CPB Netherlands Bureau for Economic Policy Analysis

CPB Discussion Paper | 386

Can we measure banking sector competition robustly?

Andrei Dubovik
Natasha Kalara

Can we measure banking sector competition robustly?*

Andrei Dubovik[a,b] and Natasha Kalara[a]

[a]CPB Netherlands Bureau for Economic Policy Analysis
[b]Erasmus University Rotterdam

October 26, 2018

Abstract

We discuss existing measures of banking competition along with their advantages and disadvantages. For the Panzar and Rosse H-statistic, we further investigate the robustness of its estimates. Specifically, we consider how the estimates vary with respect to modelling and data choices along the following dimensions: i) bank types, ii) consolidation codes, iii) time periods, iv) outliers, and v) econometric models. We construct a robust H-statistic estimate following a modified DerSimonian and Laird procedure. We find that no robust conclusions can be drawn regarding the relative competitiveness of the banking industries in European countries, nor regarding the development of the aggregate level of competition in Europe over the past twenty years. This finding illustrates why there is little consensus about the H-statistic estimates despite numerous publications on the topic. Additionally, we check which dimensions are most important in driving the differences between the estimates and find that the choice of model specification plays the largest role.

JEL Codes: G21, L13, L80.

Keywords: banks, competition, Panzar and Rosse, robustness.

*The authors are thankful to Moniek Tulen for research assistance and to Jan Boone and authors' colleagues at CPB for valuable suggestions. The authors further note that the research has been started in 2016 and uses BvD Bankscope data, which have been available at the time at Erasmus University Rotterdam. As of 2017, this data source is discontinued but the authors have chosen to keep their initial analysis, because this paper investigates earlier studies of bank competition and those are predominantly based on Bankscope data.

1 Introduction

A healthy level of competition between banks is essential for financial systems. A lack of competition in the banking sector can have negative effects on financial stability (Beck, 2008) as well as on access to finance (Beck et al., 2004; Love and Martínez Pería, 2014). The latter is especially important in Europe, where bank credit is the most important source of external financing (Allen et al., 2004; Kalara and Zhang, 2018). An accurate measurement of competition is therefore necessary for risk monitoring and for designing optimal policies. A number of indicators that proxy bank competition have been proposed in the literature. With one exception, these indicators are general competition indicators that do not pertain specifically to the banking industry. Therefore, our theoretical discussion is general as well, and we focus on the banking industry primarily in the empirical part of our paper.

Broadly speaking, the competition indicators can be split into two groups: structural and conduct indicators. Bain (1956) argued that the structure of the market determines the conduct of the firms and as a result affects their performance. A corollary to this result is that concentration measures can be used as proxies for the level of competition. The Structure-Conduct-Performance (SCP) paradigm consequently formed the cornerstone of the Harvard School of antitrust analysis. However, while concentration measures are relatively easy to compute since they do not require data on the costs of firms, they are valid only insofar as the SCP paradigm holds. The latter has received criticism through the years from various researchers, e.g. from Stigler (1983), Bothwell et al. (1984), and Smirlock (1985) to name just a few. In particular, the theory of the contestable markets (Baumol, 1982) argues that even a highly concentrated market can be competitive if the costs of entry and exit are small. Additionally, Peltzman (1977) suggests that, in fact, concentration signals efficiency rather than collusion. Efficient firms will gain a higher market share and this will lead to a more concentrated market.

The weaknesses in the concentration measures led to an alternative assessment of competition. According to the New Empirical Industrial Organization (NEIO), competition can be directly measured through the conduct assessment of firms (Bresnahan, 1989). NEIO is not a radical departure from SCP but rather an expansion of the latter. Similar with SCP, the behaviour of firms in NEIO follows from structural assumptions. Contrary to SCP, the structural assumption are reacher and cover issues such as demand elasticities or market dynamics. Importantly, NEIO admits that some of the knowledge, e.g. firms' costs, is unobservable by the researcher and needs to be inferred from the data. Popular conduct measures of competition include the Lerner index, the Panzar and Rosse H-statistic (Panzar and Rosse, 1987), the Bresnahan-Lau indicator (Bresnahan, 1982; Lau, 1982) and the Boone indicator (Boone, 2008).

The NEIO measures come with several advantages over the concentration measures, albeit they are not without shortfalls themselves. These measures require a substantial amount of data on prices and costs, which in many cases, especially in developing and emerging markets, are difficult to find. Then, individual measures have their own specific limitations. Both the Lerner index and Boone indicator require the use of marginal costs, which are difficult to estimate for the banking industry. The H-statistic is possibly non-monotone with respect to the level of competition. The H-statistic and Boone indicator cannot, in practice, differentiate between the different markets of the banking sector, e.g. loans and deposits, and thus summarize only the total activities of banks. Finally, most NEIO competition indicators do not take into account the specifics of the banking industry such as the relationship between the Central Bank's interest rate and commercial banks' rents on loans and deposits. A detailed overview of these and other shortfalls can be found in Section 2.

NEIO instruments have been used to measure competition in a variety of industries, including the banking industry. While the latter received noticeable attention, at least judging by the number of publications, the conclusions have not been unanimous. For example, De Bandt and Davis (2000), Bikker and Haaf (2002), and Claessens and Laeven (2004) calculate the H-statistic for European countries, for the years 1992–1996, 1988–1998, and 1994–2001 respectively. De Bandt and Davis find that monopolistic competition prevails in Germany and France while Italy is characterized by monopoly power. Bikker and Haaf, and Claessens and Laeven, on the other hand, conclude that all these countries are characterized by monopolistic competition. The latter two papers, while agreeing qualitatively, obtain different quantitative results. Claessens and Laeven calculate the highest H-statistic for France among the three countries, while Bikker and Haaf calculate the highest H-statistic for Italy. There is further no clear agreement between the studies that use different competition indicators. For example, Van Leuvensteijn et al. (2011), using the Boone indicator, measure the competition in the loan markets of the five biggest Eurozone economies for the years 1994–2004. They find that Germany and Italy have the most competitive banking sectors, which contradicts the findings of Claessens and Laeven (2004), who find that France has the most competitive banking sector in roughly the same period. Many researchers have examined the level of bank competition for specific countries. For example, Toolsema (2002) investigates the degree of competition in the Dutch consumer credit market with the use of the Bresnahan-Lau method and finds that the market is characterized by perfect competition, a conclusion that contradicts the other studies mentioned in this paragraph.

Besides the anecdotal evidence presented above, there has been research aimed specifically at examining the correlation between various competition measures (Carbó et al., 2009; Beck and Casu, 2016). Those results show that the measures are only weakly correlated. Furthermore, the World Bank publishes two indicators of competition, namely the H-statistic and the Boone Indicator. The two indicators as published by the World Bank have a correlation of -0.12.[1] If policy makers were to base their conclusions on one or the other indicator, they would arrive at different, even opposing, policy recommendations. The same warning holds for the researchers who include the level of banking competition as a regressor in their models.

In summary, we observe that the estimates of bank competition differ across the papers that use the same competition measure as well as across the papers that use different competition measures. This observation leads to our research question. Can we reliably compare the level of bank competition between European countries? As a starting point in answering this general question, we focus on the H-statistic and on the differences in its estimates that result from the differences in the estimation methodologies employed in the existing literature. Specifically, we compute separate H-statistics for each combination of choices along the following dimensions: i) bank types, ii) consolidation codes, iii) time periods, iv) outliers, and v) economic models. We then use a modified DerSimonian and Laird (1986) procedure to compute a robust H-statistic together with a robust estimate of its variance.

