DEPARTMENT OF ECONOMICS AND FINANCE

SCHOOL OF BUSINESS AND ECONOMICS

UNIVERSITY OF CANTERBURY

CHRISTCHURCH, NEW ZEALAND

A Replication of "Bank Competition and Financial Stability: Much Ado About Nothing?" (*Journal of Economic Surveys*, 2016)

Samangi Bandaranayake
Kuntal K. Das
W. Robert Reed

WORKING PAPER

No. 18/2018

Department of Economics and Finance
School of Business
University of Canterbury
Private Bag 4800, Christchurch
New Zealand

WORKING PAPER No. 18/2018

A Replication of "Bank Competition and Financial Stability: Much Ado About Nothing?" (*Journal of Economic Surveys*, 2016)

Samangi Bandaranayake[1]
Kuntal K. Das[1]
W. Robert Reed[1†]

October 2018

Abstract: This study replicates Zigraiova and Havranek's (2016) meta-analysis of banking competition and financial stability. It performs three types of replications: a "Reproduction" replication where Z&H's data and code are verified to reproduce the results of their study; a "Repetition" replication where the studies used by Z&H are independently recoded and then re-analyzed; and an "Extension" replication, where 35 additional studies on banking competition and stability are examined to see if they confirm Z&H's findings. Our replication analyses confirm Z&H's main finding that competition in the banking sector is unrelated to financial stability. However, we do not confirm their finding of publication selection in the "Extension" replication, and both "Repetition" and "Extension" replications fail to confirm most of Z&H's findings about the effects of data, estimation, and study characteristics on estimated competition effects. While our "Repetition" replication produced a high match rate with Z&H's data, we found that the meta-regression analysis was substantively impacted by discrepancies between the original and recoded data. We conclude that meta-regression analyses should be viewed with caution. We also demonstrate that "best practice" estimates can be precarious and highly dependent on variable specification and how "best practice" is defined. They likewise should be viewed with caution. Our study highlights the value-added that replication brings to meta-analyses. While some of Z&H's results were found to not be robust, the fact that all three types of replications confirmed Z&H's main finding about competition and stability provides strong evidence that this finding can be regarded with confidence.

Keywords: Bank competition, financial stability, Bayesian model averaging, meta-analysis, publication selection, replication.

JEL Classifications: C41, G21, G28, L11

Acknowledgements: The authors wish to thank participants at the 2018 MAER-Net Colloquium in Melbourne, Australia for helpful comments and suggestions. Special thanks goes to Tomas Havranek for making his data and code available and answering queries about his research. This research was supported by funding from the Czech Science Foundation, Grant 18-02513S.

[1] Department of Economics and Finance, University of Canterbury, NEW ZEALAND
[†] Corresponding author is Bob Reed. His email address is bob.reed@canterbury.ac.nz.

I. INTRODUCTION

Meta-analysis is a tool for aggregating results across studies that are estimating the same or a similar "effect". It is increasingly popular as a research tool. However, there is of late some concern that meta-analysis does not provide the resolution that its originators had hoped. Hannah Rothstein, a prominent meta-analysis scholar, former Editor-in-Chief of the journal *Research Synthesis Methods* and past president of the Society for Research Synthesis Methodology, was recently quoted as saying that while she has not "lost faith" in meta-analysis, she has "changed her expectations": "We used to make meta-analyses as objective as possible. Now, we try to make them as transparent as possible. … Anyone who disagrees with a certain decision will have to be able to redo it and see if that has an influence on the results." (Vrieze, 2018).

Accordingly, this study replicates a recently published meta-analysis in this journal: "Bank Competition and Financial Stability: Much Ado about Nothing" (Zigraiova and Havranek, *Journal of Economic Surveys*, 2016). We choose this article/topic because the question of how competition in the banking sector affects financial stability is an important one, especially given the devastating consequences of the recent global financial crisis. Z&H provide a comprehensive analysis of the subject using frontline meta-analysis techniques, and subject their analysis to a wide variety of robustness checks. They come to a number of conclusions. Given the importance of this topic, it is worthwhile to determine the reliability of those conclusions.

A second motivation for this research is to examine the robustness of meta-analysis as a tool for aggregating empirical research. The data in a meta-analysis consist of estimates from a set of original studies, along with various characteristics of the studies and estimation equations. Assembling these data is a tedious, time-consuming process that involves evaluating a large number of studies and estimates. For example, Z&H's final sample consists of 598 estimates taken from 31 studies. Each of these needs to be manually "coded" according to a set of characteristics determined by the researcher. The large number of estimates and the manual nature of the coding make it relatively easy to make mistakes.

Further, coding involves a fair amount of subjectivity in terms of which studies to include, what characteristics to code, and which values to assign to various characteristics. And, of course, additional studies become available over time. Very little is known about how robust meta-analysis is to alternative treatments of the same subject. This study aims to expand our knowledge about this.

II. DESCRIPTION OF Z&H

The question that Z&H's meta-analysis addresses is, Does competition in the banking sector promote financial stability? Accordingly, they collect a sample of studies that all estimate something akin to the following specification:

(1) $$StabilityMeasure_{it} = \alpha + \beta CompetitionMeasure_{it} + \boldsymbol{X_{it}\theta} + \varepsilon_{it}$$

Studies in this sample consist of 21 journal articles, 9 working papers, and a book chapter (cf. APPENDIX 1). All were "published" during the period 2006-2014, where year of publication for unpublished papers equals the date listed on the respective working paper.

Z&H assign stability measures into 8 categories. The most common measure of stability is Z-score, a measure of bank-level risk. Increases in Z-score indicate a lower probability of default, and so are positively related to financial stability. The second most common measure of stability is a dummy variable that typically indicates a systemic banking crisis (if the data are national-level), or an individual bank failure (for bank-level data). This measure is negatively associated with financial stability. A third measure is the ratio of non-performing loans (*NPL*) where, again, this measure is negatively associated with financial stability. Other measures include volatility of bank return on assets or return on equity, (*profit_volat*); profitability, as measured by ROA or ROE; *capitalization*, measured by the capital adequacy ratio or the ratio of equity to total assets; *DtoD*, measured by either the KMV model's distance-to-default or the associated probability of bankruptcy; and an "other" category for all other measures of stability.

Note that some of these measures are positively associated with financial stability, while others are negatively associated.

Z&H use 6 categories to classify the different competition measures employed by studies. All are designed to operationalize the concept of market power. The most common measure is the Lerner index. The Lerner index calculates the ratio of price markup to marginal costs. Higher values indicate less competition. The Boone indicator measures the elasticity of profits with respect to costs, with higher values also indicating less competition. A related measure is the H-statistic. It calculates the responsiveness of revenues to costs, with higher values indicating greater competition.

Z&H also include two measures of market structure. *Concentration* measures the aggregate market share of either the 3- or 5-largest banks. *HHI*, the Herfindahl-Hirshman index, is a related measure that emphasizes the market share of individual firms. Z&H note that recent research prefers direct measures of performance/behavior, such as the measures above, to measures based on market structure. All other measures of competition are assigned to an "Other" category.

It should again be noted that some of the measures are positively associated with competition, while others are negatively associated. This needs to be factored into the proper signing of the estimated value of β in equation (1). For example, a negative estimate of β given the stability/competition measures *Dummies(-)/H-statistic(+)* and a positive β estimate given *NPL(-)/Lerner(-)* both indicate that increased competition in the banking sector is positively associated with financial stability. Z&H code their estimated effects so that positive values indicate a positive relationship between competition and stability.

Unfortunately, the diverse assortment of measures for stability and competition, each employing different scales and units, makes direct comparison of estimated effects impractical. It is necessary to transform the estimates into a measure that can be compared across studies.

Accordingly, Z&H follow common practice by converting estimated effects into partial correlation coefficients (PCCs):

(2) $$\text{PCC} = \frac{t}{\sqrt{t^2+df}} \ .$$

The associated standard error is given by

(3) $$\text{SEPCC} = \sqrt{\frac{(1-PCC^2)}{df}} \ .$$

This places the various estimates of the relationship between stability and competition on a common scale.

Various authors have attempted to interpret PCC values in terms of practical, or economic, significance. According to Cohen (1988), simple correlations equal to 0.10 (in absolute value) are small, 0.30 are medium, and 0.50 are large. Doucouliagos (2011) classifies partial correlations (PCCs) as small, medium, and large based upon their relative position in the distribution of estimated PCCs. Small, medium, and large are defined by the first, second, and third quartile values. Based on a sample of 22,141 PCC values calculated for a wide variety of economic subject areas, Doucouliagos (2011) defines small, medium, and large as 0.07 (in absolute value), 0.17, and 0.33, respectively. However, these can vary substantially across subject areas. We will use these latter values in interpreting the PCC values in the analysis that follows.

All the data and code for Z&H are publicly available at the website http://meta-analysis.cz/. FIGURE 1 provides a histogram of the 598 PCC values in their dataset. The solid, vertical line indicates the sample average PCC value. The two dashed lines represent the medium PCC value and the sample average for the 376 PCC values taken from peer-reviewed, journal articles. All three values are very close to zero. Nevertheless, there is a wide distribution of PCC values across studies and estimates. One of the research goals of Z&H is to determine the extent to which heterogeneity in these values is driven by the use of different stability measures, different competition measures, and other study, estimation, and data characteristics.

III. DESCRIPTION OF REPLICATIONS

Reed (2017) identifies 6 different kinds of replications depending on the nature of the data and the type of measurements and/or analyses that are undertaken. These are represented in FIGURE 2. This study implements three of them.

In the first replication ("Reproduction"), we employ Z&H's data and code to see if we can exactly reproduce the results of their paper. Utilizing their data and code should produce results identical to the ones reported in their paper. This type of replication is sometimes called "push button replication" or "verification".

In the second replication ("Repetition"), we manually recode the same studies analyzed by Z&H. This comes closest to "Same Population/Same Measurement and/or Analysis" in FIGURE 2. While we utilize the same data, estimation and study categories that Z&H use, we independently assign values to the respective variables. For example, as noted above, it can be confusing to map the estimated coefficient of β in a study to a positive or negative sign for a given pair of stability and competition measures . We make our own sign assignments and compare these with Z&H. Another example concerns the calculation of *t*-values. When *t*-values are not directly reported, we hand-calculate them from the reported coefficient estimates/standard errors/p-values reported in the original studies. In this manner, we independently assign our own values to the variables used by Z&H. Coding is a tedious and time-consuming activity that offers many opportunities for making errors, and often requires judgment calls about how to categorize specific characteristics. The "Repetition" replication will provide some insight into how important these elements can be.

The third replication ("Extension") examines a completely different set of studies, while coding them using the same variables as Z&H. This comes closest to the replication category

"Different population/Same Measurement and/or Analysis". Z&H's meta-analysis sample consists of 598 estimates from 31 studies published from 2006-2014. We construct an entirely independent sample of 762 estimates from 35 studies published from 2003-2017. These consist of 27 journal articles, 7 working papers and a Master thesis (cf. APPENDIX 2).

Slightly over half of these were published after Z&H completed their search. Seventeen were published within their search window, but were not included for whatever reason. Despite the differences in publication years, the mean "sample year" of the data, i.e., the average mid-point of the sample periods, is practically the same.[1] The "Extension" replication will provide a look at how robust the conclusions of Z&H are when the same measurements and analyses are applied to an entirely new set of estimates.

We do not attempt replications that involve different measurements and/or analyses (cf. bottom row of FIGURE 2). We could have. Z&H is somewhat distinctive in its empirical approach. Most importantly, it largely avoids using estimated standard errors to weight the respective effect sizes. Using standard errors/precision to weight estimated effects is a common empirical approach in meta-analyses. "Fixed Effects" and "Random Effects" meta-analyses (not to be confused with "fixed effects" and "random effects" as associated with panel data estimation) both rely on weights derived from effect standard errors. Instead, when Z&H weight, they typically weight based on the number of estimates reported for the respective study. In other words, a study with 50 estimates is weighted to have the same influence as a study with 10 estimates. This approach is less common. Nevertheless, we don't explore the effect of using alternative weights, or other empirical strategies, as the jury is still out regarding best practice.

