CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

Essays on Research Reproducibility in Economics

Doctoral Dissertation

Author: Nino Buliskeria
Study program: Economics and Finance
Supervisor: Jaromir Baxa, Ph.D.
Year of defense: 2024

Declaration of Authorship

The author hereby declares that she compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis in whole or in part and agrees with the thesis being used for study and scientific purposes.

Prague, May 3, 2024

Nino Buliskeria

Acknowledgments

I extend my deepest appreciation to Prof. Jaromír Baxa, whose guidance and wisdom illuminated my path throughout this journey. His invaluable feedback and unwavering support have been the cornerstone of my work. I will greatly miss our weekly meetings and discussions.

I also owe a world of gratitude to Prof. Tomáš Havránek and Prof. Zuzana Havránková, whose insightful recommendations and collaborative spirit significantly enriched my research experience. Their expertise was instrumental in shaping the contours of this dissertation.

I am profoundly grateful to my partner and husband, Ali, whose faith in me never wavered through the highs and lows. His critical mind and kind nature have greatly helped in shaping my work and PhD journey. And to my close friend, Tamar, who has always brought light and joy into my life, even when my code would not budge.

To my family, whose endless encouragement and support have been my unfailing source of strength, I owe an immeasurable debt of gratitude. To my father, for his unconditional love; to my mother, for her constant warmth and support; and to my brother, for being my source of inspiration.

I also extend special thanks to the vibrant *602 crowd*, the brightest and kindest minds of Prague, with particular gratitude to Lenka, Klára, Levan, and Marek for being constant office mates and great friends.

Heartfelt thanks to the Institute of Economic Studies for affording me the invaluable opportunities to travel the world and learn from some of the finest faculties globally. The unwavering support of the IES faculty throughout my Ph.D. journey has been instrumental in my development and success.

I am immensely grateful to Professor Abel Brodeur, Professor Bob Reed, and Professor Chishio Furukawa, the referees of this dissertation, whose insightful comments have significantly contributed to the improvement of this manuscript. Their detailed feedback has been invaluable.

While it is impossible to list all names, I wish to express my gratitude to all the researchers who engaged with my work, shared insights at conferences, and contributed through many invaluable informal discussions.

This work was generously supported by the Charles University Grant Agency (No. 1034519), the Grant Agency of the Czech Republic (No. GACR no. 20-14990S), Charles University Research Center program No.UNCE/HUM/035, and the Cooperation Program at Charles University, research area Economics. This manuscript is part of a project that has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 870245.

ABSTRACT

This dissertation contributes to the ongoing dialogue on the credibility revolution in economics, a movement initiated by Edward Leamer's critical examination of empirical research in the early 1980s. This work delves into the core issues of meta-research, replicability, and the prevalence of publication biases and *p*-hacking within the published research through a series of detailed analyses. The first study revisits Burgess and Pande's influential analysis on the impact of India's state-led bank expansion on poverty, revealing that the purported significant effect diminishes when accounting for more comprehensive policy contexts. This finding underscores the necessity of considering external policy influences to ensure the robustness of empirical results.

The second investigation analyzes the Economic Policy Uncertainty (EPU) indices of Germany, France, Italy, Spain, and the United Kingdom, finding that apparent trends in these indices stem from methodological choices rather than actual economic uncertainty. By applying alternative normalization techniques, this study demonstrates how such methodological nuances can lead to markedly different interpretations of economic phenomena, such as Brexit or the COVID-19 pandemic.

Lastly, an expansive analysis of 20,000 studies assesses the extent of publication bias and *p*-hacking in economic research. This evaluation reveals that *p*-hacking is notably more prevalent than previously understood, suggesting a significant contribution to the overall selection bias in the literature. I find that selective reporting within studies more prevalent than publication bias arising from selection among studies. This finding underscores the considerable influence of practices such as p-hacking and method-searching, suggesting that they contribute significantly to selection bias in the economic literature and could affect the perceived reliability of published findings.

This dissertation advocates for a shift within the research community towards embracing replication studies, publishing findings regardless of their significance levels, and possibly adopting pre-registration practices to guarantee publication. Such measures are proposed as essential steps to discourage *p*-hacking and enhance the credibility of economic research.

JEL Classification: A11, C13, C40 D80, E66, E32, G21, G28, O15, O16

Keywords: access to finance, finance and development, rural poverty, economic policy uncertainty, trend-cycle decomposition, reproducibility, reliability, selective reporting, publication bias, p-hacking

Title: Essays on Research Reproducibility in Economics

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Selective reporting and publishing empirical results can distort our understanding of how robust documented regularities are and give a false impression of their generalizability. Since the early 1980s, the critical examination of empirical research, initiated by Edward Leamer, has catalyzed what is now known as the credibility revolution in economics. This movement has strongly emphasized the importance of meta-research and the replicability of published work. This wave of change has influenced research beyond economics to address what is commonly referred to as the "replication crisis" (Camerer et al., 2018), affecting fields such as medicine and epidemiology with John Ioannidis at the forefront (Begley & Ioannidis, 2015; Ioannidis, 2005; Ioannidis et al., 2017), as well as psychology and social sciences. An expanding body of work explores the issues of potential publication biases within economics and various other fields (Andrews & Kasy, 2019; Ashenfelter et al., 1999; Bruns et al., 2019; De Long & Lang, 1992; Doucouliagos & Stanley, 2013; Ferraro & Shukla, 2020; Furukawa, 2019; Havránek, 2015; Ioannidis, 2005; Ioannidis et al., 2017; Leamer, 1983; Miguel et al., 2014; Stanley, 2005, 2008). In their influential meta-analyzes, Card and Krueger (1995) addressed the pivotal question: Does raising the minimum wage reduce employment? Contrary to standard economic theory, their findings suggested that studies supporting a negative correlation between higher minimum wages and job availability could be compromised by specification-searching and publication biases. This meta-study contributed to a body of work that earned David Card the 2021 Nobel Prize in Economics.

The credibility of empirical research is the cornerstone of scientific progress, yet it remains vulnerable to the influences of p-hacking and publication biases.

Publication bias arises when editorial teams and reviewers prefer studies that demonstrate statistically significant results. Meanwhile, the perception that publication bias is prevalent can lead researchers to abandon studies with unexpected or unpromising results, exacerbating publication bias. On the other hand, p-hacking involves various tactics researchers use, sometimes unintentionally, to achieve more favorable p-values, including "specification search," "p-hacking," or "data dredging" (Brodeur et al., 2020, 2023; Lang, 2023; Mathur, 2022). These tactics can include collecting data until the results appear significant, adjusting econometric models, or setting specific sample criteria to reach desired levels of statistical significance. The urge to engage in p-hacking can come from the perceived importance of statistical significance for the probability of publication (Andrews and Kasy, 2019).

This dissertation consists of three articles on the reproducibility of published research and the extent of preferences of researchers and journals to consistently and overwhelmingly report and publish significant results.

The first paper, co-authored with Jaromir Baxa and titled "Do rural banks matter that much? Burgess and Pande (2005) reconsidered," revisits Burgess and Pande's (AER, 2005, 2070 citations until March 2024) influential analysis showing the significant effect of bank expansion on poverty reduction. The second chapter shows that the significant impact of India's state-led bank expansion on poverty disappears after considering the additional changes in the bank licensing policy. This finding underscores the necessity of considering external policy influences to ensure the robustness of empirical results.

The second paper, "Uncertain Trends in Economic Policy Uncertainty," takes a critical lens to the European Economic Uncertainty Index, one of the most popular proxies for uncertainty currently used. This paper is co-authored with Jaromir Baxa and Tomáš Šestořád. It shows that the long-term upward trends in EPU of European countries shown by Baker, Bloom, and Davis (QJE, 2016) are not necessarily driven by intrinsic trends in uncertainty and are sensitive to seemingly minor adjustments in the index's construction and disappear if alternative normalization of uncertainty article counts is used. Consequently, the uncertainty associated with recent policy events appears much lower than the original index. By applying alternative normalization techniques, this study demonstrates how such methodological nuances can lead to markedly different interpretations of economic phenomena, such as Brexit or the COVID-19 pandemic.

The third paper titled, "Disentangling p-Hacking and Publication Bias," contributes to the lively debate on the journal's desire to publish studies with significant & large results and the researcher's desire to increase the probability of publication by seeking such results through different approaches such as selective reporting, method searching, and other. The results suggest that the extent of such practices, here referred to as p-hacking, is much more prevalent than previously thought and even more important than selective publication by journals, here referred to as publication bias. This finding underscores the considerable influence of practices such as p-hacking and method-searching, suggesting that they contribute significantly to selection bias in the economic literature and could affect the perceived reliability of published findings.

In this dissertation, I demonstrate the need to develop methodologies for detecting and correcting p-hacking. More importantly, however, we should make greater efforts toward preventing such practices. Potential approaches include preregistration of studies, journal policies that encourage the submission of studies with nonsignificant results, the provision of replication packages, and the promotion of study replication, such as through replication games hosted by the Institute of Replication.

References

Andrews, I., & Kasy, M. (2019). Identification of and correction for publication bias. *American Economic Review*, *109*(8), 2766–94.

Ashenfelter, O., Harmon, C., & Oosterbeek, H. (1999). A review of estimates of the schooling/earnings relationship, with tests for publication bias. *Labour Economics*, *6*(4), 453–470.

Begley, C. G., & Ioannidis, J. P. (2015). Reproducibility in science: Improving the standard for basic and preclinical research. *Circulation Research*, *116*(1), 116–126.

Brodeur, A., Carrell, S., Figlio, D., & Lusher, L. (2023). Unpacking p-hacking and publication bias. *American Economic Review*, *113*(11), 2974–3002.

Brodeur, A., Cook, N., & Heyes, A. (2020). Methods matter: P-hacking and publication bias in causal analysis in economics. *American Economic Review*, *110*(11), 3634–3660.

Bruns, S. B., Asanov, I., Bode, R., Dunger, M., Funk, C., Hassan, S. M., Hauschildt, J., Heinisch, D., Kempa, K., König, J., et al. (2019). Reporting errors and biases in published empirical findings: Evidence from innovation research. *Research Policy*, *48*(9), 103796.

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., et al. (2018). Evaluating the replicability of social science experiments in nature and science between 2010 and 2015. *Nature Human Behaviour*, *2*(9), 637–644.

Card, D., & Krueger, A. B. (1995). Time-series minimum-wage studies: A meta-analysis. *The American Economic Review*, *85*(2), 238–243.

De Long, J. B., & Lang, K. (1992). Are all economic hypotheses false? *Journal of Political Economy*, *100*(6), 1257–1272.

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

Ferraro, P. J., & Shukla, P. (2020). Feature—is a replicability crisis on the horizon for environmental and resource economics? *Review of Environmental Economics and Policy*.

Furukawa, C. (2019). Publication bias under aggregation frictions: From communication model to new correction method. *Unpublished Paper, Massachusetts Institute of Technology*.

Havránek, T. (2015). Measuring intertemporal substitution: The importance of method choices and selective reporting. *Journal of the European Economic Association*, *13*(6), 1180–1204.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLOS Medicine*, *2*(8), e124.

Ioannidis, J. P. A., Stanley, T. D., & Doucouliagos, H. (2017). The Power of Bias in Economics Research. *The Economic Journal*, *127*(605), F236–F265.

Lang, K. (2023). *How credible is the credibility revolution?* (Tech. rep.). National Bureau of Economic Research.

Leamer, E. E. (1983). Let's take the con out of econometrics. *The American Economic Review*, *73*(1), 31–43.

Mathur, M. (2022). Sensitivity analysis for p-hacking in meta-analyses. *OSF Preprints*.

Miguel, E., Camerer, C., Casey, K., Cohen, J., Esterling, K. M., Gerber, A., Glennerster, R., Green, D. P., Humphreys, M., Imbens, G., et al. (2014).

Promoting transparency in social science research. *Science*, *343*(6166), 30–31.

Stanley, T. D. (2005). Beyond publication bias. *Journal of Economic Surveys*, *19*(3), 309–345.