We have chosen to do robustness tests on the H-statistic primarily because it have received the most attention in the literature. Besides that, there nothing specific about our choice of the H-statistic as the statistic to test for robustness. The same methodology can be used to test the Boone indicator or any other statistic for robustness; furthermore, it is possible to combine several statistics into one robust estimator. While such a comprehensive comparison falls outside the scope of our current research, we do think that additional insights can be gained from it. For instance, if the "between" statistics variance is smaller than the "within" variance, that would intuitively suggest that different statistics, while being imprecise themselves, measure the same latent variably, presumably the level of competition.

Individual H-statistic estimates are informative, that is they are significantly different from one another for at least some of the European countries. However, the respective standard errors do not account for the uncertainty associated with the choice of estimation methodology. In contrast, our robust H-statistic estimator, which explicitly accounts for the methodological uncertainty, is not informative. That is we cannot conclude that banking sector competitiveness differs significantly between any two European countries. This lack of robustness illustrates why there is little consensus in the literature regarding the H-statistic. We also measure how different choices influence the H-statistic estimates so as to understand, which factors contribute the most to the estimation differences. We conclude that model specification plays the largest role, which is in line with the critique

[1] All countries, years 2010–2015.

by Bikker et al. (2012), while the choice of time period is the second most important factor. The interpretation of these findings naturally differs. The importance of model specification is problematic, because it is likely that one model is correct and others are not. The importance of the time period simply highlights that there was noticeable change in the banking sector over the years. As for the remaining dimensions, data related factors such as what types of banks to use or how to handle different types of financial statements, while also important, play a less prominent role in explaining the differences in the H-statistic estimates. In view of these results, further theoretical rather than empirical research is needed to improve the accuracy of the measures of bank competition.

The paper is structured as follows. Section 2 provides an overview of commonly used bank competition measures along with a discussion of their shortcomings. Section 3 discusses the dataset. Section 4 introduces a methodology robust H-statistic estimator, and further explores what methodological differences are most crucial in explaining why the robust H-statistic estimator is not informative. Section 5 concludes.

2 Indicators Overview

In this section we enumerate existing competition indicators and briefly discuss their relative advantages and disadvantages. The list of existing indicators is given in Table 1. The table further lists relative disadvantages (issues) associated with each particular indicator. It should be noted that these issues, with a few exceptions, are not bank specific and arise equally well when measuring competition in other industries. Let us discuss these issues in turn.

Measures structure and not conduct. As has been discussed in the introduction, the SCP paradigm, developed by Bain (1956), argues that structural characteristics of a market influence the conduct of firms. A common corollary to this paradigm is that by measuring the concentration of a market we can measure the level of competition. However, such conclusion is problematic. Firstly, it takes a narrow view on what constitutes a market structure. Indeed, there are many factors besides concentration that can have an impact on the level of competition. If firms produce their goods in advance and then compete in prices, we obtain Cournot-type competition (Kreps and Scheinkman, 1983) and higher concentration implies lower competition. On the other hand, if the goods can be produced on demand, for example following an auctioning process, then Bertrand-type competition ensures and the level of concentration becomes an insufficient statistic to judge the level of competition. Barriers to entry and exit are still other factors that need to be included when measuring market structure, because concentrated markets can be competitive if these barriers are low (Baumol,

Table 1: Overview of Competition Indicators

Issue ("Yes" is bad, "no" is good)	C3	HHI	Lerner Index	H-statistic	Bresnahan-Lau	Boone Indicator	Interbank rate pass through
Measures structure and not conduct	Yes	Yes	No	No	No	No	No
Effectively, takes only larger banks into account	Yes	Yes	No	No	No	No	No
Additional data improves estimates	No	No	Yes	Yes	Yes	Yes	Yes
Requires knowledge of marginal costs	No	No	Yes	No	No	Yes	No
Possibly non-monotone w.r.t. the level of competition	Yes	Yes	No	Yes	No	No	No
Possibly inconsistent due to endogeneity	No	No	No	Yes	Yes	Yes	Yes
In practice, cannot split loans and deposits	No	No	No	Yes	No	Yes	No
Ignores specifics of the banking industry	Yes	Yes	Yes	Yes	Yes	Yes	No
Requires proprietary data	No	No	No	Yes	Yes	Yes	No
Estimation power relies on non-linear effects	No	No	No	No	Yes	No	No
Tally (less is better)	4	4	3	6	5	6	2

1982). Secondly, the SCP paradigm overlooks market dynamics. Strong competition can result in higher market share for more efficient firms, which would contradict the conclusion that high concentration means low competition (Demsetz, 1973; Peltzman, 1977; Boone, 2008). The competition indicators that measure the conduct of firms alleviate these concerns to some degree.

Effectively, takes only larger banks into account. By definition, the concentration ratio takes only the largest firms in the market into account. The number of firms considered usually varies between three and five. While the Herfindahl-Hirschman Index uses the market shares of all firms present in the market, the larger firms receive higher weights and so the shares of the smaller firms have a negligible influence on the final result (Bikker and Haaf, 2002). To the extent to which the behaviour of smaller firms reflects the level of competition in the market, ignoring smaller firms results in less efficient measures of competition.

Additional data improves estimates. Descriptive statistics like C3, HHI, Lerner Index (when marginal costs are available) are, in a certain sense, robust measures of competition. While exact geographic or product market definitions can be argued about, there is usually no disagreement about whether, say, HHI has been computed correctly. This is one of the reasons why these measures are widely used in practice, among others by the European Commission. In contrast, conduct measures are always econometric measures, and their accuracy improves with bigger and better data. The opposite is also true. Little data can result in poor measures, not only for the point estimates themselves but also the estimates of their variances. The rest of our paper is devoted precisely to this issue. Namely, how robust are the econometric estimates that we find in the existing literature?

Requires knowledge of marginal costs. Some of the competition measures require knowledge of marginal costs. Measuring marginal costs directly is difficult, and especially so in the banking industry, where there is a big portfolio of products that differ in their riskiness. In such industries proxies are used in place of marginal costs and thus the indicators that require the use of marginal costs become subject to additional errors. One possible approximation is to use average costs as in, e.g., Schaeck and Cihák (2014). However, this approximation is valid only as long as marginal costs are constant. There is big literature arguing that economies of scale play an important role in the banking sector (see, e.g., Beccalli et al., 2015), which suggests that the average costs approximation might not be valid in this industry. Another possibility is to explicitly estimate the production function and derive the marginal costs from there, as in, e.g., Van Leuvensteijn et al. (2011). This approach is potentially even more problematic. Firstly, there is an ongoing debate on how best to incorporate the specifics of the banking industry when estimating production functions (Hughes and Mester, 2013). Secondly, hardly any paper explicitly accounts for the errors-in-variables when estimated marginal costs are used as regressors in a computation of that or another competition indicator. It is therefore unclear whether any of the competition indicators that rely on marginal costs are in any way reliable.

Possibly non-monotone w.r.t. the level of competition. Non-monotonic competition measures make it difficult to compare various geographic markets and to investigate the development of markets over time. Therefore, when measuring competition, we want an indicator that is continuous and monotonically increasing with respect to the level of competition. Not all indicators have this property. Namely, non-monotonicity is a possible issue for the concentration indicators (Boone, 2001), and for the H-statistic (Shaffer, 1983, 2004). Still, the H-statistic is often treated in empirical work as though it is a monotonic measure (Bikker et al., 2012). In this paper we primarily focus on the robustness of the H-statistic estimates, and in doing so we treat this measure as monotonic as well, but we are aware that non-monotonicity is a potential issue.