TABLE 1 reports 15 results from Z&H. These are taken verbatim from the conclusion of their paper. The first result is the null result that competition is not related to financial stability.

[1] The average mid-year value of the sample periods for the Z&H data is 2001.015. It is 2001.075 for the "Extension" data.

The second result is that publication bias exists in the literature, with a bias against findings indicating that competition enhances financial stability. The next 13 results all refer to a meta-regression analysis in which various data, estimation, and study characteristics are associated with PCC values. Our replications will examine each of these results.

IV: REPLICATION: "Reproduction" and "Repetition"

IVA. Comparison of the Samples

As noted above, this study undertakes three types of replications: "Reproduction", "Repetition", and "Extension" (cf. FIGURE 2). With respect to the "Reproduction" replication, we note that using the data and code provided by the authors, we were able to exactly reproduce virtually all the results in their paper.[2] Consequently, we reference the results from the "Reproduction" replication by "Z&H" to emphasize the equality with their results.

For the "Repetition" replication, we revisited the studies examined by Z&H and independently assigned values to the respective variables. TABLE 2 provides a direct comparison between Z&H's values and those from the "Repetition" replication. Following Z&H, we report both unweighted ("Mean") and weighted ("Wmean") sample means, where weights consist of the inverse of the number of estimates per study.

In most cases the values are similar or even identical. For example, with respect to the main variable of interest, PCC, sample (unweighted) means for Z&H and the "Repetition" are -0.001 and 0.009, respectively. There are three reasons why we did not obtain identical PCC values to Z&H, each relating to different issues associated with calculating the *t*-statistic. First, as noted above, the assignation of signs (positive or negative) to the competition effect can be confusing, as it varies across specific stability-competition pairs. In a few cases, we calculated signs opposite to Z&H.

[2] The only discrepancy we found concerned the BMA results in their Table 10, where we obtained slightly different estimates. However, this involved one of Z&H's robustness checks, which we do not discuss here.

Second, sometimes the original studies did not allow a direct calculation of the *t*-statistic because neither a *t*-statistic nor standard error was reported, though information about p-values was provided. In those cases, the *t*-value had to be calculated from the p-value. This led to several types of discrepancies. Given a p-value, one can work backwards using the inverse of the *t*-distribution to obtain a *t*-statistic. In a few cases, when we did this, we obtained different values than Z&H.

Sometimes only stars (*-10%, **-5%, ***-1%) were reported, and not the p-values themselves. In these cases, we assumed p-values of 0.075, 0.025, or 0.005, respectively, and then worked backwards from the inverse of the t-distribution. In contrast, Z&H used several different methods to obtain assign *t*-values. A related problem arose when coefficients were insignificant, so that no stars were reported. When this happened, we set the p-value to 0.50 and then worked backwards to obtain a *t*-statistic. In Z&H's database, a variety of p-values were assigned in these cases, producing different *t*-statistics for insignificant coefficients.

A final source of discrepancies for PCC values arose when working with quadratic variables. To calculate the standard error for the linearized competition estimate, one requires the standard errors of the two quadratic coefficients (cf. Equation 6 on page 948 of Z&H). In some cases, only p-values were reported. Thus one had to convert the p-values to *t*-values, and then the *t*-values to coefficient standard errors. We sometimes obtained somewhat different values from this procedure than Z&H.

Overall, the match rate of PCC values in Z&H and the "Repetition" samples was relatively high. Approximately 88 percent of the 598 PCC values were within 5 percent of each other across the two samples.

With respect to the other variables, one of the biggest differences between Z&H's data and our "Repetition" sample related to country categorization. Z&H categorized the countries included in a study depending on whether they exclusively consisted of OECD countries. If all the countries examined by a study were OECD countries, Z&H categorized them as "developed". If none were OECD countries, Z&H categorized them as "developing and transition". The remaining category was studies that included both OECD and non-OECD countries.

Our match rates for the three categories were 82.6%, 93.3%, and 84.6%, respectively. This is reflected in the share of observations assigned to each of the three country categories. For example, Z&H calculated 33.6% of estimates were associated with studies on "developed" countries. We calculated that only 24.9% of estimates were associated with "developed" countries. A similar discrepancy arose with the categorization of "developing and transition" countries.

Another source of discrepancies is subjectivity in assigning values. An example is provided by the calculation of the sample period. For example, an original study may have data for the years *T1* to *T2*. Suppose the explanatory variables include a lagged dependent variable, as in Agoraki et al. (2011). Z&H calculate the sample period as (*T2-T1+1*). In contrast, we only include the years for which the dependent variable is available in the estimated regression equation. In the case of a single lagged, dependent variable, we calculate the sample period as (*T2-T1*). There is no right or wrong answer here, but different approaches produce different values for the variables *T* and *Sampleyear*.[3] This difference in practice resulted in a particularly low match rate of 56.4%.

In many cases, we obtained exactly the same values as Z&H. We classified stability and competition measures identically to Z&H. Other variables were very similar (e.g., *Quadratic, Logit, Regulation, Firstpub,* and *Reviewed journal*). For the variables *Citations* and *IFRecursive*, we copied the values from Z&H because those are the values that existed at the

[3] Another difference arise with the variable *Sampleyear*. Z&H calculate this as deviations from 1992.5. However if the mean of the sample period occurred before 1992, they assign it a value of zero. For example, the sample period for Beck, Demirgüç-Kunt, & Levine (2006) is 1982-1997, with mean sample period value < 1992.5, so Z&H assign a *sampleyear* value of zero to all the estimates in this study.

time they did their study and we were not able to recreate them given the later date of this study.

Of the 25,116 values in Z&H's database, we closely matched 23,825, for a successful overall match rate of 94.9%.[4] A spreadsheet that identifies all the discrepancies between Z&H and our recoded, "Repetition" database, along with all the data and code for this study, can be downloaded from Dataverse.[5] In the next section, we repeat Z&H's analysis to see whether we obtain the same qualitative results with the recoded data.

IVB. Comparison of Results

As noted in the discussion of TABLE 1, Z&H make three types of claims. Firstly, they conclude that the "mean reported estimate of the relationship" between competition and financial stability is "close to zero" (item 1 in TABLE 1). Secondly, they find "evidence for publication selection against positive results" (item 2 in TABLE 1) And thirdly, their analysis of the determinants of PCC identifies a number of relationships (and non-relationships) with various data, estimation, and study characteristics (items 3-15 in TABLE 1). We investigate each of these below.

Mean competition effect. Z&H's conclusion regarding the mean competition effect is primarily based on three sets of estimates: (i) the unconditional sample average of PCC values, (ii) the publication bias-adjusted sample average of PCCs (associated with the Precision Effect Test, or PET), and (iii) a "best practice" estimate of PCC. These are taken up in the three panels of TABLE 3.

Panel A of TABLE 3 reports unconditional mean PCC values for the full sample of 598 estimates. Z&H report unweighted and weighted means of -0.001 and -0.012. Using the

[4] The only other study we are aware of that calculates match rates for a replication of a meta-analysis is Fogarty and Giles (2018), who report an overall match rate below 50%.
[5] All the data and code to reproduce the results from this study, along with supplementary materials, are available here:
https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi%3A10.7910%2FDVN%2F9UQTEY.

"Repetition" sample, we obtain mean values of 0.009 and -0.001. If PCC values equal to 0.07 are defined as "small" following Doucouliagos (2011), these values are small indeed.

Panel B reports the results of a weighted least squares regression where PCC is regressed on a constant term and the standard error of PCC (*SEPCC*). The inclusion of the standard error term controls for publication bias, so that the constant term represents the "publication bias-adjusted mean of PCC" (what Z&H call the "effect beyond bias").[6] They report the results of three different estimation procedures applied to two different samples.

The three procedures are panel/study fixed effects estimation where the individual observations are weighted according to the inverse of the number of estimates per study (FE-Weight1); IV estimation where the log of the number of observations is used as an instrument for *SEPCC*, again weighted by the inverse of the number of estimates per study (IV-Weight1); and panel/study fixed effects estimation where the observations are weighted by the number of estimates per study and the precision of the PCC estimate (FE-Weight2).

These three estimation procedures are applied to (i) the full sample of 598 estimates, and (ii) a restricted sample of 376 estimates consisting of those estimates that appear in peer-reviewed journals ("Published"). This leads to six different estimates of mean PCC, adjusted for publication bias. The Z&H estimates range from 0.005 to 0.065, with three of the six estimates significant at the 5 percent level. Applying the same estimation procedures to the recoded, "Repetition" sample produces estimates ranging from 0.005 to 0.062 -- very close to the Z&H estimates -- though only one of the six estimates is significant at the 5 percent level. Similarly, all these values would be considered "small" using the Doucouliagos (2011) standard.

Z&H's final exercise estimates a "best practice" value for PCC. This was obtained by first determining a "best specification" via a Bayesian Model Averaging (BMA) analysis of the

[6] Z&H report a version of the specification where the variables have all been pre-multiplied by the inverse of the standard error variable. Accordingly, in their reported results, 1/SE represents the "publication bias-adjusted mean of PCC" (effect beyond bias) and the constant term estimates publication bias.

35 explanatory variables in TABLE 2. Once the variables were selected, PCC was regressed on the selected subset of variables. The associated coefficient estimates were then used to predict PCC given "best practice" values for the respective explanatory variables. For example, "best practice" values for sample size, impact factor, and citations was set at the sample maximum values. "Best practice" was also associated with estimates from peer-reviewed journal articles, studies that included regulation measures among their control variables, and studies that used the Boone index and something other than dummy variables for the competition and stability measures, among other criteria.[7]

This procedure was carried out using both weighted and unweighted values of the respective variables. Z&H calculated "best practice" estimates of 0.022 and 0.038, but only the latter was significant at the 5 percent level. The "Repetition" analysis produced "best practice" estimates of 0.039 and 0.006, with neither being statistically significant. Yet again, all four values are less than "small" based on the Doucouliagos (2011) guidelines.

In summary, our "Repetition" results confirm Z&H's conclusion that "the mean reported estimate of the [competition-stability] relationship is close to zero, even after correcting for publication bias and potential misspecification problems." Our estimates are approximately the

[7] Z&H first determine a "best variable" specification as one in which the Post-Inclusion Probability (PIP) in their BMA is greater than 0.5. They then estimate a regression in which this subset of "best" variables are chosen for explanatory variables. Using the resulting coefficient estimates, they then obtain a predicted value for PCC by setting the respective variables equal to what they consider the values likely to produce the "best" estimate of PCC. This is the explanation they gave for the variable values they used: "We plug in the sample maxima for the size of the dataset, the recursive impact factor, and the number of citations. We also prefer if the study is published in a peer-reviewed journal, if the estimation controls for regulation measures, as a higher degree of restrictions on banks' activities and barriers to bank entry is linked to systemic banking distress (Barth *et al.*, 2004; Beck *et al.*, 2006a,b), and if the researcher uses the Boone index, a relatively novel approach to measuring competition arising from the industrial organization literature. Because our focus rests primarily on the most precise competition coefficient estimates, we plug in the value 0 for the standard error of the PCC of the estimate (similarly as in Section 4, this approach corrects for publication bias). We also prefer if OLS is not used for the estimation of the competition-stability nexus, because it does not account for potential endogeneity. We prefer if a continuous variable is used as a proxy for stability, and if simple logit is not used for the estimation (again, because it does not allow for addressing endogeneity). We plug in zero for the dummy variable that corresponds to the assumed quadratic relation between competition and stability; in this case we have to linearize the estimates, which might induce a bias. We prefer if the *H*-statistic is not used in the estimation, because, as we have mentioned, it imposes restrictive assumptions on a bank's cost function that are only valid when the market in question is in equilibrium (Beck, 2008). We plug in sample means for all the other variables (Z&H, pages 963f.)." We follow a similar approach for setting variable values when our BMA selects some variables that were not selected by Z&H.

same (small) size as those reported by Z&H, and we find even fewer that are statistically significant.