Stanley, T. D. (2008). Meta-regression methods for detecting and estimating empirical effects in the presence of publication selection. *Oxford Bulletin of Economics and Statistics*, *70*(1), 103–127.

Chapter 2

Do rural banks matter that much? Burgess and Pande (2005) reconsidered

Burgess and Pande (2005) Reconsidered[1]

Nino Buliskeria **Jaromir Baxa**

Abstract

We reproduce Burgess and Pande's (American Economic Review, 2005, 95 (3), 780–795) analysis of the effect of India's state-led bank expansion on poverty. The authors instrument rural bank branch expansion by its trend reversal explained by the 1977 licensing rule and find that the bank expansion decreased poverty. However, the authors do not consider other licensing rule amendments and concurrent policies. Thus, their instrument is not necessarily exogenous to poverty. We show that the significant effect of bank expansion on poverty disappears after summarizing the trend reversal with more breaks linked to the bank licensing policy.

JEL: G21, G28, 015, 016

Keywords: Bank expansion, rural poverty, finance

Acknowledgments: This work is supported by the Grant Agency of the Czech Republic (GACR), Grant No. 20-14990S. Furthermore, Nino Buliskeria acknowledges the support by the Charles University Grant Agency, Grant No. 1034519, and Charles University Research Centre program No. UNCE/HUM/035. Jaromir Baxa acknowledges the support by the Charles University Research Centre program No. 035. We are grateful to Robin Burgess, Rohini Pande, Victor Quintas-Martinez, Andreas Mittag, Stepan Jurajda and Ali Elminejad

[1] This paper is published in the Journal of Applied Econometrics. Buliskeria, N., & Baxa, J. (2022). Do rural banks matter that much? Burgess and Pande (2005) reconsidered. Journal of Applied Econometrics, 37(6), 1266–1274. https://doi.org/10.1002/jae.2922.

See the replication package at http://qed.econ.queensu.ca/jae/datasets/buliskeria001 and full online appendix at https:onlinelibrary.wiley.comactiondownloadSupplement?doi=10.1002%2Fjae.2922&file=jae2922-sup-0001-OnlineAppendix_merged%281%29.pdf

for helpful comments and suggestions. The responsibility for all remaining errors and omissions rests solely on us.

2.1 Introduction

Over the years, easing access to financial services has become a widely accepted approach to reducing poverty and enhancing overall economic growth. Although Banerjee and Duflo (2011) have questioned the transformative power of microfinance, a focus on access to finance has been incorporated into many development strategies. Several alternative policies have already been carried out worldwide, including credit subsidies for the poor, microfinance, and bank expansion programs (Robinson, 2001; Khandker, 2005; Gertler et al., 2009). India was a pioneer in implementing large-scale state-led financial expansion programs at the end of the 20th century to expand virtually nonexistent rural banking and decrease poverty (Burgess and Pande, 2005; Kochar, 2011; Panagariya, 2008).

Burgess and Pande (2005) is one of the most influential papers studying the effect of easier access to credit and saving facilities on poverty. The authors study India's state-led bank expansion from 1961 to 2000 and find that bank expansion decreased rural poverty. Their paper has been acknowledged in policy publications (i.e., Honohan and Beck, 2007; Claessens and Feijen, 2007; Jahan and McDonald, 2011) and is mentioned in the Handbook of Development Economics (Rodrik and Rosenzweig, 2010). However, due to the identification issues of the causal effect between bank expansion and poverty, Panagariya (2008), Kochar (2011), and Fulford (2013) remain skeptical of the findings presented in Burgess and Pande (2005).

Fulford (2013) points to different effects of credit on poverty across different time horizons. Contrary to Burgess and Pande, Fulford finds that the Indian bank expansion initially decreasing poverty and increasing consumption, but increasing poverty and decreasing consumption in the long run, when the debts need to be repaid. Consequently, the increase in branches per capita in rural areas caused a slight increase in rural poverty.

The main challenge for identification is the potential endogeneity between where banks are willing to expand and poverty levels. Since banks prefer areas with increasing business opportunities, bank expansion is not exogenous to poverty. Burgess and Pande (2005) addressed the causal inference by instrumenting the bank expansion with the imposition and removal of the 1977 bank branch licensing rule, which required banks to open four branches in unbanked locations for each newly opened branch in a banked location (1:4 rule). However, Kochar (2011) and Panagariya (2008) point out that the rules linking rural and urban financial development were introduced at the beginning of the 1960s and amended multiple times, including once in 1977. The authors argue that the 1977 amendment was not significantly different from the past rules (see Panagariya, 2008, pg. 224) and was updated only two years later in 1979 to bank-to-population targeting. Hence, for most of the treatment period of 1977-1989, the bank licensing rules were different from the 1977 update. Kochar (2011) further asserts the importance of concurrent credit subsidizing policies, while Burgess and Pande do not discuss them thoroughly. However, such policies could have influenced trend reversal in the rural bank expansion while targeting poverty. Therefore, the selection of the 1977 update of licensing policy to instrument the bank expansion appears unclear, and its exogeneity to poverty is questionable.

In this paper, we successfully reproduce the empirical results of Burgess and Pande (2005) first, and next analyze their identification strategy. Since the authors use 1977 trend break in rural bank branch openings as the instrument in the original identification strategy this trend reversal should be significantly driven by the trend break in 1977, (imposition and removal of the 1977 bank branch licensing rule). After switching 1977 with other hypothetical trend break years, we find significant trend reversals almost every year between 1970 and 1984. The significance of those cut-years increases suspicion of the importance of different poverty targeting policies for the dynamics of the banking network. Some of these years coincide with the introduction of other programs targeting poverty reduction rather than bank expansion.

Next, we analyze the impact of inclusion of additional trend reversal along with the one in 1977 to control for the introduction of other amendments of bank licensing policy and other policy interventions. These results show that with an additional cut-year around 1985, when the government reduced the pace of rural expansion, the effect of bank expansion on poverty decreases and becomes statistically insignificant. The impact of bank expansion on poverty also becomes insignificant when using a more trend reversals of rural branch openings as the instrument. Therefore, the main finding of Burgess and Pande that the state-led bank expansion decreased poverty is not robust to sensible modifications of the instrumental variable. The main reason is that the trend reversal, summarized by the imposition and removal of the 1977 rule, is neither an accurate description of exogenous shifts in bank licensing policies nor exogenous to the confounding variables simultaneously driving poverty and bank expansion. These findings cast doubt on the estimated effects of bank expansion on poverty provided in the original paper.

The remainder of this paper is organized as follows. Section two reviews the history of Indian poverty-alleviating policies. The third section presents the reproduction of Burgess and Pande (2005). Section four discusses the instrument's relevance and exogeneity and finally section five concludes.

2.2 The Indian policies of the late 20th century.

The Indian government initiated numerous policies for expanding banking sector outreach and formal credit access across the country with the aim to decrease poverty. Under the Nationalization Act of 1969, the Reserve Bank of India (RBI) took over the 14 largest commercial banks and launched an extensive bank expansion program aiming at equal access to financial services for the poor (Burgess and Pande, 2005; Banerjee et al., 2004; Kochar, 2011).

Between 1969 and 1979, the banks were expected to comply with the required urban to rural bank branch opening ratios.[2] Although the ratio was modified multiple times, the bank expansion rules of before 1979 were considered inefficient at reducing poverty (Kochar, 2011 and Panagariya, 2008). Therefore in 1979, the Government of India took over and nationalized six additional banks (Banerjee et al., 2004; Cole, 2009) and updated the bank licensing policy to target the population-to-rural bank branch ratio instead of the rural-to-urban branch opening ratio (Panagariya, 2008). The highest rate of branch expansion since the 1960s was reached over the next five years from 1980 to 1985.

However, with the first wave of liberalization in 1985, the speed of rural branch openings fell dramatically. After 1985, the government aimed to strengthen the rural banking system and reduce the pace of branch expansion, which caused increasing losses (Mohanty & Acharya, 2006). The government started to introduce programs that would ease access to credit using existing banking networks. Therefore, while the expansion of branches decelerated, deposits and advances increased considerably relative to the previous periods.

[2] In 1970, the RBI adopted a 1:3 ratio of banked to unbanked branch openings. In September 1971, the rule was updated to a (1+1):3 ratio, i.e., banks had to open 3 rural branches for every branch opened in metropolitan and urban locations. Since the "1:4" rule was, in reality, the (1+1):4 rule, it was less demanding than the rule introduced in 1970 and only slightly more demanding than the (1+1):3 rule of 1971 (see Panagariya (2008) for a more detailed description).

Along with the bank branch expansion rules, other poverty elimination policies were adopted. In 1978, the government introduced the Integrated Rural Development Programme (IRDP) to increase productive assets in rural locations via credit subsidies. By the end of 1980, this program was one of the prominent tools in fighting countrywide poverty. According to Kochar (2011), the IRDP affected both the poverty and potential patterns of bank expansion, making it nearly impossible to disentangle the poverty-alleviating effects of credit subsidizing and bank expansion programs. The National Rural Employment Programme was also introduced in 1980 to address both poverty and unemployment in rural areas by providing wage employment during the slack agricultural season. While the bank expansion program was gradually abandoned from the mid-1980s, it was finally discontinued in 1990, as the RBI stopped interfering in the banks' expansion decisions (Kochar, 2011; Panagariya, 2008; Burgess and Pande, 2005). Nevertheless, the decrease in poverty accelerated in the 1990s even without bank expansion policies (Figure A3, Online Appendix A).

2.3 Reproduction of the work

Data. Following Burgess and Pande, we use Indian National Survey Data (Datt et al., 1996) on bank branch opening dates, locations, and state characteristics (poverty, wages, expenditure, land reform, and population) for 1961 to 2000 and 16 Indian states. The authors categorize the branch openings into two classes: openings in rural unbanked and banked locations. The first classification refers to branches that opened in a previously unbanked rural location; the latter refers to openings in areas with one or more branches. The number of branches per capita in each state in 1961 is used as the proxy for initial financial development, while branches per capita in each state over time measures the level of branch expansion. Finally, Burgess and Pande (2005) use a poverty headcount ratio (the share of the population below the poverty line) from national household expenditure surveys to measure poverty.

Methodology. To evaluate the state-led bank expansion effect on poverty, Burgess and Pande first address the potential endogeneity problem. Without limitations, banks are willing to expand into wealthier states. Since richer states are more successful in tackling poverty, the impact of bank expansion on poverty could be overestimated. In contrast, if the RBI successfully forced banks into opening branches in poorer states, the effect of branch expansion on poverty could be underestimated.

Burgess and Pande address the identification problem by instrumenting rural expansion with the imposition and removal of the 1:4 branch licensing policy, introduced in 1977. This policy should have caused speedier branch openings in rural versus urban locations from 1977 - 1990 relative to other periods. The validity of this instrument requires significant trend reversal in bank branch expansion caused by the exogenous policy change.

Bank expansion and initial financial development. To test the validity of the instrumental variable, the authors first study the effect of initial financial development on the trend of bank branch opening[3]:

$$B_{it}^R = \alpha_i + \beta_t + \sum_{k=1961}^{2000} (B_{i1961} * D_k)\gamma_k + \sum_{k=1961}^{2000} (X_{i1961} * D_k)\delta_k + \varepsilon_{it}, \qquad (2.1)$$

where B_{it}^R is the number of banks opened in rural areas in state i and year t. B_{i1961} measures the 1961 financial development level in state i; this variable enters the regression interacting with time fixed effects D_k, which equal one if $k = t$ and zero otherwise. The set of coefficients γ_k show the year-specific effect

[3] In this section, we used the methodology in line with the Stata code accompanying Burgess and Pande (2005). In fact, the equations reflect the working paper (Burgess & Pande, 2003) rather than the version published in the AER.

of the initial financial development on branch openings (Figure 1(a), solid line). X_{i1961} is a vector of initial state conditions — log real state income per capita, population density, and the number of rural locations per capita, all measured for 1961. These values enter the regression interacting with year dummies and year-specific coefficient δ_t.

