vector matrices) and there is a danger of creating weak trees as a consequence of how random forest is trained. Random forest trains each individual tree on random subspace of features and then decides on class assignment with majority vote (Breiman et al., 1984). However, in case of sparse matrices, many features will be insignificant, leading to possibility of random forest producing weak classifiers. What is more, random forest is not linear black-box model, which makes it difficult to understand how the decision to assign observation to one class was made (Breiman et al., 1984).

Classification is supervised machine learning technique, since it requires an example of input-output in order to learn features of the data and to correctly classify new inputs. Since the purpose of the task is to minimise the human bias and since we lack the linguistic expertise needed, the following application will be trained using the Wikipedia's sentence set described in section 3.2. This will change our approach into semi-supervised learning, since final classification of sentences in the sample of banking competition papers will be made based on different data set than the one used for training.

4.6. Application: sentence classification

The goal of sentence classification is to order sentences into factual and not factual groups and to see if there are more vague sentences in central bank's articles, as compared to the other group.

In order to prepare the data for classification, all sentences from the articles as well as the Wikipedia's sentences were cleaned and normalised as described in section 4.3.1. For the training and prediction samples, sentences that contained more than 20 characters were chosen. After applying this selection, the training sample (Wikipedia's set) consisted of 4956 uncertain and 13788 factual sentences. Three models were trained and tested: Multinomial Naïve Bayes, Random Forest with 200 trees and SVM with linear kennel. Since text data is usually linearly separable, using different kennel for SVM model would not improve its performance (Joachims, 1998).

The performance of the models cannot be directly assessed, since a semi-supervised technique was used, including different training and prediction sets of sentences. In order to validate models' performance despite of that, the Wikipedia's set of sentences was split into training and testing set. The results are reported in the Appendix C in Tables 15 and 16. The SVM model outperformed Random Forests having higher values of both precision and recall. As expected, Naïve Bayes did not produce any predictions due to the sample space that included new features. Random Forest and SVM classified respectively 508 and 526 out of 16903 sentences as uncertain. The prediction results as a share of total sentences for each model are presented in Table 8 below. Each of the models was estimated ten times in order to investigate the consistency of the predictions. As a result of the training process of the Breiman trees, discussed in section above, random forest is a weaker classifier with inconsistent predictions. For that reason, the SVM will be used in the subsequent analysis.

Group	Random Forest	SVM
All Articles	0.0301	0.0311
Central Banks	0.0281	0.0372
Others	0.0305	0.0291

Table 8: Share of uncertain sentences in article by group

The two classifiers, SVM and Random Forest are predicting similar share of the vague sentences across all papers but they do show a divergence of opinions when comparing groups with each other. By reason presented in the previous paragraph, SVM is chosen as a superior model. SVM's prediction in Table 8 shows that there is a difference in language used by the two groups, which is in line with the hypothesis of this study.

As pointed out in the literature, longer sentences might be more convoluted, hence uncertain (Ganter and Strube, 2009), therefore the next step of the analysis is to investigate if there is a correlation between the number of words used in a sentence and its vagueness.

Table 9: Count of words in a sentence per group

Group		Words in a sentence
All Articles	certain	15.5
	uncertain	17.7
Central Banks	certain	16.3
	uncertain	19.88
Others	certain	15.28
	uncertain	16.95

Table 9 is based on predictions made by the SVM classifier. The results confirm that on average, uncertain sentences are longer than the factual ones. This difference is even more prevalent in the group of papers associated with central banks.

As a final step in sentence classification, a ranking of the articles with highest number of vague sentences is created (Appendix B). Four central bank articles are among the top 20 papers with highest number of uncertain sentences. In the group of central bank papers, the article written by Philip Davis and Karim (2013) is still considered the most uncertain work, followed by the study from Claessens and Laeven, where vague language is used even in the title of their article: *"What Drives Bank Competition? **Some** International Evidence"*. Study written by Bikker (1998), which was the second central bank affiliated study that used most hedge words, drops to 20th place in this ranking.

In the group of non-central bank papers, study classified as most vague is Delis (2010). What is interesting, the same study was only ranked on the 40th place in the uncertainty index ranking. Nathan and Neave (1989) ranked on the 1st place when using count of the hedging cues, drops to the fourth place in this ranking. Carbó et al. (2009) is the 3rd most uncertain non-central bank paper. This article was also the 3rd most uncertain non-central bank paper in the uncertainty index ranking, but its overall ranking dropped from 3rd to 5th position among all papers in the sample.

This ranking shows that there are differences between the simple frequency count of the hedging cues and more complex sentence classification method, however their results do not contradict one another. A robust analysis should include both methods. Raw frequency count shows how likely a researcher is to use hedge cues and the sentence classification validates if a hedge cue contributes to the uncertainty of the whole sentence.

4.7. Measuring Similarity

The purpose of this section is to investigate whether the set of articles can be split into separate homogeneous groups based on text similarity. Two unsupervised clustering methods, which split the data into groups are presented. Since the two methods include with very little human intervention, they are considered as objective.

The popular measure of document similarity is cosine similarity. Cosine similarity calculates an angle between pair of term-frequency vectors (x, y) of two articles (d_1, d_2). Cosine similarity is normalised to be in range from -1 (exactly opposite) to 1 (exactly the same) through division of the product of term-frequency vectors by the Euclidean distance[10].

$$cosine(d_1, d_2) = \frac{\sum_{i=1} x_i y_i}{\sqrt{\sum_{i=1} x_i^2}\sqrt{\sum_{i=1} y_i^2}} \tag{7}$$

Once the cosine similarity is calculated, it can be used in hierarchical clustering to find groups of comparable documents. Agglomerate hierarchical clustering algorithm starts with each document in its own cluster. Next, pairs of clusters, which are closest to each other, are merged into one cluster. The algorithm continues to combine pairs of clusters until all observations are included in a single cluster.

Second approach, which can be used to compare documents, is topic models. Topics models describe similarities and differences in the existing corpus of documents based on the words used (Boukus and Rosenberg, 2006). Latent semantic analysis (LSA) is one example of the topic models and was already applied by Boukus and Rosenberg (2006) to analyse central banks communicates. The author analysed the information included in Federal Open Market Committee minutes and showed existence of a correlation between the minutes' content and the future economic conditions.

LSA applies singular value decomposition (SVD) to the term-document matrix. Singular value decomposition is a form of principal component analysis, which factors the term-document matrix (X) into the product of three matrices: $X = USV^T$. Matrix U can be interpreted as a collection of pseudo-documents, where each column of U characterises themes of documents in our sample. The elements i and j of matrix U are a contribution of term i to the theme j. The V matrix relates documents to themes, and i and j element of V^T defines the contribution of theme i to document j, hence how strongly theme i is expressed within a document (Boukus and Rosenberg, 2006).

[10]When applied to the word counts, cosine similarity ranges from 0 to 1, because word counts cannot be negative.

