

NBER WORKING PAPER SERIES

FIFTY SHADES OF QE:
COMPARING FINDINGS OF CENTRAL BANKERS AND ACADEMICS

Brian Fabo
Martina Jančoková
Elisabeth Kempf
Luboš Pástor

Working Paper 27849
<http://www.nber.org/papers/w27849>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2020, Revised February 2021

The paper previously circulated under the title "Fifty Shades of QE: Conflicts of Interest in Economic Research." The views in this paper are those of the authors and do not represent the views of the European Central Bank, the National Bank of Slovakia, any other institution, or the National Bureau of Economic Research. Pástor is a member of the bank board of the National Bank of Slovakia. We are grateful for comments from Wenxin Du, Mara Faccio, Pavel Gertler, Daniel Gros, Urban Jermann, Sujit Kapadia, Anil Kashyap, Luc Laeven, Ludovít Ódor, Raghuram Rajan, Kasper Roszbach, Martin Šuster, and Tomasz Wieladek, as well as to seminar participants at the European Central Bank, Košice Economic Forum, National Bank of Slovakia, Purdue University, Toulouse School of Economics, University of Chicago, University of Geneva, and University of Miami. We are also grateful to Livia Amato, Yuliia Kazmina, Fulin Li, Eva Štulrajterová and, especially, Bianca He for excellent research assistance. This research was supported by the Fama-Miller Center for Research in Finance at Chicago Booth.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2020 by Brian Fabo, Martina Jančoková, Elisabeth Kempf, and Luboš Pástor. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Fifty Shades of QE: Comparing Findings of Central Bankers and Academics
Brian Fabo, Martina Jančoková, Elisabeth Kempf, and Ľuboš Pástor
NBER Working Paper No. 27849
September 2020, Revised February 2021
JEL No. A11,E52,E58,G28

ABSTRACT

We compare the findings of central bank researchers and academic economists regarding the macroeconomic effects of quantitative easing (QE). We find that central bank papers find QE to be more effective than academic papers do. Central bank papers report larger effects of QE on output and inflation. They also report QE effects on output that are more significant, both statistically and economically, and they use more positive language in the abstract. Central bank researchers who report larger QE effects on output experience more favorable career outcomes. A survey of central banks reveals substantial involvement of bank management in research production.

Brian Fabo
Narodna banka Slovenska
Imricha Karvasa 1
Bratislava 81325
Slovakia
brian.fabo@nbs.sk

Elisabeth Kempf
Booth School of Business
University of Chicago
5807 South Woodlawn Avenue
Chicago, IL 60637
elisabeth.kempf@chicagobooth.edu

Martina Jančoková
European Central Bank
Sonnemannstraße 20
Frankfurt am Main 60314
Germany
martina.jancokova@gmail.com

Ľuboš Pástor
University of Chicago
Booth School of Business
5807 South Woodlawn Ave
Chicago, IL 60637
and NBER
lubos.pastor@chicagobooth.edu

1. Introduction

Since the 2008 financial crisis, central banks around the world have deployed unconventional monetary policy tools such as quantitative easing, forward guidance, and long-term refinancing operations. The popularity of these tools has grown since the outbreak of the COVID-19 pandemic. For example, the Federal Reserve, the European Central Bank, and the Bank of England all announced new large-scale asset purchases in March 2020.

The effectiveness of unconventional monetary policy (“QE”) has been a subject of intense debate in both academic and policy circles. A significant part of the research on QE originates in central banks (Martin and Milas (2012)). This research, which is widely cited in the media, often finds QE to be effective.¹ However, it has an aspect of self-assessment: when central banks evaluate QE, they are judging their own policy. Whether this aspect has any bearing on research output is an empirical question that we address in this paper.

We compare the findings of central bank researchers (“central bankers”) and academic economists (“academics”) regarding the effectiveness of QE. We construct a dataset comprising 54 studies that analyze the effects of QE on output or inflation in the U.S., UK, and the euro area. For each study, we record its baseline estimates of the effects of QE on the level of GDP and the price level, along with their significance. We also collect a variety of other study-specific information, such as publication status and methodology used, as well as detailed biographical information of the 116 different authors. We then compare the findings of studies written by central bankers with those written by academics.

We find that central bank papers report larger effects of QE on both output and inflation. Central bank papers are also more likely to report QE effects on output that are significant, both statistically and economically. For example, while *all* of the central bank papers report a statistically significant QE effect on output, only *half* of the academic papers do. In addition, central bank papers use more favorable language—more positive adjectives and, to a lesser extent, fewer negative adjectives—in their abstracts. Overall, central bank papers find QE to be more effective than academic papers do.

¹For example: “The good news is that, by most accounts, QE appears to have succeeded at boosting growth and lifting inflation. Martin Weale, a member of the BoE’s interest-rate setting Monetary Policy Committee, found asset purchases worth 1% of national income boosted UK gross domestic product by about 0.18% and inflation by 0.3%. A study by John Williams, president of the San Francisco Federal Reserve, concluded that asset purchases had reduced the US unemployment rate by 1.5 percentage points by late 2012 and helped the economy avoid deflation.” The Financial Times (2015).

We also uncover differences in methodological choices. For example, central bank papers are more likely to use dynamic stochastic general equilibrium (DSGE) models rather than vector autoregression (VAR) models. Yet our main result—that central bankers are more optimistic than academics in their assessments of QE—continues to hold even when we control for model choice. Differences in research quality are also unlikely to explain our results because the gap between central bankers and academics is very similar when we condition on published papers only, as well as when we weight each paper by its citations. Our results are robust to the inclusion of various controls, and they are not driven by central bankers from any single country, nor by QE programs in any single country.

To explore a possible mechanism, we relate central bankers' research findings to their subsequent career outcomes. We collect employment histories for all central bank authors and convert their job titles to numerical ranks on a six-point scale. For each author-paper pair, we measure the author's subsequent career outcome by the first change in the author's rank following the paper's first public release. We find that authors whose papers report larger effects of QE on output experience more favorable career outcomes. A one standard deviation increase in the estimated effect is associated with a career improvement of about half a rank, such as moving halfway from *Economist* to *Senior Economist*. This evidence suggests a potential role for career concerns in explaining our results.

These concerns appear to be stronger for senior central bankers because for them, we find a stronger relation between the estimated QE effects and subsequent career outcomes. Motivated by this finding, we look whether the gap between the findings of central bankers and academics is larger for papers whose authors are more senior. We find that it is, though only marginally so. Our results are consistent with the idea that senior central bankers report larger effects of QE because they have a stronger incentive to do so.

Not all central bankers face the same incentives. Top management of the German Bundesbank has taken a critical view of QE, especially in the context of the European Central Bank (ECB). Former Bundesbank officials Axel Weber and Jürgen Stark reportedly quit their ECB positions in protest over QE, and the current Bundesbank president, Jens Weidmann, has also publicly opposed it. Mindful of their bosses' views, Bundesbank researchers could potentially face career concerns very different from those of their colleagues at other central banks. Indeed, we show that studies co-authored by Bundesbank employees find QE to be *less* effective at raising output compared to academic studies. While this evidence is

weak statistically, it is suggestive of managerial influence on research outcomes.

To shed more light on this influence, we survey heads of research at the world’s leading central banks. We have received responses from 24 central banks employing over 750 research economists in total. These responses reveal substantial involvement of bank management in research production. In most banks, management participates in the selection of research topics, typically by negotiating with the researcher. Direct topic assignments occur “sometimes” (“often”) in 50% (21%) of the responding banks. In most banks, research papers are reviewed by management prior to public distribution; such reviews happen “always” (“often”) in 38% (21%) of the responding banks. Management also approves papers for public distribution: typically by the head of research (“always” in 67% of the banks), but sometimes also by the bank board (at least “sometimes” in 33% of the banks). Unlike central bankers, academics face little if any managerial interference in their research.

As we note earlier, one possible mechanism behind our results is that central bankers face career concerns. A central bank economist may worry that the nature of her findings could threaten her employment status or rank. Such a concern could affect research outcomes even if it is completely unfounded as long as the economist perceives a nonzero probability of such a threat. This channel could operate at multiple levels because not only researchers but also their superiors want to get promoted. For example, a head of research may be reluctant to defend a subordinate’s inconvenient findings in front of the bank’s board.

Besides career concerns, a central bank economist may worry that bank management could block the release of studies that find the bank’s policy to be ineffective, or to have undesirable side effects. A recent example, from a different public institution, is the controversial release of Andersen et al. (2020). That study finds that World Bank payouts of foreign aid are followed by jumps in the recipient countries’ deposits in financial havens, suggesting leaks to the pockets of the countries’ elites. According to *The Economist* (2020), after the study passed an internal peer review at the World Bank, it was “blocked by higher officials.” After a substantial delay, the study was eventually released in February 2020.²

A central bank economist may also care about the bank’s reputation, favoring conclusions

²In their evaluation of World Bank research, Banerjee et al. (2006) argue: “Internal research that was favorable to Bank positions was given great prominence, and unfavorable research ignored... there was a serious failure of the checks and balances that should separate advocacy and research.” Similarly, in their independent review of the Bank of International Settlements’ research, Allen et al. (2016) note “a tendency for analysis and research to be slanted to support the ‘house view’, especially in regard to monetary policy.”

that validate the bank's actions. The economist may even care about her own reputation if she is senior enough to have participated in the formation of the bank's policy. For example, Bernanke (2020) offers a strong endorsement of QE. Given his unique experience, Ben Bernanke is exceptionally qualified to assess the effectiveness of QE. At the same time, QE is an important part of his legacy as it was adopted while he was Fed Chair.

A more benign explanation for our results, one that does not involve incentives, relies on differences in prior beliefs combined with selection. Researchers who believe in the power of policy interventions could self-select into policy institutions such as central banks, whereas policy skeptics could end up in academia. Researchers could then favor evidence supporting their priors. Moreover, their priors could be reinforced during the research process, either through the confirmation bias (Nickerson (1998)) or through the feedback received from colleagues, whose priors may be similar.

A by-product of our work is a meta-analysis of the macroeconomic effects of QE. Averaging across all 54 studies and standardizing QE program size to 1% of the country's GDP, QE increases the output level (price level) by 0.24% (0.19%) at the peak. The average cumulative effect on output (prices) is 58% (63%) of the peak effect. About 88% (84%) of studies estimate statistically significant effects on output (prices). Across the three regions studied, QE is the most effective in the U.S., in terms of raising both output and prices.

Our study is related to the literature inspecting the credibility of scientific research (e.g., Ioannidis (2005); Fanelli (2009)). It is well known that industry-sponsored scientists may act in the interests of their sponsors. This problem has been extensively documented in biomedical research. Many studies find that research sponsored by the pharmaceutical industry tends to draw pro-industry conclusions (e.g., Bekelman et al. (2003)). A similar bias could potentially affect central bankers; in fact, while academic medical researchers are merely *sponsored* by industry, central bank economists are directly *employed* by central banks. Central banks evaluating their own policies is not unlike pharmaceutical firms evaluating their own drugs. Both have skin in the game. The problem is particularly acute for central banks that view their research as part of their own policy, because research supportive of a policy could potentially enhance the policy's effectiveness. On the other hand, alleviating this problem is the strong desire of central banks to protect their reputation.