Next, Burgess and Pande summarize the dynamics of the impact of initial financial development on bank expansion using the linear trend break model.

$$\begin{aligned} B_{it}^R = \alpha_i + \beta_t + \gamma_1(B_{i1961}^R[t-1961]) \\ + \gamma_2(B_{i1961}^R[t-1976]P_{1977}) \\ + \gamma_3(B_{i1961}^R[t-1990]P_{1990}) \\ + \gamma_4(B_{i1961}^R P_{1977}) + \gamma_5(B_{i1961}^R P_{1990}) + F(X_{i1961}) + \varepsilon_{it} \end{aligned} \tag{2.2}$$

State and year fixed effects α_i and β_t account for differences in state and time-specific characteristics. Linear time trends $[t-1961]$, $[t-1977]$, and $[t-1990]$ switch on in 1961, 1977, and 1990 and enter the regression interacted with the measure of the state's initial financial development, B_{i1961}. P_{1977} and P_{1990} are dummy variables that equal one from 1977 and 1990 to 2000, respectively.[4] The inclusion of additional controls X_{i1961} ensures that observed trend reversal in B_{i1961} does not reflect trend breaks in a state's economic and demographic characteristics. The standard errors are clustered by state to account for possible serial correlations.

We successfully reproduce all estimations in Burgess and Pande (2005) (Online Appendix A). To summarize, the trend reversals in 1977 and 1990 are statisti-

[4]The corresponding equation (3) in Burgess and Pande (2005) does not include pulse dummies P_{1977} and P_{1990} in the interaction terms with the trends, although they are included in the authors' code. These dummies ensure that the trends affect only the respective time periods: without them, the trends would be negative for preceding periods.

cally significant. There is a significant downward sloping trend in the effect of initial financial development on bank expansion after 1977, so more bank branches opened in previously unbanked locations, and the trend disappears after 1990 when the bank licensing policy was removed. These results are confirmed by the F-tests of restrictions $\gamma_1 + \gamma_2 = 0$ and $\gamma_1 + \gamma_2 + \gamma_3 = 0$ (F-test 1 & 2, Table A1, Online Appendix A). [5]

Impact of bank branch expansion on poverty. To analyze the effect of bank branch expansion on India's rural poverty, the authors estimate equation (3).

$$\begin{aligned} y_{it} = \alpha_i + \beta_t + \phi B_{it}^R + \mu_1([t-1961]B_{i1961}) \\ + \mu_2(P_{1977}B_{i1961}) + \mu_3(P_{1990}B_{i1961}) + u_{it} \end{aligned} \tag{2.3}$$

They first run a simple OLS regression (columns 1 and 2 of Table 1) and next proceed to the two-stage IV estimation. The first stage regression coincides with equation (2), and the second stage corresponds to equation (3) based on the fitted values of B i t R from equation (2).

Columns 3 to 5 of Table 1 present the reproduced IV estimates for poverty outcomes as shown in Burgess and Pande (2005). A one-point increase in per capita branch opening in rural unbanked locations explains a 4.74-percent reduction in rural poverty (column 3), which evaluated at the sample average implies a 17-percent decrease in the poverty headcount ratio. Moreover, this

[5]Following Burgess and Pande's exogeneity check, we confirm the absence of significant trend reversals in credit flows to priority sectors and primary agricultural cooperatives (columns 5 and 6, Table A1). Analogously, the main state economic and policy variables that influence rural poverty do not exhibit a trend reversal similar to that shown in Figure 1(a). Therefore in Burgess and Pande (2005), the imposition and removal of the 1:4 rule are assumed to be valid instruments since the trend breaks of the effect of initial financial development on rural branch expansion are significant, and there are no structural breaks in other political and policy variables.

process had no impact on urban poverty (column 4). Aggregate poverty in rural locations decreases by 4.10 percentage points, with every additional branch opening in a rural location per 100,000 persons (column 5).

Furthermore, we verify that the main results given in column 3 of Table 1 are robust even after controlling for time-varying political and policy variables (Table A4, Online Appendix A). Thus, the negative and significant relationship between rural branch expansion and rural poverty persists even after controlling for increased land reform and development spending, which is known to reduce rural poverty (Besley & Burgess, 2000).

Table 2.1: Bank branch expansion and poverty: IV evidence

	Headcount ratio				
	Rural			Urban	Aggregate
	OLS		IV	IV	IV
	(1)	(2)	(3)	(4)	(5)
Number of bank branches opened in rural unbanked locations per capita	2.09**	1.16	-4.74**	-0.66	-4.10**
	(0.785)	(1.024)	(1.790)	(1.066)	(1.464)
Number of bank branches per capita 1961·(1961–2000) trend		-0.43**	-0.48*	-0.26*	-0.46*
		(0.165)	(0.269)	(0.134)	(0.226)
Number of bank branches per capita in 1961·Post-1976 dummy†		-0.31	-1.42	-2.06	-1.39
		(1.229)	(2.297)	(1.654)	(2.034)
Number of bank branches per capita in 1961·Post-1989 dummy†		5.38**	-1.08	-0.47	-1.55
		(2.468)	(2.334)	(1.015)	(1.759)
State and year dummies	YES	YES	YES	YES	YES
Other controls	YES	YES	YES	YES	YES
Overidentification test			[0.99]	[0.99]	[0.99]
Adjusted R-squared	0.807	0.834	0.760	0.915	0.818
Observations	627	627	627	627	627

Source: This table is a replication of Table 3 in Burgess and Pande (2005), page 789. For replication we used data and methodology provided by the authors. †Original paper contains Post-1976 dummy·(1977–2000) trend and Post-1989 dummy·(1990–2000) trend instead, which is not consistent with the text and the stata code. Therefore, we have changed the variable names accordingly. Note: The IV estimates correspond to equation (3) for different dependent variables. Other controls include state population density, log state income per capita, log rural locations per capita, all measured in 1961. The over-identification is tested using the conventional Sargan test. Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

2.4 Sensitivity check: Different cut-years

The results and the policy implications of Burgess and Pande (2005) are heavily conditional on the assumption that their instrument, imposition and removal of the 1977 bank licensing policy, provides a credible source of exogenous variation in rural bank expansion. Thus, the trend reversal in bank expansion should be driven solely by the 1:4 rule and not by any other simultaneous policy interventions targeting rural poverty.[6]

Nevertheless, Panagariya (2008) and Kochar (2011) assert that the poverty-alleviating effect of state-led bank branch expansion cannot be evaluated without considering coexisting credit subsidizing policies and other amendments of the bank branch expansion rules. Since these policies affected poverty and enhanced access to credit, they may have created additional incentives for the expansion, aside from licensing rule requirements. Therefore, these concurrent policies make the exogeneity of the imposition and removal of the 1:4 rule questionable.

We start addressing the instrument validity concerns by re-estimating equation (2) with different hypothetical trend break years (instead of 1977) as the instruments for bank expansion in equation (3). These alternative trend breaks often coincide with other changes in the bank licensing rules or implementations of other poverty-alleviating policies, most notably the IRDP. Next, as if the most prominent policies affecting bank expansion were unknown, we test for which years the F-statistics of these trend breaks are significant and maximized.

Different cut-year than 1977. We find that the trend reversals are significant for virtually all cut-years in the sample, and there is little difference between the F-tests of 1977 and surrounding years (Figure 1(b) and Table B1.1, Online Appendix B). It can be argued that the significance of trend breaks around 1977 can be driven by the policy reversal of 1977 due to autocorrelation in the series of rural branch expansion. However, the other years are linked to different policies, so the contrary may also be true - the significance of the trend break in 1977 could be driven by the success of other policies. Additionally, the importance of other policies is supported by the F-statistics, that peaks

[6] As mentioned in Burgess and Pande (2005), the requirements for IV validity are "no direct effect on poverty outcomes" and the significance of the trend reversal.

in 1979 and 1980, when the IRDP was introducted. The significance of those cut-years increases suspicion of the importance of different poverty targeting policies for the dynamics of the banking network. Interestingly, although these trend reversals are not necessarily exogenous to the poverty rate, they lead to virtually the same implications as the estimates based on the cut-year of 1977.[7]

Since the rule of 1977 was an update of existing policy, the 1977 trend break merely summarizes the long run dynamics of rural bank expansion (Figure 1(a)). From this perspective, and in the absence of the credit subsidizing policies of the early 1980s, any cut-year effectively summarizing the trend reversal of the rural bank expansion could be an acceptable instrument, particularly when linked to other changes of licensing rules. Despite some uncertainty about the coefficient's value, the estimated effect of bank expansion on poverty would remain conceivable with different cut years as well (Figure 2(a) and Table B3.1, Online Appendix B). However, the credit subsidizing and employment policies of the early 1980s (summarized in Copestake, 1996) were conceptually different from the bank expansion program. Through these policies, the Indian government provided paid labor and cheap credit to poor households, thus making treated unbanked areas more attractive for expansion. Therefore, these credit subsidizing policies may have affected the trend reversal of rural branch expansion, even without the expansion rules, casting doubt on the validity of the original instrument.

One additional cut-year along with 1977 & 1990. We add another

[7] The quantitative results for different cut-years are shown in Tables B1.1 to B5.2, Online Appendix B, while the plots of the trend reversals are presented in Figure A6 in Online Appendix A. We also analyzed the effects of trend reversals for different cut-years on credit flows to the priority and cooperative sectors, and we do not identify significant trend reversals. According to Burgess and Pande (2005), this should be an indication of the exogeneity of specific trend breaks. As it appears, an absence of similar trend reversals in credit flows and other variables is not a sufficient condition for the exogeneity of the trend break in bank expansion with respect to poverty. These results are available upon request.

cut-year to summarize the trend of rural branch openings better (solid line in Figure 1(a)) and to check the robustness of the original results. If the 1:4 rule of 1977 was prominent for the rural bank expansion as suggested by Burgess and Pande, the inclusion of the additional cut-year, associated with other bank licensing rules, should neither influence the significance of 1977 cut-year nor change the final results for the effect of bank branch expansion on poverty.

The inclusion of additional cut-years before 1977 affects the significance of the 1977 trend reversal, but the final result of the poverty-alleviating effect of rural branch expansion remains similar. The trend breaks in 1973 to 1975 are significant, and the inclusion of the 1975 cut-year makes one in 1977 insignificant. These results indicate that the trends in unbanked branch expansion already changed before introducing the 1977 rule (Online Appendix C).

On the other hand, the inclusion of additional cut-years from 1977 onwards had diverse effects. While the 1977 break year remains statistically significant, the significance of trend reversal in the effect of initial financial development on poverty disappears[8] (Tables C1.1 & C2.1, columns 10 - 24, Online Appendix C). Notably, after including an additional cut-year between 1984 and 1987, the poverty-alleviating effect of rural branch expansion weakens and becomes statistically insignificant (see Table C3.1, columns 20-24, Online Appendix C). These cut-years coincide with the Seventh Five-Year Plan of 1985 and the first wave of liberalization.

Additional cut-years according to historical events. According to Panagariya (2008) and Mohanty and Acharya's (2006) detailed analysis of Indian supply-driven policies, bank licensing rule amendments happened in 1967,

[8]The significance of this trend reversal in the case of poverty is recognized as additional evidence for the trend break being a relevant instrument in Burgess and Pande.

1972, 1977, 1980 and 1985, and ended in 1990.[9] Therefore, we summarize the trend reversal in rural branch opening using these five cut-years (similarly to regression 2). Since these cut-years are reasonably exogenous, we use them as an instrument to estimate the impact of branch expansion on poverty (based on equation 3). As a result of this exercise, the poverty-alleviating effect of state-led bank expansion in rural areas becomes positive but statistically insignificant.[10] Similar results were obtained with 1968, 1973, 1979, 1984, and 1989 cut-years reccomended by the Bai and Perron (2003) test for multiple unknown breaks - no significant effect of the branch expansion on poverty.[11]

[9]In 1967, Gandhi's Ten-Point Programme was announced, paving the path towards priority sector lending, the nationalization of commercial banks, and other interventions. The bank licensing policy was applied in the same year, requiring a specific ratio of new rural bank branches per every urban branch opening. The 1971 war between Pakistan, partnered by the United States and India, supported by the Soviet Union, pushed India towards "socialist" interventionist policies. The bank licensing rules were updated and tightened in February 1970 and September 1971. Later, in 1974, the RBI introduced additional guidelines for banks to provide a minimum of one-third of aggregate advances to priority sectors, covering small-scale industry. Next, in 1977, the 1:4 expansion rule, highlighted by Burgess and Pande (2005), was implemented, while in 1980, the Integrated Rural Development Programme was fully extended, and the National Rural Employment Programme was launched. The year 1985 marks the start of the Seventh Five-Year Plan with strengthened though not yet systematic liberalization efforts and the year in which the low profitability of rural bank branches started to be addressed. Finally, a major liberalization plan was proposed in 1990, just before the balance of payments crisis of 1991 became apparent.