As a result, by analysing matrices V^T and U we can investigate each component (topic) to see which terms (words) are given the highest weight by this component. What is more, it is possible to check how much weight each document assigns to a component and therefore, in which topic category the document is. Later, given that information, documents can be clustered based on weights they assign to each topic.

The clustering method used to group topic data is k-means clustering, a popular unsupervised algorithm based on centroids. K-means finds an optimal split of the data using predetermined number of clusters. It is an iterative algorithm that picks the initial cluster centres randomly, then finds all observations that are the closest to these centroid and averages values of all closest observation to derive new centroid value of the group. The algorithm continues to split the data given new cluster centres. The most common distance metric used in k-means is Euclidean norm, which calculates "straight line" distance between two points p and q.

$$d(p,q) = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2} \tag{8}$$

Having established all methods necessary to analyse documents' similarity, next section will continue with an empirical application.

4.8. Application: Measuring Similarity

In order to measure similarity, firstly the cosine similarity measure is calculated among each pair of the documents in the sample. The cosine similarity matrix is a symmetric matrix, with values ranging from 0 to 1 (section 4.7) and diagonal values equal to 1. Figure 9 in appendix D is a visual representation of the similarity matrix, with article labels on both axis (central bank articles labels are written in bold). Calculated similarity matrix does not provide many insights, since 80% of the values are below 0.2, signalling low similarity. Two most interesting cases are articles with labels *p031* and *p036*, (Pawłowska (2011), Pawłowska (2012)) and *p021* and *p043* (Masood and Sergi (2011), Kang Park (2016)), which have similarity of 0.74 and 0.44 respectively. The first pair (*p031* and *p036*) is a group of articles written by the same author, on exactly the same topic but using different date range. Second pair of studies (*p021* and *p043*) has high similarity due to the country, China, which both articles analyse.

Using already computed similarity matrix documents can be clustered together using hierarchical clustering. Figure 10 visualises the results. The optimal number of clusters derived is six. Most of the clusters correspond to the regions that the each group of papers analyses:

- Orange: developing countries, regions: Tunisia, India, Jamaica, Africa
- Green: Turkey
- Blue: catch-all-cluster with many of the documents covering European Union or European Countries
- Yellow: Poland
- Red: no clear single topic, mainly banking efficiency and competition drivers
- Black: China

The initial natural split of the sample is therefore based on the country analysed. This research however, is rather focused on finding any natural structure to the data and the question if within this structure, central banks are grouped together. In order to achieve that, after discovering that regional differences are a key factor for splitting the articles, the country names are excluded from the set of words to investigate the existence of any other defining features.

The hierarchical clustering is repeated on the new sample with country names excluded (Figure 11). The result of second hierarchical clustering, are three clusters. It is difficult to interpret the key theme in the first largest cluster (green). However, it is possible with the two remaining clusters. Papers in the blue cluster are discussing topics related to competition and its relationship to concentration of the market, and all of them use Herfindahl-Hirschman index to measure concentration. The leading topic in the third cluster (black) is the relationship between efficiency of the banks and the competition. Even though the achieved results show groups of papers that are more similar to each other, the defining characteristics of the clusters are not related to the affiliation of the authors, and hence they undermine the hypothesis.

The next application will focus on topic models. It will be investigated if papers can be clustered together based on the topic they discuss and if the established topics are similar for the central bank papers.

Topic model used on the sample is Latent Semantic Indexing. In order to analyse how many components (topics) data should be split into, elbow curve graph is plotted with varying number of components[11]. Elbow curve shows what percentage of the variance is explained with the increasing number of topics.

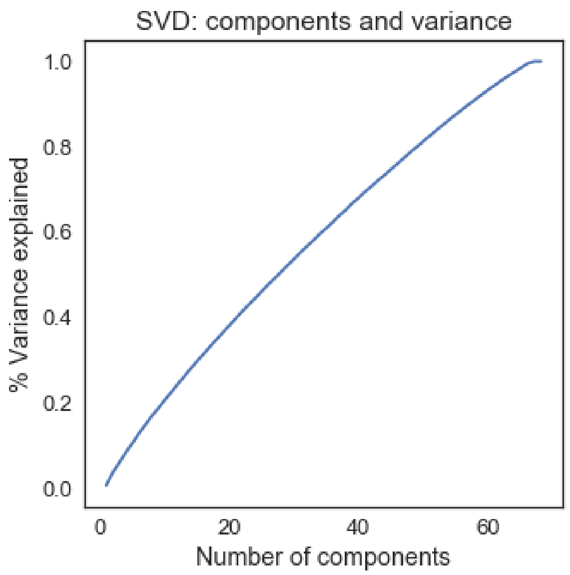

Figure 5: Elbow curve SVD

Linear trend on the elbow curve shows that variance explained increases almost linearly with the number of SVD components. Such trend suggests that it is always beneficial to add extra component but undermines the idea of topic modelling, since there is no gain in grouping papers together. There is no single component (topic) that adds significantly more marginal value than any other.

In order not to over-fit the data, 30 components are chosen, which explains approximately 50% of the variance in our sample. The resulting first 15 components with 20

[11]Maximum number of components in the case of this study is 69, as 69 papers are analysed

strongest terms are presented on Figure 12 in Appendix D, which correspond to the matrix U described in the methodology section. The resulting components point to very similar conclusion as the initial hierarchical clustering, since the strongest terms are country names.