Publication bias. Z&H next claim to find "evidence for publication selection against positive results". The first piece of evidence for this comes from the same set of regressions discussed in Panel B of TABLE 3. In those regressions, PCC was regressed on a constant term and *SEPCC* using weighted least squares to correct for heteroskedasticity. TABLE 3 reported the estimates of the constant term in that regression. The coefficient on the standard error variable also provides useful information as it estimates the extent of publication bias. In fact, a common test for publication bias known as the Funnel Asymmetry Test (FAT) consists of testing the significance of the *SEPCC* coefficient. The associated estimates are reported in Panel A of TABLE 4.

Using the six estimation approaches previously described in the discussion of TABLE 3/Panel B, Z&H estimate values for the *SEPCC* coefficient ranging from -0.757 to -4.339. The negative sign indicates that publication selection works against estimates that indicate that competition in the banking sector enhances financial stability. Four of the six estimates are significant at the 5 percent level, rejecting the null hypothesis of no publication bias. The "Repetition" results are similar. Estimates range from -0.491 to -4.190, with three of the six estimates significant at the 5 percent level. According to Doucouliagos and Stanley (2013), FAT estimates between 1 and 2 in absolute value and that are statistically significant can be considered evidence of substantial selection bias. Estimates larger than 2 constitute severe publication bias.

The second piece of evidence starts with Bayesian Model Averaging (BMA). Conceptually, BMA estimates linear regressions of all possible combinations of variable specifications. Given 35 explanatory variables, that amounts to 2^{35} variable combinations, which is too many to actually estimate. Accordingly, BMA employs Markov Chain Monte Carlo (MCMC) sampling to construct a representative subset of variable specifications from the set of all possible combinations. The estimates from each regression are weighted based on the value of the corresponding likelihood function, producing a "weighted average" for each variable coefficient.

Panel B of TABLE 4 reports three outcome measures of BMA: the posterior mean, the posterior standard deviation, and the posterior inclusion probability (*PIP*). The posterior mean is the weighted average of the respected *SEPCC* coefficients from the MCMC-sampled regression specifications. Likewise, the posterior standard deviation is the corresponding weighted, coefficient standard errors from those regressions.

Each variable appears in half of the total 2^{35} possible variable combinations. The *PIP* is the weighted probability associated with that subset of regressions. Thus, if each regression had an equal probability of being true (equal likelihood value), the *PIP* would be 0.50. Values greater than 0.50 indicate that the regressions including the variable have a higher probability of being true than the regressions that do not include the variable.

Evidence of publication bias is given by (i) a relatively large value for the posterior mean (larger than 1 in absolute value), (ii) a posterior standard deviation that is no more than approximately half the posterior mean (roughly corresponding to a *t*-ratio of 2), and (iii) a *PIP* value that is substantially greater than 0.50. Using these criteria, it is clear from the left-hand side of Panel B that Z&H's BMA results provide strong support for the existence of publication bias.

They also go one step further. Z&H take all the variables that have a *PIP* > 0.50 and estimate an OLS regression with that variable specification. The right-hand side of Panel B reports the estimates for the *SEPCC* coefficient from that regression. These are consistent with negative publication bias, but not quite as compelling. The estimated coefficient is closer to 1 in absolute value and is not significant at the 5 percent level, though it is at the 10 percent level.

As before, the "Repetition" results are very similar. Taken together, the "Repetition" results confirm Z&H's claim that the literature is biased against estimates showing a positive relationship between competition and financial stability.

Data, estimation, and study characteristics. Z&H make a number of other observations about how various data, estimation, and study characteristics are related (or not) to estimated competition effects. These are listed as items 3-15 in TABLE 1. They arrive at their conclusions following a process similar to the one represented by Panel B of TABLE 4. If the respective variable has a *PIP* value that is substantially greater than 0.50, and is significant in the associated OLS regression, they conclude there is a systematic relationship. If the PIP value is less than 0.50, they conclude no relationship exists.

APPENDIX 3 reports the complete analyses, not just for the items of TABLE 1, but for all 35 data, estimation, and study characteristics. TABLE 5 summarizes the "Repetition" results as they pertain to items 3-15 in TABLE 1. In particular, they indicate whether the underlying analysis supports or does not support Z&H's conclusions.

While six of the 13 items are confirmed in the "Repetition" replication, four are decidedly not confirmed, and two are mixed. The four items listed as "not supported" in the table all had *PIPs* less than 0.50. The two mixed items had *PIPs* greater than 0.50, but the respective coefficients were not significant in the corresponding OLS regression.

The item listed as "other" is puzzling. Z&H conclude that studies that address endogeneity tend to produce larger competition effects. However, in their BMA analysis, the endogeneity variable has a *PIP* of 0.2371, so clearly less than 0.50. Our "Repetition" also produces a *PIP* less than 0.50, so in this case, we would actually count this as supporting Z&H's results, though not supporting their stated conclusion.

In summary, unlike the analysis for mean effect and publication bias, we find that recoding the data resulted in a substantial modification of Z&H's conclusions. We will have more to say about this below, but one takeaway from this is that BMA analysis, along with multivariate regression, is likely to be more sensitive to recoding than estimates of overall effects and publication bias.

V: REPLICATION: "Extension"

VA. Comparison of the Samples

Whereas the "Repetition" replication uses the exact same papers and estimates as the original Z&H study, the "Extension" replication searches out an entire new set of studies. As noted above, while some of the studies in the "Extension" sample were published after Z&H did their literature search, many appeared in the same time frame in which Z&H did their search. This highlights how different researchers, all searching the same topic, can produce different samples.[8] The "Extension" replication will help us to appreciate the degree to which differences in search strategies across researchers can affect a meta-analysis study.

TABLE 6 compares unweighted and weighted means for the Z&H and "Extension" samples, as TABLE 2 did for "Repetition" sample. Here we are not looking to match the Z&H sample, but to identify noteworthy differences. The "Extension" sample includes more studies/estimates that focus on developed countries. 43.7% of the estimates in the "Extension" sample derive from studies that focus on OECD countries, versus only 33.6% in Z&H's sample. Fewer estimates focus on developing and transition countries.

Another difference is that a much higher proportion of estimates are derived from micro studies in the "Extension" sample (cf. *Macro* in TABLE 6). This emphasis on micro studies is also reflected in the fact that fewer studies employ country-level averages aggregated up from bank-level data (c.f. *Averaged*). Noteworthy for its similarity rather than difference is the share

[8] According to Z&H, they used four keyword pairs in searching the literature: "competition & stability", "competition & fragility", "concentration & stability", and "concentration & fragility". In contrast, we used six keyword combinations: "concentration & stability", "competition & (fragility OR financial crisis)", "concentration & stability", "concentration & fragility", "market power & stability", and "market power & fragility".

of estimates that address endogeneity. Roughly two-thirds of the estimates in both the Z&H and "Extension" samples attempt to correct for endogeneity (0.635 and 0.693, respectively).

In terms of stability measures, a larger percent of the "Extension" regressions rely on dummy variables, *NPL*, and "other" measures of stability. With respect to competition measures, Lerner remains the most popular, but a much larger share of estimates employ competition measures different from the ones listed in TABLE 6 (c.f. *Other competition*). Further, while there is a greater share of *Concentration* estimates and a lower share of *HHI* estimates, together these two concentration measures account for about the same percent of estimates as they do in Z&H's sample. Having said that, it is worth reiterating that while Z&H included these variables as measures of competition, they acknowledged that these market structure variables are generally viewed as inferior to market behavior measures such as *H-statistic*, *Boone*, and *Lerner*.

Together, GMM and TSLS account for a somewhat smaller percent of estimates in the "Extension" sample. Note that the sum of these two means is less than the mean value of *Endogeneity*, which indicates that more than half of the *Endogeneity* estimates address endogeneity via the use of lagged variables.[9] The "Extension" sample has roughly 50% more estimates that derive from panel fixed effects, with this increase coming mostly at the expense of *Logit* and (pooled) *OLS*.

A markedly fewer number of estimates come from regressions that control for regulatory and supervisory variables such as "capital stringency, supervisory power, the investor protection index, economic and banking freedom, the share of market entry restrictions or governance" (cf. *Regulation*). Finally, the studies in the "Extension" sample have a more recent date of publication (*Firstpub*). This accounts for their lower citation count. While a greater share of estimates come from peer-reviewed journals, these journals are, in general, of lower quality, at least based on their impact factor (*IFRecursive*).

[9] See Reed (2015) for how this is not a valid way to address endogeneity.

VB. Comparison of Results

Mean competition effect. As noted above, the case for Z&H's claim of a near zero competition effect is based on three pieces of evidence, the unconditional PCC mean, the publication bias-adjusted mean of PCC, and "best practice" estimates. From Panel A of TABLE 7, we see that the unconditional unweighted and weighted mean values of -0.009 and -0.016 for the "Extension" sample are both very small in size and close to the Z&H values.

The publication bias-adjusted mean values in Panel B are, with one exception, also very close to the Z&H values. The six estimation strategies employed by Z&H produced a set of mean PCC values ranging from 0.005 to 0.065. Excepting the FE(Weight2) estimate, the corresponding range for the "Extension" estimates is -0.013 to 0.053. These are all less than the 0.07 value that Doucouliagos (2011) sets as the reference point for a "small" effect.

The exception is the panel fixed effect estimate that weights on both number of estimates per study and the precision of the estimate (*FE-Weight2*). Here the estimate is 0.150 which comes close to Doucouliagos' (2011) standard of a medium-sized effect. However, the estimate is statistically insignificant at even the 10 percent level. Given the generally small size of the estimates in Panel B, along with their statistical insignificance, and the small size of the unconditional mean values in Panel A, we find that these results provide strong confirmation of Z&H's conclusion that the relationship between banking sector competition and financial stability is small to non-existent.

The fly in the ointment here is Panel C of TABLE 7. The "best practice" estimates for the "Extension" sample are 0.147 and -0.215 using the weighted and unweighted values of the variables. These stand out for several reasons. First, they are substantially larger than previous "best practice" estimates, roughly attaining to Doucouliagos' standard of a medium effect size. Further, they are highly significant – both are significant at the 1 percent level. And, lastly, and somewhat troublingly, they switch signs in going from weights to no weights.

It should be noted that these "best practice" estimates come from a BMA analysis that selects a different set of variables than Z&H. Z&H used 14 variables in the "unweighted" specification underlying their "best practice" estimates (*SEPCC, Samplesize, Developed, Developing and transition, Dummies, Profitability, H-statistic, Logit, FE, TSLS, Ownership, Citations, Firstpub, and IFRecursive*). The unweighted, "Extension" "best practice" estimate is based on 8 variables (*Sampleyear, Developed, Developing and transition, NPL, DtoD, Boone, HHI, and IFRecursive).* The associated "best practice" estimate is almost entirely accounted for by just two variables, *Boone* and *DtoD.*[10] Each variable only accounts for about 5 percent of the total sample of PCC estimates, and they rarely occur in the same regression in the original studies. Further, if one had instead set "best practice" stability and competition measures equal to *NPL* and something other than *Boone* and *HHI*, the corresponding estimate would have been -0.008 rather than -0.215.

To further show what a difference variable specification makes, if we use Z&H's variable specification with the "Extension" data, the resulting "best practice" estimate is 0.002. The corresponding weighted "best practice estimate" is 0.015. Both are statistically insignificant. The above highlights the precarious nature on which best practice estimates are based. Not only are they driven by the specific variables included in the regression, but they are heavily dependent on what one subjectively decides is "best practice."[11]

<u>Publication bias</u>. TABLE 8 reports the results of investigating publication bias in the "Extension" sample. Panel A reports the FAT results. We recall from above that Z&H's FAT results provided strong evidence of publication bias in that the size of the *SEPCC* coefficients were relatively large, and four of the six coefficients were significant at the 5 percent level. In contrast, the "Extension" results are weaker. None of the six estimates are significant. And while

[10] These two variables account for -0.180 of the total estimate of -0.216.
[11] Bandaranayake, Das, and Reed (2018) highlight the difficulty in determining "best practice" for competition measures by demonstrating that competition measures often are not correlated in the data in the manner theory predicts.

four of the estimates are relatively large in terms of size, another two are close to zero. These relatively weak results are echoed in the BMA analysis for the *SEPCC* variable. Not only is the size of the posterior mean very small, but the *PIP* doesn't even attain 0.03. Taken together, we conclude that the "Extension" sample is not characterized by publication bias.