Possibly inconsistent due to endogeneity. All competition indicators that are based on econometric estimates are subject to possible endogeneity. For example, Schaeck and Cihák (2014) argue that the Boone indicator is endogenous, because more fragile banks may indulge in riskier behaviour and the latter can be viewed as a sign of increased competition. Another example is Apergis et al. (2016), who point at the possible endogeneity when estimating the H-statistic due to the exclusion of common covariates. In this paper we do not address the robustness of the H-statistic to the choice of the estimation procedure, that is we do not compare OLS with the more endogeneity-robust estimation methods. However, our approach allows for inclusion of this type of robustness checks.

In practice, cannot split loans and deposits. All competition measures that require the knowledge of revenues or profits cannot, in practice, differentiate between loans and deposits markets, because such granular data are not available. This property is a disadvantage, because there is no prior reason to assume that the competitive situation in the loans and deposits markets is the same. For example, there is evidence that distance matters for competition for loans (Degryse and Ongena, 2005), while it seems unimportant for competition for deposits, which would imply that the same banks have higher market power in the loans market than in the deposits market. Let us also note that in the case of the Boone indicator, market shares can be substituted in place of profits so as to compute the indicator for specific product markets, as is argued in Van Leuvensteijn et al. (2011).

Ignores specifics of the banking industry. All but one competition indicators are taken directly from the general IO theory. They do not account for the particularities of the industry under consideration. This generality can be problematic in any industry, e.g. not accounting for the localized nature of competition in retail industries can both underestimate and overestimate the level of competition (Dubovik, 2018). Arguably, however, not accounting for industry specifics is most problematic when it comes to the banking industry, where factors such as risk taking and liquidity transformation play pivotal roles. Only the interest rate pass-through indicator originates in the theoretical banking literature, starting with the Monti-Klein model, albeit this indicator also misses some features of the industry. There is empirical evidence that the pass-through indicator is positively correlated with other competition measures (Kopecky and van Hoose, 2012; Van Leuvensteijn et al., 2013).

Requires proprietary data. Open data on financial institutions becomes more widely available. For instance, ECB publishes aggregated balance sheets and profit and loss statements of banks. Availability of open data helps in replicating and critically assessing earlier studies, and results in more robust statistics. Therefore, whenever an indicator requires the use of proprietary data such as Bankscope or Bank Focus from Bureau van Dijk, we count it as a disadvantage.

Estimation power relies on non-linear effects. With the exception of Bresnahan-Lau indicator, all competition measures are either descriptive statistics or can be estimated using a linear regression. The Bresnahan-Lau depends on the nonlinearities in the demand function, i.e. the estimation is not possible with a linear demand. This is a known disadvantage of this statistic, see, e.g. Shaffer (2004). Second-order effects are likely to be less visible in noisy data, and therefore the Bresnahan-Lau statistic is likely to be less robust.

3 Data Overview

As we mention earlier, we have chosen to focus on the H-statistic for our robustness tests. Most papers published to date use the Bankscope database from Bureau van Dijk to estimate the H-statistic. We therefore do the same. We would like to note that, given the Bankscope is discontinued, future papers are likely to use its successor, Bank Focus. Another possible dataset is the transparency exercises conducted by the European Banking Authority. While the dataset is relatively short, it is also publicly available and that can facilitate academic dialogue.

We restrict our analysis to the years 1995–2014 and to the European Union countries with the omission of Malta and Ireland. All the variables used are in million USD and have not been adjusted for inflation.

Different data cleaning procedures can potentially lead to different H-statistic estimates. Since investigating this sensitivity is a focus of our analysis, we postpone the customary discussion of the data cleaning procedures till the next section.

Table 2: Bankscope Variable Labels

Variable	Label
v_{11350}	Total Assets
v_{10040}	Gross Interest and Dividend Income
v_{10140}	Total Non-Interest Operating Income
v_{10070}	Total Interest Expense
v_{11650}	Total Funding
v_{10150}	Personnel Expenses
v_{10160}	Other Operating Expenses
v_{11840}	Total Equity
v_{11090}	Net Loans

Notes: the table lists all Bankscope data variables that are used in the analysis.

Based on the Bankscope data, we construct the following variables:

$$\begin{aligned} A &= v_{11350} && \text{(total assets)}, \\ R &= v_{10040} + v_{10140} && \text{(total revenue)}, \\ r &= v_{10070}/v_{11650} && \text{(funding costs)}, \\ w &= v_{10150}/A && \text{(personnel costs)}, \\ k &= v_{10160}/A && \text{(capital costs)}, \\ e &= v_{11840}/A && \text{(equity share)}, \\ l &= v_{11090}/A && \text{(loans share)}, \end{aligned}$$

where the respective labels are given in Table 2.

In total, we test 48 different estimation scenarii, where the data used for estimation can vary from scenario to scenario. Consequently, overall descriptive statistics do not exist in our case. Instead, for any given descriptive statistic Table 3 gives the range of values that this statistic takes across all scenarii.

4 Analysis

H-statistic can be estimated in different ways. Let Ω_i denote possible choices across dimension i. For example, Ω_i can be model choices or time period choices. As we will see below, there are natural definitions of dimensions so that the dimensions are independent. That is to say that any combination of choices is permissible. Then $\Omega = \prod_i \Omega_i$ is the set of all possible scenarii in which H-statistic can be estimated.

We consider 5 dimensions: 1) bank types, 2) consolidation codes, 3) time periods, 4) outliers, 5) econometric models. Mostly, we populate each dimension with choices from the existing literature. Where appropriate we

Table 3: Descriptive Statistics

Variable	Count	Min	Max	Avg	S.d.
A (total assets)	3,577–65,647	1.14–31.1	118,617–3,126,269	2,602–45,962	7,204–214,710
R (total revenue)	3,577–65,647	0.046–1.36	4,724–145,386	144–1,840	374–8,203
r (funding costs)	3,577–65,647	0.000–0.005	0.054–103	0.015–0.056	0.007–1.76
w (personnel costs)	3,577–65,647	0.000–0.001	0.053–0.533	0.013–0.015	0.005–0.022
k (capital costs)	3,577–65,647	0.000–0.001	0.088–2.08	0.010–0.017	0.006–0.035
e (equity share)	3,577–65,647	0.001–0.019	0.640–0.986	0.082–0.110	0.046–0.100
l (loans share)	3,577–65,647	0.000–0.033	0.926–1.00	0.489–0.599	0.161–0.263

Notes: in each column, the table gives the minimum and the maximum of the corresponding statistic across all scenarii; for example, there is a scenario with as few as 3,577 observations and a scenario with as many as 65,647 observations.

sometimes argue for additional choices not explicitly mentioned in the literature.

First, consider bank types. Some papers use all available banks in their analysis, namely commercial banks, cooperative banks, and savings banks (e.g., Claessens and Laeven, 2004; Bikker et al., 2012). The rationale is that all these banks compete on the same loans and deposits markets. Moreover cooperative banks and savings banks might constitute a substantial share of all banks in certain countries. Omitting specific types of banks leads therefore to less efficient estimates. However, a counterargument can also be given. Cooperative banks and savings banks might have a goal function that is different from profit maximization, in which case there is no theoretical basis for the H-statistic, and thus including these banks might lead to biased estimates. Consequently, some papers use only commercial banks (e.g., Andrieş and Căpraru, 2014). We define

$$\Omega_1 = \{\text{All banks},\ \text{Commercial banks}\}. \tag{1}$$

Second, consider consolidation codes. A bank may report consolidated statements of its financial position, in which case consolidated statements of this bank's subsidiaries are added to the financial statements of the bank itself, on an item per item basis. Alternatively, a bank may report unconsolidated statements, in which case the information on the subsidiaries is either disregarded or included as an aggregate item in the bank's statements.[2] Some banks report both types of financial statements.