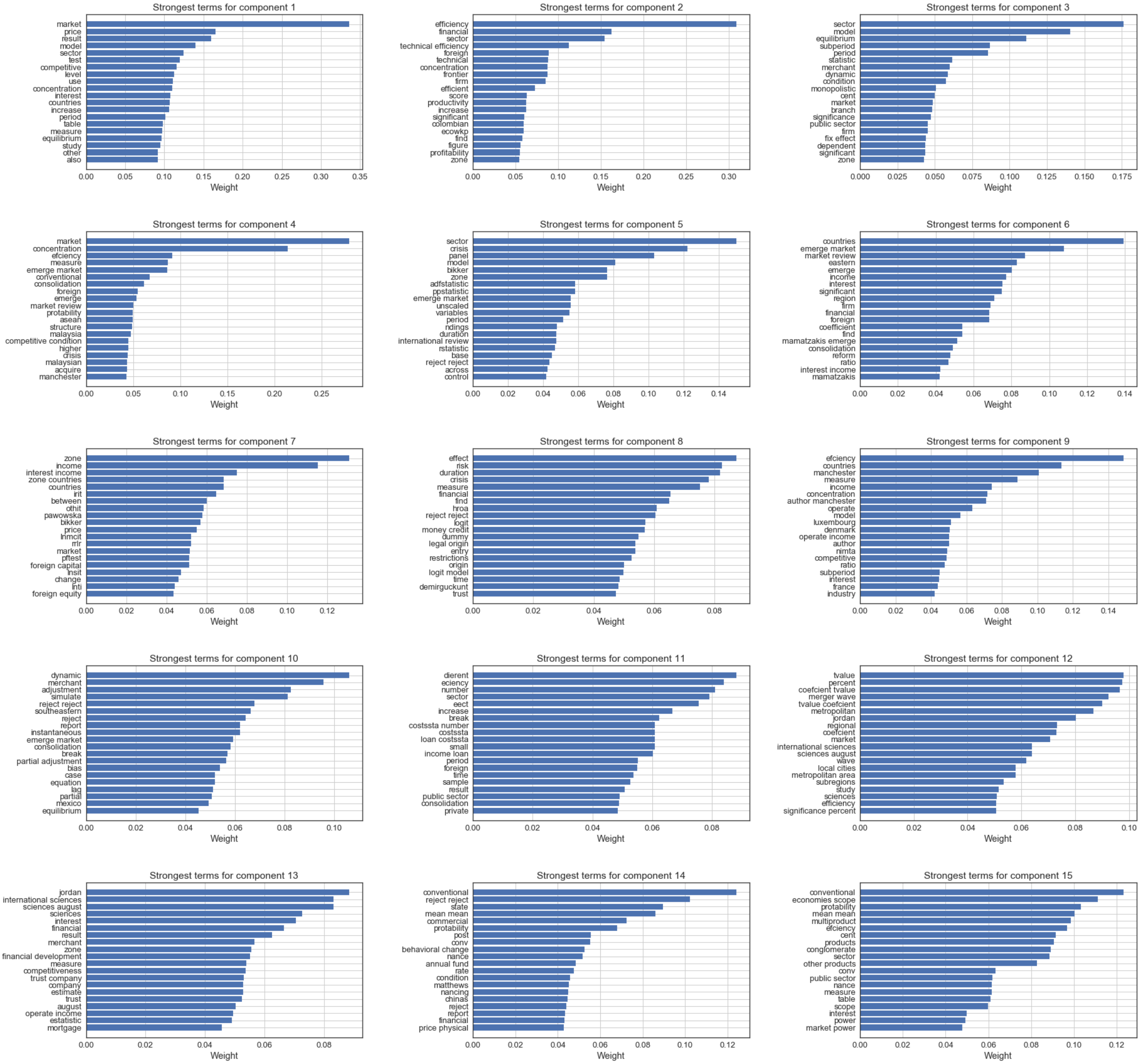

Figure 6: LSA components, excluding countries from the corpus

As before, analysis is replicated excluding the country names. Again, the same problem with choosing optimal number of components is encountered, since there is no number of components that significantly increases the data fit. In order to keep the reporting consistent, 30 components are selected (Figure 6). Topics created by the LSA on the sample without country names are more difficult to interpret, since they all look very similar. One of the reasons for that, is that the papers analysed are all related to one subject, the Panzar-Rosse competition model.

There is one topic (12) that is worth investigating further, since it includes words related to statistical testing, such as *t-value, t-value coefficient, coefficient, reject*. There is only one article that assigns the highest weight to this component, which is the Abu Khalaf et al. (2015) work, a non central bank paper.

Final approach, k-means clustering that builds on the knowledge gained during LSA, examines if papers can be split into homogeneous topic groups. K-means is applied on the low dimensional space of the LSA components to check, which papers assign similar weights to each of the LSA topics.

K-means using 2 and 10 clusters is applied. The decision to choose 10 clusters is based on the elbow curve, which measures class dispersion with varying number of clusters (Appendix D, Figure 13). The method is also repeated with 2 clusters, since the initial hypothesis stated that central bank and the non-central bank papers can be separated.

The resulting cluster split on the data with country names excluded, is presented in Figures 8 and 7[12]. Triangles represent the central bank and dots the remaining articles.

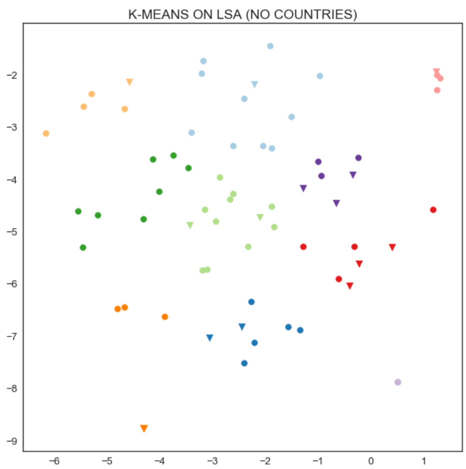

Figure 7: LSA K-means with 10 clusters

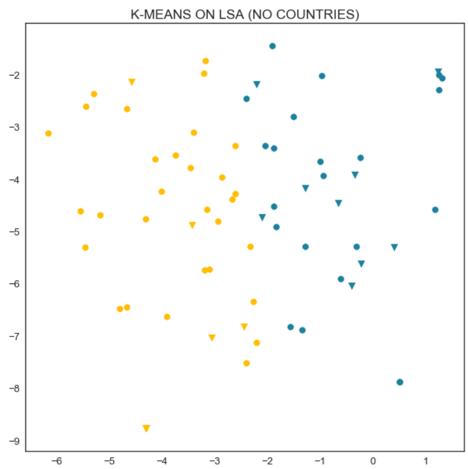

Figure 8: LSA K-means with 2 clusters

The combination of the LSA and dimensional reduction makes the interpretation of the clusters very difficult. The purpose of this method was to find an optimal split between papers, without prior assumptions and to check if central bank papers can be distinguished from other articles. This cannot be done with neither 10 nor 2 clusters, since central bank papers are mixed into almost all groups. The results for the data including country names is reported in the Appendix D, Figures 15, 16 and 14. As expected, 10 cluster split including country names divides papers based mainly on the country covered.

[12]For visualisation purposes, t-distributed stochastic neighbour embedding is used.

5. Conclusions

The purpose of this article was to introduce a set of methods that can be applied to study language and content of the banking competition studies. Methods discussed allow to measure text uncertainty, text similarity and to compare reported competition measures among two groups of researchers, affiliated with central bank and academia.

The main advantage of the presented methods is that they can be easily extended to a larger sample of the banking competition papers, and are relatively simple to implement. The simplicity of the methodology is also its disadvantage, since it might be too simplistic to find the true underlying effect.

All things considered, the empirical analysis did not find any strong support for the hypothesis that central bank researchers are more vague in reporting their results. On the contrary, the evidence suggests that central bank research reports significant results more frequently. Furthermore, there is no significant difference, in the frequency of using hedge words, between the two analysed groups. Finally, analysed central bank articles do not share any common, unique feature, that would allow to group them together and to distinguish them from other studies.

The main contribution of this paper is that it applies the natural language processing methodology to the topic that was not analysed before, the banking competition papers. Although the empirical result did not show any strong evidence supporting the hypothesis, the application was narrowed to a single competition model what is a clear limitation of this study. Extending sample of the articles and analysing language across broader spectrum of competition studies could lead to more conclusive results and is therefore recommended for the future research.

References

Abu Khalaf, B., Yaseen, H., and Omet, G. (2015). The Evolution of Bank Competition: Have Conditions Changed in the Jordanian Banking Sector? *International Journal of Academic Research in Business and Social Sciences*, 5(8):286.