Data, estimation, and study characteristics. As foreshadowed in the "best practice" estimates, BMA analysis of the "Extension" data produces very different results than Z&H. Appendix 4 gives the detailed BMA analysis, along with corresponding OLS results, for the full set of 35 control variables. As the goal here is to examine the results cited by Z&H in their conclusion, we focus our discussion on the summary of results reported in TABLE 9.

Of the thirteen results concerning data, estimation, and study characteristics mentioned in Z&H's conclusion, only one is confirmed with the "Extension" data: Studies that are more heavily cited tend to report evidence more favorable to the competition-stability hypothesis (item 14). Arguably, we confirm a second Z&H result in that we find no evidence that studies that address endogeneity produce competition effects that are any different from those that don't address endogeneity (cf. item 12 in TABLE 9). This is the result that arises from the empirical analyses in Z&H's tables, and which our results confirm, though it is at variance with what Z&H say in their conclusion. Other than that, none of the remaining 11 results are supported when we replicate their analyses with our "Extension" data.

VI. DISCUSSION AND CONCLUSION

This study performs three types of replications of Zigraiova and Havranek's (2016) meta-analysis on banking sector competition and financial stability. Using the nomenclature in Reed (2017), we perform a "Reproduction" replication, a "Repetition" replication, and an "Extension" replication. Meta-analyses, while bringing quantitative rigor to systematic reviews of empirical literature, also create numerous opportunities for errors and subjective evaluations to affect the analysis. Each of these replications is designed to investigate one aspect of this.

A “Reproduction” replication is basically a verification exercise to determine whether the data and code provided by the meta-analyst are able to exactly reproduce the results in the paper. Failure of a “Reproduction” replication indicates that the results may not be accurately reported, or that the version of the data and code made available by the authors and the version that produced the results in the study are not the same. It alerts that there may be a problem with the results reported in the published manuscript. Our replication task was facilitated by the fact that the data and code provided by Z&H allowed us to exactly reproduce all their key results.

A “Repetition” replication samples from the same population as the original study. It is used to determine whether the relationships identified in the original study are robust across different samples. In the context of a meta-analysis, the “population” can be thought of as the set of studies chosen by the original authors. The sample consists of the information extracted from these studies. Recoding the original studies from first principles can in turn be thought of as extracting another “sample” from the population. Just as there is sampling error associated with drawing observations from a population, so there is sampling error in coding studies to produce the data for the meta-analysis. There can be errors in transferring numerical information from the studies to a database, errors in making calculations (such as calculating partial correlation coefficients from coefficient estimates/standard errors/p-values), and subjectivity in ascribing values to the variables that numerically represent a study’s data, estimation, and study characteristics. These combine to produce a different meta-analysis “sample” from an identical population of original studies.

The last type of replication explored in this study is an “Extension” replication. A whole new set of studies are collected that address the same topic as the original meta-analysis. These constitute a related, but different population, than the population of studies assembled for the original meta-analysis. The purpose of conducting this type of replication is that it allows one to assess the robustness of a meta-analysis when applied to a different set of studies that conceptually address the same subject as the original.

What do our results contribute to an understanding of banking sector competition and financial stability? We generally confirm the main result of Z&H, that the competitive character of a country's banking sector bears little relationship to that country's financial stability. Both unconditional mean PCC values and publication bias-adjusted mean PCC values evidence little relationship between competition and stability. This result is robust across a variety of estimation procedures and all three types of replications.

The sole empirical exception to this is the "best practice" estimates from the "Extension" replication. However, not only are the results highly unstable across weighted and unweighted estimates, but closer examination reveals the precarious basis for "best practice" estimates. These can change dramatically depending on the particular set of variables included in the "best practice" regression, and the set of data, estimation, and study characteristics that are subjectively determined to be "best practice". As a result, we downweight the importance attached to "best practice" estimates.

The other results in Z&H do not fare as well upon replication. While we confirm Z&H's finding of publication selection in the "Repetition" replication, this result disappears in the "Extension" replication. Further, of the 13 findings concerning data, estimation, and study characteristics highlighted by Z&H in their conclusion, only one is confirmed across both "Repetition" and "Extension" replications: Studies that report a positive relationship between competition and stability are more highly cited. The evidence for the other findings is either mixed or outright does not support Z&H's results.

What do our results contribute to an understanding of the practice of meta-analyses? We find that the combination of coding errors and subjective assessments concerning data, estimation, and study characteristics can have an important influence on the results of meta-regression analyses. This likely follows from the interdependent nature of multivariate regressions. Given the inevitable multicollinearity that exists across data, estimation, and study characteristics, substantial changes in values of some of the variables affects the coefficient estimates of all the variables. Even though our "Repetition" replication had a 95% match rate with the original data, only six of the 13 results highlighted in Z&H's conclusion were confirmed with the "Repetition" data. Further, virtually none of the results were confirmed in the "Extension" replication when an entirely different set of studies was analyzed. Together, this suggests that results from meta-regression analyses should be viewed with caution.

Our study highlights the value-added that replication brings to meta-analyses. While some of Z&H's results were found to not be robust, the fact that our replication supports Z&H's main conclusion about competition and financial stability, and that this support was confirmed for all three types of replications, provides strong evidence that this finding can be regarded with confidence.

REFERENCES

Agoraki, M. E. K., Delis, M. D., & Pasiouras, F. (2011). Regulations, competition and bank risk-taking in transition countries. *Journal of Financial Stability, 7*(1), 38-48. doi: 10.1016/j.jfs.2009.08.002

Ak Kocabay, S. (2009). *Bank Competition and Banking System Stability: Evidence from Turkey.* Middle East Technical University, Turkey The Graduate School of Social Sciences

Akins, B., Li, L., Ng, J., & Rusticus, T. O. (2016). Bank Competition and Financial Stability: Evidence from the Financial Crisis. *Journal of Financial and Quantitative Analysis, 51*(01), 1-28. doi: 10.1017/s0022109016000090

Ali, M. S. B., Intissar, T., & Zeitun, R. (2016). Banking Concentration and Financial Stability. New Evidence from Developed and Developing Countries. *Eastern Economic Journal, 44*(1), 117-134. doi: 10.1057/eej.2016.8

Amidu, M. (2013). The effects of the structure of banking market and funding strategy on risk and return. *International Review of Financial Analysis, 28,* 143-155. doi: 10.1016/j.irfa.2013.03.001

Amidu, M., & Wolfe, S. (2013). Does bank competition and diversification lead to greater stability? Evidence from emerging markets. *Review of Development Finance, 3*(3), 152-166. doi: 10.1016/j.rdf.2013.08.002

Andries, M. A., & Capraru, B. (2010). *How does EU banking competition impact financial stability?* Lessons Learned from the Financial Crisis Proceedings of 13th International Conference on Finance and Banking, 3-11.

Anginer, D., Demirguc-Kunt, A., & Zhu, M. (2014). How does competition affect bank systemic risk? *Journal of Financial Intermediation, 23*(1), 1-26. doi: 10.1016/j.jfi.2013.11.001

Ashraf, D., Ramady, M., & Albinali, K. (2016). Financial fragility of banks, ownership structure and income diversification: Empirical evidence from the GCC region. *Research in International Business and Finance, 38,* 56-68. doi: 10.1016/j.ribaf.2016.03.010

Bandaranayake, S., Das, K.K., and Reed, W.R. (2018). Are competitive banking systems really more stable? *Critical Finance Review*, forthcoming.

Bazzana, F., & Yaldiz, E. (2010). The effect of market power on bank risk taking in Turkey. *Financial Theory and Practice*, 34(3), 297–314.

Beck, T., De Jonghe, O., & Schepens, G. (2013). Bank competition and stability: Cross-country heterogeneity. *Journal of Financial Intermediation, 22*(2), 218-244. doi: 10.1016/j.jfi.2012.07.001

Beck, T., Demirgüç-Kunt, A., & Levine, R. (2003). *Bank Concentration and Crises.* NBER Working Paper Series (9921).

Beck, T., Demirgüç-Kunt, A., & Levine, R. (2006). Bank concentration, competition, and crises: First results. *Journal of Banking & Finance, 30*(5), 1581-1603. doi: 10.1016/j.jbankfin.2005.05.010

Beck, T., Demirgüç-Kunt, A., & Levine, R. (2007). Bank concentration and fragility: Impact and mechanics. In R. a. M. C. Stulz (Ed.), *The Risks of Financial Institutions*. University of Chicago Press: National Bureau of Economic Research.

Berger, A. N., Klapper, L. F., & Turk-Ariss, R. (2009). Bank Competition and Financial Stability. *Journal of Financial Services Research, 35*(2), 99-118. doi: 10.1007/s10693-008-0050-7

Boyd, J. H., De Nicolo, G., & Jalal, A. M. (2006). *Bank Risk-Taking and Competition Revisited: New Theory and New Evidence*. IMF Working Paper, (06/297).

Bretschger, L., Kappel, V., & Werner, T. (2012). Market concentration and the likelihood of financial crises. *Journal of Banking & Finance, 36*(12), 3336-3345. doi: 10.1016/j.jbankfin.2012.07.016

Bushman, R. M., Hendricks, B. E., & Williams, C. D. (2016). Bank Competition: Measurement, Decision-Making, and Risk-Taking. *Journal of Accounting Research, 54*(3), 777-826. doi: 10.1111/1475-679x.12117

Cihak, M., & Hesse, H. (2010). Islamic banks and financial stability: An empirical analysis. *Journal of Financial Services Research*, 38, 95–113.

Cipollini, A., & Fiordelisi, F. (2009). *The impact of bank concentration on financial distress: The case of the European banking system*. Centro Studi di Banca e Finanza (CEFIN) (Center for Studies in Banking and Finance) 09021, Universita di Modena e Reggio Emilia, Facolt`a di Economia "Marco Biagi".

Cohen, J. (1988). *Statistical Power Analysis in the Behavioral Sciences*. 2nd ed. Hillsdale: Erlbaum.

Deltuvaite, V. (2012). The concentration–stability relationship in the banking system: An empirical research. *Economics and Management* 17(3), 869–878.

De Nicolo, G., & Loukoianova, E. (2007). *Bank Ownership, Market Structure and Risk*. IMF Working papers, (07/215).

Diallo, B. (2015). Bank competition and crises revisited: New results. *Economics Letters, 129*, 81-86. doi: 10.1016/j.econlet.2015.02.015

Doucouliagos, H. (2011) How large is large? Preliminary and relative guidelines for interpreting partial correlations in economics. Economics Series 2011/5, Deakin University, Faculty of Business and Law, School of Accounting, Economics and Finance.

Doucouliagos, H. and Stanley, T.D. (2013). Are all economic facts greatly exaggerated? Theory competition and selectivity. *Journal of Economic Surveys*, *27*(2), 316‒339.

Dushku, E. (2016). Bank risk-taking and competition in the Albanian banking sector. *South-Eastern Europe Journal of Economics 2*, 187-203.

Fernandez, R. O., & Garza-Garciab, J. G. (2012). *The relationship between bank competition and financial stability: A case study of the Mexican banking industry*. Centre for Global Finance Working Paper (03/12).

Fiordelisi, F., & Mare, D. S. (2014). Competition and financial stability in European cooperative banks. *Journal of International Money and Finance, 45*, 1-16. doi: 10.1016/j.jimonfin.2014.02.008

Fogarty, J. & Giles, M. (2018). Recidivism and education revisited: evidence for the USA. Paper presented at the 2018 MAER-Net Colloquium, Melbourne, Australia.

Fu, X., Lin, Y., & Molyneux, P. (2014). Bank competition and financial stability in Asia Pacific. *Journal of Banking & Finance, 38*, 64-77. doi: 10.1016/j.jbankfin.2013.09.012

Fungáčová, Z., & Weill, L. (2009). How market power influences bank failures: Evidence from Russia. BOFIT Discussion Papers 12/2009, Bank of Finland, Institute for Economies in Transition.

