DEPARTMENT OF ECONOMICS AND FINANCE

SCHOOL OF BUSINESS AND ECONOMICS

UNIVERSITY OF CANTERBURY

CHRISTCHURCH, NEW ZEALAND

Another Look at "Bank Competition and Financial Stability: Much Ado about Nothing?"

NOTE: This paper is a revision of University of Canterbury WP No. 2018/18

Samangi Bandaranayake
Kuntal K. Das
W. Robert Reed

WORKING PAPER

No. 8/2019

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

WORKING PAPER No. 8/2019

Another Look at “Bank Competition and Financial Stability: Much Ado about Nothing?”

Samangi Bandaranayake[1]
Kuntal K. Das[2]
W. Robert Reed[2†]

August 2019

Abstract: This study replicates Zigraiova and Havranek’s (2016) meta-analysis of banking competition and financial stability. It performs multiple 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; an “Extension” replication where additional studies on banking competition and stability are analyzed; and a “Robustness Analysis” where we check Z&H’s results using an alternative empirical procedure. Our analysis strongly confirms Z&H’s main finding that competition in the banking sector has an economically negligible effect on financial stability. This result is consistently confirmed across a variety of replication analyses. Most impressively, we confirm their finding even when we analyze a completely independent set of 35 studies not included in Z&H’s meta-analysis. Our results for Z&H’s other findings are less supportive. As the first comprehensive replication of a meta-analysis, this study also provides insights into the robustness of meta-analysis. We find that meta-regression analysis, where estimated effects are related to data, estimation, and study characteristics, is sensitive to how data are coded and to the choice of estimation procedure; and that this sensitivity extends to “best practice” estimates.

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

JEL Classifications: C41, G21, G28, L11

Acknowledgments: W. Robert Reed acknowledges financial support from the Czech Science Foundation, Grant 18-02513S.

[1] Department of Finance, University of Colombo, SRI LANKA
[2] Department of Economics and Finance, University of Canterbury, NEW ZEALAND

†Corresponding author: W. Robert Reed, email: 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, some have noted 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. Zigraiova and Havranek (Z&H) provide an extensive analysis of the subject using frontline meta-analysis techniques, including 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 numerical descriptions of the associated data, estimating equations, and studies. Collecting and recording this information is a tedious, time-consuming process. Z&H's final dataset consisted of 25,116 numerical values. Each of these had to be manually entered into a spreadsheet. The large number of values 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? 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 their 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 eight 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 (*dummies*) 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 six 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. This 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.

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. We use these benchmarks to interpret the PCC values in the analysis that follows.

III. DESCRIPTION OF REPLICATIONS

Reed (2017) identifies six 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 TABLE 1. This study implements aspects of all of them.

We begin by implementing a "Reproduction" replication. We utilize Z&H's data and code to see if we can exactly reproduce the results of their paper. This should produce results identical to the ones reported in their paper. This type of replication is sometimes also called "push button replication" or "verification". We can confirm straightaway that we are able to reproduce virtually all the results in their paper.[1]

We then undertake a "Repetition" replication in which we manually recode the same studies analyzed by Z&H. This comes closest to "Same Population/Same Measurement and/or Analysis" in TABLE 1. While we utilize the same data, estimation and study categories that Z&H use, we independently assign values to the respective variables.

We follow-up that by performing an "Extension" replication in which we examine a completely different set of studies, while following the same coding scheme used by 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. APPENDIX2). 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.

Lastly, we conduct a "Robustness Analysis" replication where we use a different estimator than Z&H. It is quite common in meta-analyses to use a weighted estimator that weights observations by the inverse of the variance of the estimated effects. In contrast, Z&H adopt a somewhat distinctive empirical approach. They employ a variety of estimators, including panel fixed effects, instrumental variables, and weighted least squares that weights on the number of estimates per study. However, they do not use the conventional inverse variance estimator. Our robustness analysis checks whether this makes a difference for Z&H's results.

[1] 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.

TABLE 2 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. 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 analysis examines each of these 15 results. Because we conduct many different types of replications, we merge the datasets from our "Repetition" and "Extension" analyses and focus our discussion on a replication using this combined sample.[2] We simply call this the "Replication" analysis. Results from the individual "Repetition" and "Extension" replications are available in a supplementary document.[3]

IV: REPLICATION

IVA. Comparison of the Samples

This section compares Z&H's original sample with the comprehensive dataset used in our subsequent replication. The latter combines Z&H's dataset with 762 additional observations from our "Extension" replication, producing a combined "Replication" sample of 1,360 observations. It should be noted that the Z&H observations in the combined Replication sample are not exactly the same as the observations in Z&H's original sample. In the process of recoding, we uncovered some mistakes and made some different judgement calls when it came to coding specific variables. However, the correspondence between the two Z&H samples is quite high. 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]

---

[2] See Appendices 1 and 2 for listings of the individual studies.

[3] This supplementary document, along with data and code to reproduce all the results in our analysis, are publicly available in the Harvard Dataverse archive: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi%3A10.7910%2FDVN%2F9UQTEY.

[4] A spreadsheet that identifies all the discrepancies between Z&H and our recoded, "Repetition" database, are available in the Dataverse archive noted above.