Bhattacharya, S. and Boot, A. (1998). The Economics of Bank Regulation. *Journal of Money, Credit and Banking*, 30(4):745–770.

Bikker, J. A. (1998). Competition and Concentration in the EU Banking Industry. Technical report.

Bikker, J. A., Shaffer, S., and Spierdijk, L. (2012). Assessing competition with the panzar-rosse model: The role of scale, costs, and equilibrium. *Review of Economics and Statistics*.

Bikker, J. A. and Spierdijk, L. (2017). *Handbook of Competition in Banking and Finance*. Elgar.

Blinder, A., Ehrmann, M., Fratzscher, M., and Jansen, D.-J. (2008). Central Bank Communication and Monetary Policy: A Survey of Theory and Evidence. *Journal of Economic Literature*, 46(4):910–945.

Boukus, E. and Rosenberg, J. V. (2006). The Information Content of FOMC Minutes. *Ssrn*.

Breiman, L., Friedman, J., Stone, C., and Olshen, R. (1984). *Classification and Regression Trees*. Taylor & Francis.

Carbó, S., Humphrey, D., Maudos, J., and Molyneux, P. (2009). Cross-country comparisons of competition and pricing power in European banking. *Journal of International Money and Finance*.

Claessens, S. and Laeven, L. What Drives Bank Competition? Some International Evidence. Technical Report 3.

Delis, M. D. (2010). Competitive conditions in the Central and Eastern European banking systems. *Omega*.

Ehrmann, M. (2005). How should central banks communicate?

Ganter, V. and Strube, M. (2009). Finding Hedges by Chasing Weasels: Hedge Detection Using Wikipedia Tags and Shallow Linguistic Features. *Nlp.Csie.Ncnu.Edu.Tw*, (August):173–176.

Greenspan, A. (2003). Symposium: Monetary Policy under Uncertainty. Technical report, Federal Reserve Bank of Kansas City.

Harvey, C. R. (2017). Presidential Address: The Scientific Outlook in Financial Economics. *Journal of Finance*, 72(4):1399–1440.

Hyland, K. (1998). Hedging in Scientific Research Articles.

Issing, O. (2005). Communication, Transparency, Accountability: Monetary Policy in the Twenty-First Century. pages 65–84.

Joachims, T. (1998). Text Categorization with Support Vector Machines: Learning with Many Relevant Features. *Machine Learning*, 1398(LS-8 Report 23):137–142.

Kahveci, E. and Odaba, A. (2016). Central Banks' Communication Strategy and Content Analysis of Monetary Policy Statements: The Case of Fed, ECB and CBRT. *Procedia - Social and Behavioral Sciences*, 235(October):618–629.

Kang Park, B. H. (2016). How Competitive and Stable is the Commercial Banking Industry in China after Bank Reforms? Technical Report 1.

Light, M., Qiu, X. Y., and Srinivasan, P. (2004). The Language of Bioscience : Facts , Speculations , and Statements in Between. *BioLink 2004 – Proceedings of the Workshop on Linking Biological Literature, Ontologies and Databases*, pages 17–24.

Loughran, T. and Mcdonald, B. (2016). Textual Analysis in Accounting and Finance: A Survey. *Journal of Accounting Research*, 54(4):1187–1230.

Masciandaro, D. and Romelli, D. (2016). From Silence to Voice: Monetary Policy, Central Bank Governance and Communication.

Masood, O. and Sergi, B. S. (2011). China's banking system, market structure, and competitive conditions. *Frontiers of Economics in China*.

Medlock, B. and Briscoe, T. (2007). Weakly supervised learning for hedge classification in scientific literature. *Annual Meeting of Assosiation of Computational Linguistics*, 45(June):992.

Muchlinski, E. (2011). *Central Banks and Codec Language*. Paigrave Macmillan.

Nathan, A. and Neave, E. H. (1989). Competition and Contestability in Canada's Financial System: Empirical Results. Technical Report 3.

Panzar, J. and Rosse, J. (1987). Testing For " Monopoly " Equilibrium. *Journal, The Economics, Industrial*, 35(4):443–456.

Pawłowska, M. (2011). Competition in the Polish banking market prior to recent crisis for the period 1997-2007-empirical results obtained with the use of three different models. Technical Report 5.

Pawłowska, M. (2012). Competition, concentration and foreign capital in the Polish banking sector (prior and during the financial crisis). Technical report.

Philip Davis, E. and Karim, D. (2013). Exploring the sort and long-run links from banking competition to risk-reconciling conflicting hypotheses? Technical report.

Stajner, S., Glavas, G., Ponzetto, S. P., and Stuckenschmidt, H. (2017). Domain Adaptation for Automatic Detection of Speculative Sentences. *Proceedings - IEEE 11th International Conference on Semantic Computing, ICSC 2017*, pages 164–171.

Stanley, T. D., Doucouliagos, H., Giles, M., Heckemeyer, J. H., Johnston, R. J., Laroche, P., Nelson, J. P., Paldam, M., Poot, J., Pugh, G., Rosenberger, R. S., and Rost, K. (2013). Meta-analysis of economics research reporting guidelines. *Journal of Economic Surveys*, 27(2):390–394.

Velldal, E. (2011). Predicting speculation: A simple disambiguation approach to hedge detection in biomedical literature. *Journal of Biomedical Semantics*, 2(5).

Appendix A Analysed Articles

Label	Article Title
p001	Assessing Competition with the Panzar-Rosse Model: The Role of Scale, Costs, and Equilibrium
p002	Assessing Banking Sector Competition in Zimbabwe Using a Panzar-Rosse Approach
p003	Conduct in a Banking Monopoly
p004	Assessing competition in the banking industry: A multi-product approach
p005	Bank Competition and Efficiency in the FYR Macedonia
p007	Assessing Competition With the Panzar-Rosse Model: An empirical analysis of European Union banking industry
p008	Testing for competition in the Spanish banking industry: The Panzar-Rosse approach revisited
p009	Competition, Concentration and the Relevance of New Banks Entry in the Indian Banking System
p010	Banking competition in Africa: Subregional comparative studies
p011	Evaluating the state of competition of the Greek banking industry
p012	New evidence on assessing the level of competition in the European Union banking sector: A panel data approach
p013	Competitive conditions in the Jamaican banking market 19982009
p014	Competition in fragmented markets: New evidence from the German banking industry in the light of the subprime crisis
p015	Empirical assessment of the competitive conduct of Nigerian banks in a post-consolidation era
p016	The Evolution of Competition of the Tunisian Banking Sector: An Empirical Analysis
p018	Estimation of the competitive conditions in the Czech banking sector
p019	Competitive conditions in Islamic and conventional banking: A global perspective
p020	The Evolution of Bank Competition: Have Conditions Changed in the Jordanian Banking Sector?
p021	China's Banking System, Market Structure, and Competitive Conditions
p022	Measurement of Competitiveness Degree in Tunisian Deposit Banks: An Application of the Panzar and Rosse Model
p023	Competition in Indian Banking An Empirical Evaluation
p024	A Formal Test of Competition in the Banking Sector of Pakistan: An Application of PR-H Statistic

p025 How Banking competition Changed over Time

p026 Competitive conditions in the Central and Eastern European banking systems

p027 Competitive Conditions in the Turkish Banking Systems

p028 The evolution of competition in banking in a transition economy: an application of the Panzar-Rosse model to Armenia

p029 Efficiency And Competition In Romanian Banking System: Empirical Evidences

p030 Consolidation and Competition in Emerging Market: An Empirical Test for Malaysian Banking Industry

p031 Competition in the Polish banking market prior to recent crisis for the period 1997-2007 – empirical results obtained with the use of three different models

p032 Exploring the Short- and Long-Run Links from Bank Competition to Risk – Reconciling Conflicting Hypotheses?