[10]The estimated coefficient is 1.66 (Table D3, column 3, Online Appendix D) while Burgess and Pande's estimate is -4.74, as reported in Table 1.

[11]Notably, initial conditions (in 1961) play an essential role in Burgess and Pande (2005) - the first-stage regression, equation (2), estimates how the branch openings depend on initial financial development and several other initial conditions, like rural locations, population, and income. Moreover, Ravallion and Datt (2002) find that non-farm growth benefits the poor more in states with initially higher literacy, farm productivity, rural living standards, and lower landlessness and infant mortality. Burgess and Pande (2005) consider only urbanization rate and initial output as controls for the level of initial financial development. Once the set of controls is extended with the initial levels of infant mortality, literacy, crop production per farm worker, and percent of laborers of rural farmworkers, the estimated effect of initial financial development on branch openings (Figure 1a) changes dramatically. The trend reversal appears at the beginning of the 1980s, and the estimated impact of bank expansion on rural poverty remains negative but becomes statistically insignificant. The results of this additional exercise are available in Online Appendix E.

2.5 Conclusion

We reproduced the study on the impact of bank branch expansion in previously unbanked areas on poverty by Burgess and Pande (2005). Using trend reversals in 1977 & 1990 as instruments, we successfully identified significant trend reversals in bank branch expansion and obtained the same negative effect of new branch openings on poverty as in the original study. Then, to assess validity of identification of the effect of bank branch expansion on poverty with trend reversals, we review the history of the Indian banking policies and repeated Burgess and Pande's exercise with other hypothetical policy introduction years of 1970 – 1984. Our results imply that any cut-year from 1974 to 1981 leads to similar results to those with a cut-year in 1977. These results hold even though some of the hypothetical breaks are also associated with policies targeting poverty and not bank expansion. Therefore, the trend reversal in rural bank expansion of 1977 could be caused by other policies rather than by the 1977 bank licensing rule itself, which casts doubt on the instrument validity used for identification of the effect of bank branch expansion on poverty.

Next, as additional robustness checks, we introduced additional trend breaks, along with that of 1977, based on a sequence of primary policy shifts or Bai and Perron's test for multiple unknown breaks. Notably, upon adding a break in the middle of the 1980s (i.e., the years when liberalization efforts increased and the rate of rural bank branch expansion decreased), the impact of rural branch expansion on poverty decreases, and while remaining positive, it loses statistical significance.

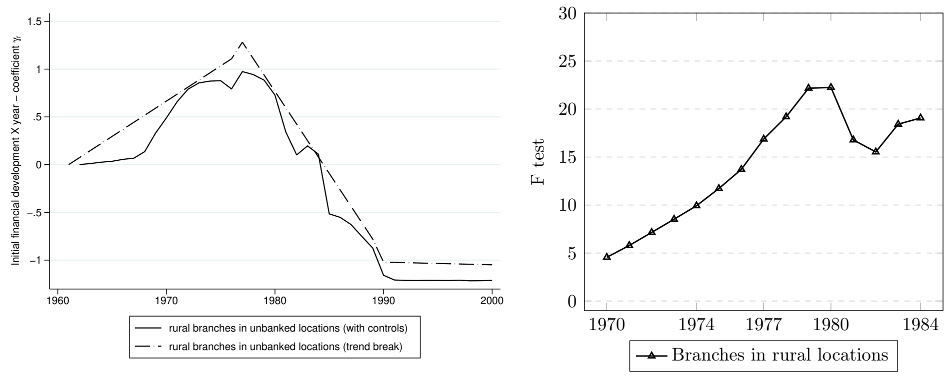
We conclude that Burgess and Pande's central challenge of disentangling the effect of rural bank branch expansion from the effects of other policies on poverty has not been reliably addressed. The prominence of the 1979-1980 thresholds in the rural branch expansion and sensitivity of the final results to different instrumental variable specifications suggests that it is not possible to identify the poverty-alleviating impact of the bank branch expansion separate from the effect of the credit subsidizing programs (i.e., the IRDP). Consequently, situating banks in unbanked locations is unlikely to have results similar to those given in Burgess and Pande (2005) without further subsidizing credit and aiding with other policies. In conclusion, the 1977 and 1990 trend reversals' validity as the instrumental variable of branch expansion is highly debatable due to its potential correlation with poverty outcomes caused by policies different from the 1977 bank licensing policy. Therefore, our results imply that the final effect of bank branch expansion on poverty is not properly estimated.

From a policy perspective, our results further amplify doubts expressed in the working paper version of Burgess and Pande (2005) (Burgess & Pande, 2003) on the effectiveness and cost-efficiency of the state-led bank branch expansion relative to potential alternatives. Despite the inclination towards bank branch expansion in fighting poverty, critical lessons from the Indian experience, including high default rates when using credit as a redistribution tool, must not be overlooked. Therefore, development strategies should consider more efficient policies in mitigating poverty and not rely on easier access to banking alone.

Figure 2.1: Initial financial development and rural bank branch expansion

(a) Bank branch expansion

(b) Trend breaks in rural bank expansion

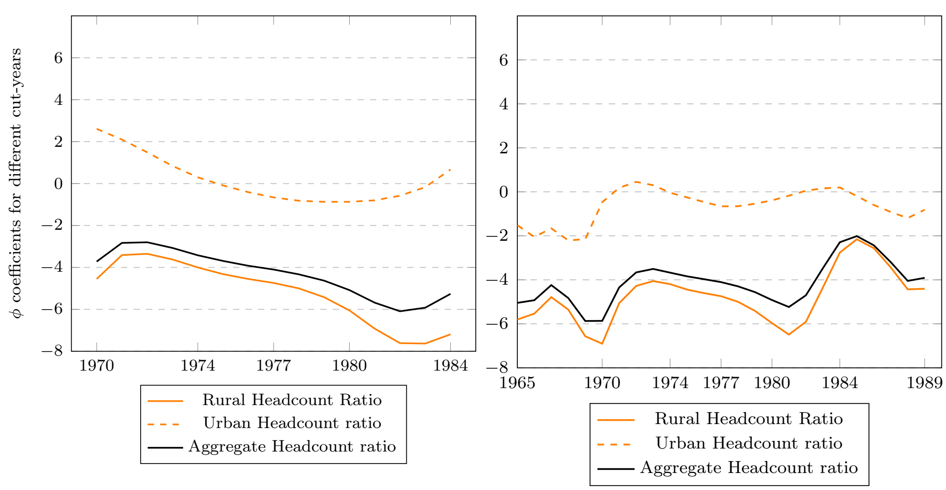


Note: **Figure 1a.** The series "rural branches in unbanked locations (with controls)" shows the annual coefficients of the effect of initial financial development on branch expansion from the equation (1). The series "rural branches in unbanked locations (trend break)" graphs the trends obtained from equation (2), and correspond to the results reported in Table A1, column 1 in Appendix. Burgess and Pande (2005) disregard the impact of pulse dummies in their Figure 1. The reference year is 1961. The figure correspond to Figures 1 in Burgess and Pande (2005), p.784. **Figure 1b.** The figures show F-statistics of a restriction that in the equation (2) the coefficients $\gamma_1 + \gamma_2 = 0$ (F-test 1 in Table B1.1, Online Appendix B) in any of the sample years. For all years, the signs of the coefficients switch as well, so the trend reversal is statistically significant. The coefficients and the F-statistics shown in figures 4a & 4b are presented in the tables B1.1 - B1.4 & B2.1 - B2.3, Online Appendix B.

Figure 2.2: Impact of bank branch expansion on headcount poverty, instrumental variable approach

(a) Different cut year T instead of 1977

(b) Multiple cut years: 1977 and year T



Note: **Figure 2a** shows the coefficient ϕ from equation (3) for different cut-year T instead of 1977, whereas **Figure 2b** shows the coefficient ϕ when including additional cut-year T along with 1977 in equation (3). In both cases, the cut-year in 1990 is included as well. The coefficients shown in figure 2 are presented in tables C3.1 - C3.3, Online Appendix C.

References

Bai, J., & Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, *18*(1), 1–22.

Banerjee, A., Cole, S., & Duflo, E. (2004). Banking reform in India. *India Policy Forum*, *1*(1), 277–332.

Banerjee, A., & Duflo, E. (2011). *Poor economics: A radical rethinking of the way to fight global poverty*. Public Affairs.

Besley, T., & Burgess, R. (2000). Land reform, poverty reduction, and growth: Evidence from India. *The Quarterly Journal of Economics*, *115*(2), 389–430.

Burgess, R., & Pande, R. (2003). Do rural banks matter? evidence from the Indian social banking experiment. *BREAD Working Paper: No. 037*.

Burgess, R., & Pande, R. (2005). Do rural banks matter? Evidence from the Indian social banking experiment. *American Economic Review*, *95*(3), 780–795.

Claessens, S., & Feijen, E. (2007). Financial sector development and the millennium development goals. *World Bank Working Paper No. 89.*

Cole, S. (2009). Fixing market failures or fixing elections? agricultural credit in India. *American Economic Journal: Applied Economics*, *1*(1), 219–50.

Copestake, J. G. (1996). The resilience of IRDP: Reform and perpetuation of an Indian myth. *Development Policy Review*, *14*(1), 51–68.

Datt, G., Ozler, B., & Ravallion, M. (1996). A database on poverty and growth in India. *World Bank Development Research Group, Washington DC*.

Fulford, S. L. (2013). The effects of financial development in the short and long run: Theory and evidence from India. *Journal of Development Economics*, *104*, 56–72.

Gertler, P., Levine, D. I., & Moretti, E. (2009). Do microfinance programs help families insure consumption against illness? *Health Economics*, *18*(3), 257–273.

Honohan, P., & Beck, T. (2007). *Making finance work for Africa*. World Bank.

Jahan, S., & McDonald, B. J. (2011). A Bigger Slice of a Growing Pie: Developing the financial sector accelerates economic growth and can enhance income equality. *Finance & Development*, *48*(003).

Khandker, S. R. (2005). Microfinance and poverty: Evidence using panel data from Bangladesh. *The World Bank Economic Review*, *19*(2), 263–286.

Kochar, A. (2011). The distributive consequences of social banking: A microempirical analysis of the Indian experience. *Economic Development and Cultural Change*, *59*(2), 251–280.

Mohanty, A., & Acharya, S. (2006). *Operational analysis of regional rural banks*. Gyan Publishing House.

Panagariya, A. (2008). *India: The Emerging Giant*. Oxford University Press.

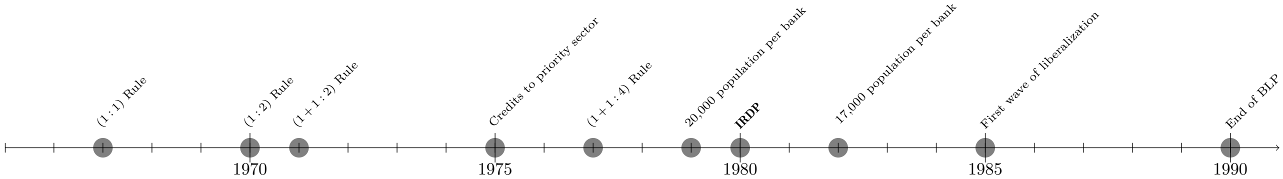
Ravallion, M., & Datt, G. (2002). Why has economic growth been more pro-poor in some states of India than others? *Journal of Development Economics*, *68*(2), 381–400.