Fungáčová, Z., & Weill, L. (2013). Does competition influence bank failures? *Economics of Transition, 21*(2), 301-322. doi: 10.1111/ecot.12013

Goetz, M. (2016). *Competition and Bank Stability*. CFS Working Paper, (559). Germany.

Hope, J. C., Gwatidzo, T., & Ntuli, M. (2013). Investigating the effect of bank competition on financial stability in 10 African countries. *International Business & Economics Research Journal*, 12(7) 755–768.

Huljak, I. (2015). Market power and stability of CEE banks. *Business Systems Research Journal, 6*(2). doi: 10.1515/bsrj-2015-0013

Iskenderoglu, O., & Tomak, S. (2013). Competition and stability: An analysis of the Turkish banking system. *International Journal of Economics and Financial Issues,* 3(3) 752-762.

Jeon, J. Q., & Lim, K. K. (2013). Bank competition and financial stability: A comparison of commercial banks and mutual savings banks in Korea. *Pacific-Basin Finance Journal, 25*, 253-272. doi: 10.1016/j.pacfin.2013.10.00

Jiang, L., Levine, R., & Lin, C. (2017). *Does competition affect bank risk?* NBER Working Paper Series.

Jiménez, G., Lopez, J. A., & Saurina, J. (2013). How does competition affect bank risk-taking? *Journal of Financial Stability, 9*(2), 185-195. doi: 10.1016/j.jfs.2013.02.004

Kasman, A., & Kasman, S. (2016). Bank size, competition and risk in the Turkish banking industry. *Empirica, 43*(3), 607-631. doi: 10.1007/s10663-015-9307-1

Kasman, S., & Kasman, A. (2015). Bank competition, concentration and financial stability in the Turkish banking industry. *Economic Systems, 39*(3), 502-517. doi: 10.1016/j.ecosys.2014.12.003

Kick, T., & Prieto, E. (2015). Bank Risk and Competition: Evidence from Regional Banking Markets. *Review of Finance, 19*(3), 1185-1222. doi: 10.1093/rof/rfu019

Labidi, W., & Mensi, S. (2015). *Does banking market power matter on financial (in)stability? Evidence from the banking industry MENA region*. Working Paper Series, (908). University of Manouba.

Leroy, A., & Lucotte, Y. (2017). Is there a competition-stability trade-off in European banking? *Journal of International Financial Markets, Institutions and Money, 46*, 199-215. doi: 10.1016/j.intfin.2016.08.009

Liu, H., Molyneux, P., & Nguyen, L. H. (2012). Competition and risk in South East Asian commercial banking. *Applied Economics, 44*(28), 3627-3644. doi:10.1080/00036846.2011.579066

Liu, H., Molyneux, P., & Wilson, J. O. S. (2010). *Competition and Stability in European Banking: A Regional Analysis*. Working Papers, (10019). Bangor Business School, Prifysgol Bangor University (Cymru/Wales).

Liu, H., & Wilson, J. O. S. (2013). Competition and risk in Japanese banking. *The European Journal of Finance, 19*(1), 1-18. doi: 10.1080/1351847x.2011.633614

Mirzaei, A., Moore, T., & Liu, G. (2013). Does market structure matter on banks' profitability and stability? Emerging vs. advanced economies. *Journal of Banking & Finance, 37*(8), 2920-2937. doi: 10.1016/j.jbankfin.2013.04.031

Okumus, H. S., & Kibritciartar, O. (2012). Islamic Banks And Financial Stability in the GCC: An Empirical Analysis. *Istanbul Commerce University Journal of Social Sciences 11*(21), 147-164.

Pak, O., & Nurmakhanova, M. (2013). The Effect of Market Power on Bank Credit Risk-Taking and Bank Stability in Kazakhstan. *Transition Studies Review, 20*(3), 335-350. doi: 10.1007/s11300-013-0297-z

Pawlowska, M. (2016). Does the size and market structure of the banking sector have an effect on the financial stability of the European Union? *The Journal of Economic Asymmetries, 14*, 112-127. doi: 10.1016/j.jeca.2016.07.009

Pino, G., & Araya, I. (2013). Impact of the Heterogeneity in Market Power on the Relationship Between Risk Taking and Competition: Case of the Chilean Banking Sector. *Emerging Markets Finance and Trade, 49*(4), 98-112. doi: 10.2753/ree1540-496x490405

Reed, W. R. (2015). On the practice of lagging variables to avoid simultaneity. *Oxford Bulletin of Economics & Statistics, 77*(6), 897-905.

Reed, W. R. (2017). Replications in labor economics. *IZA World of Labor*, 2017: 413.

Ruiz-Porras, A. (2008). *Banking Competition and Financial Fragility: Evidence from Panel-Data*. Economic Studies, (23,1).

Samantas, I. (2013). *Bank competition and financial (in)stability in Europe: A sensitivity analysis*. MPRA (51621).

Sarkar, S., & Sensarma, R. (2016). The relationship between competition and risk-taking behaviour of Indian banks. *Journal of Financial Economic Policy, 8*(1), 95-119. doi: 10.1108/jfep-05-2015-0030

Schaeck, K., & Cihak, M. (2008). *How does competition affect efficiency and soundness in banking? New empirical evidence*. Working Paper Series, (932).

Schaeck, K., & Cihak, M. (2014). Competition, Efficiency, and Stability in Banking. *Financial Management, 43*, 215 - 241.

Schaeck, K., Cihak, M., & Wolfe, S. (2009). Are Competitive Banking Systems More Stable? *Journal of Money, Credit and Banking, 41*(4), 711-734.

Sinha, P., & Sharma, S. (2016). *Relationship of Financial Stability and Risk with Market Structure and Competition: Evidence from Indian Banking Sector*. Munich Personal RePEc Archive, (72247). Germany

Soedarmono, W., Machrouh, F., & Tarazi, A. (2011). Bank market power, economic growth and financial stability: Evidence from Asian banks. *Journal of Asian Economics, 22*(6), 460-470. doi: 10.1016/j.asieco.2011.08.003

Soedarmono, W., Machrouh, F., & Tarazi, A. (2013). Bank competition, crisis and risk taking: Evidence from emerging markets in Asia. *Journal of International Financial Markets, Institutions and Money, 23*, 196-221. doi: 10.1016/j.intfin.2012.09.009

Soedarmono, W., & Tarazi, A. (2016). Competition, Financial Intermediation, and Riskiness of Banks: Evidence from the Asia-Pacific Region. *Emerging Markets Finance and Trade, 52*(4), 961-974. doi: 10.1080/1540496x.2015.1018039

Tabak, B. M., Fazio, D. M., & Cajueiro, D. O. (2012). The relationship between banking market competition and risk-taking: Do size and capitalization matter? *Journal of Banking & Finance, 36*(12), 3366-3381. doi: 10.1016/j.jbankfin.2012.07.022

Tan, Y., & Floros, C. (2014). Risk, Profitability, and Competition: Evidence from the Chinese Banking Industry. *The Journal of Developing Areas, 48*(3), 303-319. doi: 10.1353/jda.2014.0054

Troug, H. A., & Sbia, R. (2015). The Relationship between Banking Competition and Stability in Developing Countries: The Case of Libya. *International Journal of Economics and Financial Issues 5*(3), 772-779.

Turk-Ariss, R. (2010). On the implications of market power in banking: Evidence from developing countries. *Journal of Banking & Finance, 34*(4), 765-775. doi: 10.1016/j.jbankfin.2009.09.004

Uhde, A., & Heimeshoff, U. (2009). Consolidation in banking and financial stability in Europe: Empirical evidence. *Journal of Banking & Finance, 33*(7), 1299-1311. doi: 10.1016/j.jbankfin.2009.01.006

Vrieze. J. (2018). Meta-analyses were supposed to end scientific debates. Often, they only cause more controversy. *Science*. doi: 10.1126/science.aav4617.

Wang, X., Zeng, X., & Zhang, Z. (2014). The influence of the market power of Chinese commercial banks on efficiency and stability. *China Finance Review International, 4*(4), 307-325. doi: 10.1108/cfri-07-2013-0096

Yeyati, E. L., & Micco, A. (2007). Concentration and foreign penetration in Latin American banking sectors: Impact on competition and risk. *Journal of Banking & Finance, 31*(6), 1633-1647. doi: 10.1016/j.jbankfin.2006.11.003

Zigraiova, D., & Havranek, T. (2016). Bank Competition and Financial Stability: Much Ado About Nothing? *Journal of Economic Surveys, 30*(5), 944-981. doi: 10.1111/joes.12131

TABLE 1
15 Results from Z&H

1	*"Our results suggest that the mean reported estimate of the [competition-stability] relationship is close to zero, even after correcting for publication bias and potential misspecification problems."*
2	*"We find evidence for publication selection against positive results; that is, some authors of primary studies tend to discard estimates inconsistent with the competition-fragility hypothesis."*
3	*"Researchers who use heterogeneous samples of countries (including both developed and developing economies) tend to obtain smaller estimates.*
4	*"The effect of competition on stability is larger in developed countries, but even there the positive effects do not seem to be strong."*
5	*"…accounting for potential nonlinearities in the effect of competition on stability is important and typically yields smaller estimates of the competition-stability nexus."*
6	*"We also find that, in general, researchers who have more data at their disposal tend to report smaller estimates."*
7	*"…it does not seem to matter for the results whether the authors of primary studies use micro or macro data."*
8	*"…if dummy variables (usually indicating financial crises) are used as a proxy for stability, the authors tend to report much larger estimates than when a continuous measure of financial stability is used."*
9	*"Studies using the H-statistic tend to report larger estimates of the competition-stability nexus."*
10	*"…studies that employ the Boone index usually show smaller estimates."*
11	*"…we find no evidence of systematic differences between the results of the studies that use competition measures and the studies that use concentration as a proxy for competition."*
12	*"…if the researchers ignore the endogeneity problem in regressing financial stability on bank competition, they tend to underestimate the effect."*
13	*"…controlling for supervisory and regulatory conditions in regressions usually decreases the reported estimates."*
14	*"…studies that receive more citations … tend to report larger estimates of the competition-stability nexus."*
15	*"…studies that … are published in journals with a high impact factor tend to report larger estimates of the competition-stability nexus."*

SOURCE: Zigraiova and Havranek (2016, pages 974f.)

TABLE 2
Comparison of Sample Means for Z&H and Repetition

Variable	*Z&H*		*Repetition*	
	Mean	*WMean*	*Mean*	*WMean*
Data characteristics				
PCC	-0.001	-0.012	0.009	-0.001
SEPCC	0.027	0.029	0.027	0.028
Samplesize	7.835	7.760	7.844	7.776
T	2.224	2.264	2.183	2.212
Sampleyear	8.889	9.340	9.253	9.739
Countries examined				
Developed	0.336	0.366	0.249	0.275
Developing and transition	0.324	0.376	0.258	0.247
Other countries*	0.339	0.258	0.493	0.479
Design of the analysis				
Quadratic	0.119	0.217	0.120	0.218
Endogeneity	0.635	0.713	0.574	0.692
Macro	0.256	0.133	0.171	0.140
Averaged	0.120	0.085	0.027	0.056
Treatment of stability				
Dummies	0.142	0.129	0.142	0.129
NPL	0.050	0.095	0.050	0.095
Z-score	0.452	0.537	0.452	0.537
Profit volatility	0.075	0.039	0.075	0.039
Profitability	0.043	0.045	0.043	0.045
Capitalization	0.069	0.040	0.069	0.040
DtoD	0.065	0.047	0.065	0.047
Other stability*	0.104	0.069	0.104	0.069

Variable	Z&H		Repetition	
	Mean	WMean	Mean	WMean
Treatment of competition				
H-statistic	0.090	0.098	0.090	0.098
Boone	0.075	0.108	0.075	0.108
Concentration	0.157	0.147	0.159	0.148
Lerner	0.360	0.414	0.360	0.414
HHI	0.266	0.197	0.266	0.197
Other competition*	0.052	0.037	0.050	0.035
Estimation method				
Logit	0.172	0.161	0.172	0.161
OLS	0.137	0.115	0.166	0.138
FE	0.229	0.136	0.251	0.163
RE	0.067	0.043	0.072	0.050
GMM	0.182	0.309	0.167	0.295
TSLS	0.149	0.110	0.122	0.077
Other method*	0.064	0.126	0.050	0.115
Control variables				
Regulation	0.239	0.282	0.237	0.272
Ownership	0.166	0.271	0.186	0.289
Global	0.794	0.764	0.844	0.781
Publication characteristics				
Citations	2.045	1.790	2.045	1.790
Firstpub	6.453	6.677	6.406	6.677
IFrecursive	0.243	0.205	0.243	0.205
Reviewed journal	0.629	0.677	0.629	0.677

Note. The table presents summary statistics of regression variables. The left side of the table presents summary statistics from data published by Zigraiova and Havranek (2016). The right side presents summary statistics from data re-coding. SD is the standard deviation and WMean is the mean weighted by the inverse of the number of estimates reported per study. An asterisk indicates the reference category.