TABLE 3 reports both unweighted and weighted means for the Z&H and Replication samples. We follow Z&H in weighting individual observations by the inverse of the number of estimates per study. As one would expect, compared to Z&H's sample, the studies in the Replication sample rely on more recent data (*Sampleyear*), have more citations (*Citations*), and were published more recently (*Firstpub*).[5] There are other differences between the two samples: the Replication sample has a greater percent of studies that measure stability using *NPL* or "*Other*" stability measures, and a smaller percentage of studies that use *Z-score* or the bankruptcy measure *DtoD*. Likewise, the Replication sample has a smaller percent of estimated effects that measure competition using the *H-statistic* or the *Boone* indicator, but a larger percent that use *Concentration* or some "other" measure of competition. Other differences are that the Replication sample includes a smaller share of estimates based on data from *Developing and transition* countries, and that come from *Logit* or *TSLS* estimates; and a larger share that are published in peer-*Reviewed journals*. Nevertheless, the differences do not appear to be substantial.

IVB. Comparison of Results

As noted in the discussion of TABLE 2, Z&H report three kinds of results. Firstly, they conclude that the "mean reported estimate of the relationship" between competition and financial stability is "close to zero" (item 1 in TABLE 2). Secondly, they find "evidence for publication selection against positive results" (item 2 in TABLE 2) 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 2). We investigate each of these below.

<u>Mean competition effect</u>. 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

[5] Citations and recursive impact factors (*IFRecursive*) for all observations were updated to make them current (collected in July 2019).

Test, or PET), and (iii) a "best practice" estimate of PCC. These are taken up in the three panels of TABLE 4.

Panel A of TABLE 4 reports unconditional mean PCC values for the original Z&H dataset of 598 estimates and the comprehensive Replication dataset of 1,360 estimates. Unweighted and weighted PCC mean values for Z&H are -0.001 and -0.012. Following their empirical approach with the Replication dataset, we obtain mean values of -0.001 and -0.009. We also calculate the mean value of PCC using the inverse variance estimator ("WLS-Weight3"). This produces an overall mean value of -0.002 for PCC. If "small" is defined following Doucouliagos (2011) as PCC values equal to 0.07 in absolute value, these values are all very small.

Panel B reports the results from using a variety of estimators to estimate a 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] Z&H use 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); panel/study fixed effects with 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). As a robustness check, we also include the inverse variance estimator (WLS-Weight3).

[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.

These four estimation procedures are applied to (i) the full sample of all estimated effects, and (ii) a restricted sample consisting of those estimates that appear in peer-reviewed journals ("Published"). This leads to eight 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 expanded Replication sample produces estimates ranging from 0.009 to 0.054. None are significant at the 5 percent level. Further, when we use the inverse variance estimator to estimate the publication bias-adjusted mean of PCC, the corresponding estimates are -0.003 and -0.018, both of which are statistically insignificant. Again, all these values would be considered "small" using the Doucouliagos (2011) standard.

Z&H's final exercise estimated a "best practice" value for PCC. This was obtained by first determining a "best specification" via a Bayesian Model Averaging (BMA) analysis of the 35 explanatory variables in TABLE 3.[7] Once the variables were selected, PCC was regressed on the selected subset of variables ("Frequentist Regression"). 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 were 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 indicator and something other than dummy variables for the competition and stability measures, respectively.[8]

[7] The variables for the "best specification" are those which have Post-Inclusion Probability (PIP) greater than 0.5 in the BMA analysis.

[8] Z&H explain their criteria for selection values for the "best practice" regression as follows: "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

Z&H carry out this BMA analysis using both unweighted and weighted (inverse number of estimates per study) values of the respective variables. They calculate “best practice” estimates of 0.038 and 0.022, though only the former is significant at the 5 percent level. Applying exactly the same procedure as Z&H, our Replication analysis produces “best practice” estimates of -0.014 and 0.078. Neither is significant at the 5 percent level. Further, when we use inverse variance estimates, we obtain a best practice estimate of -0.020, which is again statistically insignificant.

In summary, our Replication analysis produces evidence that strongly supports 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 same (small) size as those reported by Z&H, and we find even fewer that are statistically significant.

Publication bias. Z&H next report “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 4. In those regressions, PCC was regressed on a constant term and *SEPCC* using panel fixed effects and various procedures to address endogeneity and differences in the number of estimates per study. TABLE 4 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

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.

Funnel Asymmetry Test (FAT) consists of testing the significance of the *SEPCC* coefficient. The associated estimates are reported in Panel A of TABLE 5.

In their study, 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. To aid in interpreting these results, Z&H note that Doucouliagos and Stanley (2013) identify that "the literature suffers from substantial selectivity" if the estimated *SEPCC* coefficient is (i) statistically significant and (ii) has an absolute value greater than 1.

Using the same estimation procedures on the Replication sample produces similarly sized, but statistically weaker, coefficient estimates. Estimates of the publication bias term vary from -0.819 to -3.679. However, only one of the six coefficients is significant at the 5 percent level. Further, when the inverse variance estimator is used to estimate the regression, the coefficients are positive, insignificant, and small in size. Thus the Replication results mostly satisfy the second of Doucouliagos and Stanley (2013)'s guidelines, but not the first.

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 5 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.

In the spirit of Doucouliagos and Stanley (2013)'s guidelines, evidence of substantial selectivity is given by (i) a posterior mean value 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 side of Panel B that Z&H's BMA results provide evidence in support of the existence of negative publication bias.

They also go one step further. Z&H take all the variables that have a PIP > 0.50 and estimate a "frequentist regression" with that variable specification. The right 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.