Robinson, M. (2001). *The microfinance revolution: Sustainable finance for the poor*. World Bank Publications.

Rodrik, D., & Rosenzweig, M. (Eds.). (2010). *Handbook of development economics* (1st ed., Vol. 5). Elsevier.

2.A Additional figures and tables

Figure 2.A.1: Chronology of the Bank Licensing Policies



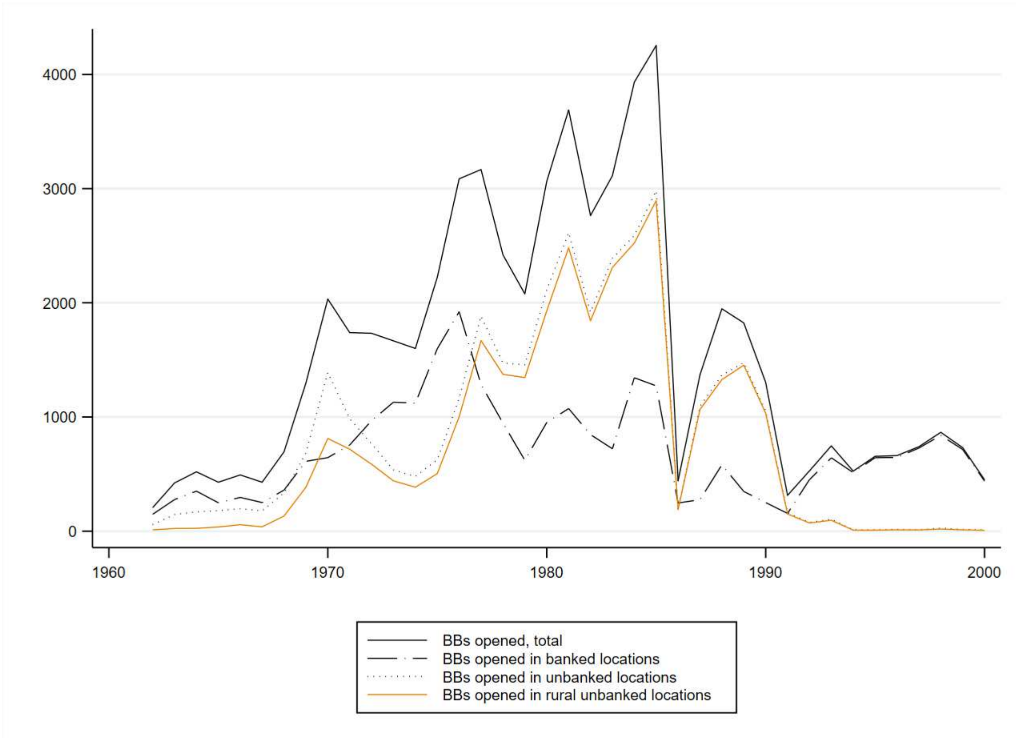
Note: The chronology of events is presented as described by Kochar (2011), Panagariya (2008) and Banerjee et al. (2004).

History: The Government of India initiated numerous policies expanding the banking sector outreach and formal credit access all over the country with the aim to decrease poverty. Starting from 1949, the Reserve Bank of India (RBI) was in control of the banking network expansion as the commercial banks needed a license from the RBI to open a new branch. Later, under the Nationalization Act of 1969, the RBI took over the 14 largest commercial banks and launched a massive bank expansion program aiming at equal access to financial services for the poor (Burgess and Pande, 2005; Banerjee et al., 2004; Kochar, 2011).

Between 1969 and 1979, the banks were expected to comply with the required urban to rural bank branch opening ratio. In 1970, the RBI adopted 1:3 between banked and unbanked branch openings. In September 1971 the rule was updated to a (1+1):3 ratio, i.e., banks had to open 3 rural banks for every bank opened in metropolitan and urban locations, one in each. Although the ratio was modified multiple times, the bank expansion rules before 1979 were considered inefficient in reducing poverty (Kochar, 2011 and Panagariya, 2008).

Therefore in 1979, the Government of India took over and nationalized six additional banks (Banerjee et al., 2004; Cole, 2009), and updated the bank licensing policy to target the population-to-rural bank ratio instead of rural-to-urban branch opening ratio (Panagariya, 2008). The highest speed of branch expansion, since the 1960s, was reached during the next five years, from 1980 to 1985. However, with the first

Figure 2.A.2: Annual Bank Branch Openings in India



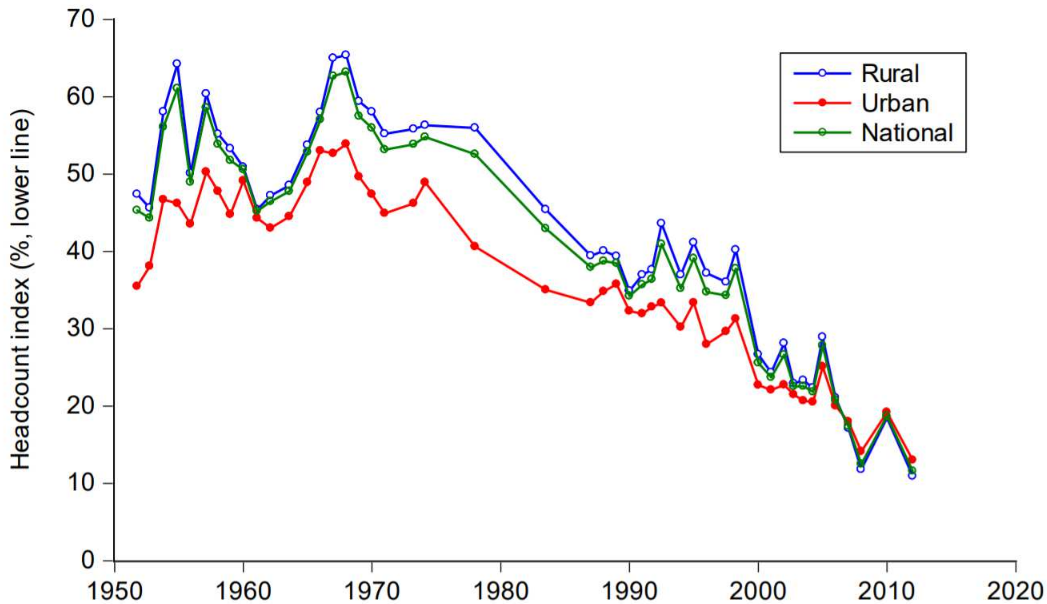
Note: This figure shows how many bank branches (BBx) were opened in India each year, further distringuishing between openings in banked locations (dot-dashed line), unbanked locations (dotted line), and rural unbanked locations (orange line).

wave of liberalization in 1985, the speed of rural branch openings dramatically fell (Figure tba).

Mohanty and Acharya (2006) point out that after 1985, the government aimed to strengthen the rural bank system and reduce the pace of branch expansion, causing increasing losses. The government aimed to increase the functionality of those branches. Therefore, as the expansion speed decreased, the deposits and advances increased considerably compared to the previous periods. The government started to introduce programs that would ease access to credit using existing banking networks.

In 1978, the government introduced the Integrated Rural Development Programme (IRDP) to increase productive assets in rural locations via credit subsidies. By the end of 1980, this program became one of the most prominent tools in fighting countrywide poverty. The National Rural Employment Programme was also introduced in 1980 to address both poverty and unemployment in rural areas by providing wage employment during the slack agricultural season. While the bank expansion program was gradually abandoned since the mid-1980s, it was finally halted in 1990, as the RBI stopped interfering in the banks' expansion decisions (Kochar, 2011; Panagariya, 2008; Burgess and Pande, 2005). Nevertheless, the decrease of poverty accelerated in the 1990s even without bank expansion policies (Figure A1).

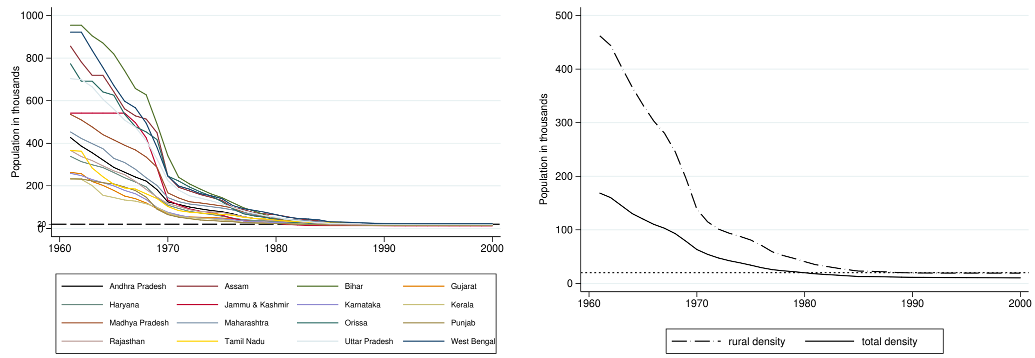
Figure 2.A.3: Evolution of Indian Poverty



Note: Decrease of rural poverty in India accelerated in the 1990s, after the bank expansion policy was abandoned. Note, however, that over the course of the 1970s the data on poverty are based on surveys conducted in five years intervals, so the turning points in poverty could have appeared in other years than shown in Figure A3. Souce:Datt et al. (2016).

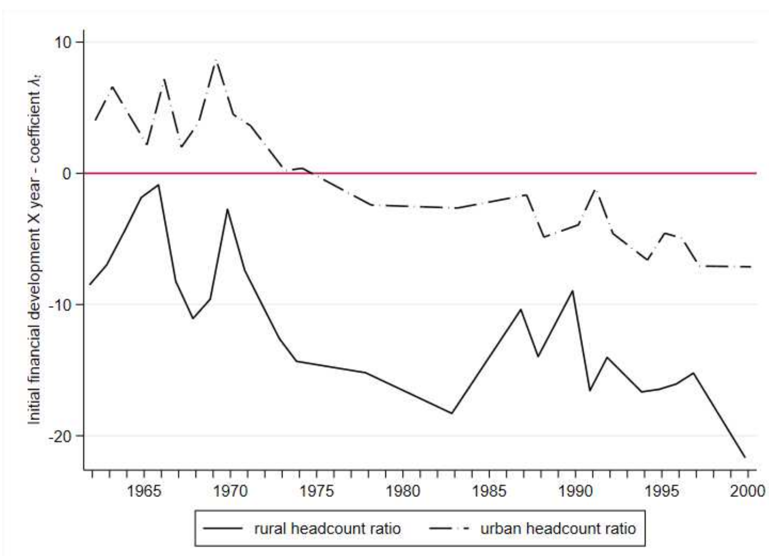
Figure 2.A.4: Population-to-rural bank ratio total for India and by state

(a) Population-to-rural bank ratio by state (b) Population-to-rural bank ratio, India



Note: The population-to-rural bank ratio (by state and country average) significantly decreases after the 1960s and converges to around 20000 people per bank, even in rural areas before 1990. The figures were generated based on data provided by Burgess and Pande (2005)

Figure 2.A.5: Initial financial development and poverty



Note: Figure A5 graphs the impact of initial financial development λ_t from equation (4) in Burgess and Pande (2005) for both rural and urban poverty over the period 1961 - 2000. It shows that the initial financial development is negatively correlated to rural poverty. This figure replicates Figure 3 in Burgess and Pande (2005), p.787.