Using only unconsolidated statements to estimate the H-statistic prevents double counting, albeit only partially as "Total assets" might still include the assets of the subsidiaries. The second advantage of using only unconsolidated statements comes from the observation that many banks have

[2] E.g. "Assets of subsidiaries" could be added to the parent's balance sheet statement, while the profits of the subsidiaries could be disregarded in the parent's profit and loss statement.

subsidiaries abroad. Whenever that is the case, the H-statistic estimates based on the consolidated statements will mix the competitive behaviour on the considered market with the competitive behaviour on some other markets. Using unconsolidated statements partially prevents this mixing issue. For these reasons some papers use only unconsolidated statements (e.g., De Bandt and Davis, 2000; Weill, 2013; Andrieş and Căpraru, 2014). The main disadvantage of using unconsolidated statements is the lack of data: most banks report consolidated statements but do not report unconsolidated ones. Consequently, H-statistic estimates will be inefficient. Many papers trade possible bias for higher efficiency and consider consolidated bank statements when available and unconsolidated bank statements otherwise (e.g., Claessens and Laeven, 2004). It would seem logical to consider unconsolidated statements first and consolidated second, but nobody does so. We follow the literature and define

$$\Omega_2 = \{\text{Uncons.},\ \text{Cons. + uncons.}\}. \tag{2}$$

Third, consider the time period. Every paper chooses a different time period that best suits its research question. For instance, De Bandt and Davis (2000) choose 1992–1996 to asses the banking competition on the eve of EMU, while Weill (2013) chooses 2002–2010 to assess the evolution of the competition following the tighter integration of the EU banking markets. Further, there is no consensus on what the optimal period length is. For instance, Weill (2013) uses cross-section estimates, while the World Bank uses panel estimates based on all available years, starting from 1985, till present day.[3] Too short a period will yield inefficient estimates. Too long a period will overlap with major changes in the banking industry and will also fail to yield an efficient estimate of the more recent competitive situation. We summarize the existing approaches by looking the last 5, 10, and 20 years. Namely,

$$\Omega_3 = \{2010\text{–}2014,\ 2005\text{–}2014,\ 1995\text{–}2014\}. \tag{3}$$

Fourth, consider outliers. Most papers routinely drop the bottom 0.5% and the top 0.5% of all variables. In some countries this approach leads to mistakes. For instance, the Netherlands has 3 major banks, with ING being the largest. Suppose the estimation period is three years. If ING has highest assets in one year, highest interest income in another year, and maybe highest equity in yet third year, then these three years will be deleted, essentially removing the largest Dutch bank before estimating the competition in the Dutch banking sector. We therefore also consider an estimation option where no outliers are removed. So,

$$\Omega_4 = \{\text{Remove outl.},\ \text{Keep outl.}\}. \tag{4}$$

[3]We have explicitly clarified this point with the World Bank.

Finally, consider econometric models. Most papers use the following specification:

$$\ln R_{it} = \alpha_i + \beta_t + \gamma_r \ln r_{it} + \gamma_w \ln w_{it} + \gamma_k \ln k_{it} + \\ + \delta_e \ln e_{it} + \delta_l \ln l_{it} + \delta_A \ln A_{it} + \epsilon_{it}, \quad (5)$$

where i is bank, t is year, and α, β are respectively bank and time fixed effects. The H-statistic is then estimated as:

$$H = \hat{\gamma}_r + \hat{\gamma}_w + \hat{\gamma}_k. \quad (6)$$

The existing papers differ in what control variables they include, i.e. $\ln e$ and $\ln l$ or something else. They also differ in whether the dependent variable is $\ln R$ or $\ln(R/A)$, but that does not matter for the H-statistic as long as $\ln A$ is one of the control variables. If that is the case, then the specification is called scaled. However, including $\ln A$ as a control variable is not theoretically sound. By definition, the Panzar-Rosse H-statistic is the sum of revenue elasticities with respect to factor prices. If $\ln A$ is included as a control variable, or equivalently, if $\ln(R/A)$ is used as the dependent variable, then what gets estimated is the sum of price elasticities with respect to factor prices. The latter is not theoretically related to the degree of competition. Bikker et al. (2012) explore these arguments further and suggest that an unscaled model should be estimated instead:

$$\ln R_{it} = \alpha_i + \beta_t + \gamma_r \ln r_{it} + \gamma_w \ln w_{it} + \gamma_k \ln k_{it} + \\ + \delta_e \ln e_{it} + \delta_l \ln l_{it} + \epsilon_{it}. \quad (7)$$

Whether estimating the unscaled model alleviates the aforementioned concerns is still unclear, because the bank fixed effects from the unscaled model will be strongly correlated with $\ln A$. In any case, we include both specifications in our robustness tests:

$$\Omega_5 = \{\text{Scaled}, \ \text{Unscaled}\}. \quad (8)$$

As mentioned earlier, we define $\Omega = \prod_{i=1}^{5} \Omega_i$, which gives 48 possible estimation scenarii ($|\Omega| = 48$).

We ask two question. First, can we compute an H-statistic which is in a certain sense robust? Second, which of the data cleaning and modeling choices matters most?

4.1 Robust H-statistic

One of the more common approaches in meta-analysis of summarizing the results of independent studies is DerSimonian and Laird (1986) method, DSL for brevity, that is based on the random effects model. We will base our robust H-statistic on this method. However, DSL and other common meta-analysis methods are not directly applicable to our case due to the dependency between our scenarii. If we do not account for this dependency, we will grossly underestimate the variance of any robust estimator. Indeed, suppose we duplicate all our scenarii by considering an extra dimension with two dummy choices. In this case we want the variance of our robust estimator to remain the same. However, if independence is assumed, as it is in DSL, then the variance will decrease twofold. To resolve this issue we introduce a correction to DSL method. Namely, we assume that extra scenarii do not add any extra information that can improve estimates, and so we multiply the variance of the DSL estimator by the total number of scenarii. This modification that we propose is ad-hoc but the rigorous alternative is to allow for a fully flexible correlation between 48 scenarii, which would imply estimating 1128 coefficients and is thus infeasible. And if we estimate a random effects model with a restricted correlation matrix, then it is likely that we will be artificially lowering the variance of our robust estimator due to implicit independence assumptions.

If we choose a specific scenario $\omega \in \Omega$, we can compute the corresponding H-statistic $H(\omega)$ and its variance $\mathrm{Var}(H(\omega))$. Following DSL method and our proposed correction, we can then compute a robust H-statistic H^* and its variance $\mathrm{Var}(H^*)$ as follows (see, e.g., Normand, 1999). Let

$$v_\omega = \frac{1}{\mathrm{Var}(H(\omega))}, \quad \bar{H} = \frac{\sum_\omega v_\omega H(\omega)}{\sum_\omega v_\omega}, \quad Q = \sum_\omega (H_\omega - \bar{H})^2, \tag{9}$$

$$\tau^2 = \max\left(0, \frac{Q-(n-1)}{\sum_\omega v_\omega - \frac{\sum_\omega v_\omega^2}{\sum_\omega v_\omega}}\right), \quad v_\omega^* = \frac{1}{\mathrm{Var}(H(\omega)) + \tau^2}. \tag{10}$$