Using the same empirical approach as Z&H, but on the expanded Replication sample, we obtain posterior mean values very close to Z&H (-1.8207 versus -1.7883). However, the results from the frequentist regression are not as strong. The estimated coefficient for *SEPCC* is -0.722 versus Z&H's estimate of -1.194, and is not even significant at the 10 percent level (p-value = 0.500).

Similar results are produced by the inverse variance estimator. The absolute value of the BMA posterior mean estimate is smaller, though still larger than 1. On the other hand, the estimated coefficient for *SEPCC* in the frequentist regression is of the correct sign and larger in absolute value. However it is not significant at the 10 percent level.

We take these replication results as providing mixed support for Z&H's claim of negative publication bias. We generally, but not always, obtain negative estimates of the publication bias variable that are of a size indicating "substantial selectivity". However, of the eight regressions in Panel A, only one is significant at the 5-percent level (FE_Published-Weight2). Further, neither of the frequentist regressions associated with the BMA analysis in Panel B are statistically significant at even the 10-percent level.

<u>Data, estimation, and study characteristics</u>. 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 2. They arrive at their conclusions following a process similar to the one represented by Panel B of TABLE 5. If the respective variable has a PIP value that is substantially greater than 0.50, and is significant in the associated frequentist regression, they conclude there is a systematic relationship. If the PIP value is less than 0.50, they conclude no relationship exists.

Our assessment of Z&H's findings is complicated by the fact that we have two sets of replication analyses. One set of replication exercises exactly reproduces Z&H's empirical procedures. The other uses inverse variance weights in its replication analyses. This creates four sets of estimates to be considered in determining whether our Replication results provide "evidence for", "evidence against", or "mixed evidence": (i) two sets of BMA results, and (ii) two sets of frequentist regression estimates. Note also that Z&H have two categories of outcomes. One category finds evidence of a "signed effect"; i.e., a given data, estimation, or study characteristic is positively (negatively) associated with estimated competition effects. The other set consists of a "null effect" where the given characteristic is concluded to not affect the estimated competition estimates.

TABLE 6 gives the criteria we use to determine whether our replication analysis supports Z&H's conclusions. For a signed effect, we expect the BMA analysis to find that the associated variable has a weighted probability of being in the "true" regression of at least 50% (PIP > 0.50). We also expect the posterior mean BMA values to have the same sign that Z&H found. Lastly we expect at least one of the coefficients in the associated frequentist regression to be statistically significant at the 5 percent level with the correct sign.

For null effects, the PIPs in both of the BMA analyses should not be close to one (PIP < 0.90). Further, if the respective variable is included in the frequentist regression, its estimated coefficient should be statistically insignificant. The criteria for "evidence against" and "mixed evidence" follow similarly. We note that when Z&H report a finding of a signed effect and our replication results are consistent with a null effect, we classify that as "evidence against".

APPENDIX 3 reports detailed results from BMA analyses. These are summarized in TABLE 7.[9] Of the 13 data, estimation, and study items listed in TABLE 2, our replication analysis finds "evidence for" two of them, "mixed evidence" for four more, and "evidence against" six others.

For example, Z&H conclude that estimated competition effects tend to be larger when primary studies use the H-statistic. Evidence in favor of this is that the PIPs for the *H-statistic* variable in the two replication exercises are 0.993 and 1.000. Both posterior mean BMA values are positive, and both coefficients are positive and statistically significant in the associated frequentist regressions.

We also find evidence in support of Z&H's finding that the use of micro or macro data has a "null effect" on estimates of competition effects. In our two replication BMA analyses, the

[9] APPENDIX 4 reports Replication results for the other variables.

PIPs for the *Macro* variable are both less than 0.90 (0.769 and 0.770). Further, both coefficients are insignificant in the associated frequentist regressions (p-values of 0.586 and 0.778, respectively).

An example of a result for which we found mixed evidence is Z&H's finding that the effect of competition on stability is larger in developed countries. Consistent with this, BMA analysis using inverse variance weights produces a positive posterior mean with a PIP of 0.985. However, the coefficient for *Developed* is insignificant in the associated frequentist regression (p-value = 0.216). Further, in the replication where we reproduce Z&H's procedures on the expanded Replication sample, the posterior mean BMA value is negative with a PIP of 0.046.

In contrast, our BMA analyses of the *Quadratic* and *Dummies* variables are examples where we find evidence against Z&H's findings. Z&H report that studies that use a nonlinear (quadratic) specification to model the effect of competition tend to produce smaller estimates. In the replication exercise where we exactly reproduce Z&H's empirical procedures, we obtain a PIP of 0.091 for the *Quadratic* variable. Further, when we use inverse variance weights, the mean BMA value is positive with a PIP of 1.000.

We similarly find evidence against Z&H's finding that primary studies that use dummy variables to proxy stability find larger effects of competition. In this case, both of our two replication exercises produce PIPs less than 0.90 (0.772 and 0.785), with our replication using inverse variance weights producing a wrong-signed BMA posterior mean.

We note that 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 0.237, so clearly less than 0.50. Our Replication 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.

IVC. Summary of Replication Results

Summary of preceding results. Z&H present 15 findings in their meta-analysis of the literature on banking competition and financial stability. The preceding discussion has focused on two sets of replication analyses. The first set of replication analyses reproduces Z&H's empirical procedures. The second set follows the more common approach of using inverse variance weights. Both sets of analyses are applied to an expanded dataset that combines estimates from Z&H's 31 studies with those from an additional 35 studies identified in our "Extension" replication.