Academic economists who judge central bank policies may not have skin in the game, but they have their own incentive to find strong results because they face the pressure to

publish. Academics' career concerns are commonly summarized as "publish or perish." These concerns seem weaker for central bankers, who can often substitute policy work for journal publications. The need to publish creates a pressure for academics to find significant results because of the well-known publication bias: journals are more likely to publish positive results than negative ones (e.g., Fanelli (2010a)). This bias is particularly strong in economics (e.g., De Long and Lang (1992); Fanelli (2010b)). Some authors do not submit null findings (Franco, Malhotra and Simonovits (2014)); others inflate the values of just-rejected tests by choosing "significant" specifications (Brodeur et al. (2016)). Ioannidis et al. (2017) argue that many results in the economic literature are exaggerated.

The publication bias is not the only thorn in the side of economic research. Academics do not always disclose their private financial affiliations (Carrick-Hagenbarth and Epstein (2012)). Zingales (2013) discusses how academic research could be corrupted by economists' outside employment opportunities or their desire to gain access to proprietary data. Reported estimates of policy-relevant parameters, such as fiscal multipliers, reflect the authors' national backgrounds and political orientation (Asatryan et al. (2020); Jelveh et al. (2018)). Other problems include scientific misconduct (List et al. (2001)) and the inability to replicate economic findings (Christensen and Miguel (2018)).

Our study is also related to the literature on career concerns. This large literature finds evidence of such concerns not only for private-sector workers such as analysts and executives, but also for public-sector workers such as banking regulators (Lucca et al. (2014)) and federal government employees (Blanes i Vidal et al. (2012)). In contrast, there is little work on the incentives of central bankers. That work focuses mostly on the voting members of a central bank's monetary policy committee (e.g., Sibert (2003); Hansen et al. (2018)). We are not aware of any prior work on the incentives or biases of central bank research economists.

2. Data

We construct a dataset comprising studies of the effects of unconventional monetary policy on output and inflation. We aim to cover all studies, published and unpublished, that analyze the policy effects for at least one of three economies: the United States (US), the United Kingdom (UK), and the euro area (EA). For ease of exposition, we refer to these economies, including EA, as "countries." We include papers studying the EA as a whole, but not papers studying individual European countries. We focus on papers containing a

quantitative analysis, either model-based or empirical, of the effects on output, inflation, or both. We restrict our attention to output and inflation because these macroeconomic variables are of primary interest to central banks. We do not consider studies of the effects of policy on asset prices unless they also analyze the effects on output or inflation.

To identify the papers, we began by manually searching for 40 keywords in the Google Scholar and RePEc IDEAS databases. The keywords are listed in the Appendix, which is available on the authors' websites. These keywords cover not only quantitative easing but also other unconventional monetary policy tools that operate through central bank balance sheets, such as long-term refinancing operations. Since about 80% of the papers in our sample study quantitative easing, we refer to all papers as "QE" studies, for brevity.

We conducted the search in July and August of 2019. To allow for a lag between a paper's public release and its indexing in the databases, we included only papers released prior to July 2018. For each of the 40 keywords, we selected all papers on the first 10 pages of search results in both portals, with each page containing about 10 papers. All of these papers were independently evaluated for inclusion by two economists. Both of them read each paper's title, abstract, and introduction to assess whether the paper contains the kind of analysis described earlier. This analysis produced the first tier of papers in our sample.

To construct the second tier, we proceeded in three steps. First, we selected all papers cited in the references of all first-tier papers. Second, we selected all papers that cite any of the first-tier papers, using the Google Scholar functionality to query for all documents citing a specified paper. Third, for each paper selected in the first two steps, we read its title, abstract, and introduction to determine its eligibility. Those papers deemed eligible constitute the second tier of papers in our sample. We repeated this procedure one more time, creating a third tier. While it is possible for our procedure to have missed a relevant paper, such a paper would have to escape our deep, 40-keyword, two-portal search, not cite any of our sample papers, and not be cited by any of our sample papers.

We include only papers that contain original quantitative estimates, thus excluding review papers. We exclude papers not written in English, as well as master's and bachelor's theses. We include all other types of papers: journal publications, working papers, book chapters, and policy papers. Our final sample consists of 54 papers written by 137 authors, 116 of whom are unique. All 54 papers are listed in the Appendix.

For each paper, we collect information on the year of first public distribution, year of

journal publication (if any), publication outlet, authors' names, and the methodology used (e.g., a DSGE or VAR model). We obtain impact factors and article influence scores from Clarivate Analytics Web of Science for the year of the paper's publication.

We record the effects of QE on the level of GDP and the price level as implied by the authors' baseline model. We distinguish four estimated effects: (i) the *peak* effect of the QE program studied (*Total Peak Effect*); (ii) the *cumulative* effect of QE, defined as the effect at the end of the time period studied by the authors (*Total Cumulative Effect*); (iii) the *peak* effect after standardizing QE program size to 1% of GDP (*Standardized Peak Effect*); and (iv) the *cumulative* effect after the same standardization (*Standardized Cumulative Effect*). The construction of all four variables is described in detail in Section 2.2. We code all variables based on the published version of the paper if such a version is available at the time of our search. If the paper is unpublished at the time of our search, then we code all variables based on the paper's most recent version that we can find online.

We also record the authors' assessments of the statistical and economic significance of their estimated effects of QE. Whenever available, we use the authors' own verbal assessment of statistical significance. If unavailable, we infer statistical significance from confidence intervals reported in the corresponding figure or table, using the peak effect. We also gather the confidence level used by authors to assess significance (e.g., 95%, 90%, or 68%). When statistical significance is not discussed and no standard errors or p -values are reported, we set the variable to missing. For economic significance, we always use the authors' own verbal assessment. For example, if a study states the effect of QE is "negligible", we code economic significance as zero; if the effect is "sizable", we code it as one. For ambiguous cases, we code economic significance as 0.5. Examples include studies stating that the effect of QE is positive upon impact but disappears quickly, or that it is positive but sensitive to model specification. When economic significance is not discussed, we set it to missing.

Finally, we manually collect information on the employment history, job titles, and educational background for the 116 authors by using online searches and information from public LinkedIn pages. To determine author affiliation, we use the author's main employer at the time of the paper's first public distribution, as determined by our search in the summer of 2019. We categorize all authors whose primary affiliation is a central bank as central bankers. We classify authors from the Bank of International Settlements (BIS) as 0.5 central

bankers due to the close ties between the BIS and the central banking community.³ We refer to all other authors as academics.

To maximize the quality of our dataset, two teams of two economists independently went through all 54 studies and constructed the key quantities of interest, including the estimated effects of QE and their significance. The teams then compared notes and discussed all controversial cases before reaching convergence. Furthermore, to facilitate replication by other researchers, we disclose our full paper-level dataset in the Appendix.

2.1. Summary Statistics

In this section, we briefly summarize selected features of our data. In the Appendix, we provide a more detailed data description in the form of tables and bar charts.

The papers in our sample appear in each year between 2010 and 2018, with three to ten papers per year. They do not have to contain estimates for both output and inflation, but about 90% of them do. More than 57% of the papers are published in peer-reviewed journals. The average impact factor of those journals at the time of the paper’s publication is 1.42. 35% of the papers use DSGE models. The average paper is written by 2.54 authors and it studies the effects of QE in 1.26 countries. The euro area receives the most attention, but each of the three countries is studied by at least 13 papers.

As for the authors, 60% of them are primarily affiliated with a central bank. Central banks employing the most authors are the Bank of England (21), EA national central banks (20), the ECB (17), and the Federal Reserve (16). Academics are employed mostly at universities in Europe (18), UK (10), and the U.S. (9). 17% of the authors are female and 89% hold a PhD degree. Most authors have earned their PhD’s at prestigious universities in the U.S. and UK such as Princeton (8) and Cambridge (6). The average author experience (i.e., the number of years since earning the highest educational degree) is 11 years.

2.2. The Effects of QE on Output and Inflation

For each paper and country studied, we record the estimated effects of QE on output (i.e., real GDP or industrial production) and inflation (i.e., CPI) based on the authors’ baseline specification. As a rule, we record the effects on the *level*—the level of output and

³In the Appendix, we show that our main results are robust to classifying BIS authors as full central bankers. They are also robust to classifying researchers at the International Monetary Fund and the World Bank as 0.5 central bankers, although such an alternative classification seems harder to justify.

the price level. Letting Y denote the actual level of the outcome variable (i.e., with QE) and \hat{Y} denote its counterfactual level (i.e., without QE), we are interested in the percentage difference, $(Y - \hat{Y})/\hat{Y}$. If the paper reports the effect of QE on the level of output or prices, we record the peak and cumulative effects as displayed in Figure 1. If the paper reports only the effects on the growth rate, we sum up the individual growth estimates to determine the impact on the level. We describe the details of this conversion, and list the estimated effects for each paper-country pair in our sample, in the Appendix.

We focus on the effect most prominently advertised in the paper, ignoring estimates from robustness checks, alternative specifications, and extensions. We standardize the effects to a QE program size equal to 1% of the respective country's GDP at the time QE was first introduced. For the U.S. and UK, we use 1% of the annualized 2009 Q1 GDP, consistent with Weale and Wieladek (2016). For the EA, whose asset purchase programs started in 2015 Q1, we use 1% of the annualized 2015 Q1 GDP. We obtain GDP estimates from the FRED database. Performing the standardization also requires the size of aggregate asset purchases for each QE program. We report our estimates of these sizes in the Appendix. Following Weale and Wieladek (2016), we include Treasury purchases for the U.S. programs, and all securities purchased under the Asset Purchasing Facility for the UK programs. For the EA, program size includes all securities purchased under the Asset Purchase Program, because asset-backed securities are a small fraction of the overall program size.

We show the means, medians, and standard deviations of the estimated QE effects in the Appendix. The average (median) paper in our sample estimates that QE increases output by 1.57% (1.25%) at the peak. Standardized to a QE program size equivalent to 1% of GDP, the average (median) peak effect on output is 0.24% (0.16%). As for inflation, the average (median) study finds that QE raises the price level by 1.42% (0.93%) at the peak. Standardizing to 1% of GDP, the average (median) effect on the price level is 0.19% (0.11%). For both output and inflation, the average standardized cumulative effect is equal to about 60% of the average standardized peak effect, indicating that about 40% of the peak effect vanishes by the end of the period studied.

There is substantial heterogeneity in the effectiveness of QE across the three countries. Focusing on standardized effects, which are easier to compare across countries due to differences in QE program sizes, QE is most effective at raising output in the U.S., followed by the EA and UK. For inflation, QE is again the most effective in the U.S.