p033 Efficiency and Contestability in the Colombian Banking System

p034 Market Structure and Competitive Conditions in the Arab GCC Banking System

p035 An empirical analysis of competition in the Indian Banking Sector in dynamic panel framework

p036 Competition, concentration and foreign capital in the Polish banking sector (prior and during the financial crisis)

p037 Competition, Concentration, Efficiency, and their Relationship in the Polish Banking Sector

p038 Competition in the banking sector: New evidence from a panel of emerging market economies and the financial crisis

p039 Does higher bank concentration reduce the level of competition in the banking industry? Further evidence from South East Asian economies

p040 Bank Competition in Sub-Saharan African Countries: Has Anything Changed in the Light of 2007–2008 Global Financial Crisis?

p041 Research Of Competition In Deposit Market Of Ukraine Based On The Panzar-Rosse Model

p042 Banking Competition and Efficiency: Empirical Analysis on the Bosnia and Herzegovina Using Panzar-Rosse Model

p043 How Competitive and Stable is the Commercial Banking Industry in China after Bank Reforms?

p044 State Aid and Competition in Banking: the Case of China in the Late Nineties

p045 Concentration and Competition in the Banking Sector of Turkey

p046 Effect of Liberalization on Banking Competition

p047 Testing for competition in the South African banking sector p048 Competition and Contestability in Canada's Financial System: Empirical Results

p049 Market structure and performance in Spanish banking

p051 Competition and Market Contestability in Japanese Commercial Banking

p052 Assessing competitive conditions in the Greek banking system

p053 Competition and concentration in the EU banking industry

p054 Competition, contestability and market structure in European banking sectors on the eve of EMU

p055 Consolidation and market structure in emerging market banking systems

p057 Banking competition and macroeconomic conditions: a disaggregate analysis

p058 The competitive nature of the Arab Middle Eastern banking markets

p059 What Drives Bank Competition? Some International Evidence

p061 Competition and concentration in the banking sector of the South Eastern European region

p062 Competition and Contestability in Transition Banking: An Empirical Analysis

p063 Bank Competition, Concentration and Efficiency in the Single European Market

p064 Competition in the Turkish banking industry

p065 Competition and Concentration in the New European Banking Landscape

p066 Competitive conditions among the major British banks

p067 Restructuring, consolidation and competition in Latin American banking markets

p068 On the measurement of market power in the banking industry

p069 Market Competition before and after Bank Merger Wave: A Comparative Study of Korea and Japan

p070 Competition Tests with a Non-Structural Model:The Panzar-Rosse Method Applied to Germanys Savings Banks

p071 Competition in banking: A disequilibrium approach

p072 Are Competitive Banking Systems More Stable?

p073 Market power in local banking monopolies

p074 Cross-country comparisons of competition and pricing power in European banking

Appendix B Uncertainty Scores

Table 11: Uncertainty score non central bank articles

Rank	Article label	Uncertainty Index
2	p048	0.0821
3	p074	0.0776
5	p070	0.0690
6	p003	0.0666
7	p057	0.0659
8	p034	0.0650
9	p014	0.0631
10	p051	0.0615
11	p073	0.0609
12	p005	0.0604
13	p019	0.0596
16	p071	0.0575
18	p033	0.0571
19	p064	0.0569
21	p045	0.0562
22	p020	0.0554
23	p010	0.0544
24	p030	0.0540
25	p012	0.0537
26	p062	0.0534
27	p046	0.0532
28	p044	0.0528
29	p068	0.0524
30	p063	0.0519
31	p061	0.0517
32	p066	0.0513
33	p035	0.0508
34	p042	0.0507
36	p007	0.0499

37	p072	0.0499
38	p040	0.0499
40	p026	0.0494
41	p021	0.0493
42	p018	0.0488
43	p055	0.0475
45	p015	0.0473
46	p011	0.0473
47	p027	0.0473
48	p002	0.0458
49	p029	0.0458
52	p058	0.0441
53	p009	0.0440
54	p038	0.0433
55	p013	0.0432
56	p043	0.0424
57	p016	0.0396
58	p047	0.0376
59	p039	0.0362
62	p022	0.0342
64	p004	0.0305
65	p069	0.0304
66	p067	0.0286

Table 12: Uncertainty score central bank articles

Rank	Article label	Uncertainty Index
1	p032	0.0895
4	p053	0.0736
14	p065	0.0587
15	p054	0.0580
17	p001	0.0574
20	p024	0.0569
35	p059	0.0500
39	p008	0.0498
44	p052	0.0474
50	p023	0.0453
51	p025	0.0450
60	p036	0.0357
61	p031	0.0342
63	p041	0.0331
67	p037	0.0283

Table 13: Sentence uncertainty score non central bank articles

Rank	Article	No. uncertain sentences	Total sentences	Uncertainty Index
1	p026	15	174	0.0862
4	p048	16	260	0.0615
5	p074	20	344	0.0581
6	p062	15	259	0.0579
7	p034	10	181	0.0552
8	p057	10	183	0.0546
10	p051	6	114	0.0526
11	p007	9	175	0.0514
12	p029	6	117	0.0513
13	p011	18	355	0.0507
14	p020	5	100	0.0500
15	p009	9	200	0.0450
16	p045	10	225	0.0444

17	p067	7	158	0.0443
18	p027	7	171	0.0409
19	p018	5	126	0.0397
21	p063	13	337	0.0386
22	p069	4	108	0.0370
23	p003	8	235	0.0340
25	p042	6	181	0.0331
27	p068	9	284	0.0317
29	p061	7	227	0.0308
30	p073	10	326	0.0307
31	p071	8	264	0.0303
33	p047	7	238	0.0294
34	p039	12	430	0.0279
35	p072	10	359	0.0279
36	p058	3	108	0.0278
37	p055	7	260	0.0269
38	p010	9	346	0.0260
41	p016	5	201	0.0249
42	p015	5	203	0.0246
43	p038	3	122	0.0246
44	p033	13	533	0.0244
46	p066	5	213	0.0235
48	p002	7	313	0.0224
49	p044	5	226	0.0221
50	p013	3	136	0.0221
52	p030	4	182	0.0220
53	p012	6	278	0.0216
54	p005	5	233	0.0215
56	p070	5	257	0.0195
57	p022	4	212	0.0189
58	p040	4	229	0.0175
59	p021	3	177	0.0169
60	p043	4	246	0.0163