Our replication analyses confirm Z&H's claim that competition in the banking sector is unrelated to financial stability. We find mixed evidence to support their finding of negative publication bias. The remaining 13 results rely on BMA analysis of the effects of various data, estimation, and study characteristics on estimated competition effects. We find evidence for two of their findings, mixed evidence for four, and no support for six, with another result difficult to classify.

The BMA results are summarized in TABLE 7, with the more detailed BMA analyses provided in APPENDIX 3. The two results for which our BMA replications produce evidence to support Z&H's claims are: (i) "Studies using the H-statistic tend to report larger estimates of the competition-stability nexus.", and (ii) "…it does not seem to matter for the results whether the authors of primary studies use micro or macro data."

Summary of results from the "Repetition" and "Extension" replications. As noted above, we also did "Repetition" and "Extension" replications. The Repetition replication independently coded the Z&H data directly from the primary studies. It then applied Z&H's empirical procedures to this recoded data. The Extension replication also reproduced Z&H's empirical procedures, but applied them to a new set of data gleaned from 35 additional studies not included in Z&H. The detailed results of these replications are reported in a supplementary file accompanying this study.[10]

We briefly summarize the main findings from these other replication exercises. Both confirm that competition has an economically negligible impact on financial stability.[11] With respect to negative publication bias, the Repetition results are very similar to those of Z&H. In contrast, the Extension results do not provide corroborating support. Finally, the two types of replication exercises produce substantially different results regarding the various data, estimation, and study characteristics. Even though the match rate was quite high between Z&H data and our recoding of their data, the Repetition BMA analysis only corroborated six of their 13 results, with another two results finding mixed evidence. The Extension BMA analysis was only able to corroborate one of their 13 findings.

V. CONCLUSION

It is common in empirical economics for studies to produce conflicting results. Meta-analysis provides a way to aggregate findings from different studies to allow an overall conclusion. However, meta-analyses can also produce conflicting results. This can occur because meta-analyses include different studies. Or because they characterize studies differently. Or because they use different empirical procedures.

This study is the first attempt to investigate the robustness of meta-analysis by undertaking an extensive replication of a previously published meta-analysis. We replicate Zigraiova and Havranek's (*Journal of Economic Surveys,* 2016) meta-analysis of the effect of competition in the banking sector on financial stability. Our extensive replications include "Reproduction", "Repetition", and "Extension" replications, along with a "Robustness Analysis" that uses a different estimation procedure (Reed, 2017).

---

[10] The Supplement, along with the data and code to reproduce its results, is available in the Dataverse archive mentioned above.

[11] The one exception is the "best practice" results in the Extension replications. However, these we attribute to the inherent instability associated with the way "best practice" estimates are calculated.

We chose to replicate Z&H because it addresses a very important subject. Gaining a better understanding of the determinants of systemic instability is critical if economies are to avoid future global financial crises. We also chose their study because the authors provided exceptional access to their data and code, enabling us to faithfully replicate their work. We have two sets of findings: those that apply to the subject of competition and financial stability, and those that apply to the implementation of meta-analysis.

With respect to the role of competition on financial stability, our replication strongly confirms Z&H's main finding that competition in the banking sector has an economically negligible effect on financial stability. This result is consistently confirmed across a variety of replication analyses. Most impressively, we confirm their finding even when we analyze a completely independent set of 35 studies not included in Z&H's meta-analysis. Our results for Z&H's other findings are less supportive. Our replication analyses produce mixed evidence to support their finding of negative publication bias in the literature. Further, we find that estimated effects of data, estimation, and study characteristics vary widely across different types of replications.

Our study also provides insights into the robustness of meta-analyses. The black box of meta-analyses is how researchers code studies to characterize various data, estimation, and study characteristics. In this respect, we can confirm that our recoding of Z&H's studies produced a high rate of agreement. Of the 25,116 values in Z&H's database, we closely matched 23,825, for an overall match rate of 94.9%. However, even given this high match rate, the BMA analyses, and associated frequentist regressions, sometimes produced different results. For example, using our recoded version of their data and applying the exact same empirical procedures as Z&H, we were only able to corroborate six of their 13 results, with another two results finding mixed support. We interpret this finding as highlighting the sensitivity of meta-regression analyses, where estimated effects are regressed on data, estimation, and study characteristics. This sensitivity extends to “best practice” estimates, as these depend critically on the variables included in the underlying regression.

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

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
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)

TABLE 2
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	*"Studies using the H-statistic tend to report larger estimates of the competition-stability nexus."*
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	*"…it does not seem to matter for the results whether the authors of primary studies use micro or macro data."*
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 3
Comparison of Sample Means for Z&H and Replication