TABLE 3
Comparison of Results (Repetition): Overall Mean Effect

A. Unconditional Mean of PCC

Type of Replication	*Mean*	*WMean*
Z&H	-0.001	-0.012
Repetition	0.009	-0.001

B. PET (publication bias-adjusted mean of PCC)

Type of Replication	*FE (Weight1)*	*FE_Published (Weight1)*	*IV (Weight1)*	*IV_Published (Weight1)*	*FE (Weight2)*	*FE_Published (Weight2)*
Z&H	0.005 *[n=598]*	0.065 *[n=376]*	0.019** *[n=598]*	0.053*** *[n=376]*	0.013 *[n=598]*	0.056** *[n=376]*
Repetition	0.005 *[n=598]*	0.062 *[n=376]*	0.018* *[n=598]*	0.050*** *[n=376]*	0.013 *[n=598]*	0.054* *[n=376]*

C. "Best Practice" Estimate of PCC: All countries (full sample)

Type of Replication	*Weighted*	*Unweighted*
Z&H	0.022	0.038**
Repetition	0.039	0.006

TABLE 4
Comparison of Results (Repetition): Publication Bias

A. FAT

Type of Replication	*FE (Weight1)*	*FE_Published (Weight1)*	*IV (Weight1)*	*IV_Published (Weight1)*	*FE (Weight2)*	*FE_Published (Weight2)*
Z&H	-0.757 *[n=598]*	-4.000* *[n=376]*	-1.706** *[n=598]*	-3.344*** *[n=376]*	-1.539** *[n=598]*	-4.339** *[n=376]*
Repetition	-0.491 *[n=598]*	-3.833 *[n=376]*	-1.343** *[n=598]*	-3.152*** *[n=376]*	-1.251 *[n=598]*	-4.190** *[n=376]*

B. BMA Results: *SEPCC*

Type of Replication	*Bayesian model averaging*			*OLS*		
	Post. mean	*Post. SD*	*PIP*	*Coeff*	*Robust SE*	*p-value*
Z&H	-1.7883	0.2046	1.000	-1.1940	0.6511	0.067
Repetition	-1.8207	0.1947	1.000	-1.4576	0.8458	0.085

TABLE 5
Comparison of Results (Repetition): BMA

SUPPORTED	
4	"The effect of competition on stability is larger in developed countries, but even there the positive effects do not seem to be strong."
6	"We also find that, in general, researchers who have more data at their disposal tend to report smaller estimates."
7	"…it does not seem to matter for the results whether the authors of primary studies use micro or macro data."
9	"Studies using the H-statistic tend to report larger estimates of the competition-stability nexus."
11	"…we find no evidence of systematic differences between the results of the studies that use competition measures and the studies that use concentration as a proxy for competition."
14	"…studies that receive more citations … tend to report larger estimates of the competition-stability nexus."
NOT SUPPORTED	
8	"…if dummy variables (usually indicating financial crises) are used as a proxy for stability, the authors tend to report much larger estimates than when a continuous measure of financial stability is used."
10	"…studies that employ the Boone index usually show smaller estimates."
13	"…controlling for supervisory and regulatory conditions in regressions usually decreases the reported estimates."
15	"…studies that … are published in journals with a high impact factor tend to report larger estimates of the competition-stability nexus."
MIXED	
3	"Researchers who use heterogeneous samples of countries (including both developed and developing economies) tend to obtain smaller estimates.
5	"…accounting for potential nonlinearities in the effect of competition on stability is important and typically yields smaller estimates of the competition-stability nexus."
OTHER	
12	"…if the researchers ignore the endogeneity problem in regressing financial stability on bank competition, they tend to underestimate the effect."* ** This statement appears to be at variance with the BMA results reported in Z&H, which indicates that studies that ignore endogeneity bias are no different than those that address it. The "Repetition" replication supports the results as reported in Z&H's Table 5.*

TABLE 6
Comparison of Sample Means for Z&H and Extension

Variable	*Z&H*		*Extension*	
	Mean	*WMean*	*Mean*	*WMean*
Data characteristics				
PCC	-0.001	-0.012	-0.009	-0.016
SEPCC	0.027	0.029	0.029	0.034
Samplesize	7.835	7.760	7.779	7.247
T	2.224	2.264	2.425	2.352
Sampleyear	8.889	9.340	9.644	10.509
Countries examined				
Developed	0.336	0.366	0.437	0.324
Developing and transition	0.324	0.376	0.243	0.343
Other countries*	0.339	0.258	0.320	0.333
Design of the analysis				
Quadratic	0.119	0.217	0.060	0.137
Endogeneity	0.635	0.713	0.693	0.520
Macro	0.256	0.133	0.163	0.143
Averaged	0.120	0.085	0.012	0.004
Treatment of stability				
Dummies	0.142	0.129	0.211	0.149
NPL	0.050	0.095	0.160	0.199
Z-score	0.452	0.537	0.270	0.335
Profit volatility	0.075	0.039	0.039	0.054
Profitability	0.043	0.045	0.058	0.041
Capitalization	0.069	0.040	0.029	0.027
DtoD	0.065	0.047	0.024	0.009
Other stability*	0.104	0.069	0.209	0.185

Variable	***Z&H***		***Extension***	
	Mean	***WMean***	***Mean***	***WMean***
Treatment of competition				
H-statistic	0.090	0.098	0.039	0.056
Boone	0.075	0.108	0.038	0.040
Concentration	0.157	0.147	0.244	0.213
Lerner	0.360	0.414	0.357	0.357
HHI	0.266	0.197	0.139	0.200
Other competition*	0.052	0.037	0.182	0.133
Estimation method				
Logit	0.172	0.161	0.119	0.090
OLS	0.137	0.115	0.093	0.173
FE	0.229	0.136	0.315	0.267
RE	0.067	0.043	0.108	0.081
GMM	0.182	0.309	0.232	0.263
TSLS	0.149	0.110	0.056	0.041
Other method*	0.064	0.126	0.076	0.087
Control variables				
Regulation	0.239	0.282	0.089	0.100
Ownership	0.166	0.271	0.102	0.149
Global	0.794	0.764	0.678	0.740
Publication characteristics				
Citations	2.045	1.790	0.643	0.599
Firstpub	6.453	6.677	10.571	10.429
IFrecursive	0.243	0.205	0.108	0.085
Reviewed journal	0.629	0.677	0.785	0.771

Note. The table presents summary statistics of regression variables. The left side of the table presents summary statistics from data published by Zigraiova and Havranek (2016). The right side presents summary statistics from the “Extension” data. SD is the standard deviation and WMean is the mean weighted by the inverse of the number of estimates reported per study. An asterisk indicates the reference category.

TABLE 7
Comparison of Results (Extension): Overall Mean Effect

A. Unconditional Mean of PCC

Type of Replication	*Mean*	*WMean*
Z&H	-0.001	-0.012
Extension	-0.009	-0.016

B. PET (publication bias-adjusted mean of PCC)

Type of Replication	*FE (Weight1)*	*FE_Published (Weight1)*	*IV (Weight1)*	*IV_Published (Weight1)*	*FE (Weight2)*	*FE_Published (Weight2)*
Z&H	0.005 *[n=598]*	0.065 *[n=376]*	0.019** *[n=598]*	0.053*** *[n=376]*	0.013 *[n=598]*	0.056** *[n=376]*
Extension	0.030 *[n=762]*	0.025 *[n=598]*	-0.013 *[n=762]*	-0.008 *[n=598]*	0.150 *[n=762]*	0.053 *[n=598]*

C. "Best Practice" Estimate of PCC: All countries

Type of Replication	*Weighted*	*Unweighted*
Z&H	0.022	0.038**
Extension	0.147***	-0.215***

TABLE 8
Comparison of Results (Extension): Publication Bias

A. FAT

Type of Replication	*FE (Weight1)*	*FE_Published (Weight1)*	*IV (Weight1)*	*IV_Published (Weight1)*	*FE (Weight2)*	*FE_Published (Weight2)*
Z&H	-0.757 *[n=598]*	-4.000* *[n=376]*	-1.706** *[n=598]*	-3.344*** *[n=376]*	-1.539** *[n=598]*	-4.339** *[n=376]*
Extension	-2.409 *[n=762]*	-1.918 *[n=598]*	0.837 *[n=762]*	0.222 *[n=598]*	-8.191 *[n=762]*	-3.282 *[n=598]*

B. BMA Results: ***SEPCC***

Type of Replication	***Bayesian model averaging***			***OLS***		
	Post. mean	*Post. SD*	*PIP*	*Coeff*	*Robust SE*	*p-value*
Z&H	-1.7883	0.2046	1.000	-1.194	0.651	0.067
Extension	0.0028	0.0715	0.0260	----	----	----

TABLE 9
Comparison of Results (Extension): BMA

SUPPORTED	
14	"…studies that receive more citations … tend to report larger estimates of the competition-stability nexus."
NOT SUPPORTED	
3	"Researchers who use heterogeneous samples of countries (including both developed and developing economies) tend to obtain smaller estimates.
4	"The effect of competition on stability is larger in developed countries, but even there the positive effects do not seem to be strong."
5	"…accounting for potential nonlinearities in the effect of competition on stability is important and typically yields smaller estimates of the competition-stability nexus."
6	"We also find that, in general, researchers who have more data at their disposal tend to report smaller estimates."
7	"… it does not seem to matter for the results whether the authors of primary studies use micro or macro data."
8	"…if dummy variables (usually indicating financial crises) are used as a proxy for stability, the authors tend to report much larger estimates than when a continuous measure of financial stability is used."
9	"Studies using the H-statistic tend to report larger estimates of the competition-stability nexus."
10	"…studies that employ the Boone index usually show smaller estimates."
11	"…we find no evidence of systematic differences between the results of the studies that use competition measures and the studies that use concentration as a proxy for competition."
13	"…controlling for supervisory and regulatory conditions in regressions usually decreases the reported estimates."
15	"…studies that … are published in journals with a high impact factor tend to report larger estimates of the competition-stability nexus."
OTHER	
12	"…if the researchers ignore the endogeneity problem in regressing financial stability on bank competition, they tend to underestimate the effect."* ** This statement appears to be at variance with the BMA results reported in Z&H, which indicates that studies that ignore endogeneity bias are no different than those that address it. The "Repetition" replication supports the results as reported in Z&H's Table 5.*