61	p019	3	227	0.0132
62	p014	4	307	0.0130
64	p004	5	562	0.0089
65	p035	3	351	0.0085
66	p064	1	134	0.0075
67	p046	1	241	0.0041

Table 14: Sentence uncertainty score central bank articles

Rank	Article	No. uncertain sentences	Total sentences	Uncertainty Index
2	p032	25	302	0.0828
3	p059	5	80	0.0625
9	p065	19	355	0.0535
20	p053	14	358	0.0391
24	p023	7	207	0.0338
26	p041	4	122	0.0328
32	p001	12	396	0.0303
39	p025	7	275	0.0255
40	p054	6	238	0.0252
45	p008	10	419	0.0239
47	p024	6	259	0.0232
55	p052	4	187	0.0213

Appendix C Sentence classification

Table 15: Classification report SVM on testing sample

	Precision	Recall	F1 score	Support
0	0.89	1.00	0.94	4094
1	0.98	0.68	0.81	1530
Avg	0.92	0.91	0.90	5624

Table 16: Classification report RF on testing sample

	Precision	Recall	F1 score	Support
0	0.85	0 .98	0.91	4094
1	0.93	0.62	0.74	1530
Avg	0.88	0.87	0.86	5624

Appendix D Similarity

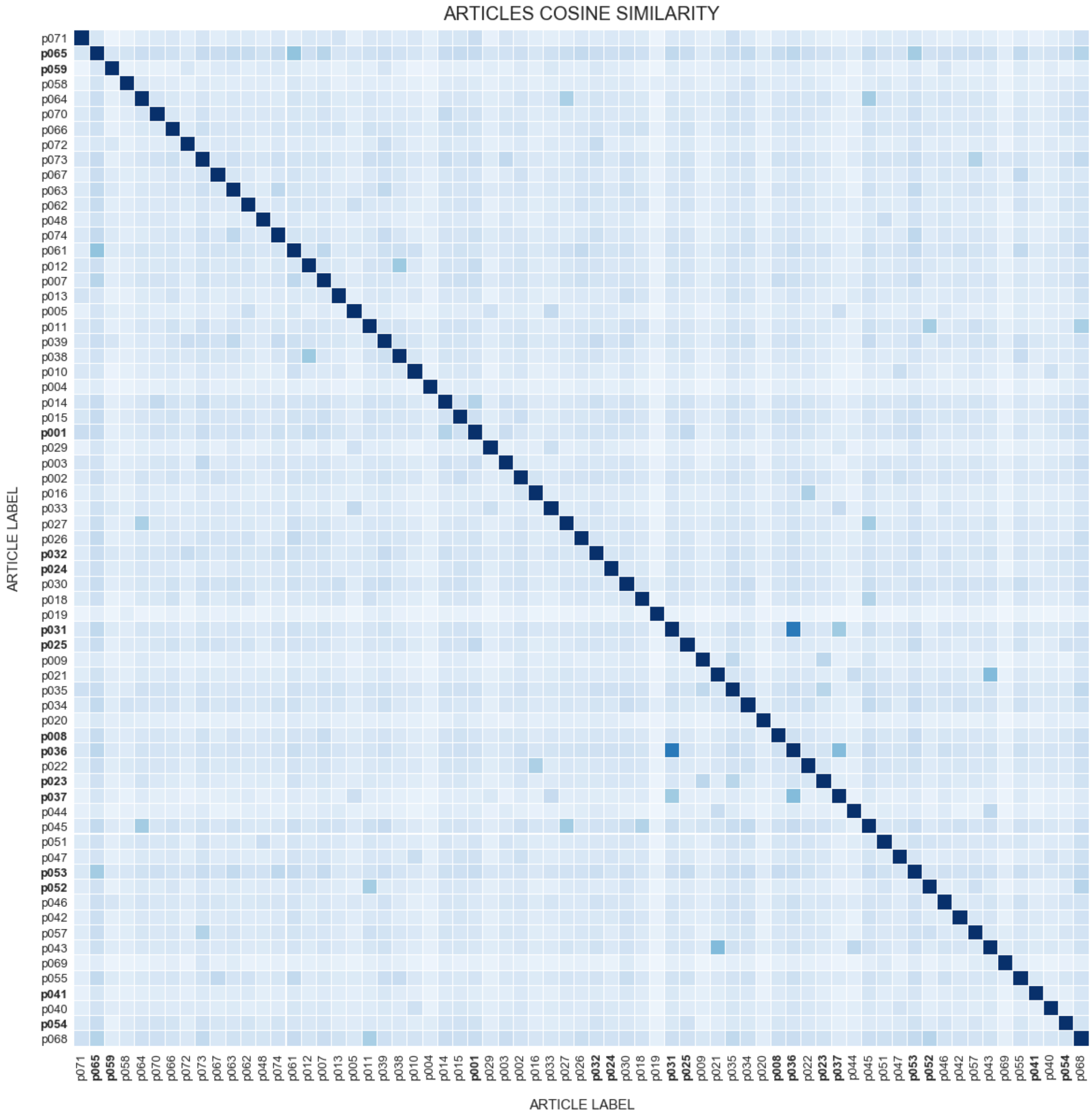

Figure 9: Cosine Similarity heat-map

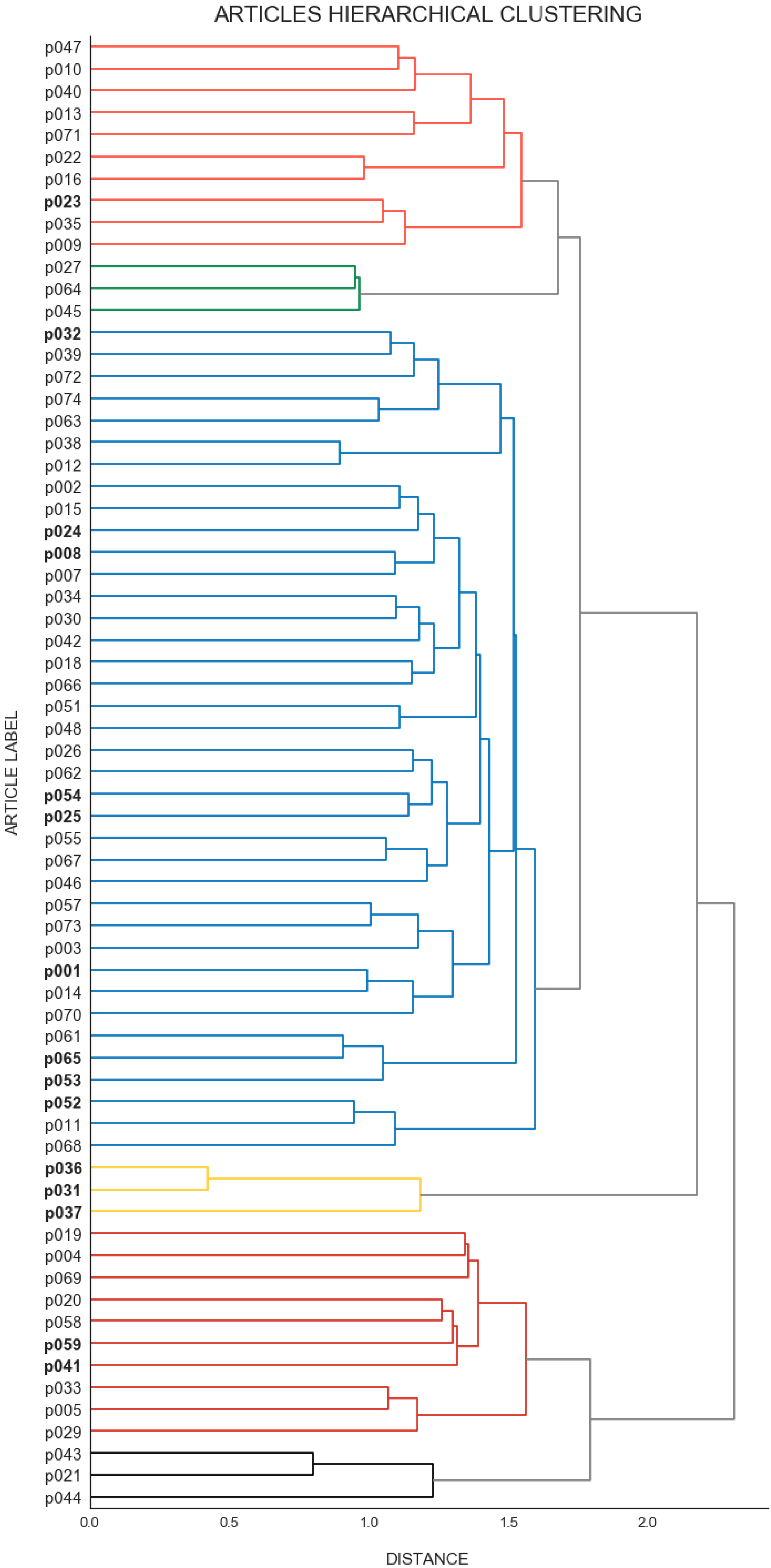

Figure 10: Hierarchical clustering