Variable	*Z&H (N = 598)*		*Replication (N = 1,360)*	
	Mean	***WMean***	***Mean***	***WMean***
Data characteristics				
PCC	-0.001	-0.012	-0.001	-0.009
SEPCC	0.027	0.029	0.028	0.031
Samplesize	7.835	7.760	7.808	7.496
T	2.224	2.264	2.318	2.286
Sampleyear	8.889	9.340	9.472	10.148
Countries examined				
Developed	0.336	0.366	0.354	0.301
Developing and transition	0.324	0.376	0.249	0.298
Other countries*	0.339	0.258	0.396	0.401
Design of the analysis				
Quadratic	0.119	0.217	0.087	0.175
Endogeneity	0.635	0.713	0.640	0.601
Macro	0.256	0.133	0.166	0.141
Averaged	0.120	0.085	0.018	0.029
Treatment of stability				
Dummies	0.142	0.129	0.181	0.140
NPL	0.050	0.095	0.112	0.150
Z-score	0.452	0.537	0.350	0.430
Profit volatility	0.075	0.039	0.055	0.047
Profitability	0.043	0.045	0.051	0.043
Capitalization	0.069	0.040	0.046	0.033
DtoD	0.065	0.047	0.042	0.027
Other stability*	0.104	0.069	0.162	0.131

Variable	*Z&H (N = 598)*		*Replication (N = 1,360)*	
	Mean	***WMean***	***Mean***	***WMean***
Treatment of competition				
H-statistic	0.090	0.098	0.062	0.076
Boone	0.075	0.108	0.054	0.072
Concentration	0.157	0.147	0.207	0.183
Lerner	0.360	0.414	0.358	0.384
HHI	0.266	0.197	0.195	0.199
Other competition*	0.052	0.037	0.124	0.087
Estimation method				
Logit	0.172	0.161	0.143	0.123
OLS	0.137	0.115	0.125	0.156
FE	0.229	0.136	0.287	0.218
RE	0.067	0.043	0.092	0.066
GMM	0.182	0.309	0.204	0.278
TSLS	0.149	0.110	0.085	0.058
Other method*	0.064	0.126	0.065	0.100
Control variables				
Regulation	0.239	0.282	0.154	0.181
Ownership	0.166	0.271	0.139	0.214
Global	0.794	0.764	0.751	0.759
Publication characteristics				
Citations	2.045	1.790	2.368	2.107
Firstpub	6.453	6.677	8.740	8.667
IFRecursive	0.243	0.205	0.201	0.163
Reviewed journal	0.629	0.677	0.716	0.727

NOTE: This table presents summary statistics for variables used in our analysis. The left side of the table presents summary statistics from data published by Zigraiova and Havranek (2016).

The right side presents summary statistics for the merged “Repetition” and “Extension” datasets. Variables are described in the text. WMean is the mean weighted by the inverse of the number of estimates reported per study. An asterisk indicates that the respective variable is the reference category in the subsequent analysis.

TABLE 4
Comparison of Results: Overall Mean Effect

A. Unconditional Mean of PCC

Type of Replication	*Mean*	*WMean*	*WLS (Weight3)*
Z&H	-0.001	-0.012	----
Replication	-0.001	-0.009	-0.002

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)*	*WLS (Weight3)*	*WLS_Published (Weight3)*
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]*	----	----
Replication	0.014 *[n=1360]*	0.040 *[n=974]*	0.009 *[n=1360]*	0.026 *[n=974]*	0.054 *[n=1360]*	0.053* *[n=974]*	-0.003 *[n=1360]*	-0.018 *[n=974]*

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

Type of Replication	*Unweighted*	*Weighted*	*WLS (Weight3)*
Z&H	0.038**	0.022	----
Replication	-0.014	0.078*	-0.020

NOTE: "Z&H" reproduces results from Zigraiova and Havranek (2016). "Replication" reports results using the combined Replication dataset ("Repetition" and "Extension" datasets). Panel A duplicates the summary statistics for PCC from TABLE 3. Panel B reports the results from estimating a regression where PCC is regressed on a constant term and the standard error of PCC (SEPCC). The values in the table are the estimates for the constant term. The first six columns reproduce Z&H's procedure and use three different estimation procedures applied to two different samples (i) panel/study fixed effects estimation where the individual observations are weighted according to the inverse of the number of estimates per study ("FE-Weight1"); (ii) panel/study fixed effect with 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"). As a robustness check, we also include the inverse variance estimator ("WLS-Weight3"). The "best practice" results of Panel C are obtained by first determining a "best specification" using Bayesian Model Averaging (BMA). Three variants are estimated: analysis with unweighted observations, analysis with weights equal to the number of estimates per study ("Weighted"), and analysis with weights equal to the inverse of the variance of the estimated effect ("WLS-Weight3"). Using this "best specification", PCC is regressed on the selected subset of variables ("Frequentist Regression"). The associated coefficient estimates are then used to predict PCC given "best practice" values for the respective explanatory variables. Standard errors are clustered at the study level.

TABLE 5
Comparison of Results: Publication Bias

A. FAT

Type of Replication	*FE (Weight1)*	*FE_Published (Weight1)*	*IV (Weight1)*	*IV_Published (Weight1)*	*FE (Weight2)*	*FE_Published (Weight2)*	*WLS (Weight3)*	*WLS_Published (Weight3)*
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]*	----	----
Replication	-1.142 *[n=1360]*	-2.734* *[n=974]*	-0.819 *[n=1360]*	-1.865 *[n=974]*	-3.501 *[n=1360]*	-3.679** *[n=974]*	0.084 *[n=1360]*	0.860 *[n=974]*

B. BMA Results: *SEPCC*

Type of Replication	***Bayesian Model Averaging***			***Frequentist Regression***		
	Post. mean	*Post. SD*	*PIP*	*Coeff*	*Robust SE*	*p-value*
Z&H	-1.7883	0.2046	1.000	-1.194	0.651	0.067
Replication	-1.8207	0.1947	1.000	-0.722	1.071	0.500
Replication (WLS-Weight3)	-1.3250	0.2228	0.9998	-2.263	1.399	0.106