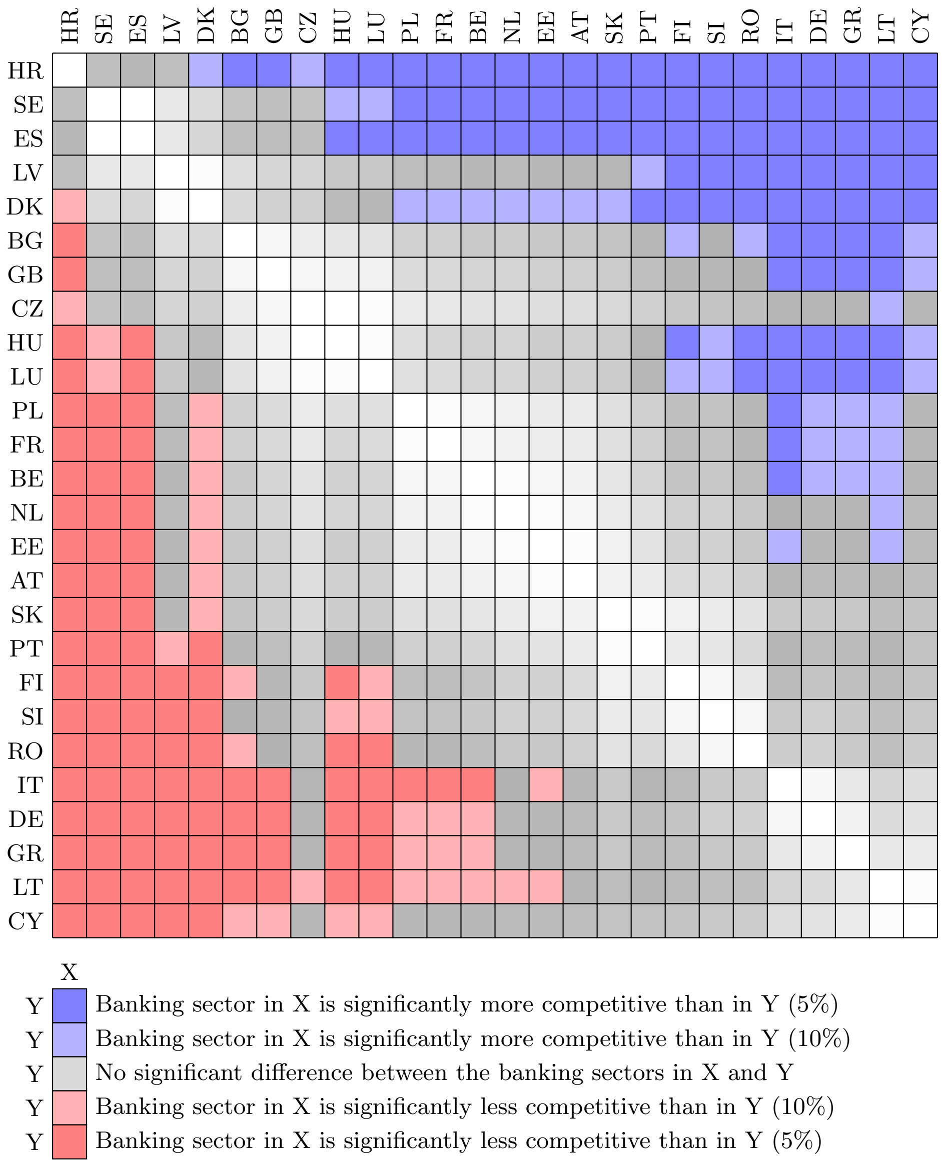
Then

$$H^* = \frac{\sum_\omega v_\omega^* H(\omega)}{\sum_\omega v_\omega^*}, \quad \mathrm{Var}(H^*) = \frac{N}{\sum_\omega v_\omega^*} \text{ (corrected variance).} \tag{11}$$

To put our consequent results into perspective, let us first consider the World Bank estimates of the H-statistic. The World Bank publishes only point estimates, but at our request they have provided us with the estimates of the variances as well. Fig. 1 presents a cross-country comparison that is based on the World Bank estimates. While for many countries there are no significant differences between them, a number of conclusions can nevertheless be drawn. For example, the banking sector in the Netherlands is significantly more competitive than the one in Spain. Having said that, there are two substantial problems with Fig. 1.

Firstly, as we do not fix the hypothesis we are testing in advance but rather choose significant differences between countries a posteriori, we need to adjust p-values accordingly (multiple comparisons problem), which is not

Figure 1: World Bank's H-statistic, 2015 estimates



Notes: for each pair of countries, the figure shows whether the difference in the respective H-statistic estimates is significantly different from zero; the countries are ordered by the H-statistic; p-values are *not corrected* for multiple hypothesis testing.

done if Fig. 1. Secondly, with a different estimation scenario the results can change significantly, which indicates, as discussed earlier, that the variances computed by the World Bank might be inconsistent. We can address the latter problem by using our robust H-statistic. The corresponding results are shown in Fig. 2. As can be seen from the figure, no conclusions can be drawn regarding the relative levels of the banking sector competitiveness in European countries. This finding illustrates why there is little consensus regarding the corresponding values of the H-statistic despite the fact that there is close to 100 peer-reviewed publications that apply this statistic to the banking industry.

Various papers, for instance Andrieş and Căpraru (2014), study the development of the H-statistic over time. We can ask similar questions. For example, has the level of banking competition changed significantly in Europe in the past twenty years? To this end, we modify our procedure as follows.

Firstly, for any given year t we consider the same choices as before with the exception of the time period dimension. For the time period, we now use rolling windows of 3 and 5 years long (these lengths are common in the literature, see, e.g., Schaeck and Cihák 2014 or Beck et al. 2013). Formally,

$$\Omega_3(t) = \{[t-2,t],\, [t-4,t]\}. \tag{12}$$

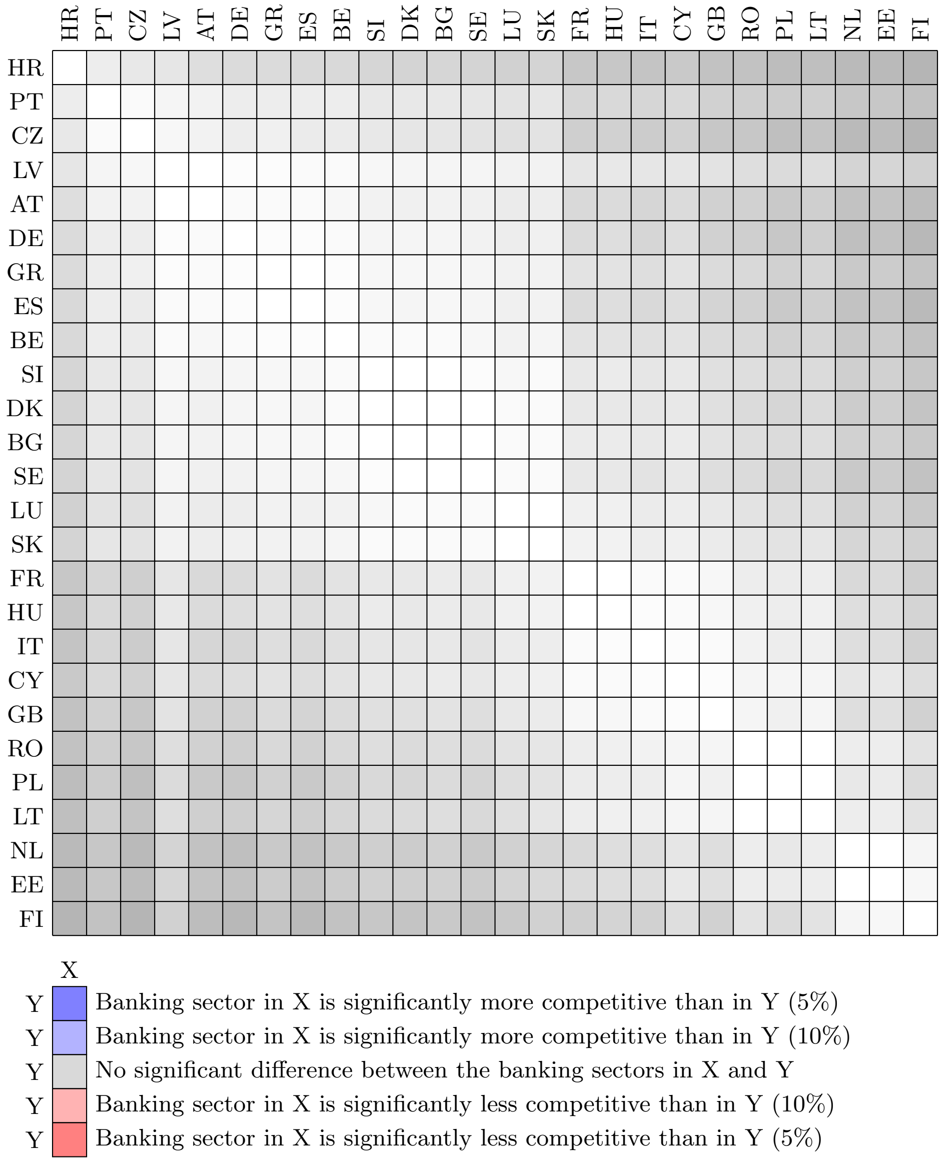
Secondly, for a given scenario ω we compute individual H-statistics $H_k(\omega)$ for every country k. We then compute a European aggregate $H(\omega)$ using the original DSL method, that is, without out proposed information correction. In other words, we assume that every extra country brings additional independent information about the level of banking competition in Europe. If anything, we therefore underestimate the variance of the aggregate H-statistic.