Figure 2.A.6: Initial financial development and rural branch expansion by cut-years

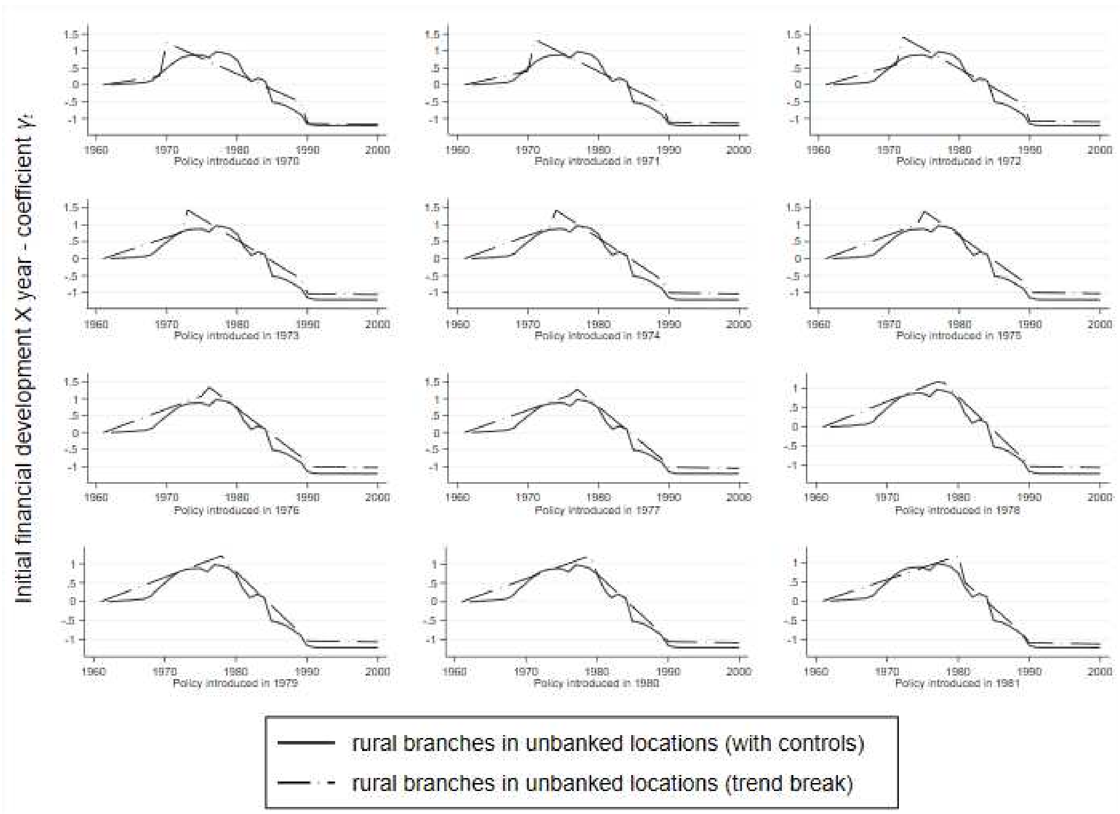


Figure A6 presents the trend reversals obtained from introducing the different hypothetical cut-years, and their fit to the coefficient of the effect of initial financial development on branch expansion (under a similar exercise to Burgess and Pande, 2005). Subfigure "Policy introduced in 1977" corresponds to Figure 1 in Burgess and Pande, 2005, p.784. Therefore, it represents the benchmark for the comparison. Note: The series "rural branches in unbanked locations (with controls)" shows the annual coefficients of the effect of initial financial development on branch expansion from equation (1). The series "rural branches unbanked locations (trend break)," dashed line, graphs the trends obtained from equation (2); the values are reported in Table B1 and correspond to Table A1, column 1.

Figure 2.A.7: Initial financial development and rural branch expansion with additional cut-year to the one in 1977

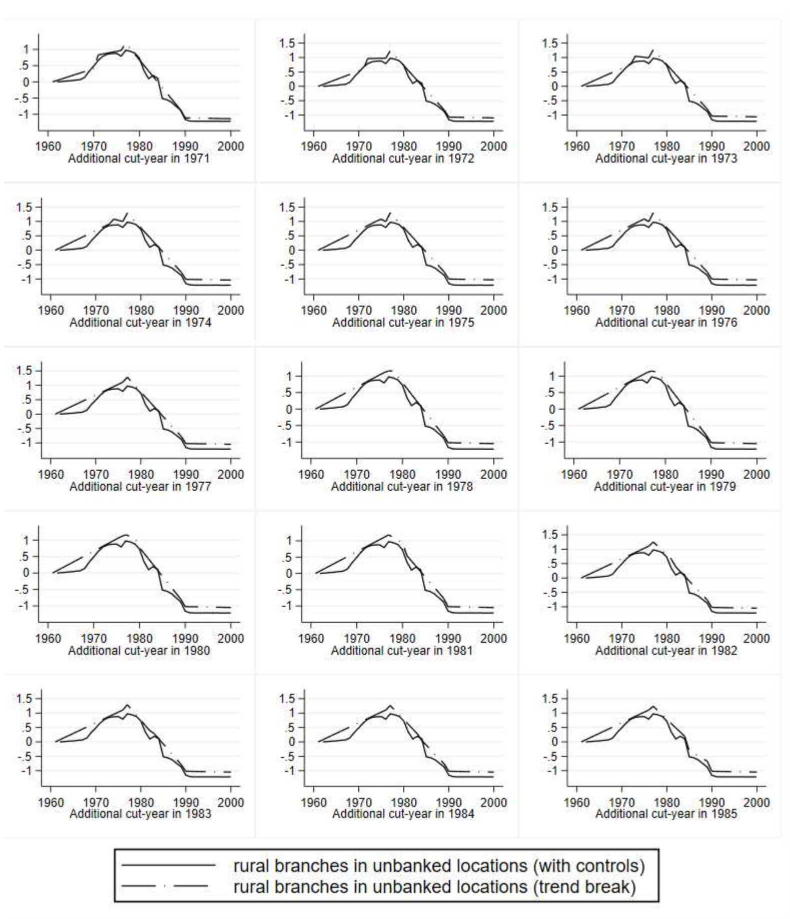


Figure A7 presents the trend reversals obtained from introducing the different hypothetical cut-years in addition to the one in 1977, and their fit to the coefficient of the effect of initial financial development on branch expansion (under a similar exercise to Burgess and Pande, 2005). Subfigure "Additional cut-year in 1977" corresponds to Figure 1 in Burgess and Pande, 2005, p.784 and is the benchmark for the comparison. Note: The series "rural branches in unbanked locations (with controls)" shows the annual coefficients of the effect of initial financial development on branch expansion from equation (1). The series "rural branches unbanked locations (trend break)," dashed line, graphs the trends obtained from equation (2); the values are reported in Table C1 and correspond to Table A1, column 1.

Figure 2.A.8: Initial financial development and rural branch expansion with additional cut-year to 1977

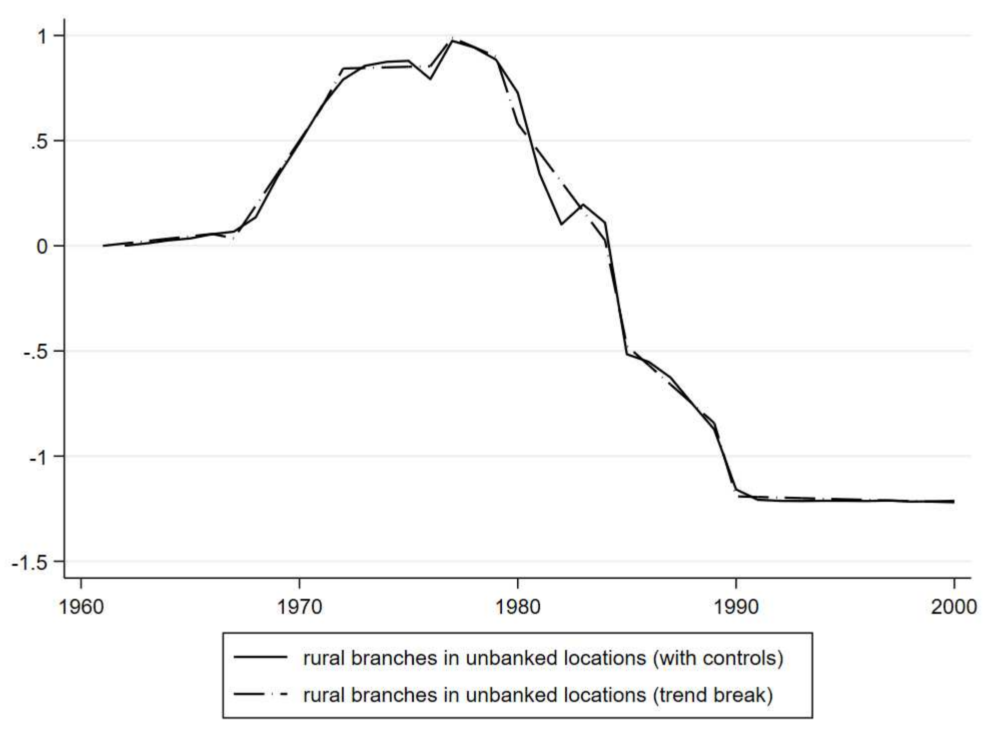


Figure A8 presents the trend reversals obtained from introducing the different hypothetical cut-years, and their fit to the coefficient of the effect of initial financial development on branch expansion (under a similar exercise to Burgess and Pande, 2005). Subfigure "Policy introduced in 1977" corresponds to Figure 1 in Burgess and Pande (2005), p.784. Therefore, it represents the benchmark for the comparison. Note: The series "rural branches in unbanked locations (with controls)" shows the annual coefficients of the effect of initial financial development on branch expansion from the equation (1). The series "rural branches unbanked locations (trend break)" graphs the trends obtained from equation (2) and correspond to Table (1), column (1).

Table 2.A.1: Banking as a function of initial financial development

	Branches in rural unbanked locations	Rural bank		Branches in banked locations	Credit share	
		Credit share	Savings share		Priority sector	Cooperative
	(1)	(2)	(3)	(4)	(5)	(6)
Number of bank branches per capita in 1961*(1961–2000) trend	0.07**	0.18	-0.03	0.14***	-0.08	0.42
	(0.028)	(0.209)	(0.235)	(0.012)	(0.626)	(0.337)
Number of bank branches per capita in 1961*(1977–2000) trend	-0.25***	-1.09**	-0.82***	-0.07***	0.08	0.02
	(0.030)	(0.434)	(0.252)	(0.020)	(0.865)	(0.416)
Number of bank branches per capita in 1961*(1990–2000) trend	0.17***	0.87***	0.43*	0.10**	-0.18	-0.18
	(0.042)	(0.263)	(0.229)	(0.041)	(0.333)	(1.013)
Number of bank branches per capita in 1961*Post-1976 dummy†	0.34	-0.30	-0.17	0.53**	-3.37	-3.80
	(0.251)	(1.495)	(0.777)	(0.187)	(2.402)	(2.237)
Number of bank branches per capita in 1961*Post-1989 dummy†	-0.24	1.95	0.44	-0.40***	-0.05	-3.32
	(0.152)	(1.490)	(0.533)	(0.103)	(1.858)	(2.803)
State and year dummies	YES	YES	YES	YES	YES	YES
Other controls	YES	YES	YES	YES	YES	YES
Adjusted R-squared	0.963	0.879	0.870	0.981	0.863	0.806
F-test 1	16.87	12.80	25.67	8.975	0.000	5.484
	[0.001]	[0.003]	[0.000]	[0.009]	[0.988]	[0.033]
F-test 2	0.491	0.099	9.000	27.22	1.785	0.060
	[0.494]	[0.757]	[0.009]	[0.000]	[0.201]	[0.810]
Observations	636	512	512	636	512	494

Source: This table is a replication of Table 1 in Burgess and Pande (2005), page 785. For replication we used data and methodology provided by the authors.†Original paper contains Post-1976 dummy*(1977–2000) trend and Post-1989 dummy*(1990–2000) trend instead, which is not consistent with the text and the stata code accompanying the paper. Therefore, we have changed the variable names accordingly. Note: p-values of tests in brackets. Coefficient estimates from regressions in the form of equation (2). Other controls include state population density, log state income per capita, log rural locations per capita, all measured in 1961. F-test 1 and F-test 2 test the linear restriction that the coefficients in the first two and three rows sum to 0. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 2.A.2: Bank branch expansion and poverty: reduced form evidence