NOTE: "Z&H" reproduces results from Zigraiova and Havranek (2016). "Replication" reports results using the combined Replication dataset ("Repetition" and "Extension" datasets). Panel A are the results from estimating a regression where PCC is regressed on a constant term and the standard error of PCC (SEPCC). The values in the table are the estimates for the SEPCC coefficient. The first six columns reproduce Z&H's procedure and use three different estimation procedures applied to two different samples (i) panel/study fixed effects estimation where the individual observations are weighted according to the inverse of the number of estimates per study ("FE-Weight1"); (ii) panel/study fixed effect with 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 (iii) 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"). As a robustness check, we also include the inverse variance estimator ("WLS-Weight3"). Panel B reports the results of BMA analysis. The left side is the BMA results for the SEPCC variable. "Post. Mean" is posterior mean, "Post. SD" is posterior standard deviation, and "PIP" is posterior inclusion probability. The right side reports the results of a frequentist regression where PCC is regressed on the "best specification" variables, where "best specification" variables are those that have a PIP greater than 0.5 in the associated BMA analysis. The values are the estimates of the SEPCC coefficient in that regression. "Z&H" reproduces Z&H's estimates, weighting observations by the inverse of the number of estimates per study. "Replication" applies the same analysis to the combined Replication dataset. "Replication (WLS-Weight3)" weights observations according to the inverse of the variance of the estimated effect. Standard errors are clustered at the study level.

TABLE 6
Criteria for Determining Whether Replication Results Support Z&H's Results

Z&H Result	***Replication Results Provide:*** ***Evidence For***	***Evidence Against***	***Mixed Evidence***
A signed effect (positive/negative)	• Both PIPs > 0.50 *AND* • Both posterior mean BMA values have the correct sign *AND* • At least one of the coefficients in the frequentist regressions is significant and has the correct sign	• Both PIPs < 0.90 and both coefficients are insignificant in the frequentist regressions *OR* • At least one posterior mean BMA value has the incorrect sign with a PIP > 0.90 *OR* • At least one of the coefficients in the frequentist regression is significant with the incorrect sign	Anything else
A null effect	• Both PIPs < 0.90 *AND* • Coefficients in the frequentist regressions are insignificant whenever they are included	• At least one of the coefficients in the frequentist regressions is significant *OR* • Both posterior mean BMA values have the same sign and have PIPs > 0.90	Anything else

TABLE 7
Results of Replication Analysis: BMA

EVIDENCE FOR	
7	"Studies using the H-statistic tend to report larger estimates of the competition-stability nexus."
9	"…it does not seem to matter for the results whether the authors of primary studies use micro or macro data."
MIXED EVIDENCE	
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."
6	"We also find that, in general, researchers who have more data at their disposal tend to report smaller estimates."
15	"…studies that … are published in journals with a high impact factor tend to report larger estimates of the competition-stability nexus."
EVIDENCE AGAINST	
5	"…accounting for potential nonlinearities in the effect of competition on stability is important and typically yields smaller estimates of the competition-stability nexus."
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."
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."
14	"…studies that receive more citations … 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.*

NOTE: The left column allows one to match the respective result in TABLE 2. Results are categorized following a BMA/frequentist regression analysis where the replication results use the criteria in TABLE 6 to assign outcomes to one of three categories: "Evidence For", "Mixed Evidence", and "Evidence Against". The detailed results of the BMA/frequentist regression are reported in APPENDIX 3. Both BMA and frequentist regression analyses weight observations by the inverse of the variance of the estimated effect.

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)	Journal	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
Replication Results for Z&H's BMA Analysis: TABLE 2 Variables

Type of Replication	Bayesian Model Averaging			Frequentist Regression		
	Post. mean	Post. SD	PIP	Coeff	Robust SE	p-value
EVIDENCE FOR						
Treatment of competition: H-statistic						
Z&H	0.108	0.022	1.000	0.114	0.018	0.000
Replication	0.087	0.021	0.993	0.111	0.040	0.005
Replication (WLS-Weight3)	0.042	0.014	1.000	0.044	0.018	0.015
Design of the analysis: Macro						
Z&H	0.002	0.012	0.070	----	----	----
Replication	-0.052	0.034	0.769	-0.017	0.031	0.586
Replication (WLS-Weight3)	0.007	0.016	0.770	0.011	0.038	0.778
MIXED EVIDENCE						
Countries examined: Developed						
Z&H	0.201	0.022	1.000	0.176	0.029	0.000
Replication	-0.001	0.003	0.046	----	----	----
Replication (WLS-Weight3)	0.012	0.006	0.985	0.013	0.011	0.216
Countries examined: Developing and transition						
Z&H	0.107	0.017	1.000	0.099	0.026	0.000
Replication	0.000	0.002	0.030	----	----	----
Replication (WLS-Weight3)	0.015	0.007	0.984	0.016	0.013	0.227
Data Characteristics: Samplesize						
Z&H	-0.037	0.003	1.000	-0.024	0.009	0.007
Replication	-0.024	0.004	1.000	-0.011	0.016	0.477
Replication (WLS-Weight3)	-0.016	0.004	1.000	-0.017	0.012	0.142
Publication characteristics: IFRecursive						
Z&H	0.106	0.053	0.875	0.096	0.048	0.043
Replication	0.146	0.031	0.999	0.074	0.053	0.166
Replication (WLS-Weight3)	-0.002	0.011	0.750	-0.001	0.036	0.976