Thirdly, having defined the aggregate H-statistic for a given year and a given scenario, we employ our original procedure to compute a robust aggregate H-statistic for that year over all scenarii. The results of these calculations for years 1995–2014 are presented in Fig. 3. As can be seen from the figure, if we employ our robust H-statistic, then no conclusions can be drawn regarding the development of banking competition in Europe over the past twenty years.

4.2 Choice Importance Classification

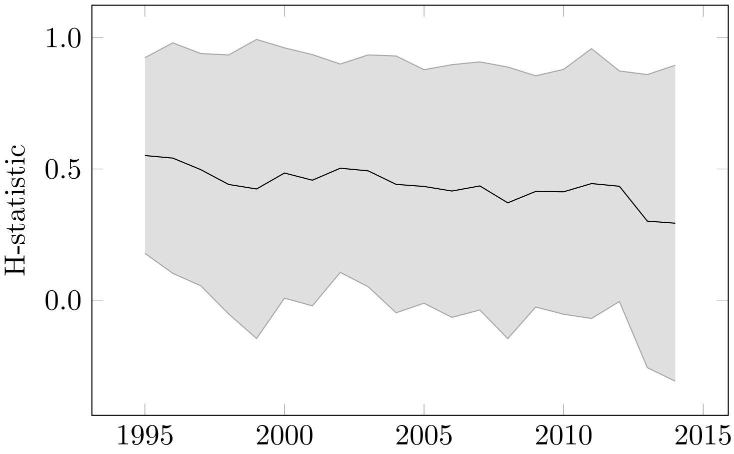
We have seen that if we treat all scenarii as equally likely to deliver consistent estimates, then it is not possible to compare the competitiveness of the banking industry across European countries. If we want to arrive at more efficient estimates, then we need to do further research on what choices are more theoretically and empirically sound and what choices are less sound. However, where is it best to start? Do we need to put more effort into the

Figure 2: Robust H-statistic



Notes: for each pair of countries, the figure shows whether the difference in the respective H-statistic estimates is significantly different from zero; the countries are ordered by the H-statistic; p-values are robust to different estimation scenarii but are *not corrected* for multiple hypothesis testing.

Figure 3: Robust H-statistic for EU-15



Notes: the confidence interval is at 95% level; p-values are robust to different estimation scenarii.

ownership structure of banks so as to figure out which consolidation codes are best to use? Or should we instead investigate how the estimation window influences the results and whether it is possible to choose an estimation window that is optimal in some sense? Answering these questions can help guide future research effort.

We can approach this puzzle formally by measuring how different choices influence the H-statistic estimates. Let K denote the set of European countries that we consider, and let $H_k(\omega)$ denote the H-statistic computed for country k, with $k \in K$. For any two scenarii a and b we define the distance between them as an L2 distance over the corresponding cross-country estimates of the H-statistic:

$$d(a, b) = \sqrt{\frac{1}{|K|} \sum_{k \in K} (H_k(a) - H_k(b))^2}. \tag{13}$$

For any $A \subset \Omega$ we define the diameter of A, denoted by $s(A)$, as follows:

$$s(A) = \max_{a \in A, b \in A} d(a, b). \tag{14}$$

That is, if we have a group of scenarii A, then $s(A)$ measures how much uncertainty over the H-statistic remains among those scenarii.

For any $a \in \Omega_j$ we define the corresponding cylinder set:

$$\xi_j(a) = \{\omega \in \Omega : \omega_j = a\}. \tag{15}$$

That is, $\xi(a)$ is a collection of all scenarii, where the choice along dimension j is fixed at a. E.g., if $j = 2$ and $a =$ Uncons., than $\xi_j(a)$ consists of all the estimation scenarii where only unconsolidated statements are used.

For any $A \subset \Omega$ and for any dimension j we define the unimportance of the choices along that dimension, denoted by $\phi_j(A)$, as the average diameter of the subsets generated by those choices and restricted to A:

$$\phi_j(A) = \frac{1}{|\Omega_j|} \sum_{\omega \in \Omega_j} s(A \cap \xi_j(\omega)). \tag{16}$$

If $\phi_j(A) < \phi_i(A)$, then splitting the scenarii from A along dimension j results in subsets with smaller uncertainty over the H-statistic than if we split those scenarii along dimension i. Hence, the choices along dimension j are more important.

We proceed to classify the importance of all the choices as follows: 1) we start with $A = \Omega$; 2) we then choose $j = \arg\min_i \phi_i(A)$, the choices along dimension j are thus the most important ones; 3) then, for each $a \in \Omega_j$ we set $B_j = A \cap \xi_j(a)$; 4) finally, for each B_j we search for the next most important dimension, that is we repeat the procedure starting from 2) with $A = B_j$. The results are given in Fig. 5.