Most studies conclude that the estimated effects of QE on output and prices are positive and statistically significant, consistent with prior reviews of this literature (e.g., Dell’Ariccia et al. (2018)). However, there is also substantial heterogeneity in point estimates. Understanding whether some of this heterogeneity is related to the institutional environment in which authors operate is the goal of the following sections.

3. Research Outcomes and Central Bank Affiliation

This section conducts a systematic comparison of the research findings of central bankers and academics regarding the effectiveness of QE. These findings include the estimated effects of QE on output and inflation, the statistical and economic significance of these effects, and the tone of the language used to summarize the paper’s results.

3.1. The Effect of QE on Output

Figure 2 reports histograms for the estimated effects of QE on the output level, separately for studies with at least one central bank author (“CB”) and those with no such authors (“Not CB”). The four panels correspond to the four measures introduced previously. For all of them, the distributions of central bank papers are shifted visibly to the right, indicating that such papers find systematically larger effects of QE on output.

The same result follows from Panel A of Table 1, which compares the means and medians of the estimated effects of QE on output across papers with and without at least one central bank author. Both types of papers find QE to be successful at raising output, on average, but central bank papers find substantially larger effects. This is true based on both means and medians, indicating that the gap is not driven by outliers. Moreover, the outliers in Figure 2 do not seem to be low-quality papers, judging by their publication success. Among the five papers finding the largest effects on output, the publication rate is 100%.

Table 2 confirms the result based on regression evidence. We regress the estimated output effect on the share of central bank authors, *CB Affiliation*, defined as the share of authors who are affiliated with a central bank at the time of the paper’s first public distribution. In the strictest specifications, shown in columns (3) and (6), we also include country fixed effects and controls for the number of authors and average author experience:

$$y_{ij} = \alpha_j + \beta [\text{CB Affiliation}]_i + \gamma' X_i + \epsilon_{ij} , \tag{1}$$

where y_{ij} is the effect of QE on output estimated by study i for country j 's QE, α_j is a fixed effect for the country in which QE takes place, and X_i are the two controls. The latter control is $\log(3 + \text{average author experience})$. We add three to ensure the logarithm is always well defined, given that the minimum value of author experience in our sample is -2 , for an author who wrote their paper two years prior to earning a Ph.D.

Columns (3) and (6) show that changing the share of central bank authors from zero to 100% is associated with a 0.723 percentage points larger peak effect and a 0.512 percentage points larger cumulative effect on output (Panel A). These are sizable magnitudes relative to the unconditional means of 1.57% and 0.87%, respectively. The results based on standardized effects, reported in Panel B, are also economically large. Going from zero to 100% central bank authors corresponds to a 0.152 percentage points larger standardized peak effect: an increase by two thirds of the unconditional mean. For the standardized cumulative effect, the difference is 0.122 percentage points, equivalent to 87% of the unconditional mean. These results show that the differences in research findings observed in Panel A are not due to central bankers studying larger QE programs.

To assess statistical significance, we report t -statistics based on standard errors clustered at the paper level. By clustering in this way, we allow the residuals ϵ_{ij} in equation (1) to be correlated within papers that analyze multiple QE programs. Such papers make various choices, methodological and expositional, that can affect their assessments of all QE programs. Therefore, we model ϵ_{ij} as having a group structure: $\epsilon_{ij} = v_i + \eta_{ij}$, where v_i is a random component specific to paper i and η_{ij} are mean-zero and uncorrelated.

A potential concern about inference based on cluster-robust standard errors is that the variance estimator converges to the true value as the number of clusters goes to infinity, whereas we have at most 54 clusters. To address this concern, we implement a wild cluster bootstrap procedure (Cameron et al. (2008)). To compute the p -values, we use the post-estimation command *boottest* developed by Roodman et al. (2019), assuming the null hypothesis and using Webb weights and 10,000 replications. We report these p -values in square brackets. We generally use these p -values, which tend to be more conservative than t -statistics, to assess statistical significance. In Panel A of Table 2, the coefficient on *CB Affiliation* is significant at either the 5% level or the 10% level for the peak effect, but it is insignificant for the cumulative effect. In Panel B, the coefficient is significant at the 5% level in all columns except for column (3), where it is significant at the 10% level.

3.2. The Effect of QE on Inflation

Figure 3 reports histograms for the estimated effects of QE on the price level, analogous to the histograms for output plotted in Figure 2. Just like for output, the distributions of central bank papers are shifted to the right relative to academic papers, indicating that central bank papers tend to find QE to be more effective at raising prices. The same conclusion follows from Panel B of Table 1, which compares the means and medians of the estimated effects of QE on inflation across papers with and without central bank authors.

In Table 3 we repeat the regression analysis from Table 2, but with a different dependent variable: y_{ij} in equation (1) is now the estimated effect on inflation rather than output. According to Panel A, columns (3) and (6), changing the share of central bank authors from zero to 100% corresponds to a 1.279 percentage points larger peak effect and a 1.394 percentage points larger cumulative effect on prices. These effects are large relative to the unconditional means of 1.42% and 0.89%, respectively. In Panel B, the coefficients on *CB Affiliation* are even larger than the corresponding unconditional means: 0.201 percentage points for the peak effect and 0.190 percentage points for the cumulative effect. All of the coefficients in Table 3 are statistically significant at the 5% level.

3.3. Significance

Next, we examine the statistical and economic significance of the effects of QE on output and inflation. Our main interest is in whether studies by central bankers and academics differ in their assessments of this significance. One advantage of looking at significance is that it is directly comparable across studies with no need for any standardization or conversion in the construction of the peak and cumulative effects of QE.

We first compute the shares of studies that find a statistically significant effect of QE on output, separately for central bankers and academics. The difference is striking: while *half* of the academic papers find a significant effect, *all* of the central bank papers do. The difference in the assessments of economic significance is almost equally large. For inflation, the differences are smaller but still notable; for example, while 75% of the academic papers find a statistically significant effect, about 90% of the central bank papers do.

Table 4 shows the extent to which these differences are statistically significant and robust to the inclusion of control variables. We estimate the regression specification in equation (1), with y_{ij} redefined to denote either statistical or economic significance, for the effects of QE on either output or inflation. For output (Panel A), the coefficient on *CB Affiliation* is

always positive and significant at the 5% level, whether the dependent variable is statistical or economic significance. The magnitude of the effect is also large: the estimate in Panel A, column (3), implies that increasing the share of central bank authors from zero to 100% corresponds to a 36.6 percentage points higher likelihood of the study finding a statistically significant effect of QE on output. The magnitude is even larger, 39.9 percentage points, for economic significance of the QE effect on output. For inflation (Panel B), we also find economically large effects, but they are not statistically significant.

3.4. *Alternative Specifications*

We consider various modifications of our baseline regression (1), as analyzed in Sections 3.1 through 3.3. We summarize the results here and show the tables in the Appendix.

Recall that the main independent variable in regression (1), *CB Affiliation*, is the fraction of the paper’s authors who are affiliated with a central bank. Our results in Tables 2 through 4 hold also when we replace this granular measure by an indicator we call *Discrete*, which is equal to one if at least one of the authors is affiliated with a central bank or the BIS, and zero otherwise. In addition, we replace *CB Affiliation* by two zero/one indicators: *Mixed*, which is equal to one if the share of central-bank-affiliated authors is strictly between zero and one, and *Pure CB*, which equals one if all of the authors are central bankers. We find positive point estimates of the coefficients on both indicators in all 36 specifications considered in Tables 2 through 4. Moreover, the estimated slope on *Pure CB* exceeds that on *Mixed* in 33 of the 36 specifications, suggesting that central bankers tend to find larger effects of QE when they have no academic coauthors.

Different central banks may have different research-vetting policies. Motivated by this possibility, we separate central bank authors by the country of the bank they work for. We replace *CB Affiliation* in equation (1) by four zero/one indicators: *EA CB* is equal to one if at least one of the authors is affiliated with the ECB or a national central bank in the euro area, *UK CB* equals one if at least one author is affiliated with the Bank of England, *US CB* equals one if at least one author is at the Federal Reserve, and *Other CB* equals one if at least one author is at another central bank or the BIS. The omitted group is academics. We find positive point estimates of the slopes on all four indicators, suggesting that our results in Tables 2 to 4 are not driven by authors from any single country. Fed researchers tend to find the largest effects of QE on output, whereas Bank of England researchers tend to find the largest effects on inflation. Euro area central bankers find relatively weak QE

effects, largely due to the weaker effects reported by Bundesbank researchers (Section 4.3). However, the differences across central banks are not statistically significant.

Taking a different country-by-country perspective, we focus on the country in which QE takes place. We observe that the point estimates of β in equation (1) are generally positive for all three countries, though their statistical significance is mixed. Looking at the effects of QE on output, the β estimate is the largest, and significant, for QE conducted in the U.S. In other words, the gap between the output effects reported by central bankers and academics is largest when they analyze U.S. QE. For the effects of QE on inflation, the β estimates are large, and typically significant, in both the U.S. and UK.

Motivated by the strong results we find for the U.S., we dig deeper into them by considering the three main QE programs in the U.S. separately. The point estimates of β are positive and large for all three programs, but they are often insignificant because of small sample sizes: we have 12 studies of the output effects of QE1, 12 studies for QE2, and only 4 studies for QE3. The estimates are similar across the three programs, indicating that our results are not driven by any individual U.S. program.

We do not control for the specific QE program in our baseline regressions because the choice of which program to study is made by the authors. For example, if QE1 is perceived to have been more effective than QE2 or QE3, an author aiming to report stronger QE effects can choose to analyze the first round of QE rather than its later rounds. Nevertheless, we also report results when controlling for QE program dummies, thus comparing central bankers and academics analyzing the same QE program. For papers studying more than one program, more than one dummy is switched on at the same time. Adding QE program dummies tends to reduce the statistical significance of the results: the estimate of β is significant in 17 (22) specifications at the 5% (10%) level, out of all 36 specifications considered in Tables 2 through 4, whereas in those tables, β is significant in 25 (28) specifications at the 5% (10%) level. The decline in statistical significance is unsurprising because we can only consider QE programs studied by at least two papers. Nevertheless, even with QE program dummies, the point estimates have the same signs as their counterparts in Tables 2 through 4 in all specifications, and their magnitudes are economically significant.

We further explore whether central bankers are more optimistic when they study QE launched by their own central bank. We add to equation (1) a variable capturing the share of authors who are affiliated with the central bank of the QE program studied. We also include

paper fixed effects, thus comparing the effects of QE in different countries as estimated by the same paper. We do not find support for the hypothesis. Naturally, with paper fixed effects, statistical power is limited because few papers study multiple QE programs. Recall that the average (median) paper in our sample studies QE in only 1.26 (1) countries.