Type of Replication	Bayesian Model Averaging			Frequentist Regression		
	Post. mean	Post. SD	PIP	Coeff	Robust SE	p-value
			EVIDENCE AGAINST			
			Design of the analysis: Quadratic			
Z&H	-0.053	0.012	0.997	-0.044	0.013	0.001
Replication	-0.001	0.005	0.091	----	----	----
Replication (WLS-Weight3)	0.041	0.010	1.000	0.044	0.024	0.061
			Treatment of stability: Dummies			
Z&H	0.211	0.028	1.000	0.184	0.019	0.000
Replication	0.063	0.041	0.772	0.025	0.039	0.527
Replication (WLS-Weight3)	-0.004	0.008	0.785	-0.006	0.019	0.744
			Treatment of competition: Boone			
Z&H	-0.071	0.031	0.897	-0.058	0.023	0.010
Replication	-0.003	0.012	0.106	----	----	----
Replication (WLS-Weight3)	0.012	0.007	0.960	0.012	0.010	0.234
			Treatment of competition: Concentration			
Z&H	-0.018	0.023	0.474	----	----	----
Replication	0.131	0.018	1.000	0.090	0.027	0.001
Replication (WLS-Weight3)	-0.014	0.010	0.934	-0.017	0.016	0.306
			Control variables: Regulation			
Z&H	-0.032	0.020	0.798	-0.036	0.014	0.010
Replication	0.000	0.002	0.029	----	----	----
Replication (WLS-Weight3)	0.000	0.004	0.737	0.000	0.005	0.964
			Publication characteristics: Citations			
Z&H	0.050	0.009	1.000	0.046	0.009	0.000
Replication	0.055	0.006	1.000	0.036	0.017	0.033
Replication (WLS-Weight3)	-0.009	0.005	0.969	-0.010	0.014	0.454

Type of Replication	Bayesian Model Averaging			Frequentist Regression		
	Post. mean	**Post. SD**	**PIP**	**Coeff**	**Robust SE**	**p-value**
			OTHER			
			Design of the analysis: Endogeneity			
Z&H	0.010	0.021	0.237	----	----	----
Replication	0.042	0.012	0.990	0.016	0.021	0.435
Replication (WLS-Weight3)	-0.008	0.006	0.917	-0.009	0.009	0.269

NOTE: The left side of the table reports BMA results for the respective variable. "Post. Mean" is posterior mean, "Post. SD" is posterior standard deviation, and "PIP" is posterior inclusion probability. The right side reports the results of a frequentist regression where PCC is regressed on the "best specification" variables, where "best specification" variables are those that have a PIP greater than 0.5 in the BMA analysis. The values in the frequentist regression are the estimates of the given variable's coefficient. "Z&H" reproduces Z&H's estimates, weighting observations by the inverse of the number of estimates per study. "Replication" applies the same analysis to the combined Replication dataset. "Replication (WLS-Weight3)" weights observations according to the inverse of the variance of the estimated effect. Standard errors are clustered at the study level. "----" in the frequentist regression indicates that the variable was not included in the "best specification" and hence not included in the associated frequentist regression. Coefficient standard errors in the frequentist regression are clustered at the study level. Categorization into "Evidence For", "Mixed Evidence", and "Evidence Against" follows the criteria listed in TABLE 6.

APPENDIX 4
Replication Results for Z&H's BMA Analysis: Other Variables

Type of Replication	Bayesian Model Averaging			Frequentist Regression		
	Post. mean	Post. SD	PIP	Coeff	Robust SE	p-value
Data Characteristics: SEPCC						
Z&H	-1.788	0.205	1.000	-1.194	0.651	0.067
Replication	-1.325	0.223	1.000	-0.722	1.071	0.500
Replication (WLS-Weight3)	-2.096	NA	1.000	-2.263	1.399	0.106
Data Characteristics: T						
Z&H	0.001	0.004	0.052	----	----	----
Replication	0.000	0.003	0.035	----	----	----
Replication (WLS-Weight3)	-0.002	0.004	0.774	-0.003	0.007	0.693
Data Characteristics: Sampleyear						
Z&H	0.000	0.001	0.046	----	----	----
Replication	0.005	0.002	0.973	0.007	0.002	0.003
Replication (WLS-Weight3)	0.002	0.001	0.999	0.003	0.002	0.182
Design of the analysis: Averaged						
Z&H	-0.000	0.005	0.040	----	----	----
Replication	0.001	0.005	0.032	----	----	----
Replication (WLS-Weight3)	0.027	0.014	0.981	0.031	0.023	0.187
Treatment of stability: NPL						
Z&H	0.002	0.006	0.132	----	----	----
Replication	0.000	0.002	0.026	----	----	----
Replication (WLS-Weight3)	0.025	0.008	1.000	0.027	0.017	0.115
Treatment of stability: Z-score						
Z&H	-0.000	0.003	0.063	----	----	----
Replication	0.000	0.002	0.034	----	----	----
Replication (WLS-Weight3)	0.021	0.006	1.000	0.022	0.014	0.122
Treatment of stability: Profit volatility						
Z&H	0.001	0.005	0.037	----	----	----
Replication	0.001	0.007	0.049	----	----	----
Replication (WLS-Weight3)	0.030	0.007	1.000	0.032	0.015	0.032