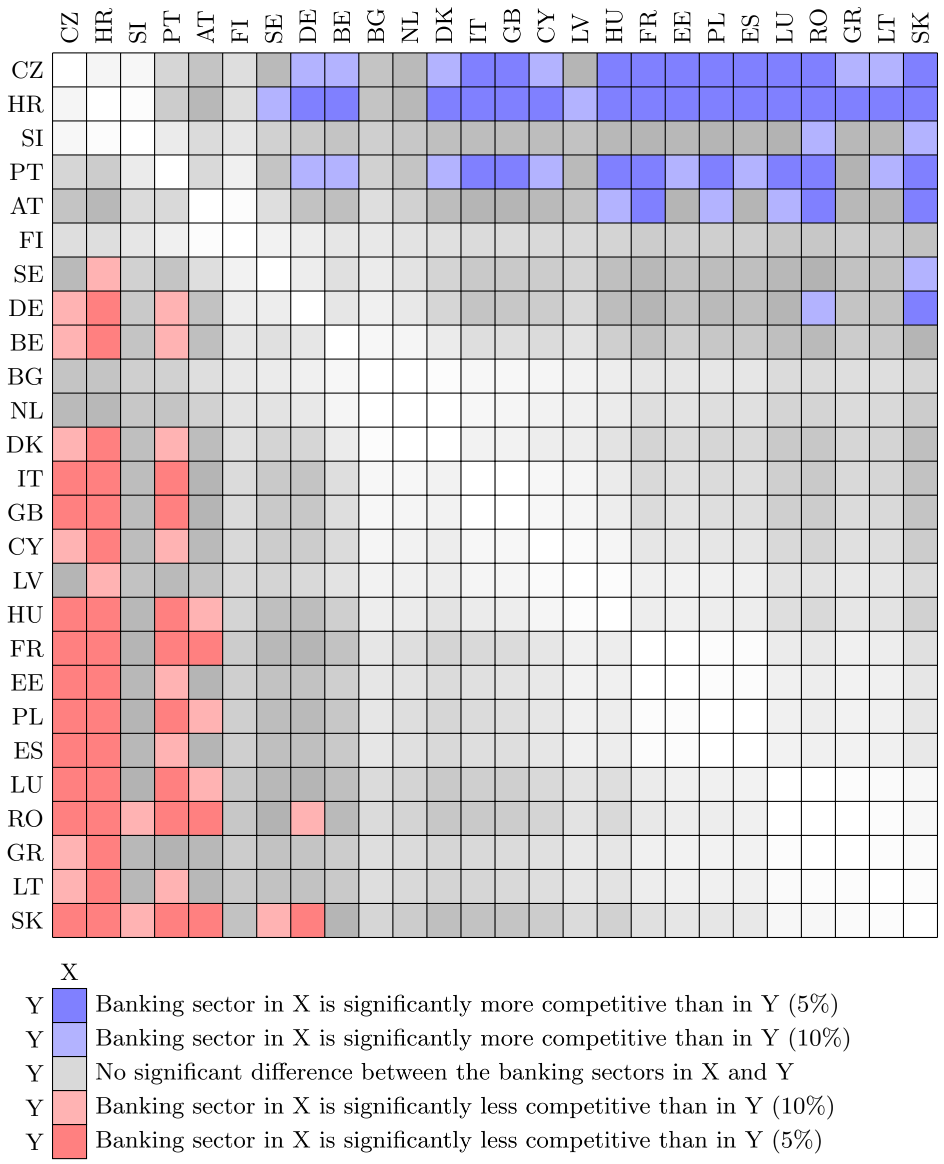
We find that model specification plays the largest role. This finding corroborates Bikker et al. (2012), who were one of the first to suggest that scaled and unscaled models lead to substantially different results. After model specification, estimation period plays the most important role. Then sometimes financial statement types and sometimes bank types are most important. Handling of outliers is least important.

In this paper we do not aim to judge which estimation choices are the correct ones. Nevertheless, it can be informative to see just how much we could say about banking competition were we to restrict the set of scenarii with specific estimation choices. Fig. 4 shows a cross-country comparison using robust H-statistic estimates, when the scenarii are restricted to the unscaled specification and the period 2010–2014 (thus, we fix the two dimensions that introduce most uncertainty). We obtain some significant differences across the countries. However, the p-values have not been corrected for multiple hypothesis testing, and were we to do so, the results would be less significant still. Secondly, regarding the comparison of the Dutch banking sector to that of other countries, no significant conclusions can be drawn in the first place.

5 Discussion

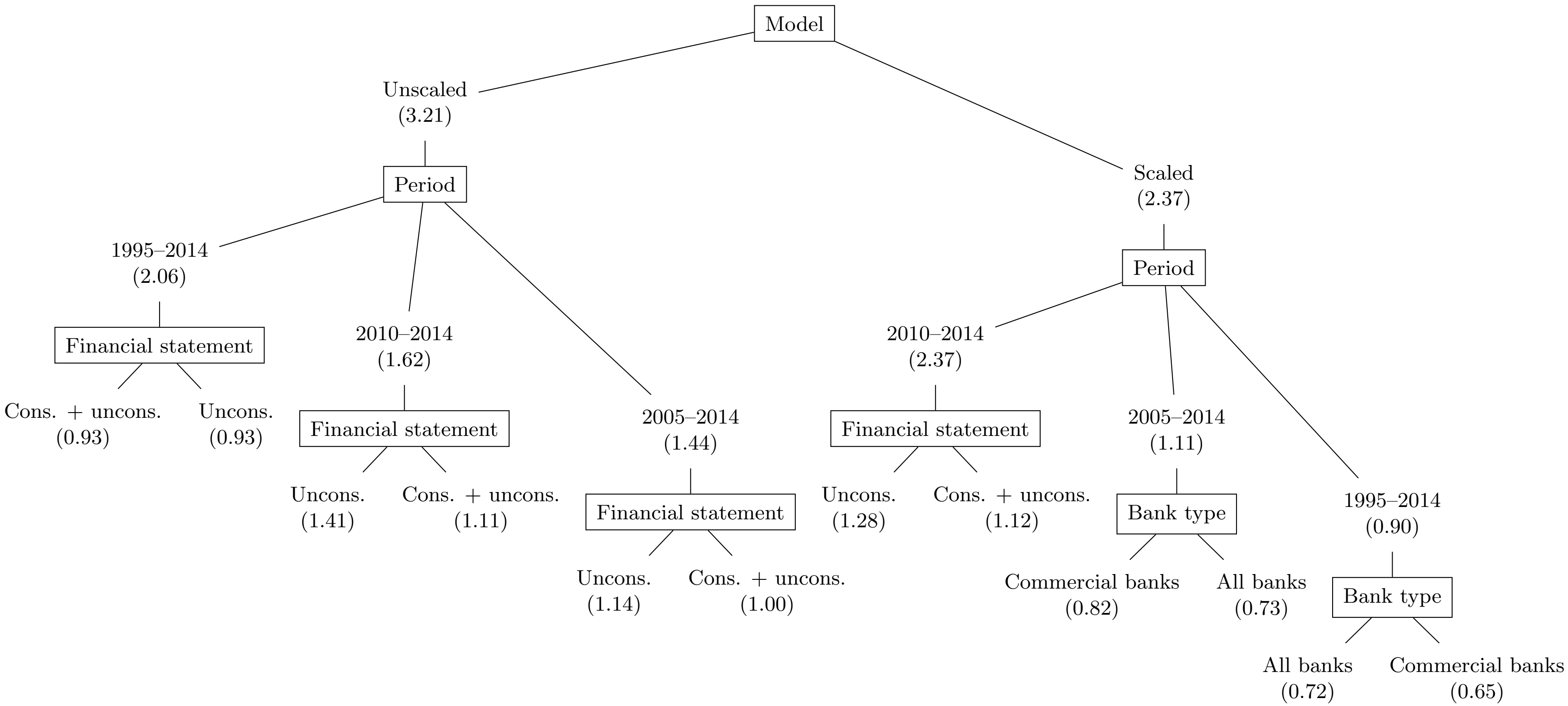
There are some 7 measures of competition that has been computed for the banking industry. From our understanding of the literature, the H-statistic is the most commonly used measure, albeit it is also one of the most problematic. We have studied the H-statistic measure in detail to understand how robust it is, and what are the most crucial factors that influence its estimates. We have found the measure to be not robust and have concluded

Figure 4: Restricted Robust H-statistic



Notes: for each pair of countries, the figure shows whether the difference in the respective H-statistic estimates is significantly different from zero; the countries are ordered by the H-statistic; p-values are robust to different estimation scenarii but are *not corrected* for multiple hypothesis testing; the scenarii considered are restricted to the unscaled specification and the period 2010–2014.