As noted earlier, our sample contains papers studying not only quantitative easing but also other unconventional monetary policy programs, such as long-term refinancing operations. When we exclude those other programs from the analysis, keeping only QE narrowly defined, we find results similar to those in Tables 2 through 4. We also find similar effects when we control for the time gap between the QE program studied and the year of the paper’s first release. This alleviates the concern that the reported differences could be driven by differences in the timing of studies by academics and central bankers.

3.5. *Tone*

We now compare the tone of the language that central bankers and academics use when they assess QE. We focus on the paper’s abstract, which summarizes the paper’s main findings in a non-technical manner. As we are unaware of any lexical sentiment model trained on the economic research literature, we create our own lexicon, which we show in the Appendix. We consider adjectives such as “significant,” “sizable,” and “large” as positive, conveying the message that QE is effective, and adjectives such as “small,” “negligible,” and “weak” as negative. We compute the shares of positive and negative adjectives out of all adjectives in the abstract. The abstract’s “sentiment score” is the share of positive adjectives minus the share of negative adjectives. We estimate the model

$$y_i = \delta_i^{US} + \delta_i^{UK} + \delta_i^{EA} + \beta [\text{CB Affiliation}]_i + \gamma' X_i + \epsilon_i, \quad (2)$$

where y_i is the sentiment score for the abstract of study i ; δ_i^{US} , δ_i^{UK} , and δ_i^{EA} are indicators equal to one if the study analyzes QE in the U.S., UK, or EA, respectively, and zero otherwise; and X_i are the same controls as in equation (1). If a paper studies QE in multiple countries, then multiple indicators are switched on.

Panel A of Table 5 shows that central bankers use more positive language than academics when describing their results. Column (3) shows that a 100 percentage point increase in the share of central bank authors is associated with an increase in the sentiment score of 0.056, which is equal to 85% of one standard deviation of the sentiment score. The result

is significant at the 5% level. In Panels B and C, we decompose the sentiment score into the shares of positive and negative adjectives, and we run the analysis separately for both shares. We find that central bank studies use both more positive adjectives and fewer negative adjectives. The finding based on positive adjectives is economically larger and, unlike the one based on negative adjectives, it is statistically significant.

That central bankers use more favorable language is not surprising given our results in Sections 3.1 through 3.3. Nonetheless, we find it reassuring that our main result is robust to using a different type of measurement, one based on text rather than numbers. Moreover, the estimates of β in equation (2) remain similar when we add controls for the magnitudes of the reported effects on output and inflation. Thus, central bankers use more favorable language than academics even when describing effects that are equally large. However, this result is weaker compared to Table 5. Recall that the β estimate in column (3) of Panel A of Table 5 is easily significant at the 5% level. The estimate remains significant at the 5% level when we add controls for output effects, standardized or not, but the p -value rises to about 0.1 when we control for inflation effects. See the Appendix for details.

We also analyze the text of the papers' conclusions. The results are similar to those based on the abstract—the point estimates of β are positive when the left-hand side variable is either the sentiment score or the fraction of positive adjectives, and they are negative for the fraction of negative adjectives—but the results are not statistically significant, as we show in the Appendix. Compared to the abstract, the conclusions usually contain more discussion unrelated to the paper's core contribution, such as directions for future work.

The Appendix also shows results from the analysis that computes the abstract's sentiment score based on two alternative dictionaries: the Harvard IV4 semantic dictionary and the Loughran and McDonald (2011) financial dictionary. We do not find significant differences between central bankers and academics based on those dictionaries. However, we find the results based on our simple dictionary far more credible because we designed it specifically for economic research. In contrast, the Harvard IV4 dictionary is designed for use in a variety of contexts outside economics, and the Loughran-McDonald dictionary is designed for the analysis of 10-K reports of publicly-traded companies. The positive and negative labels assigned to words in these dictionaries do not reflect the meaning of these words in the economics literature. As a result, these dictionaries do not contain key adjectives that clearly indicate positive language in our context, such as “significant,” “large,” and “con-

siderable.” Moreover, they do contain many words that are irrelevant and potentially even misleading in our context. For example, the words classified as negative in the Loughran and McDonald (2011) lexicon under “A” include “abnormal,” “absence,” “against,” “aftermath,” “antitrust,” “anomaly,” and even “argue” and “argument,” thereby casting the whole economic literature in a rather negative light (incorrectly, we believe).

3.6. Methodological Choices

Researchers analyzing the same phenomenon by using different methodologies can arrive at different conclusions. We examine two methodological choices that authors make: which model to use and how to report statistical significance. We find differences between central bankers and academics in both dimensions. We summarize the results here and show the tables in the Appendix.

First, we explore differences in model choice. The most commonly used models in our sample, by far, are DSGE and VAR models. We redefine y_i to be an indicator equal to one if the study uses a DSGE model and zero if it uses a VAR model. We then regress y_i on the share of central bank authors, again using a linear probability model following equation (2). We find that central bankers are more likely to use DSGE models. With country dummies and controls, papers with 100% central bank authors are 31.9 percentage points more likely to use a DSGE model than papers with no central bank authors. This result is highly economically significant, though it is statistically significant only at the 10% level.

In our baseline regression (1), we do not control for the model chosen by the authors because model choice could be strategic. Both DSGE and VAR models give their users some flexibility: DSGE models require a variety of structural choices, whereas VAR models rely on a specific econometric specification. Either way, a user aiming for a particular outcome can pull on multiple levers to get closer to that outcome. Nonetheless, we rerun our baseline regressions after adding controls for model fixed effects (DSGE, VAR, or other). We find that the estimated β coefficients remain positive in all 36 specifications considered in Tables 2 through 4, and they are statistically significant at the 5% (10%) level in 23 (29) specifications. The magnitudes of these coefficients become larger in 27 specifications and smaller only in 9 specifications. Our main results are thus robust to controlling for model choice.

Second, we test whether central bankers and academics are equally likely to disclose the confidence level (or, alternatively, the standard error) used to assess statistical significance. When this level is not disclosed, it is more difficult to corroborate the author’s verbal as-

assessment of statistical significance. We estimate a linear probability model that regresses an indicator equal to one if the paper does not disclose the confidence level, and zero if it does, on the share of central-bank-affiliated authors. This matches the regression specification in equation (2), with y_i denoting an indicator for nondisclosure. We restrict the sample to studies that comment on the statistical significance of either output or inflation.

We find that central bankers are somewhat less likely to disclose the confidence level. With country dummies and controls, papers with 100% central bank authors are 15.8 percentage points less likely to report the width of the confidence interval than papers with no central bank authors. However, this relation is not statistically significant.

Finally, we ask which studies assess significance by using more conservative confidence intervals—ones constructed at the 95% confidence level, rather than lower levels, such as 90% or 68%. A study using a 95% confidence interval is less likely to find significance. We redefine y_i in equation (2) to be an indicator equal to one if the study uses a 95% confidence interval, and zero otherwise. With country dummies and controls, papers with 100% central bank authors are 20.2 percentage points less likely to use a 95% confidence interval than papers with no central bank authors. This result is economically sizable, but it is statistically significant only at the 10% level. The magnitudes are similar if we add a control for a dummy variable indicating whether the study uses Bayesian or frequentist inference.

4. Mechanism

This section explores potential reasons why central bankers are more optimistic than academics in their assessments of QE. One possible mechanism is career concerns. In principle, bank management could make promotion decisions in a way that encourages bank employees to assess the bank’s policies favorably. We test this hypothesis in Section 4.1. Bank management can also directly influence research outcomes at various stages of the research production process, from topic assignment, through internal review, to the approval for public distribution, as we show in Section 4.2. If bank management is skeptical of QE, the bank’s research tends to be skeptical as well, as we show in the context of the German Bundesbank in Section 4.3. The evidence in Sections 4.1 through 4.3 is consistent with managerial influence on central bank research outcomes. In Section 4.4, we discuss other potential mechanisms, such as differences in prior beliefs and differences in research quality.

4.1. Career Concerns

To examine the potential for career concerns influencing central bank research, we relate the research findings of central bankers to the bankers' subsequent career outcomes. After manually collecting employment histories for the central bank authors in our sample, we convert their job titles to numerical ranks. We create these ranks on a six-point scale for central bankers, and a four-point scale for academics, as described in the Appendix. We restrict the analysis to authors who remain affiliated with a central bank and experience at least one career update within five years following the paper's first public distribution. We impose the first filter because it is unclear whether transitions to academia or the private sector should be treated as promotions or demotions. The purpose of the second filter is to reduce the noise induced by stale CV information, since authors may not regularly update their job titles.⁴ These filters result in a sample of 33 central bankers (27 of whom are unique) and 23 papers. We then compute a new variable, career outcome, defined as the difference between the author's rank after her first career update following the paper's first distribution, and her rank at the time of that first distribution. Out of these 33 authors, 19 experience a promotion, 4 experience a demotion (3 of which are associated with a move to a different central bank), and 10 experience no change in rank.

Are central bank researchers more likely to get promoted if they find QE to be more effective? To address this question, we regress the authors' career outcomes on the reported effects on output, country fixed effects, and controls:

$$y_{aij} = \alpha_j + \beta \text{Effect}_{ij} + \gamma' X_{ai} + \epsilon_{aij} , \quad (3)$$

where y_{aij} is the difference between author a 's rank after her first career update following the first release of study i examining QE in country j , and her rank at the time of the first release. Note that y_{aij} does not vary across j for given values of a and i . In addition, Effect_{ij} is the effect of QE on output estimated by study i for country j 's QE, α_j is a fixed effect for the country in which QE takes place, and X_{ai} are controls. These controls

⁴In the Appendix, we report results obtained when we include authors with no career updates, and when we treat departures to academia and the private sector as demotions. Treating departures as demotions leads to similar conclusions. Including authors with no career update within five years of the paper's distribution leads to insignificant results. This is expected, because the absence of a career update may be due to either stale CV information or fixed review periods at central banks and, as a result, the signal-to-noise ratio for these types of career outcomes is likely to be low.

include author experience and the number of authors, as before. In addition, we control for the number of years since the author’s most recent career update and for dummy variables indicating the author’s rank at the time of the paper’s first release, because these variables are important determinants of subsequent career outcomes. The dummy variables are six indicators δ_{ai}^r , where $\delta_{ai}^r = 1$ if author a has rank r at the time of paper i ’s first release, and $\delta_{ai}^r = 0$ otherwise, for $r \in \{1, 2, \dots, 6\}$. Compared to including just one control for the author’s rank, including these six controls allows for non-linearities in the relationship between author rank and promotion outcomes.

In Panel A of Table 6, we show that reporting larger effects on output, peak or cumulative, is associated with more favorable career outcomes. The point estimate in column (3) implies that a one standard deviation increase in the peak effect is associated with a subsequent career improvement by 0.59 ranks ($= 0.485 \times 1.21$). In column (6), a one standard deviation increase in the cumulative effect on output corresponds to a subsequent career improvement by 0.57 ranks ($= 0.460 \times 1.23$). Both estimates are significant at the 5% level. For comparison, a one-unit change in rank is equivalent to moving, for example, from *Economist* to *Senior Economist*, or from *Deputy Director* to *Director*.