FIGURE 1
Empirical Distribution of PCC Values in Z&H

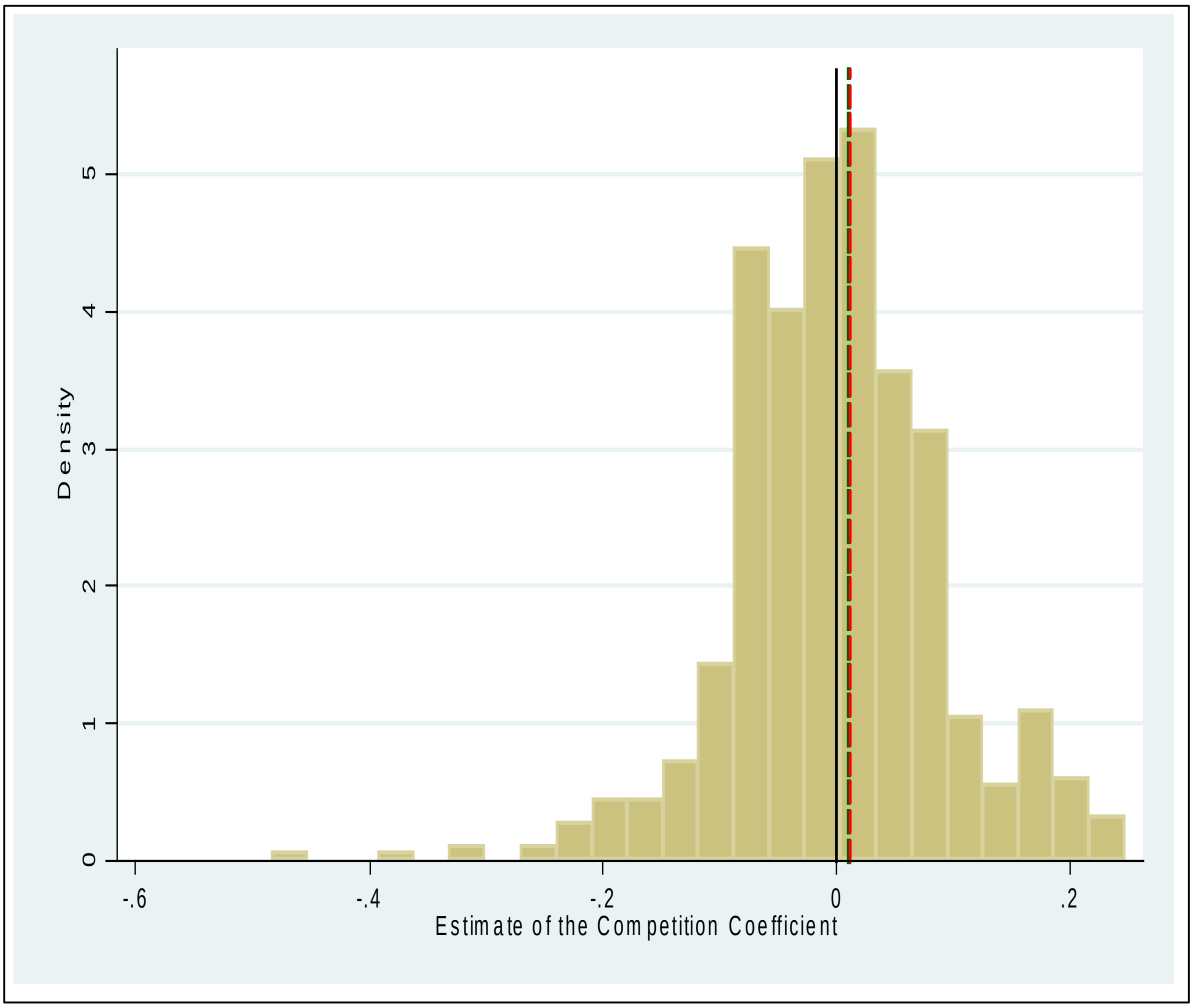

FIGURE 2
Six Different Kinds of Replications

Measurement and/or Analysis	***Source of Data***		
	Same dataset	***Same population***	***Different population***
Same	***REPRODUCTION***	***REPETITION***	***EXTENSION***
Different	*Robustness Analysis – Same Dataset*	*Robustness Analysis – Same Population*	*Robustness Analysis – Different Population*

Source: Reed (2017)

APPENDIX 1
Studies Used in Z&H's Analysis

ID	*Study*	*Publication Type*	*Number of Estimates*
1	Agoraki, Delis, & Pasiouras (2011)	Journal	10
2	Anginer, Demirguc-Kunt, & Zhu (2014)	Journal	27
3	Bazzana & Yaldiz (2010)	Journal	20
4	Beck, De Jonghe, & Schepens (2013)	Journal	19
5	Beck, Demirgüç-Kunt, & Levine (2006)	Journal	24
6	Beck, Demirgüç-Kunt, & Levine (2007)	Book chapter	24
7	Berger, Klapper, & Turk-Ariss (2009)	Journal	9
8	Boyd, De Nicolo, & Jalal (2006)	Working paper	84
9	Tabak, Fazio, & Cajueiro (2012)	Journal	4
10	Andries & Capraru (2010)	Working paper	2
11	Schaeck, Cihak, & Wolfe (2009)	Journal	20
12	Schaeck & Cihak (2014)	Journal	13
13	Cipollini & Fiordelisi (2009)	Working paper	18
14	Hope, Gwatidzo, & Ntuli (2013)	Journal	9
15	Cihak & Hesse (2010)	Journal	18
16	Schaeck & Cihak (2008)	Working paper	16
17	De Nicolo & Loukoianova (2007)	Working paper	22
18	Fernandez & Garza-Garciab (2012)	Working paper	4
19	Fu, Lin, & Molyneux (2014)	Journal	26
20	Jeon & Lim (2013)	Journal	36
21	Liu, Molyneux, & Wilson (2013)	Working paper	12
22	Liu & Wilson (2013)	Journal	7
23	Liu, Molyneux, & Nguyen (2012)	Journal	24
24	Soedarmono, Machrouh, & Tarazi (2013)	Journal	54
25	Samantas (2013)	Working paper	23
26	Turk-Ariss (2010)	Journal	6
27	Iskenderoglu & Tomak (2013)	Journal	6
28	Uhde & Heimeshoff (2009)	Journal	31
29	Fungáčová & Weill (2009)	Working paper	17
30	Yeyati & Micco (2007)	Journal	9
31	Deltuvaite (2012)	Journal	4

APPENDIX 2
Studies Used in "Extension" Analysis

ID	*Study*	*Publication Type*	*Number of Estimates*
32	Akins, Li, Ng, & Rusticus (2016)	Journal	14
33	Ali, Intissar, & Zeitun (2016)	Journal	32
34	Amidu (2013)	Journal	29
35	Amidu & Wolfe (2013)	Journal	22
36	Ashraf, Ramady, & Albinali (2016)	Journal	5
37	Beck, Demirgüç-Kunt, & Levine (2003)	Working paper	35
38	Bretschger, Kappel & Werner (2012)	Journal	40
39	Bushman, Hendricks, & Williams (2016)	Journal	20
40	Diallo (2015)	Journal	6
41	Dushku (2016)	Journal	16
42	Fiordelisi & Mare (2014)	Journal	30
43	Fungáčová & Weill (2013)	Journal	64
44	Goetz (2016)	Working paper	55
45	Hulijak (2015)	Journal	12
46	Jiang, Levine, & Lin (2017)	Working paper	35
47	Jiménez, Lopez, & Saurina (2013)	Journal	9
48	Kasman & Kasman (2015)	Journal	24
49	Kasman & Kasman (2016)	Journal	16
50	Kick & Prieto (2015)	Journal	39
51	Labidi & Mensi (2015)	Working paper	6
52	Leroy & Lucotte (2017)	Journal	60
53	Mirzaei, Moore, & Liu(2013)	Journal	28
54	Okumus & Kibritciartar (2012)	Working paper	5
55	Pak & Nurmakhanova (2013)	Journal	12
56	Pawlowska (2016)	Journal	45
57	Pino & Araya (2013)	Journal	8
58	Ruiz-Porras (2008)	Working paper	6
59	Ak Kocabay (2009)	Masters thesis	16
60	Sarkar & Sensarma (2016)	Journal	25
61	Sinha & Sharma (2016)	Working paper	6
62	Soedarmono, Machrouh & Tarazi (2011)	Journal	12
63	Soedarmono & Tarazi (2016)	Journal	12
64	Tan & Floros (2014)	Journal	16
65	Troug & Sbia (2015)	Journal	1
66	Wang, Zeng & Zhang (2014)	Journal	1

APPENDIX 3
Comparison of Results (Repetition): BMA

Type of Replication	Bayesian model averaging			OLS		
	Post. mean	Post. SD	PIP	Coeff	Robust SE	p-value
			Data Characteristics: SEPCC			
Z&H	-1.7883	0.2046	1.000	-1.194	0.651	0.067
Repetition	-1.8207	0.1947	1.000	-1.458	0.846	0.085
			SUPPORTED			
			Data Characteristics: Samplesize			
Z&H	-0.0367	0.0035	1.000	-0.024	0.009	0.007
Repetition	-0.0367	0.0051	1.000	-0.029	0.014	0.042
			SUPPORTED			
			Data Characteristics: T			
Z&H	0.0005	0.0039	0.052	----	----	----
Repetition	0.0006	0.0039	0.061	----	----	----
			SUPPORTED			
			Data Characteristics: Sampleyear			
Z&H	0.0000	0.0005	0.046	----	----	----
Repetition	0.0034	0.0024	0.740	0.003	0.003	0.324
			NOT SUPPORTED			
			Countries examined: Developed			
Z&H	0.2015	0.0219	1.000	0.176	0.029	0.000
Repetition	0.0505	0.0091	1.000	0.053	0.021	0.013
			SUPPORTED			
			Countries examined: Developing and transition			
Z&H	0.1072	0.0169	1.000	0.099	0.026	0.000
Repetition	-0.0012	0.0062	0.075	----	----	----
			NOT SUPPORTED			
			Design of the analysis: Quadratic			
Z&H	-0.0533	0.0124	0.997	-0.044	0.013	0.001
Repetition	-0.0487	0.0115	0.998	-0.026	0.017	0.124
			MIXED			

Type of Replication	Bayesian model averaging			OLS		
	Post. mean	Post. SD	PIP	Coeff	Robust SE	p-value
			Design of the analysis: Endogeneity			
Z&H	0.0100	0.0212	0.237	----	----	----
Repetition	0.0005	0.0039	0.051	----	----	----
			SUPPORTED			
			Design of the analysis: Macro			
Z&H	0.0025	0.0124	0.070	----	----	----
Repetition	0.0010	0.0099	0.058	----	----	----
			SUPPORTED			
			Design of the analysis: Averaged			
Z&H	-0.0004	0.0047	0.040	----	----	----
Repetition	0.0570	0.0247	0.905	0.0380	0.0263	0.149
			NOT SUPPORTED			
			Treatment of stability: Dummies			
Z&H	0.2115	0.0282	1.000	0.184	0.019	0.000
Repetition	-0.1562	0.0333	0.999	-0.176	0.037	0.000
			NOT SUPPORTED			
			Treatment of stability: NPL			
Z&H	0.0020	0.0060	0.132	----	----	----
Repetition	0.0001	0.0019	0.036	----	----	----
			SUPPORTED			
			Treatment of stability: Z-score			
Z&H	-0.0005	0.0027	0.063	----	----	----
Repetition	0.0000	0.0015	0.036	----	----	----
			SUPPORTED			
			Treatment of stability: Profit volatility			
Z&H	0.0006	0.0051	0.037	----	----	----
Repetition	0.0000	0.0042	0.034	----	----	----
			SUPPORTED			

Type of Replication	Bayesian model averaging			OLS		
	Post. mean	Post. SD	PIP	Coeff	Robust SE	p-value
			Treatment of stability: Profitability			
Z&H	-0.0003	0.0030	0.035	----	----	----
Repetition	-0.0006	0.0042	0.053	----	----	----
			SUPPORTED			
			Treatment of stability: Capitalization			
Z&H	0.0001	0.0029	0.027	----	----	----
Repetition	-0.0001	0.0036	0.035	----	----	----
			SUPPORTED			
			Treatment of stability: DtoD			
Z&H	-0.0013	0.0078	0.050	----	----	----
Repetition	-0.1008	0.0269	0.991	-0.0789	0.0353	0.025
			NOT SUPPORTED			
			Treatment of competition: H-statistic			
Z&H	0.1083	0.0217	1.000	0.114	0.018	0.000
Repetition	0.1021	0.0167	1.000	0.105	0.028	0.000
			SUPPORTED			
			Treatment of competition: Boone			
Z&H	-0.0709	0.0313	0.897	-0.058	0.023	0.010
Repetition	-0.0008	0.0061	0.068	----	----	----
			NOT SUPPORTED			
			Treatment of competition: Concentration			
Z&H	-0.0185	0.0226	0.474	----	----	----
Repetition	-0.0012	0.0066	0.069	----	----	----
			SUPPORTED			
			Treatment of competition: Lerner			
Z&H	0.0036	0.0130	0.122	----	----	----
Repetition	0.0318	0.0124	0.955	0.0054	0.0155	0.728
			NOT SUPPORTED			