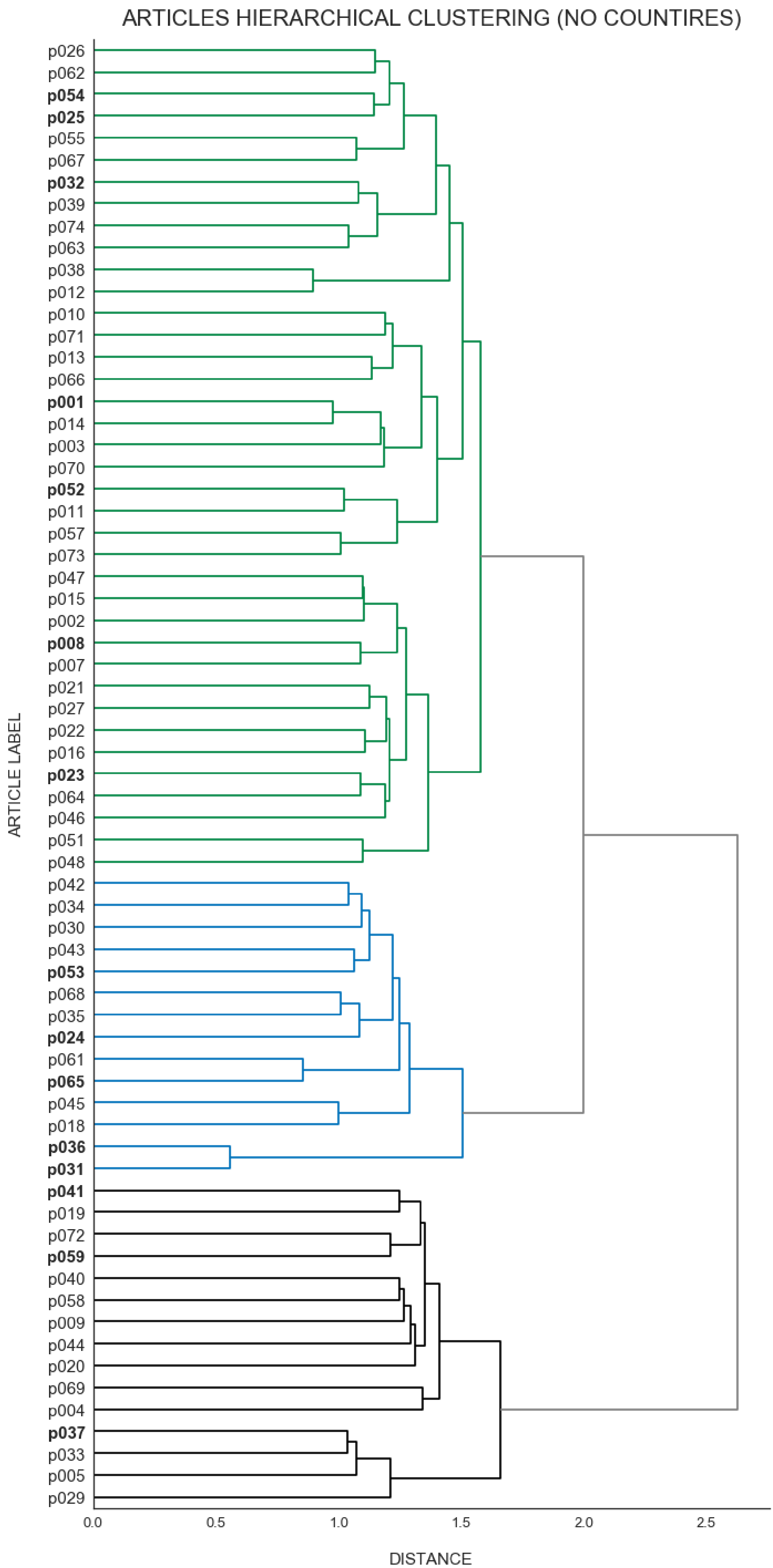

Figure 11: Hierarchical clustering, text without country names

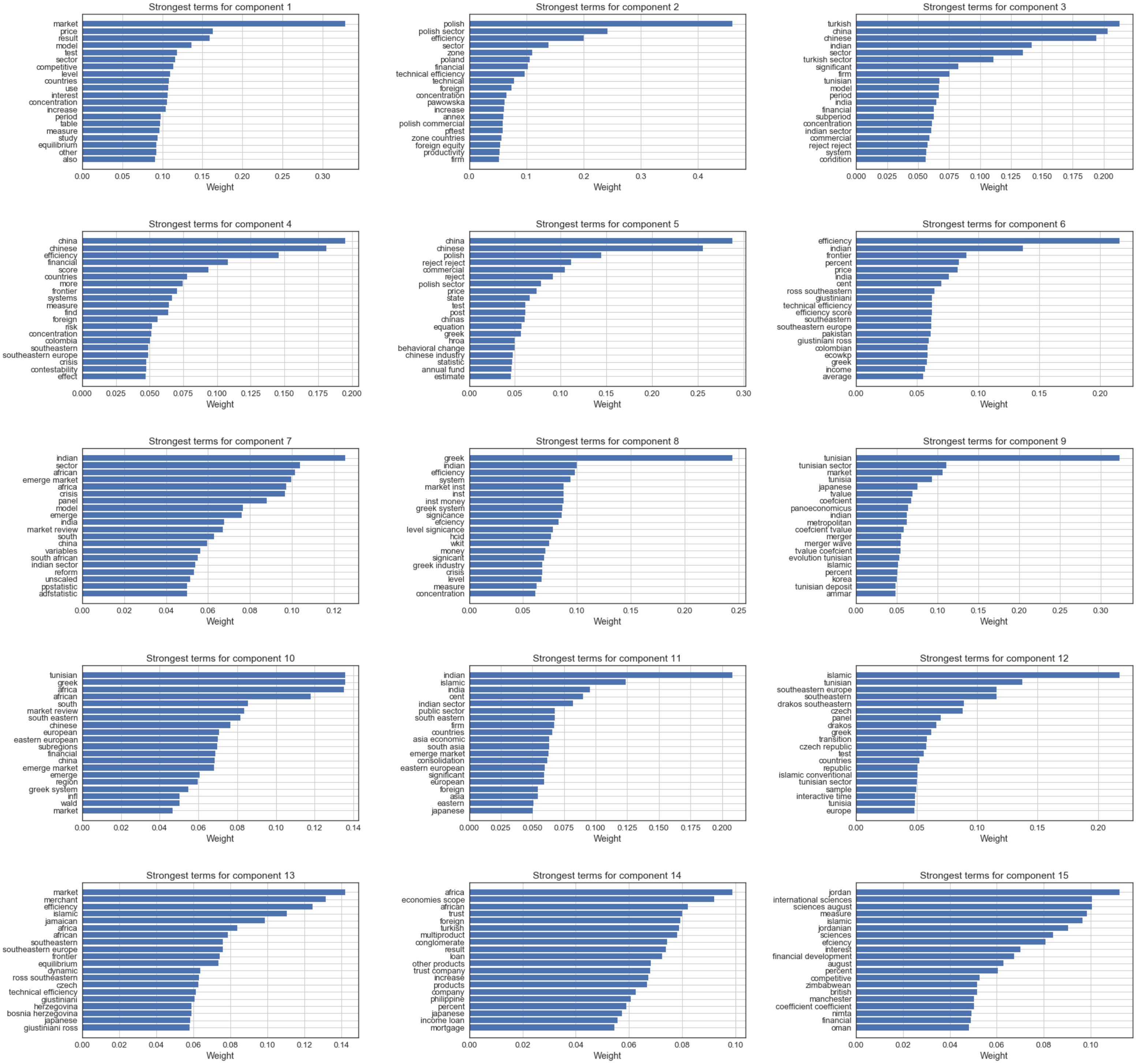

Figure 12: LSA components

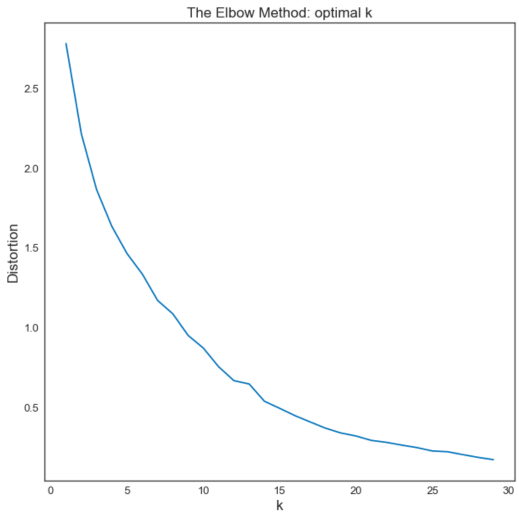

Figure 13: Elbow curve, k-means (no county names)

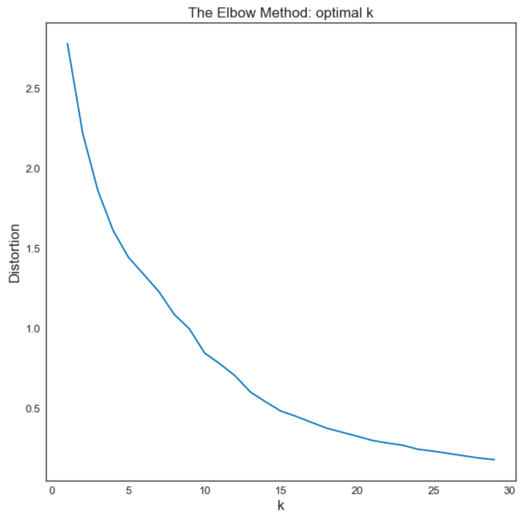

Figure 14: Elbow curve, k-means

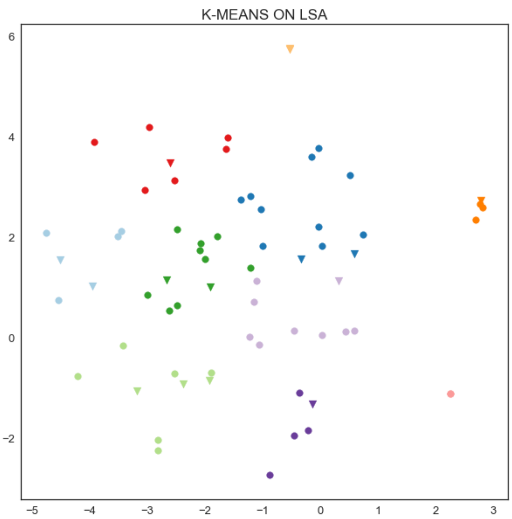

Figure 15: LSA K-means with 10 clusters

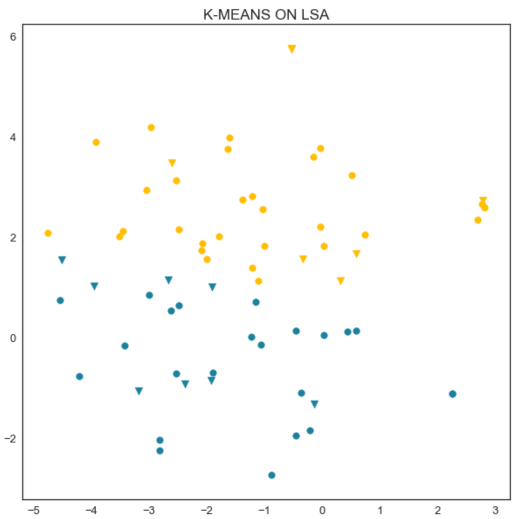

Figure 16: LSA K-means with 2 clusters