Figure 5: What Influences H-statistic Estimates?



Notes: the figure presents a decision tree for the H-statistic estimates; framed boxes are dimensions, unframed boxes are choices in the respective dimensions; the number accompanying a choice is the diameter of the following subtree; only first 3 levels of the tree are shown; top level is the most informative in explaining the resulting differences in the estimates, second level is less informative, and so on.

that no comparisons between the European countries can be made regarding the competitiveness of their banking sectors. We have also found that the most crucial factor that influences the results, and hence the first factor that needs to be studied further, is the choice of model specification.

This later finding suggests that to better our understanding of the banking competition in European countries we need not pursue further empirical research, but rather further theoretical research so as to be able to answer with a degree of certainty which models are best suited to measure competition. For example, a model of banking competition can be developed, flexible enough so that it can be calibrated to the observed distribution of bank sizes and numbers across various countries. This model further needs to allow for various degrees of banking competition. Then existing approaches to measure competition and their consistency can be rigorously compared using a simulations study. Such an exercise would reduce the discussion about which empirical paper is best at measuring competition to the discussion about which IO model is best at describing bank interactions. The latter discussion is more fruitful because any theory model yields additional testable predictions, besides those related to competition, and thus the existing data can give more information.

References

Allen, F., Chui, M. K., and Maddaloni, A. (2004). Financial systems in europe, the usa, and asia. *Oxford Review of Economic Policy*, 20(4):490–508.

Andrieş, A. M. and Căpraru, B. (2014). The nexus between competition and efficiency: The european banking industries experience. *International Business Review*, 23(3):566–579.

Apergis, N., Fafaliou, I., and Polemis, M. L. (2016). New evidence on assessing the level of competition in the european union banking sector: A panel data approach. *International Business Review*, 25:395–407.

Bain, J. S. (1956). *Barriers to new competition*. Harvard University Press.

Baumol, W. J. (1982). Contestable markets: An uprising in the theory of industry structure. *The American Economic Review*, 72(1):1–15.

Beccalli, E., Anolli, M., and Borello, G. (2015). Are european banks too big? evidence on economies of scale. *Journal of Banking & Finance*, 58:232–246.

Beck, T. (2008). Bank competition and financial stability: Friends or foes? The World Bank Policy Research Working Paper no. 4656.

Beck, T. and Casu, B. (2016). *The Palgrave Handbook of European Banking*. Palgrave Macmillan.

Beck, T., De Jonghe, O., and Schepens, G. (2013). Bank competition and stability: Cross-country heterogeneity. *Journal of Financial Intermediation*, 22:218–244.

Beck, T., Demirgüç-Kunt, A., and Maksimovic, V. (2004). Bank competition and access to finance: International evidence. *Journal of Money, Credit and Banking*, 36(3):627–648.

Bikker, J. and Haaf, K. (2002). Competition, concentration and their relationship: An empirical analysis of the banking industry. *Journal of Banking & Finance*, 26(11):2191–2214.

Bikker, J. A., Shaffer, S., and Spierdijk, L. (2012). Assessing competition with the panzar-rosse model: The role of scale, costs, and equilibrium. *The Review of Economics and Statistics*, 94(4):1025–1044.

Boone, J. (2001). Intensity of competition and the incentive to innovate. *International Journal of Industrial Organization*, 19:705–726.

Boone, J. (2008). A new way to measure competition. *The Economic Journal*, 118(531):1245–1261.

Bothwell, J. L., Cooley, T. F., and Hall, T. E. (1984). A new view of the market structure-performance debate. *The Journal of Industrial Economics*, 32(4):397–417.

Bresnahan, T. (1989). Empirical studies of industries with market power. In *Handbook of industrial organization*, volume 3, pages 1011–1057. North-Holland.

Bresnahan, T. F. (1982). The oligopoly solution concept is identified. *Economics Letters*, 10(1):87–92.

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*, 28(1):115–134.

Claessens, S. and Laeven, L. (2004). What drives bank competition? some international evidence. *Journal of Money, credit, and Banking*, 36(3):563–583.

De Bandt, O. and Davis, E. P. (2000). Competition, contestability and market structure in european banking sectors on the eve of emu. *Journal of Banking & Finance*, 24(6):1045–1066.

Degryse, H. and Ongena, S. (2005). Distance, lending relationships, and competition. *The Journal of Finance*, 60:231–266.

Demsetz, H. (1973). Industry structure, market rivalry, and public policy. *The Journal of Law and Economics*, 16(1):1–9.

DerSimonian, R. and Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7:177–188.

Dubovik, A. (2018). Mergers on networks. Working Paper.

Hughes, J. P. and Mester, L. J. (2013). Who said large banks don't experience scale economies? evidence from a risk-return-driven cost function. *Journal of Financial Intermediation*, 22:559–585.

Kalara, N. and Zhang, L. (2018). The changing landscape of firm financing in europe, the united states and japan. Working Paper.

Kopecky, K. J. and van Hoose, D. D. (2012). Imperfect competition in bank retail markets, deposit and loan rate dynamics, and incomplete pass through. *Journal of Money, Credit and Banking*, 44:1185–1205.

Kreps, D. M. and Scheinkman, J. A. (1983). Quantity precommitment and bertrand competition yield cournot outcomes. *The Bell Journal of Economics*, 14:326–337.

Lau, L. J. (1982). On identifying the degree of competitiveness from industry price and output data. *Economics Letters*, 10(1):93–99.

Love, I. and Martínez Pería, M. S. (2014). How bank competition affects firms' access to finance. *The World Bank Economic Review*, 29(3):431–448.

Normand, S.-L. T. (1999). Meta-analysis: formulating, evaluating, combining, and reporting. *Statistics in Medicine*, 18:321–359.

Panzar, J. C. and Rosse, J. N. (1987). Testing for "monopoly" equilibrium. *The Journal of Industrial Economics*, 35(4):443–456.

Peltzman, S. (1977). The gains and losses from industrial concentration. *The Journal of Law & Economics*, 20(2):229–263.

Schaeck, K. and Cihák, M. (2014). Competition, efficiency, and stability in banking. *Financial Management*, 43(1):215–241.

Shaffer, S. (1983). Non-structural measures of competition: Toward a synthesis of alternatives. *Economics Letters*, 12:349–353.

Shaffer, S. (2004). Patterns of competition in banking. *Journal of Economics and Business*, 56(4):287–313.

Smirlock, M. (1985). Evidence on the (non) relationship between concentration and profitability in banking. *Journal of Money, Credit and Banking*, 17(1):69–83.

Stigler, G. J. (1983). *The organization of industry.* The University of Chicago Press.

Toolsema, L. (2002). Competition in the dutch consumer credit market. *Journal of Banking & Finance*, 26(11):2215–2229.

Van Leuvensteijn, M., Bikker, J. A., van Rixtel, A. A., and Sørensen, C. K. (2011). A new approach to measuring competition in the loan markets of the euro area. *Applied Economics*, 43(23):3155–3167.

Van Leuvensteijn, M., Sørensen, C. K., Bikker, J. A., and van Rixtel, A. A. (2013). Impact of bank competition on the interest rate pass-through in the euro area. *Applied Economics*, 45(11):1359–1380.

Weill, L. (2013). Bank competition in the eu: How has it evolved? *Journal of International Financial Markets, Institutions & Money*, 26:100–112.

Publisher:

CPB Netherlands Bureau for Economic Policy Analysis
P.O. Box 80510 | 2508 GM The Hague
T (088) 984 60 00

November 2018