Type of Replication	Bayesian model averaging			OLS		
	Post. mean	Post. SD	PIP	Coeff	Robust SE	p-value
			Treatment of competition: HHI			
Z&H	0.0023	0.0108	0.085	----	----	----
Repetition	0.0023	0.0079	0.117	----	----	----
			SUPPORTED			
			Estimation method: Logit			
Z&H	-0.1874	0.0230	1.000	-0.160	0.019	0.000
Repetition	0.1605	0.0270	1.000	0.174	0.035	0.000
			NOT SUPPORTED			
			Estimation method: OLS			
Z&H	-0.0352	0.0244	0.756	-0.038	0.018	0.038
Repetition	0.0003	0.0038	0.047	----	----	----
			NOT SUPPORTED			
			Estimation method: FE			
Z&H	0.0113	0.0211	0.277	----	----	----
Repetition	0.0005	0.0043	0.046	----	----	----
			SUPPORTED			
			Estimation method: RE			
Z&H	0.0018	0.0115	0.058	----	----	----
Repetition	0.0888	0.0237	0.996	0.085	0.043	0.049
			NOT SUPPORTED			
			Estimation method: GMM			
Z&H	-0.0003	0.0029	0.040	----	----	----
Repetition	0.0059	0.0127	0.223	----	----	----
			SUPPORTED			
			Estimation method: TSLS			
Z&H	-0.0001	0.0030	0.032	----	----	----
Repetition	0.0004	0.0044	0.043	----	----	----
			SUPPORTED			

Type of Replication	Bayesian model averaging			OLS		
	Post. mean	Post. SD	PIP	Coeff	Robust SE	p-value
			Control variables: Regulation			
Z&H	-0.0321	0.0197	0.798	-0.036	0.014	0.010
Repetition	-0.0105	0.0157	0.370	----	----	----
			NOT SUPPORTED			
			Control variables: Ownership			
Z&H	-0.0147	0.0175	0.481	----	----	----
Repetition	-0.0318	0.0180	0.822	-0.030	0.023	0.179
			NOT SUPPORTED			
			Control variables: Global			
Z&H	-0.0017	0.0058	0.116	----	----	----
Repetition	0.0385	0.0107	0.986	0.022	0.018	0.224
			NOT SUPPORTED			
			Publication characteristics: Citations			
Z&H	0.0497	0.0092	1.000	0.046	0.009	0.000
Repetition	0.0726	0.0088	1.000	0.059	0.016	0.000
			SUPPORTED			
			Publication characteristics: Firstpub			
Z&H	0.0219	.0044	1.000	0.023	0.003	0.000
Repetition	0.0328	0.0041	1.000	0.026	0.008	0.001
			SUPPORTED			
			Publication characteristics: IFrecursive			
Z&H	0.1060	0.0528	0.875	0.096	0.048	0.043
Repetition	0.0007	0.0073	0.044	----	----	----
			NOT SUPPORTED			
			Publication characteristics: Reviewed journal			
Z&H	-0.0249	0.0186	0.725	-0.015	0.014	0.289
Repetition	-0.1172	0.0115	1.000	-0.079	0.024	0.001
			MIXED			

Note. The first and the second row of each panel presents BMA results and OLS estimation from the "Reproduction" and "Repetition" replications, respectively. "Post. Mean" is posterior mean, "Post. SD" is posterior standard deviation, and "PIP" is posterior inclusion probability. OLS estimation includes explanatory variables with PIP greater than 0.5. The standard errors are clustered at the study level. The inverse number of estimates per study is taken as the weight.

APPENDIX 4
Comparison of Results (Extension): BMA

Type of Replication	Bayesian model averaging			OLS		
	Post. mean	Post. SD	PIP	Coeff	Robust SE	p-value
Data Characteristics: SEPCC						
Z&H	-1.7883	0.2046	1.000	-1.194	0.651	0.067
Extension	0.0028	0.0715	0.026	----	----	----
NOT SUPPORTED						
Data Characteristics: Samplesize						
Z&H	-0.0367	0.0035	1.000	-0.024	0.009	0.007
Extension	-0.0003	0.0025	0.039	----	----	----
NOT SUPPORTED						
Data Characteristics: T						
Z&H	0.0005	0.0039	0.052	----	----	----
Extension	-0.0446	0.0228	0.906	-0.024	0.025	0.335
NOT SUPPORTED						
Data Characteristics: Sampleyear						
Z&H	0.0000	0.0005	0.046	----	----	----
Extension	-0.0034	0.0037	0.532	0.002	0.003	0.624
NOT SUPPORTED						
Countries examined: Developed						
Z&H	0.2015	0.0219	1.000	0.176	0.029	0.000
Extension	-0.0776	0.0311	0.933	-0.035	0.029	0.227
NOT SUPPORTED						
Countries examined: Developing and transition						
Z&H	0.1072	0.0169	1.000	0.099	0.026	0.000
Extension	0.0046	0.0151	0.116	----	----	----
NOT SUPPORTED						
Design of the analysis: Quadratic						
Z&H	-0.0533	0.0124	0.997	-0.044	0.013	0.001
Extension	0.0084	0.0205	0.184	----	----	----
NOT SUPPORTED						

Type of Replication	Bayesian model averaging			OLS		
	Post. mean	Post. SD	PIP	Coeff	Robust SE	p-value
			Design of the analysis: Endogeneity			
Z&H	0.0100	0.0212	0.237	----	----	----
Extension	-0.0180	0.0317	0.304	----	----	----
			SUPPORTED			
			Design of the analysis: Macro			
Z&H	0.0025	0.0124	0.070	----	----	----
Extension	-0.1327	0.0405	0.980	-0.087	0.034	0.011
			NOT SUPPORTED			
			Design of the analysis: Averaged			
Z&H	-0.0004	0.0047	0.040	----	----	----
Extension	0.0001	0.0243	0.017	----	----	----
			SUPPORTED			
			Treatment of stability: Dummies			
Z&H	0.2115	0.0282	1.000	0.184	0.019	0.000
Extension	0.0577	0.0604	0.549	0.041	0.032	0.208
			MIXED			
			Treatment of stability: NPL			
Z&H	0.0020	0.0060	0.132	----	----	----
Extension	0.0026	0.0118	0.074	----	----	----
			SUPPORTED			
			Treatment of stability: Z-score			
Z&H	-0.0005	0.0027	0.063	----	----	----
Extension	-0.0487	0.0222	0.903	-0.033	0.030	0.277
			NOT SUPPORTED			
			Treatment of stability: Profit volatility			
Z&H	0.0006	0.0051	0.037	----	----	----
Extension	0.0036	0.0151	0.077	----	----	----
			SUPPORTED			

Type of Replication	Bayesian model averaging			OLS		
	Post. mean	Post. SD	PIP	Coeff	Robust SE	p-value
			Treatment of stability: Profitability			
Z&H	-0.0003	0.0030	0.035	----	----	----
Extension	-0.0264	0.0483	0.276	----	----	----
			SUPPORTED			
			Treatment of stability: Capitalization			
Z&H	0.0001	0.0029	0.027	----	----	----
Extension	-0.0009	0.0087	0.027	----	----	----
			SUPPORTED			
			Treatment of stability: DtoD			
Z&H	-0.0013	0.0078	0.050	----	----	----
Extension	-0.0028	0.0274	0.027	----	----	----
			SUPPORTED			
			Treatment of competition: H-statistic			
Z&H	0.1083	0.0217	1.000	0.114	0.018	0.000
Extension	0.0030	0.0130	0.076	----	----	----
			NOT SUPPORTED			
			Treatment of competition: Boone			
Z&H	-0.0709	0.0313	0.897	-0.058	0.023	0.010
Extension	-0.0162	0.0362	0.211	----	----	----
			NOT SUPPORTED			
			Treatment of competition: Concentration			
Z&H	-0.0185	0.0226	0.474	----	----	----
Extension	0.1293	0.0218	1.000	0.1001	0.0286	0.000
			NOT SUPPORTED			
			Treatment of competition: Lerner			
Z&H	0.0036	0.0130	0.122	----	----	----
Extension	-0.0009	0.0093	0.056	----	----	----
			SUPPORTED			

Type of Replication	Bayesian model averaging			OLS		
	Post. mean	Post. SD	PIP	Coeff	Robust SE	p-value
			Treatment of competition: HHI			
Z&H	0.0023	0.0108	0.085	----	----	----
Extension	0.0637	0.0193	0.973	0.075	0.050	0.129
			NOT SUPPORTED			
			Estimation method: Logit			
Z&H	-0.1874	0.0230	1.000	-0.160	0.019	0.000
Extension	-0.1681	0.0590	0.998	-0.084	0.044	0.058
			SUPPORTED			
			Estimation method: OLS			
Z&H	-0.0352	0.0244	0.756	-0.038	0.018	0.038
Extension	-0.2274	0.0297	1.000	-0.150	0.052	0.004
			SUPPORTED			
			Estimation method: FE			
Z&H	0.0113	0.0211	0.277	----	----	----
Extension	-0.0797	0.0342	0.951	-0.036	0.029	0.217
			NOT SUPPORTED			
			Estimation method: RE			
Z&H	0.0018	0.0115	0.058	----	----	----
Extension	-0.0981	0.0442	0.945	-0.039	0.040	0.327
			NOT SUPPORTED			
			Estimation method: GMM			
Z&H	-0.0003	0.0029	0.040	----	----	----
Extension	0.0019	0.0115	0.056	----	----	----
			SUPPORTED			
			Estimation method: TSLS			
Z&H	-0.0001	0.0030	0.032	----	----	----
Extension	-0.0014	0.0111	0.039	----	----	----
			SUPPORTED			

Type of Replication	Bayesian model averaging			OLS		
	Post. mean	Post. SD	PIP	Coeff	Robust SE	p-value
Control variables: Regulation						
Z&H	-0.0321	0.0197	0.798	-0.036	0.014	0.010
Extension	-0.0003	0.0062	0.024	----	----	----
NOT SUPPORTED						
Control variables: Ownership						
Z&H	-0.0147	0.0175	0.481	----	----	----
Extension	-0.0121	0.0216	0.288	----	----	----
SUPPORTED						
Control variables: Global						
Z&H	-0.0017	0.0058	0.116	----	----	----
Extension	0.0478	0.0338	0.756	0.042	0.037	0.255
NOT SUPPORTED						
Publication characteristics: Citations						
Z&H	0.0497	0.0092	1.000	0.046	0.009	0.000
Extension	0.1456	0.0174	1.000	0.101	0.031	0.001
SUPPORTED						
Publication characteristics: Firstpub						
Z&H	0.0219	0.0044	1.000	0.023	0.003	0.000
Extension	0.0215	0.0048	1.000	0.013	0.004	0.003
SUPPORTED						
Publication characteristics: IFrecursive						
Z&H	0.1060	0.0528	0.875	0.096	0.048	0.043
Extension	0.0227	0.0670	0.138	----	----	----
NOT SUPPORTED						
Publication characteristics: Reviewed journal						
Z&H	-0.0249	0.0186	0.725	-0.015	0.014	0.289
Extension	-0.2189	0.0209	1.000	-0.150	0.048	0.002
MIXED						

Note. The first and the second row of each panel presents BMA results and OLS estimation from the "Reproduction" and "Repetition" replications, respectively. "Post. Mean" is posterior mean, "Post. SD" is posterior standard deviation, and "PIP" is posterior inclusion probability. OLS estimation includes explanatory variables with PIP greater than 0.5. The standard errors are clustered at the study level. The inverse number of estimates per study is taken as the weight.