Panel B of Table 6 shows that the positive relation between career outcomes and estimated output effects holds also for standardized effects, with statistical significance at the 5% level in column (6) and the 10% level in column (3). A one standard deviation rise in the standardized peak effect corresponds to a 0.75 rank improvement ($= 2.661 \times 0.28$).

Of course, a positive association does not establish a causal link. Career outcomes and research output could be correlated for other reasons. For example, papers reporting larger effects could be easier to publish. Publications, in turn, could lead to promotions. To address this concern, we control for an indicator equal to one if the paper came out in a peer-reviewed journal and zero otherwise. The results are similar to those in Table 6 (see the Appendix). For another example, employees who care so much about their employer that they are willing to distort their research findings could show their affection also in other ways, such as by working hard, and they could earn a promotion that way. This channel seems harder to control for. Whether central bank research is biased by career concerns is an important but messy question for which clean identification seems difficult to come by.

Among the outcome variables analyzed in Section 3, career outcomes are most closely related to estimated effects on output. For economic significance of output, we find an

economically strong relation once all control variables are included, but the relation is not statistically significant. We do not relate career outcomes to statistical significance of estimated effects because there is no variation in the subset of central bank papers (recall that they all find statistical significance). There is no significant relation between career outcomes and estimated effects on inflation, as we show in the Appendix.

4.1.1. *Seniority*

Are career concerns stronger for senior or junior central bankers? It is not obvious for whom we should expect to see a stronger relation between research findings and career outcomes. On the one hand, research output may be a more important criterion in the promotions of junior researchers, for whom research represents the bulk of their work. On the other hand, support from top management may matter more for the career advancement of senior researchers. There may also be more discretion in promotions at the senior level.

To address this question, we repeat the analysis from Table 6, but we interact $Effect_{i,j}$ from equation (3) with the author's career rank, or *Seniority*. We find that the interaction between the effect on output and *Seniority* is positive and significant. A one standard deviation increase in *Seniority* raises the sensitivity of career outcomes to the estimated effect on output by about 50%. The interaction between the effect on inflation and *Seniority* is also positive, but it is significant only at the 10% level, and only for standardized effects. In contrast, for output, the interaction is positive for both standardized and non-standardized effects, as well as for both peak and cumulative effects, and it is significant at the 5% level in three of the four cases. The table is in the Appendix.

The above results are consistent with career concerns being stronger for senior central bank researchers. If that is the case, and if seniority does not play a similar role for academics (it is not clear why it should), we should expect to see larger differences in research findings between central bankers and academics when the authors are more senior.

To test this prediction, we repeat the analysis from Tables 2 through 4, except that we interact *CB Affiliation* with the rank of the most senior author on the team, *Max Seniority*. For each of the 12 outcome variables (three tables with four variables each), the estimated coefficient on the interaction between *CB Affiliation* and *Max Seniority* is positive. Thus, the findings of central bankers and academics are further apart if there is a more senior person on the team. The interaction coefficient is not always statistically significant (it is significant at the 5% level for two variables and at the 10% level for three additional variables), but

its magnitudes are large. For example, if the rank of the most senior author increases by one standard deviation, the difference in the estimated peak effects between a study with zero central bankers and 100% central bankers increases by more than one percentage point. This is true for both output and inflation. See the Appendix for details. Our results are consistent with the hypothesis that central bankers who are more senior report larger QE effects, relative to academics, because they face stronger career concerns.

4.2. Survey of Central Banks

Independent of the potential promotion channel, bank management can influence research outcomes in a number of ways. For example, management can assign a topic to a researcher, signaling the topic's importance to the bank. Superiors can suggest methodologies, data sources, and literature. If they are not convinced by the paper's results, superiors can return the paper with suggestions for improvement. They can provide helpful guidance and valuable resources. However, besides anecdotal evidence, the economics profession knows little about the extent of management involvement in central bank research.

To fill this gap, we conduct a survey of research practices at the world's leading central banks. We organized the survey in cooperation with the National Bank of Slovakia.⁵ We reached out to 54 heads of research, covering the central banks in all OECD countries and all EU member states, including the ECB, the Federal Reserve Board, and 12 regional Feds. In return for participating, we promised to share aggregated results with the respondents. We assured them that no individual responses would be published, and that only anonymized responses could be pooled and used for research purposes. We sent out the initial invitation on July 3, 2020; a reminder went out ten days later.

The survey contains four main questions, each containing three to six multiple-choice subquestions, for a total of 18 questions. We also asked for the number of research-active economists in full-time equivalents employed by the bank, the bank's name, and an email address to which we can send summary results. We have received 24 responses, representing a response rate of 44.4%. The 24 central banks employ over 750 researchers in total.⁶

Figure 4 presents the aggregated responses to the four main questions. In response to

⁵We thank Martin Šuster, the bank's head of research, for his generous help throughout the process.

⁶We originally received 25 responses but one respondent asked to withdraw from the survey after the first public circulation of our paper. The main conclusion from the survey—that management is substantially involved in research production—is unaffected by the exclusion of this respondent.

the first question, “*How are research topics selected in your central bank?*”, 20 (15) central banks indicate that research topics are at least sometimes (often) mutually agreed by researchers and management, and 17 banks respond that topics are at least sometimes assigned by management. Responding to the second question, “*How are draft research papers reviewed/commented on in your institution, prior to their public distribution?*”, 21 (14) central banks indicate that papers are reviewed at least sometimes (often) by management. In 9 banks, this review happens for all papers. The third question is, “*How are your institution’s draft research papers approved for public distribution?*”. Bank management and, most commonly, the head of research, is frequently involved in approving papers for publication: 20 (18) banks respond that the head of research approves papers at least sometimes (often). The bank board also gets involved in the approval process, at least sometimes, in eight banks. Finally, when asked “*What criteria can lead to the paper being rejected (i.e., not approved for public distribution)?*”, most central banks list “*substandard methodology, unreliable data, deficient modeling approach*”, followed by “*results not robust or not significant*”. The latter criteria are used by 18 (10) banks at least sometimes (often).

The survey evidence reveals substantial involvement by management in the research process at most central banks. This involvement creates an opportunity for bank management to influence research outcomes, offering a potential explanation for our findings in Section 3. However, such an interpretation is subject to numerous caveats.

First and foremost, the fact that management involvement exists does not imply that it affects research outcomes as measured in our study. Management involvement is necessary but not sufficient for research outcomes to be influenced by management; the survey evidence thus supports only the necessary condition. The involvement can take different forms, many of which help improve the quality of research output without introducing any bias. For example, many research directors view their role largely as helping their staff write better papers. The first two survey questions pool research directors and senior managers into a single “management” category, masking the different roles of these two types of managers, as well as their potentially different sensitivities to “undesirable” policy messages.

Second, given the survey’s brevity, the responses cannot reveal the full range of practices across banks. For example, economists in many banks split their time between policy work and their own research. Management is likely to be more involved in the selection of topics for policy work than for individual research, yet the first survey question does not distinguish

between the two types of work. In addition, whether a study of the effectiveness of QE counts as research or policy work may differ across banks.

Finally, the set of central banks in our survey sample differs from the set of banks whose economists are in our pool of authors who have analyzed QE. While the two sets overlap, the overlap is modest. We do not know how similar the research processes in the two sets of banks are. If they are substantially different, then our survey sheds little light on the involvement of bank management in the production of studies on QE.

4.3. *The Bundesbank*

If central bankers' findings regarding QE are colored by the views of bank management, we should see weaker QE effects reported by researchers at central banks whose management has taken a critical stance towards QE. A prominent example is the German Bundesbank, whose top management has publicly criticized the ECB for its bond-buying program. According to media outlets, former Bundesbank president Axel Weber and vice president Jürgen Stark resigned from their ECB positions, allegedly in protest over QE, and the current Bundesbank president, Jens Weidmann, has also publicly opposed QE. Could these skeptical views of the bank's top brass be reflected in the writings of the bank's researchers?

We test this hypothesis by repeating the analysis in Tables 2 through 4, replacing *CB Affiliation* in regression (1) with three indicators: *German CB* is equal to one if at least one of the authors is employed by the Bundesbank, *Other EA CB* is equal to one if at least one author works at the ECB or a euro area national central bank other than the Bundesbank, and *Non-EA CB* is equal to one if at least one author is from a central bank outside the euro area or the BIS. The omitted group are academics. If a paper has authors from both the Bundesbank and another central bank, multiple indicators are switched on.

We find that Bundesbank authors find strikingly different results regarding the effectiveness of QE in raising output. Bundesbank papers report *smaller* effects of QE on output compared to academics, on average, whereas other central banks, both inside and outside the euro area, find *larger* effects. This pattern holds for all four measures of output, as we show in the Appendix. For example, the average estimated peak effect on output for Bundesbank papers is 0.88 percentage points smaller than the average peak effect for academics. In contrast, other central banks in the euro area find effects that are 0.44 percentage points larger, and banks outside the euro area find effects that are 0.69 percentage points larger, on average, compared to academics. The difference in the point estimates for *German CB*

and *Other EA CB* is mostly statistically significant based on cluster-robust standard errors but not based on bootstrapped p -values, in part because there are only four Bundesbank papers in our sample. Nevertheless, the different signs and large magnitudes of the coefficients support the managerial influence interpretation of our main results.

As for the QE effect on inflation, the differences between Bundesbank authors and other central bankers are much smaller. Moreover, the Bundesbank estimates exceed those of academics. These results are not surprising. German opposition to ECB's QE has been based largely on concerns about redistribution within the euro area, not about QE being ineffective at raising inflation. On the contrary, a popular view in Germany is that QE could be too effective in that regard. The view that printing money causes inflation is traditionally strong in Germany, whose collective memory is still scarred by the hyperinflation that took place in the Weimar Republic in the early 1920s.

4.4. Alternative Mechanisms

While our evidence in Sections 4.1 through 4.3 is consistent with managerial influence on central bank research outcomes, the evidence is not causal, and it is therefore inconclusive. In this section, we discuss other mechanisms that could potentially contribute to the observed differences between the findings of central bankers and academics.

One such mechanism is reputation concerns. These could involve concerns about the bank's reputation and, for very senior researchers, concerns about their own reputation. Like career concerns, reputation concerns reflect researchers' incentives because in both cases, a researcher derives a private benefit from reaching a particular research outcome. We have no evidence on the potential contribution of reputation concerns to our results.

Another potential mechanism is differences in prior beliefs combined with selection. Researchers with different priors about the effectiveness of policy interventions may self-select into different institutions. For example, if researchers optimistic about QE select into central banks, or the pessimists select into academia, this selection could explain the differences in research outcomes between central bankers and academics. Furthermore, any differences in priors can be reinforced during the research process, both by the researcher herself (confirmation bias) and by the feedback she receives from like-minded colleagues. This mechanism does not involve any distorted incentives. We have no evidence in favor of, or against, differences in priors, but we do have evidence on three other potential explanations.