	Annual coef. rural head-count ratio	Headcount ratio			Wage	
		Rural	Urban	Aggregate	Agricultural	Factory
	(1)	(2)	(3)	(4)	(5)	(6)
Annual coefficients for branches in rural unbanked locations	-4.71*** (1.01)					
Number of bank branches per capita in 1961*(1961–2000) trend		-0.77*** (0.235)	-0.27 (0.237)	-0.71*** (0.225)	-0.00 (0.006)	0.01 (0.019)
Number of bank branches per capita in 1961*(1977–2000) trend		1.15** (0.424)	0.15 (0.257)	0.99*** (0.332)	-0.01 (0.008)	-0.01 (0.019)
Number of bank branches per capita in 1961*(1990–2000) trend		-1.15*** (0.342)	-0.31 (0.378)	-1.04*** (0.310)	0.05* (0.023)	-0.02 (0.010)
Number of bank branches per capita in 1961*Post-1976 dummy†		-3.77* (1.940)	-2.76 (2.286)	-3.53* (1.706)	0.09* (0.049)	0.04 (0.047)
Number of bank branches per capita in 1961*Post-1989 dummy†		1.20 (2.387)	0.50 (0.964)	0.62 (1.819)	-0.03 (0.054)	0.01 (0.022)
State and year dummies		YES	YES	YES	YES	YES
Other controls		YES	YES	YES	YES	YES
Adjusted R-squared		0.835	0.913	0.875	0.901	0.701
F-test 1		1.497 (0.240)	0.373 (0.551)	1.760 (0.205)	23.95 (0.000)	0.234 (0.636)
F-test 2		2.973 (0.105)	3.948 (0.066)	4.148 (0.059)	1.884 (0.191)	6.066 (0.026)
Observations	39	627	627	627	545	553

Source: This table is a replication of Table 2 in Burgess and Pande (2005), page 788. For replication we used data and methodology provided by the authors. † Original paper contains Post-1976 dummy*(1977–2000) trend and Post-1989 dummy*(1990–2000) trend instead, which is not consistent with the text and the stata code. Therefore, we have changed the variable names accordingly. Note: p-values of tests in brackets. The first column reports the regression of the annual coefficients at the rural headcount ratio (γ_t, equation (1)) on the annual coefficients on initial financial development (λ_t, equation (3)). The other columns show estimated coefficients from regressions similar to equation (2) but with the respective headcount ratios as dependent variables. For the definition of other control variables and F-tests see Table 1. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 2.A.3: Bank branch expansion and poverty: instrumental variables evidence

	Headcount ratio								Wage	
	Rural			Urb	Aggr	Rural			Agri	Factory
						1961-1989	1977-2000	Survey years		
	OLS		IV	IV	IV	IV	IV	IV	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Number of bank branches opened in rural unbanked locations per capita	2.09**	1.16	-4.74**	-0.66	-4.10**	-4.70**	-6.84**	-4.21*	0.08*	0.05
	(0.785)	(1.024)	(1.790)	(1.066)	(1.464)	(1.821)	(2.805)	(2.263)	(0.042)	(0.083)
Number of bank branches per capita 1961*(1961–2000) trend		-0.43**	-0.48*	-0.26*	-0.46*	-0.43	-0.80*	-0.46	-0.01	0.01
		(0.165)	(0.269)	(0.134)	(0.226)	(0.264)	(0.447)	(0.281)	(0.004)	(0.013)
Number of bank branches per capita in 1961*Post-1976 dummy†		-0.31	-1.42	-2.06	-1.39	-2.13		-1.31	0.04	0.03
		(1.229)	(2.297)	(1.654)	(2.034)	(2.587)		(3.322)	(0.059)	(0.065)
Number of bank branches per capita in 1961*Post-1989 dummy†		5.38**	-1.08	-0.47	-1.55		-0.45	-0.79	0.11	-0.05
		(2.468)	(2.334)	(1.015)	(1.759)		(2.903)	(2.614)	(0.068)	(0.047)
State and year dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Other controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Overidentification test			[0.99]	[0.99]	[0.99]			[1]	[0.98]	[0.99]
Adjusted R-squared	0.807	0.834	0.760	0.915	0.818	0.804	0.807	0.734	0.868	0.699
Observations	627	627	627	627	627	460	375	375	545	554

Source: This table is a replication of Table 3 in Burgess and Pande (2005), page 789. For replication we used data and methodology provided by the authors. † Original paper contains Post-1976 dummy*(1977–2000) trend and Post-1989 dummy*(1990–2000) trend instead, which is not consistent with the text and the stata code. Therefore, we have changed the variable names accordingly. Note: The IV estimates correspond to equation (3) for different dependent variables. Other controls include state population density, log state income per capita, log rural locations per capita, all measured in 1961. The over-identification is tested using the conventional Sargan test. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 2.A.4: Rural credit and savings and poverty: instrumental variables evidence

	Headcount Ratio					
	Rural		Urban		Aggregate	
	(1)	(2)	(3)	(4)	(5)	(6)
Rural bank credit share	-1.52**		-0.67		-1.37**	
	(0.694)		(0.466)		(0.586)	
Rural bank savings share		-2.22**		-1.05		-2.01***
		(0.781)		(0.675)		(0.647)
Number bank branches per capita in 1961*(1961–2000) trend	-1.01*	-1.51**	-0.70**	-0.96**	-0.96**	-1.42***
	(0.496)	(0.538)	(0.253)	(0.343)	(0.406)	(0.437)
Number bank branches per capita in 1961*Post-1976 dummy†	-2.89	-2.05	-1.59	-1.23	-2.60	-1.84
	(1.681)	(2.340)	(1.975)	(2.554)	(1.677)	(2.518)
Number bank branches per capita in 1961*Post-1989 dummy†	4.40	2.13	2.87	1.88	3.53	1.47
	(2.644)	(2.653)	(2.345)	(1.310)	(2.352)	(1.975)
State and year dummies	YES	YES	YES	YES	YES	YES
Other controls	YES	YES	YES	YES	YES	YES
Overidentification test	[0.99]	[0.99]	[0.99]	[0.99]	[0.99]	[0.99]
Adjusted R-squared	0.686	0.602	0.903	0.879	0.746	0.669
Observations	503	503	503	503	503	503

Source: This table is a replication of Table 4 in Burgess and Pande (2005), page 791. For replication we used data and methodology provided by the authors. † Original paper contains Post-1976 dummy*(1977–2000) trend and Post-1989 dummy*(1990–2000) trend instead, which is not consistent with the text and the stata code. Therefore, we have changed the variable names accordingly. Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 2.A.5: Bank branch expansion and poverty reduction: robustness checks

	Rural Headcount Ratio		Urban Headcount Ratio	
	(1)	(2)	(3)	(4)
Number branches opened in rural unbanked locations per capita	-4.12**	-3.77**	-1.05	-0.81
	(1.544)	(1.544)	(1.061)	(0.908)
Cumulative land reform	-1.75**	-1.87**	0.41	0.27
	(0.696)	(0.678)	(0.286)	(0.302)
Health and education spending	-10.97	-3.31	23.52	23.74
	(30.908)	(28.402)	(14.531)	(14.796)
Other development spending	-40.84***	-37.32**	6.31	5.73
	(12.394)	(13.365)	(12.083)	(11.890)
Fraction legislators from:				
Congress parties		-13.07		0.22
		(8.904)		(3.138)
Janata parties		-11.62		1.62
		(6.899)		(3.184)
Hindu parties		6.15		9.61
		(12.905)		(8.361)
Hard Left parties		-14.81		1.76
		(9.074)		(3.718)
Regional parties		-15.11		-2.34
		(12.911)		(4.596)
State and year dummies	YES	YES	YES	YES
Other controls	YES	YES	YES	YES
Overidentification test	[0.99]	[0.99]	[0.99]	[0.99]
Adjusted R-squared	0.802	0.816	0.915	0.916
Observations	605	603	605	603

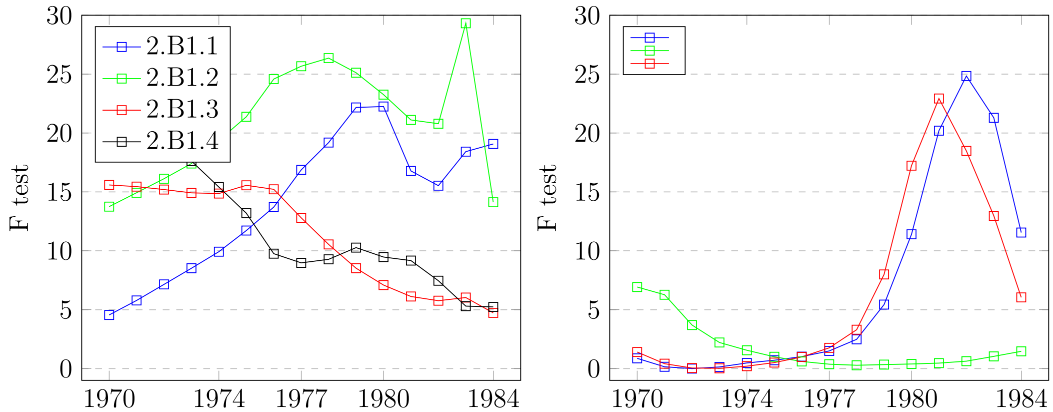
Source: This table is a replication of Table 5 in Burgess and Pande (2005), page 792. For replication we used data and methodology provided by the authors. Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

2.B Replication of Burgess and Pande (2005) with different cut-year

2.B1 Table 1 - Banking as a function of initial financial development

Figure 2.B1.1: F-test 1 from Tables 2.B1.1 - 2.B1.4 and 2.B2.1 - 2.B2.3

(a) F-test 1 from Tables 1.1 - 1.4 (b) F-test 1 from Tables 2.1 - 2.3



Note: The figures show F-statistics of a linear restriction test of no trend reversal (F-test 1) in any of the years in the sample. **2.B1.1.a.** Series 2.B1.1 shows the F-statistics of Table 2.B1.1, i.e., trend reversals in branches opened in rural regions. Series 2.B1.2 and 2.B1.3 report trend reversals in rural bank credit and savings shares. Series 2.B1.4 shows tests of trend reversals in branches opened in already banked regions. **2.B1.1.b.** Series shows the F-statistics of Table , i.e., trend reversals in Rural Headcount Ratio. report trend reversals in Aggregate Headcount Ratio. The trend reversals in Urban Headcount Ratio (Table) are insignificant. See online Appendix B.