Type of Replication	Bayesian Model Averaging			Frequentist Regression		
	Post. mean	Post. SD	PIP	Coeff	Robust SE	p-value
			Treatment of stability: Profitability			
Z&H	-0.000	0.003	0.035	----	----	----
Replication	-0.003	0.011	0.094	----	----	----
Replication (WLS-Weight3)	0.011	0.008	0.944	0.013	0.015	0.397
			Treatment of stability: Capitalization			
Z&H	0.000	0.003	0.027	----	----	----
Replication	-0.001	0.007	0.042	----	----	----
Replication (WLS-Weight3)	0.013	0.008	0.960	0.014	0.019	0.456
			Treatment of stability: DtoD			
Z&H	-0.001	0.008	0.050	----	----	----
Replication	-0.165	0.039	0.997	-0.098	0.042	0.020
Replication (WLS-Weight3)	0.037	0.012	1.000	0.039	0.027	0.141
			Treatment of competition: Lerner			
Z&H	0.004	0.013	0.122	----	----	----
Replication	-0.008	0.015	0.236	----	----	----
Replication (WLS-Weight3)	-0.024	0.006	1.000	-0.026	0.010	0.011
			Treatment of competition: HHI			
Z&H	0.002	0.011	0.085	----	----	----
Replication	0.031	0.017	0.805	0.048	0.030	0.107
Replication (WLS-Weight3)	-0.003	0.005	0.803	-0.005	0.007	0.506
			Estimation method: Logit			
Z&H	-0.187	0.023	1.000	-0.160	0.019	0.000
Replication	-0.129	0.025	1.000	-0.051	0.043	0.235
Replication (WLS-Weight3)	0.006	0.009	0.827	0.008	0.012	0.503
			Estimation method: OLS			
Z&H	-0.035	0.024	0.756	-0.038	0.018	0.038
Replication	-0.143	0.013	1.000	-0.062	0.028	0.027
Replication (WLS-Weight3)	-0.019	0.010	0.972	-0.021	0.018	0.248

Type of Replication	Bayesian Model Averaging			Frequentist Regression		
	Post. mean	Post. SD	PIP	Coeff	Robust SE	p-value
Estimation method: FE						
Z&H	0.011	0.021	0.277	----	----	----
Replication	-0.058	0.014	0.996	-0.031	0.021	0.139
Replication (WLS-Weight3)	-0.001	0.009	0.762	-0.001	0.014	0.937
Estimation method: RE						
Z&H	0.002	0.011	0.058	----	----	----
Replication	0.000	0.004	0.030	----	----	----
Replication (WLS-Weight3)	0.016	0.011	0.950	0.018	0.017	0.312
Estimation method: GMM						
Z&H	-0.000	0.003	0.040	----	----	----
Replication	-0.002	0.007	0.086	----	----	----
Replication (WLS-Weight3)	-0.038	0.010	1.000	-0.041	0.022	0.059
Estimation method: TSLS						
Z&H	-0.000	0.003	0.032	----	----	----
Replication	0.001	0.004	0.035	----	----	----
Replication (WLS-Weight3)	-0.004	0.008	0.792	-0.004	0.011	0.707
Control variables: Ownership						
Z&H	-0.015	0.017	0.481	----	----	----
Replication	-0.053	0.010	1.000	-0.038	0.025	0.119
Replication (WLS-Weight3)	-0.007	0.008	0.857	-0.009	0.011	0.418
Control variables: Global						
Z&H	-0.002	0.005	0.116	----	----	----
Replication	0.073	0.011	1.000	0.049	0.020	0.013
Replication (WLS-Weight3)	0.014	0.005	1.000	0.015	0.007	0.038
Publication characteristics: Firstpub						
Z&H	0.022	0.004	1.000	0.023	0.003	0.000
Replication	0.008	0.002	1.000	0.004	0.004	0.243
Replication (WLS-Weight3)	0.005	0.001	1.000	0.005	0.004	0.179

Type of Replication	Bayesian Model Averaging			Frequentist Regression		
	Post. mean	Post. SD	PIP	Coeff	Robust SE	p-value
	Publication characteristics: Reviewed journal					
Z&H	-0.025	0.019	0.725	-0.015	0.014	0.289
Replication	-0.153	0.010	1.000	-0.087	0.027	0.001
Replication (WLS-Weight3)	-0.012	0.007	0.959	-0.012	0.017	0.478

NOTE: The left side of the table reports BMA results for the respective variable. "Post. Mean" is posterior mean, "Post. SD" is posterior standard deviation, and "PIP" is posterior inclusion probability. The right side reports the results of a frequentist regression where PCC is regressed on the "best specification" variables, where "best specification" variables are those that have a PIP greater than 0.5 in the BMA analysis. The values in the frequentist regression are the estimates of the given variable's coefficient. "Z&H" reproduces Z&H's estimates, weighting observations by the inverse of the number of estimates per study. "Replication" applies the same analysis to the combined Replication dataset. "Replication (WLS-Weight3)" weights observations according to the inverse of the variance of the estimated effect. Standard errors are clustered at the study level. "----" in the frequentist regression indicates that the variable was not included in the "best specification" and hence not included in the associated frequentist regression. Coefficient standard errors in the frequentist regression are clustered at the study level.