One of them is that papers on QE written by central bankers and academics are of

different quality. For example, if central bank papers were of higher quality and the effects of QE were truly strong, then we would expect central bank papers to find stronger QE effects. Given the management involvement documented earlier, it is indeed possible that central banks have a more rigorous vetting process for new working papers compared to universities, allowing central banks to discard lower-quality papers. Moreover, central bankers may simply know more about QE than academics, given the nature of the subject.

However, higher research quality at central banks seems unlikely to explain our results, for five reasons. First, papers written by central bankers and academics are of comparable quality, based on three measures of quality: publication status, journal impact factor, and the article influence score. We show this, as well as the following two results, in the Appendix. Second, the gaps in research findings between central bankers and academics remain largely unchanged once we condition on published papers only. Third, we find very similar results when we replace OLS regressions with weighted least squares, weighting each paper by its Google Scholar citations as of September 2019.⁷ Fourth, central bank papers are somewhat less likely to report standard errors around their estimates (see Section 3.6). Finally, to explain the opposite results regarding output for Bundesbank authors, a story based on differential research quality would have to assume that Bundesbank research is of different quality compared to other central banks.

Another possible explanation for our results is differences in methodology. Recall from Section 3.6 that central bankers use DSGE models more often than academics do. This choice need not be strategic; it may be guided by the popularity of DSGE models in central bank policy work. If DSGE models generate stronger effects of QE compared to VAR models, this difference could potentially explain our results. However, when we add controls for model choice to regression (1), our results continue to hold (see Section 3.6).

Finally, as conveyed to us by a central banker, it is possible that academics seek sensational results—such as that vast amounts of spending by central banks were ineffective in improving macroeconomic outcomes—to boost their reputations. Two facts cast doubt on this explanation. First, academic reputation is generally judged by the publication record. Finding a null result makes a paper harder, not easier, to publish. Consistent with this view, finding larger effects of QE on output increases the odds of publication in a peer-reviewed

⁷Specifically, we weight each paper by the logarithm of one plus the number of citations for the paper, divided by the logarithm of one plus the average number of citations across all papers released in the same year. The scaling addresses the fact that older papers tend to have more citations.

journal (see the Appendix). Second, if the results were driven by academics' career concerns, then we should see stronger results among junior authors, who strive to earn tenure. In contrast, recall from Section 4.1.1 that differences between the findings of central bankers and academics are more pronounced among senior authors.

5. Conclusion

Central bank economists are more optimistic than academics in their assessments of the macroeconomic effects of QE. Based on a sample of 54 studies, studies written by central bankers report stronger effects of QE on both output and inflation. Central bank studies are also more likely to report significant QE effects on output, and their abstracts use more favorable language, compared to those written by academics.

Whose findings are closer to the truth remains unclear. Our evidence does not imply that central bank research is biased; perhaps academic research is biased toward insignificance, despite the publication bias in academic journals. However, our evidence suggests some role for career concerns at central banks. We find that central bankers whose papers report larger effects of QE on output have better career outcomes. The somewhat weaker effects found by Bundesbank researchers are also consistent with career concerns. Finally, our survey reveals that in most central banks, management influences research topics, reviews papers, and approves them for public distribution. The involvement of bank management in the production of bank research extends far beyond that of university management in academic research. Some involvement of bank management seems necessary, given the broader mission of a central bank. The extent to which this involvement affects research outcomes remains unclear, creating opportunities for future research.

Importantly, we do not argue that central bank research should be discounted. In many ways, central banks are in an excellent position to provide accurate assessments of their own policies. Like pharmaceutical firms studying their own drugs, central banks have superior information about their own products, exceptionally strong expertise in the subject matter, and an intense concern for their reputation. In addition, central banks are public institutions of the highest integrity. They understand that the effectiveness of their policy is predicated on their own credibility. We are not questioning that credibility. We simply offer novel evidence on a previously unexplored aspect of central bank research.

References

- Allen, F., Bean, C., De Gregorio, J., 2016. Independent review of BIS research: Final report. Bank of International Settlements.
- Andersen, J.J., Johannesen, N., Rijkers, B., 2020. Elite capture of foreign aid: Evidence from offshore bank accounts. Working Paper, World Bank.
- Asatryan, Z., Havlik, A., Heinemann, F., Nover, J., 2020. Biases in fiscal multiplier estimates. *European Journal of Political Economy* 63, 101861.
- Banerjee, A.V., Deaton, A., Lustig, N., Rogoff, K., Hsu, E., 2006. An evaluation of World Bank research, 1998-2005. Working Paper, World Bank.
- Bekelman, J.E., Li, Y., Gross, C.P., 2003. Scope and impact of financial conflicts of interest in biomedical research: A systematic review. *Journal of the American Medical Association* 289, 454–465.
- Bernanke, B.S., 2020. The new tools of monetary policy. *American Economic Review* 110, 943–983.
- Brodeur, A., Lé, M., Sangnier, M., Zylberberg, Y., 2016. Star wars: The empirics strike back. *American Economic Journal: Applied Economics* 8, 1–32.
- Cameron, A.C., Gelbach, J.B., Miller, D.L., 2008. Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics* 90, 414–427.
- Carrick-Hagenbarth, J., Epstein, G.A., 2012. Dangerous interconnectedness: Economists’ conflicts of interest, ideology and financial crisis. *Cambridge Journal of Economics* 36, 43–63.
- Christensen, G., Miguel, E., 2018. Transparency, reproducibility, and the credibility of economics research. *Journal of Economic Literature* 56, 920–980.
- De Long, J.B., Lang, K., 1992. Are all economic hypotheses false? *Journal of Political Economy* 100, 1257–1272.
- Dell’Ariccia, G., Rabanal, P., Sandri, D., 2018. Unconventional monetary policies in the Euro Area, Japan, and the United Kingdom. *Journal of Economic Perspectives* 32, 147–172.
- Fanelli, D., 2009. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLOS ONE* 4, 1–11.
- Fanelli, D., 2010a. Do pressures to publish increase scientists’ bias? An empirical support from US states data. *PLOS ONE* 5.
- Fanelli, D., 2010b. “Positive” results increase down the hierarchy of the sciences. *PLOS ONE* 5, 1–10.
- Franco, A., Malhotra, N., Simonovits, G., 2014. Publication bias in the social sciences: Unlocking the file drawer. *Science* 345, 1502–1505.

- Hansen, S., McMahon, M., Prat, A., 2018. Transparency and deliberation within the FOMC: A computational linguistics approach. *Quarterly Journal of Economics* 133, 801–870.
- Ioannidis, J.P.A., 2005. Why most published research findings are false. *PLOS Medicine* 2.
- Ioannidis, J.P.A., Stanley, T.D., Doucouliagos, C., 2017. The power of bias in economics research. *Economic Journal* 127, F236–F265.
- Jelveh, Z., Kogut, B., Naidu, S., 2018. Political language in economics. Working Paper, Columbia Business School.
- List, J.A., Bailey, C.D., Euzent, P.J., Martin, T.L., 2001. Academic economists behaving badly? A survey on three areas of unethical behavior. *Economic Inquiry* 39, 162–170.
- Loughran, T., McDonald, B., 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-ks. *Journal of Finance* 66, 35–65.
- Lucca, D., Seru, A., Trebbi, F., 2014. The revolving door and worker flows in banking regulation. *Journal of Monetary Economics* 65, 17–32.
- Martin, C., Milas, C., 2012. Quantitative easing: a sceptical survey. *Oxford Review of Economic Policy* 28, 750–764.
- Nickerson, R.S., 1998. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology* 2, 175–220.
- Roodman, D., Ørregaard Nielsen, M., MacKinnon, J.G., Webb, M.D., 2019. Fast and wild: Bootstrap inference in Stata using boottest. *Stata Journal* 19, 4–60.
- Sibert, A., 2003. Monetary policy committees: Individual and collective reputations. *Review of Economic Studies* 70, 649–665.
- The Economist, 2020. The World Bank loses another chief economist. February 13.
- The Financial Times, 2015. Four reasons why QE will be different in the eurozone. January 20.
- Blanes i Vidal, J., Draca, M., Fons-Rosen, C., 2012. Revolving door lobbyists. *American Economic Review* 102, 3731–3748.
- Weale, M., Wieladek, T., 2016. What are the macroeconomic effects of asset purchases? *Journal of Monetary Economics* 79, 81–93.
- Zingales, L., 2013. Preventing economists’ capture, in: Carpenter, D., Moss, D.A. (Eds.), *Preventing regulatory capture: Special interest influence and how to limit it*. Cambridge University Press, pp. 124–151.

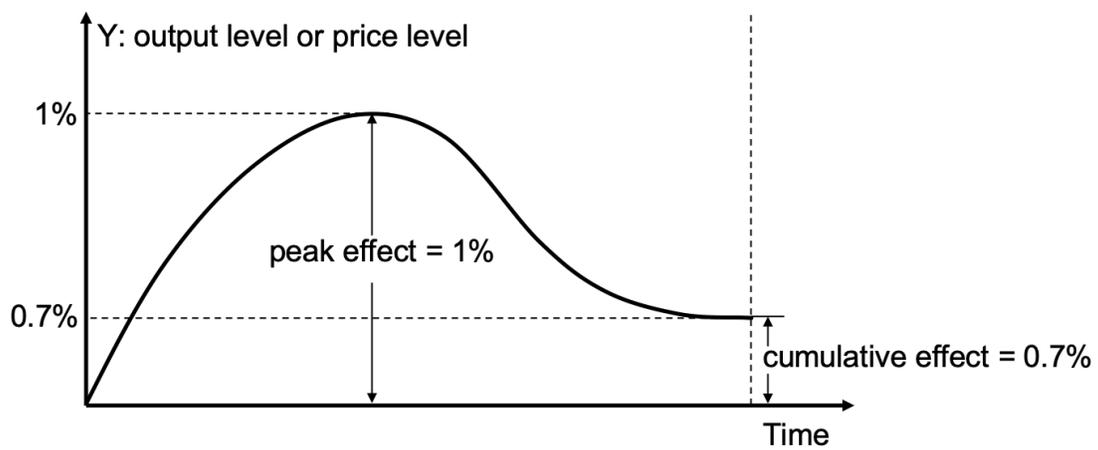


Figure 1: **Visualization of the Peak and Cumulative Effects.** The figure illustrates how we compute the peak and cumulative effects of QE on the level of the outcome variable for the most common case, in which the authors plot the effect of QE on the level of the outcome variable.

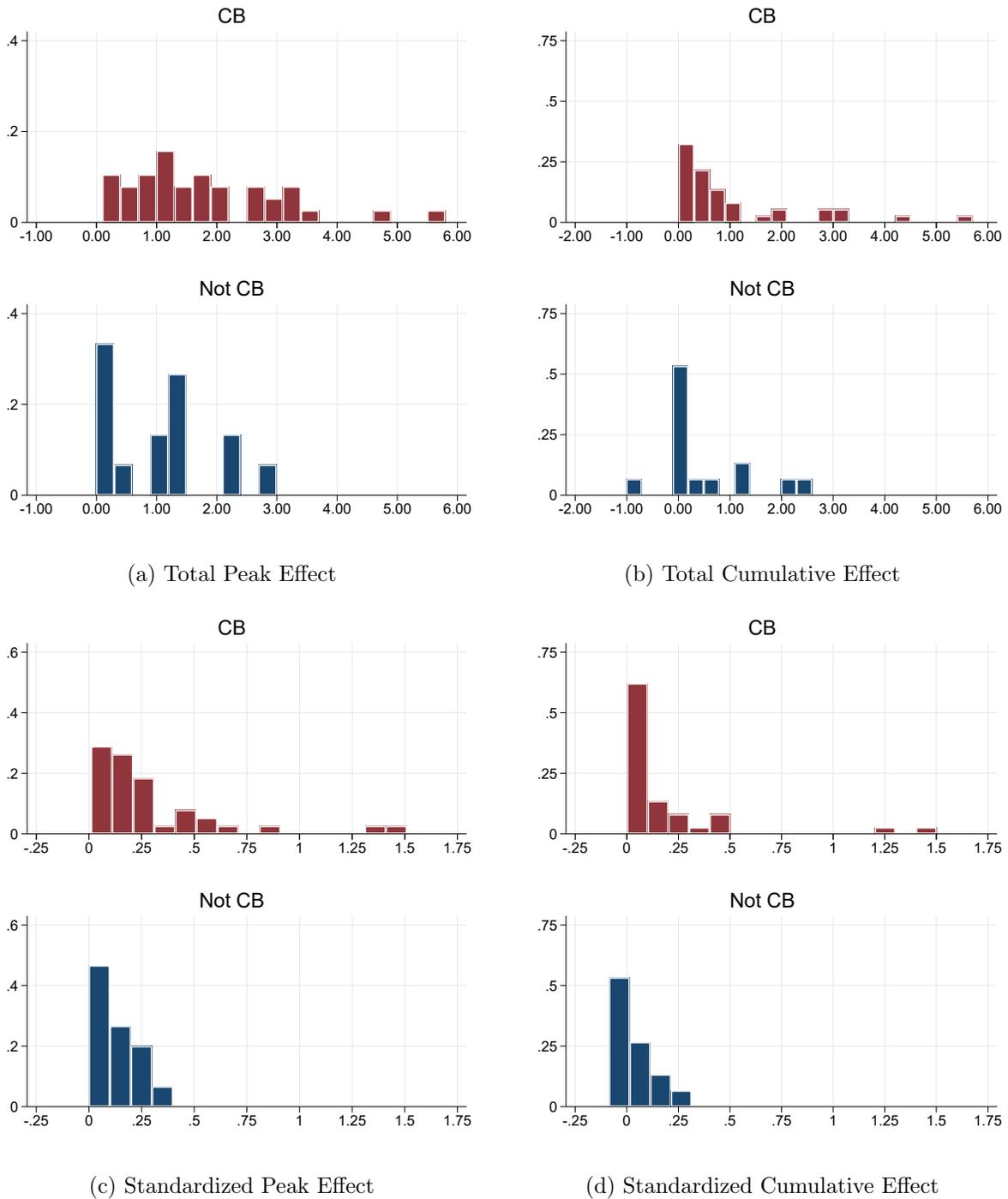


Figure 2: **Effects of QE on Output by Central Bank Affiliation.** The figure plots histograms for the estimated effects on output, separately for papers with and without CB-affiliated authors. Studies with 0.5 central bankers, but no “full” central banker, are excluded. Panels A and B show the total estimated peak and cumulative effects of the QE program studied on the level of output. Panels C and D show the estimated peak and cumulative effects of the QE program studied on the level of output, after standardizing the QE program size to 1% of GDP.

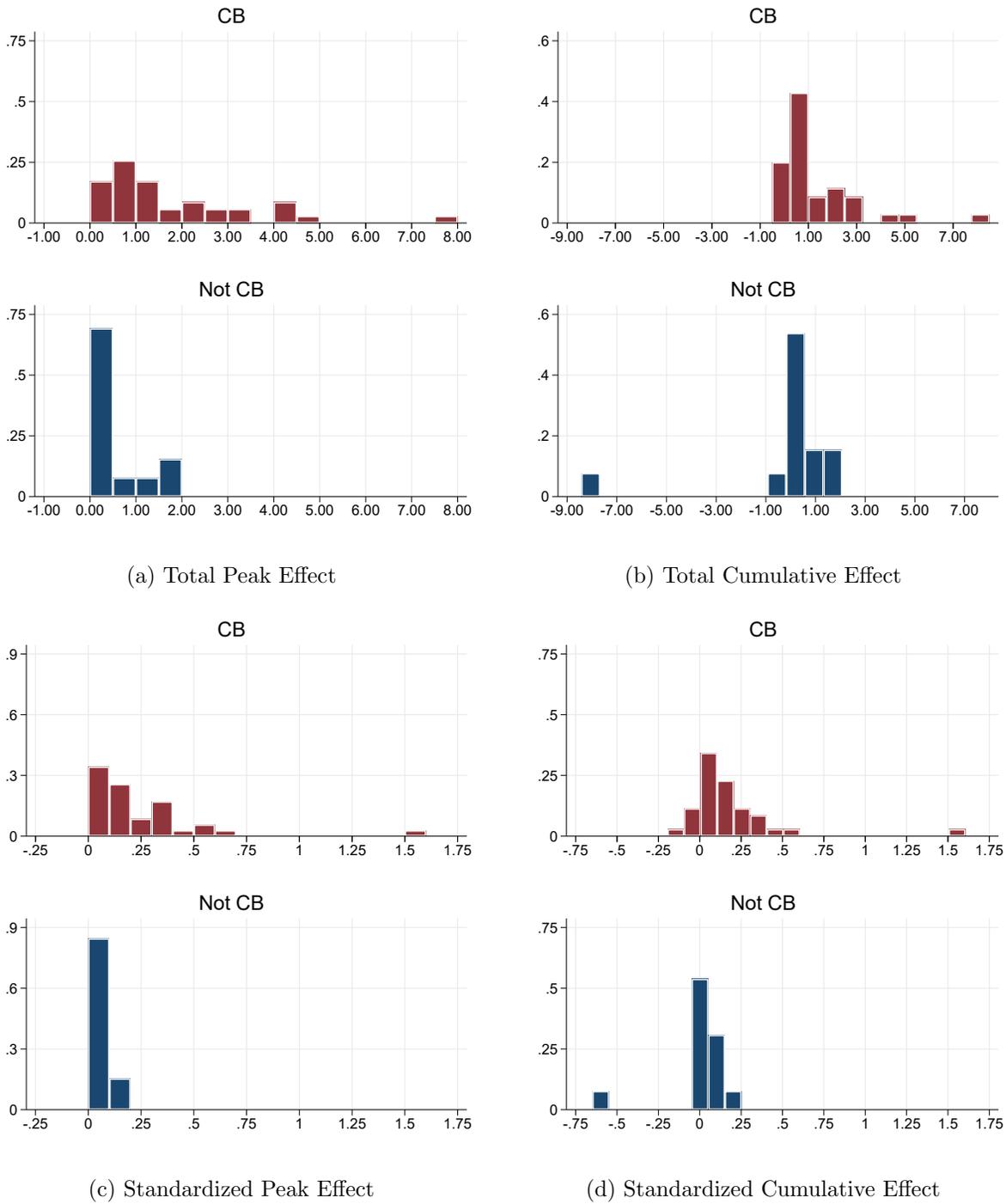
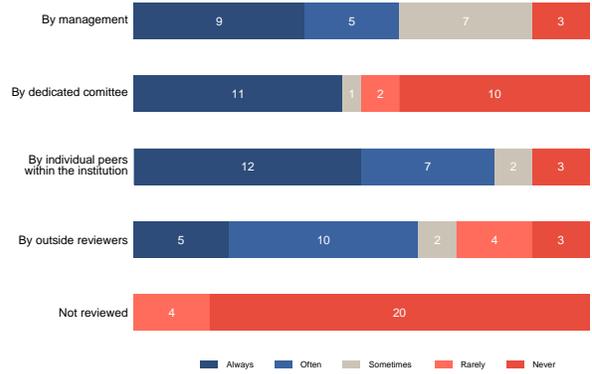
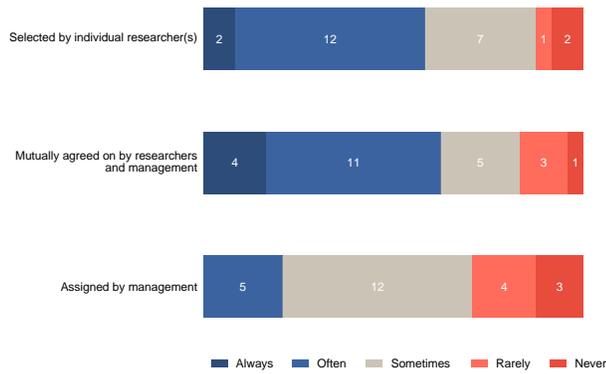
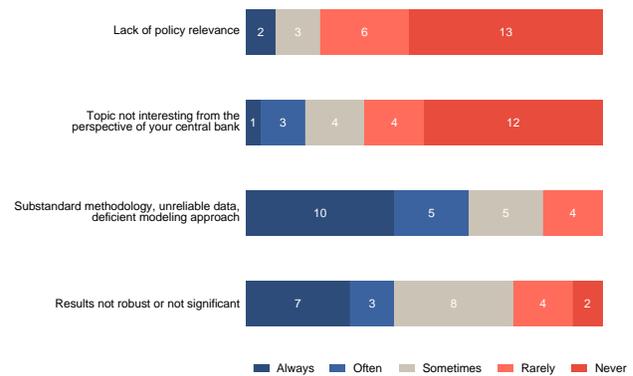
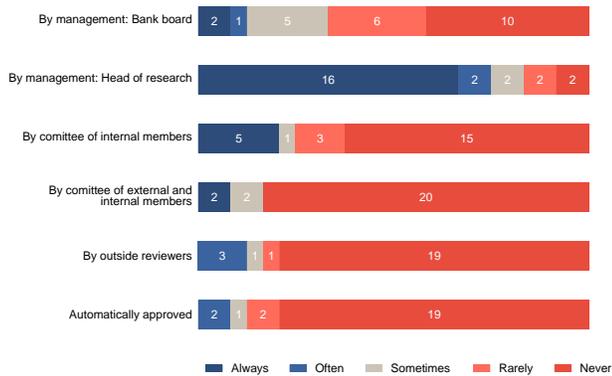


Figure 3: **Effects of QE on Inflation by Central Bank Affiliation.** The figure plots histograms for the estimated effects on inflation, separately for papers with and without CB-affiliated authors. Studies with 0.5 central bankers, but no “full” central banker, are excluded. Panels A and B show the total estimated peak and cumulative effects of the QE program studied on the price level. Panels C and D show the estimated peak and cumulative effects of the QE program studied on the price level, after standardizing the QE program size to 1% of GDP.



(A) How are research topics selected in your central bank?

(B) How are draft research papers reviewed / commented on in your institution, prior to their public distribution?



(C) How are your institution's draft research papers approved for public distribution?

(D) What criteria can lead to the paper being rejected (i.e., not approved for public distribution)?

Figure 4: **Survey of Central Banks.** The figure reports survey responses of 24 central banks.

Table 1: **Effects of QE on Output and Inflation by Central Bank Affiliation**

This table reports the means and medians (in parentheses) for the estimated effects of QE on output and inflation, as well as for indicators of statistical significance, separately for papers with and without CB-affiliated authors. Studies with 0.5 central bankers, but no “full” central banker, are excluded. We always report the effect on the output level or price level, in percent. Standardized effects refer to the effect of a QE program size equivalent to 1% of GDP. The unit of observation is the paper-country.

	All	CB	Not CB
<i>Panel A: Effect on Output</i>			
Peak effect on output	1.57 (1.25)	1.75 (1.53)	1.00 (1.00)
Standardized peak effect on output	0.24 (0.16)	0.28 (0.18)	0.11 (0.10)
Cumulative effect on output	0.87 (0.40)	1.06 (0.42)	0.48 (0.05)
Standardized cumulative effect on output	0.14 (0.04)	0.18 (0.06)	0.04 (0.01)
<i>Panel B: Effect on Inflation</i>			
Peak effect on inflation	1.42 (0.93)	1.79 (1.17)	0.54 (0.40)
Standardized peak effect on inflation	0.19 (0.11)	0.24 (0.15)	0.05 (0.04)
Cumulative effect on inflation	0.89 (0.75)	1.35 (0.82)	-0.21 (0.14)
Standardized cumulative effect on inflation	0.12 (0.08)	0.18 (0.11)	-0.01 (0.01)
<i>Panel C: Significance</i>			
Statistical significance: output	0.88 (1.00)	1.00 (1.00)	0.50 (0.50)
Statistical significance: inflation	0.84 (1.00)	0.89 (1.00)	0.75 (1.00)

Table 2: **Effects of QE on Output**

This table regresses the estimated effects of QE on output on the share of authors with central bank affiliation. In Panel A, we use the total estimated effect of the QE program studied on the level of output. Panel B uses the estimated effect on the level of output, after standardizing the QE program size to 1% of GDP. Controls include the number of authors and the logarithm of three plus the average author experience. t -statistics, reported in parentheses, are based on standard errors clustered at the paper level. p -values obtained using the wild cluster bootstrap procedure (10,000 repetitions) are reported in square brackets. The unit of observation is the paper-country.

Panel A: Total Program Effect

	Peak Effect			Cumulative Effect		
	(1)	(2)	(3)	(4)	(5)	(6)
CB Affiliation	0.789 (2.16) [0.041]	0.770 (2.17) [0.038]	0.723 (1.83) [0.085]	0.620 (1.60) [0.115]	0.526 (1.51) [0.134]	0.512 (1.36) [0.167]
Country FE		X	X		X	X
Controls			X			X
Observations	58	58	58	57	57	57
R^2	0.072	0.103	0.112	0.043	0.091	0.096

Panel B: Standardized Effect

	Peak Effect			Cumulative Effect		
	(1)	(2)	(3)	(4)	(5)	(6)
CB Affiliation	0.164 (2.38) [0.021]	0.163 (2.48) [0.018]	0.152 (2.11) [0.052]	0.140 (2.17) [0.021]	0.127 (2.17) [0.020]	0.122 (1.90) [0.049]
Country FE		X	X		X	X
Controls			X			X
Observations	58	58	58	57	57	57
R^2	0.060	0.170	0.206	0.048	0.078	0.106

Table 3: **Effects of QE on Inflation**

This table regresses the estimated effects of QE on inflation on the share of authors with central bank affiliation. In Panel A, we use the total estimated effect of the QE program studied on the price level. Panel B uses the estimated effect on the price level, after standardizing the QE program size to 1% of GDP. Controls include the number of authors and the logarithm of three plus the average author experience. *t*-statistics, reported in parentheses, are based on standard errors clustered at the paper level. *p*-values obtained using the wild cluster bootstrap procedure (10,000 repetitions) are reported in square brackets. The unit of observation is at the paper-country level.

Panel A: Total Program Effect

	Peak Effect			Cumulative Effect		
	(1)	(2)	(3)	(4)	(5)	(6)
CB Affiliation	1.409 (3.42) [0.002]	1.493 (3.33) [0.002]	1.279 (2.79) [0.011]	1.700 (2.24) [0.011]	1.687 (2.20) [0.013]	1.394 (2.04) [0.044]
Country FE		X	X		X	X
Controls			X			X
Observations	53	53	53	53	53	53
R^2	0.142	0.239	0.298	0.126	0.126	0.195

Panel B: Standardized Effect

	Peak Effect			Cumulative Effect		
	(1)	(2)	(3)	(4)	(5)	(6)
CB Affiliation	0.197 (2.61) [0.007]	0.227 (2.73) [0.005]	0.201 (2.73) [0.004]	0.205 (2.31) [0.004]	0.220 (2.31) [0.004]	0.190 (2.41) [0.007]
Country FE		X	X		X	X
Controls			X			X
Observations	53	53	53	53	53	53
R^2	0.110	0.248	0.296	0.106	0.186	0.226

Table 4: **Significance**

This table regresses the statistical and economic significance of the estimated effects of QE on output and inflation on the share of central bank affiliated authors. In Panel A (B), the dependent variable is the reported statistical and economic significance of the effect on output (inflation). Controls include the number of authors and the logarithm of three plus the average author experience. t -statistics, reported in parentheses, are based on standard errors clustered at the paper level. p -values obtained using the wild cluster bootstrap procedure (10,000 repetitions) are reported in square brackets. The unit of observation is the paper-country.

Panel A: Effect on Output

	Statistical Significance			Economic Significance		
	(1)	(2)	(3)	(4)	(5)	(6)
CB Affiliation	0.412 (2.42) [0.041]	0.388 (2.41) [0.035]	0.366 (2.20) [0.043]	0.335 (2.78) [0.019]	0.344 (2.78) [0.019]	0.399 (3.42) [0.005]
Country FE		X	X		X	X
Controls			X			X
Observations	41	41	41	66	66	66
R^2	0.233	0.280	0.298	0.139	0.145	0.250

Panel B: Effect on Inflation

	Statistical Significance			Economic Significance		
	(1)	(2)	(3)	(4)	(5)	(6)
CB Affiliation	0.202 (1.18) [0.339]	0.202 (1.25) [0.283]	0.164 (1.11) [0.372]	0.196 (1.29) [0.222]	0.207 (1.36) [0.197]	0.248 (1.86) [0.093]
Country FE		X	X		X	X
Controls			X			X
Observations	38	38	38	60	60	60
R^2	0.044	0.118	0.208	0.041	0.043	0.137

Table 5: **Tone of the Abstract**

This table regresses measures of the tone of the paper’s abstract on the share of central bank affiliated authors. In Panel A, the dependent variable is the sentiment score, computed as the difference in the percentage of positive and negative adjectives in the abstract. In Panel B (C), the dependent variable is the percentage of positive (negative) adjectives in the abstract. Controls include the number of authors and the logarithm of three plus the average author experience. *t*-statistics, reported in parentheses, are based on robust standard errors. *p*-values obtained using the wild cluster bootstrap procedure (10,000 repetitions) are reported in square brackets. The unit of observation is the paper.

Panel A: Sentiment Score			
	(1)	(2)	(3)
CB Affiliation	0.046 (2.05) [0.049]	0.053 (2.59) [0.014]	0.056 (2.60) [0.013]
Country Dummies		X	X
Controls			X
Observations	54	54	54
R^2	0.081	0.129	0.133
Panel B: Percentage of Positive Adjectives			
	(1)	(2)	(3)
CB Affiliation	0.033 (1.61) [0.125]	0.040 (2.15) [0.043]	0.043 (2.22) [0.030]
Country Dummies		X	X
Controls			X
Observations	54	54	54
R^2	0.052	0.128	0.136
Panel C: Percentage of Negative Adjectives			
	(1)	(2)	(3)
CB Affiliation	-0.013 (-1.35) [0.197]	-0.013 (-1.22) [0.252]	-0.013 (-1.16) [0.272]
Country Dummies		X	X
Controls			X
Observations	54	54	54
R^2	0.040	0.048	0.052

Table 6: Career Outcomes and Effects of QE on Output

This table regresses career outcomes on the author’s estimated effects of QE on output. The dependent variable is the difference between the author’s rank after her first career update following the paper’s first circulation, and her rank at the time of first circulation. In Panel A, we use the total estimated effect of the QE program studied on the level of output. In Panel B, the QE program size is standardized to 1% of GDP. Controls include the number of authors, the logarithm of three plus the researcher’s experience, the number of years since the author’s last career update, as well as dummy variables indicating the author’s rank at the time of the paper’s first circulation. We restrict the sample to authors who remain affiliated with a central bank and experience at least one career update after the paper’s first circulation. *t*-statistics, reported in parentheses, are based on standard errors clustered at the author level. *p*-values obtained using the wild cluster bootstrap procedure (10,000 repetitions) are reported in square brackets. The unit of observation is the author-paper-country.

Panel A: Total Program Effect

	Peak Effect			Cumulative Effect		
	(1)	(2)	(3)	(4)	(5)	(6)
Effect on output	0.264 (2.32) [0.027]	0.219 (1.85) [0.037]	0.485 (2.65) [0.018]	0.204 (1.78) [0.079]	0.204 (1.25) [0.234]	0.460 (2.12) [0.019]
Country FE		X	X		X	X
Controls			X			X
Observations	34	34	31	32	32	30
R^2	0.030	0.066	0.553	0.027	0.076	0.550

Panel B: Standardized Effect

	Peak Effect			Cumulative Effect		
	(1)	(2)	(3)	(4)	(5)	(6)
Effect on output	1.407 (1.41) [0.231]	1.009 (1.15) [0.356]	2.661 (1.86) [0.082]	2.311 (2.00) [0.041]	1.838 (1.45) [0.134]	4.095 (2.15) [0.022]
Country FE		X	X		X	X
Controls			X			X
Observations	34	34	31	32	32	30
R^2	0.044	0.062	0.553	0.051	0.081	0.569