Table 2.B1.1: Branched in rural unbanked locations

	1970	1971	1972	1973	1974	1975	1976	**1977**	1978	1979	1980	1981	1982	1983	1984
	Branched in Rural Unbanked locations														
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Number of Bank Branches per capita in 1961 *:															
(1961 - 2000) trend	0.03***	0.05***	0.06***	0.07***	0.08***	0.08***	0.08**	0.07**	0.07**	0.07**	0.07**	0.06*	0.05	0.04	0.03
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
(T - 2000) trend	-0.13**	-0.15***	-0.18***	-0.20***	-0.21***	-0.23***	-0.24***	-0.25***	-0.25***	-0.25***	-0.24***	-0.22***	-0.22***	-0.22***	-0.20***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.033)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.05)	(0.05)	(0.06)
(1990- 2000) trend	0.09*	0.10**	0.11**	0.12**	0.13***	0.14***	0.15***	0.17***	0.17***	0.18***	0.17***	0.16***	0.16***	0.18***	0.16***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.04)
Post-T dummy	1.08***	1.02***	0.92***	0.79***	0.65***	0.53**	0.41*	0.34	0.15	-0.06	-0.31	-0.54	-0.61	-0.59*	-0.77***
	(0.14)	(0.14)	(0.13)	(0.13)	(0.15)	(0.18)	(0.21)	(0.25)	(0.26)	(0.29)	(0.33)	(0.39)	(0.35)	(0.30)	(0.23)
Post-1989 dummy	-0.63***	-0.56**	-0.50**	-0.44**	-0.39*	-0.34*	-0.29	-0.24	-0.22	-0.21*	-0.23**	-0.26***	-0.25***	-0.22***	-0.25***
	(0.21)	(0.21)	(0.21)	(0.21)	(0.20)	(0.18)	(0.17)	(0.15)	(0.14)	(0.12)	(0.10)	(0.06)	(0.05)	(0.04)	(0.05)
State & year dummy	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Other controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Adjusted R^2	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.961	0.96
F-test 1	4.55	5.78	7.15	8.52	9.92	11.72	13.71	16.87	19.19	22.17	22.25	16.78	15.53	18.42	19.07
	[0.05]	[0.03]	[0.02]	[0.00]	[0.01]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
F-test 2	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49
	[0.49]	[0.49]	[0.49]	[0.49]	[0.49]	[0.49]	[0.49]	[0.49]	[0.49]	[0.49]	[0.49]	[0.49]	[0.49]	[0.49]	[0.49]
Observations	636	636	636	636	636	636	636	636	636	636	636	636	636	636	636

This table is a replication of Table 1, column 1 in Burgess and Pande (2005), page 785, for different cut-years. The 8th column presents the results from the original regression (break in 1977) and is the benchmark for the comparison. For replication, we used data and methodology provided by the authors. F-test 1 and F-test 2 test the linear restriction that the coefficients in the first two and three rows sum to 0. Standard errors clustered by state are reported in parentheses; p-values are in square brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table 2.B1.2: Rural bank credit share

	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984
	Rural Bank Credit share														
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Number of Bank Branches per capita in 1961 *:															
(1961 - 2000) trend	−0.21	0.56	0.43	0.33	0.23	0.20	0.25	0.18	0.03	−0.08	−0.14	−0.19	−0.26	−0.33	−0.35
	(0.34)	(0.56)	(0.46)	(0.39)	(0.28)	(0.26)	(0.25)	(0.21)	(0.19)	(0.20)	(0.21)	(0.22)	(0.23)	(0.24)	(0.23)
(T - 2000) trend	−0.46	−1.28**	−1.20**	−1.15**	−1.09**	−1.10**	−1.16**	−1.09**	−0.99**	−0.92*	−0.87	−0.80	−0.74	−0.71	−0.47
	(0.47)	(0.58)	(0.52)	(0.49)	(0.43)	(0.45)	(0.46)	(0.43)	(0.44)	(0.48)	(0.53)	(0.57)	(0.59)	(0.59)	(0.53)
(1990- 2000) trend	0.63***	0.68***	0.73***	0.77***	0.82***	0.86***	0.87***	0.87***	0.91***	0.95**	0.96**	0.95**	0.96**	1.00**	0.78*
	(0.17)	(0.18)	(0.20)	(0.21)	(0.23)	(0.23)	(0.24)	(0.26)	(0.30)	(0.34)	(0.38)	(0.41)	(0.44)	(0.46)	(0.42)
Post-T dummy	4.07**	3.48**	3.02**	2.59*	2.31	1.74	0.42	−0.30	−0.29	−0.24	−0.66	−1.19	−1.34	−1.31	−2.75*
	(1.62)	(1.46)	(1.35)	(1.27)	(1.51)	(1.31)	(1.44)	(1.50)	(1.64)	(1.62)	(1.42)	(1.19)	(0.86)	(1.06)	(1.53)
Post-1989 dummy	0.65	0.94	1.22	1.47	1.72	1.91	1.92	1.95	2.07	2.22	2.25	2.22	2.25	2.32	1.95
	(1.93)	(1.84)	(1.74)	(1.64)	(1.54)	(1.52)	(1.52)	(1.49)	(1.48)	(1.48)	(1.52)	(1.51)	(1.51)	(1.48)	(1.50)
State & year dum	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Other controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Adjusted R^2	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88
Observations	512	512	512	512	512	512	512	512	512	512	512	512	512	512	512

This table is a replication of Table 1, column 2 in Burgess and Pande (2005), page 785, for different cut-years. The 8th column presents the results from the original regression (break in 1977) and is the benchmark for the comparison. For replication, we used data and methodology provided by the authors. Rural bank credit share is the percentage of total bank credit accounted for by rural branches. F-test 1 and F-test 2 test the linear restriction that the coefficients in the first two and three rows sum to 0. Standard errors clustered by state are reported in parentheses; p-values are in square brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table 2.B1.3: Rural bank saving share

	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984
	Rural Bank Saving Share														
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Number of Bank Branches per capita in 1961 *:															
	0.42	0.46	0.46	0.44	0.29	0.19	0.09	−0.03	−0.12	−0.17	−0.22	−0.30	−0.36	−0.41*	−0.44*
(1961 - 2000) trend	(0.31)	(0.72)	(0.58)	(0.49)	(0.37)	(0.29)	(0.26)	(0.24)	(0.23)	(0.23)	(0.22)	(0.22)	(0.22)	(0.23)	(0.22)
	−1.08**	−1.17	−1.21*	−1.21**	−1.08**	−1.00***	−0.92***	−0.82***	−0.75***	−0.69**	−0.62**	−0.56**	−0.49*	−0.46**	−0.36*
(T - 2000) trend	(0.45)	(0.71)	(0.60)	(0.52)	(0.40)	(0.33)	(0.29)	(0.25)	(0.24)	(0.24)	(0.24)	(0.25)	(0.25)	(0.20)	(0.17)
	0.25	0.30	0.33	0.36	0.39	0.40	0.42*	0.43*	0.45*	0.45*	0.43*	0.45*	0.43*	0.46**	0.39
(1990- 2000) trend	(0.27)	(0.27)	(0.27)	(0.27)	(0.26)	(0.25)	(0.24)	(0.23)	(0.22)	(0.22)	(0.22)	(0.24)	(0.22)	(0.21)	(0.25)
	3.61**	3.05***	2.33***	1.47***	1.01*	0.46	0.00	−0.17	−0.31	−0.78	−1.22*	−1.10	−1.31**	−1.15	−1.64*
Post-T dummy	(1.41)	(0.91)	(0.56)	(0.44)	(0.57)	(0.62)	(0.85)	(0.78)	(0.63)	(0.63)	(0.62)	(0.69)	(0.54)	(0.79)	(0.90)
	−0.50	−0.24	−0.03	0.11	0.23	0.32	0.37	0.44	0.51	0.49	0.43	0.49	0.46	0.51	0.39
Post-1989 dummy	(0.47)	(0.45)	(0.45)	(0.46)	(0.47)	(0.48)	(0.51)	(0.53)	(0.56)	(0.59)	(0.63)	(0.67)	(0.69)	(0.59)	(0.55)
State & year dummy	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Other controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Adjusted R-squared	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87
Observations	512	512	512	512	512	512	512	512	512	512	512	512	512	512	512

This table is a replication of Table 1, column 3 in Burgess and Pande (2005), page 785, for different cut-years. The 8th column presents the results from the original regression (break in 1977) and is the benchmark for the comparison. For replication, we used data and methodology provided by the authors. Rural bank saving share is the percentage of total bank saving accounted for by rural branches. F-test 1 and F-test 2 test the linear restriction that the coefficients in the first two and three rows sum to 0. Standard errors clustered by state are reported in parentheses; p-values are in square brackets. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 2.B1.4: Branches in banked locations

	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984
	Branches in banked locations														
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Number of Bank Branches per capita															
in 1961 · (1961 - 2000) trend	0.09***	0.09***	0.10***	0.11***	0.12***	0.12***	0.13***	0.14***	0.15***	0.15***	0.16***	0.16***	0.16***	0.15***	0.15***
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
in 1961 · (T - 2000) trend	0.04	0.03	0.01	0.00	−0.02	−0.04*	−0.06**	−0.07***	−0.08***	−0.09***	−0.09***	−0.10***	−0.10***	−0.11***	−0.12***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
in 1961·(1990- 2000) trend	0.04	0.05	0.06	0.07*	0.08*	0.08**	0.09**	0.10**	0.10**	0.10**	0.11**	0.11**	0.12**	0.13**	0.13***
	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)
in 1961·Post-T dummy	0.41***	0.51***	0.60***	0.64***	0.66***	0.66***	0.62***	0.53**	0.40**	0.26	0.19	0.09	0.04	0.01	−0.04
	(0.11)	(0.12)	(0.14)	(0.16)	(0.18)	(0.20)	(0.20)	(0.19)	(0.17)	(0.16)	(0.16)	(0.16)	(0.15)	(0.13)	(0.12)
in 1961·Post-1989 dummy	−0.69***	−0.65***	−0.60***	−0.56***	−0.51***	−0.47***	−0.43***	−0.40***	−0.39***	−0.39***	−0.38***	−0.37***	−0.36***	−0.34***	−0.32***
	(0.16)	(0.16)	(0.15)	(0.14)	(0.13)	(0.12)	(0.11)	(0.10)	(0.10)	(0.09)	(0.09)	(0.08)	(0.08)	(0.08)	(0.07)
State & year dum	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Other controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Adjusted R^2	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
F-test 1	24.83	22.04	19.54	17.62	15.41	13.19	9.76	8.98	9.28	10.27	9.47	9.17	7.46	5.30	5.24
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.02]	[0.04]	[0.04]
F-test 2	27.22	27.22	27.22	27.22	27.22	27.22	27.22	27.22	27.22	27.22	27.22	27.22	27.22	27.22	27.22
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Observations	636	636	636	636	636	636	636	636	636	636	636	636	636	636	636

This table is a replication of Table 1, column 4 in Burgess and Pande (2005), page 785, for different cut-years. The 8th column presents the results from the original regression (break in 1977) and is the benchmark for the comparison. For replication, we used data and methodology provided by the authors. F-test 1 and F-test 2 test the linear restriction that the coefficients in the first two and three rows sum to 0. Standard errors clustered by state are reported in parentheses; p-values are in square brackets. *** p<0.01, ** p<0.05, * p<0.1.

2.B2 Table 2 - Bank branch expansion and poverty: reduced form evidence

Table 2.B2.1 : Rural Headcount Ratio

	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984
	Rural Headcount Ratio										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Number of Bank Branches per capita in 1961 x :											
	-0.58	-0.69**	-0.75***	-0.77***	-0.78***	-0.78***	-0.77***	-0.77***	-0.76***	-0.75***	-0.75***
(1961-2000) trend	(0.38)	(0.29)	(0.25)	(0.23)	(0.23)	(0.24)	(0.24)	(0.24)	(0.23)	(0.22)	(0.22)
	0.75	0.92**	1.04**	1.15**	1.26**	1.38***	1.52***	1.66***	1.77***	1.80***	1.53***
(T-2000) trend	(0.46)	(0.42)	(0.41)	(0.42)	(0.43)	(0.40)	(0.36)	(0.33)	(0.31)	(0.31)	(0.29)
	-0.95***	-1.01***	-1.07***	-1.15***	-1.25***	-1.38***	-1.52***	-1.67***	-1.79***	-1.82***	-1.56***
(1990-2000) trend	(0.30)	(0.31)	(0.32)	(0.34)	(0.35)	(0.34)	(0.35)	(0.38)	(0.42)	(0.46)	(0.47)
	-6.42**	-5.23**	-4.37*	-3.77*	-3.37*	-3.14*	-2.82	-2.3	-1.43	0.1	2.92
in 1961·Post-T D	(2.89)	(2.41)	(2.161)	(1.94)	(1.68)	(1.66)	(2.00)	(2.36)	(2.59)	(2.56)	(2.18)
	2.05	1.81	1.52	1.2	0.84	0.41	-0.01	-0.4	-0.69	-0.75	-0.32
Post-1989 D	(2.42)	(2.42)	(2.41)	(2.38)	(2.31)	(2.15)	(1.99)	(1.84)	(1.71)	(1.61)	(1.59)
State & year D	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Other controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Adjusted R^2	0.84	0.83	0.83	0.83	0.84	0.835	0.83	0.83	0.83	0.83	0.83
Observations	627	627	627	627	627	627	627	627	627	627	627

This table is a replication of Table 2, column 2 in Burgess and Pande (2005), page 788, for different cut-years. The 4th column presents the results from the original regression and is the benchmark for the comparison. For replication, we used data and methodology provided by the authors. F-test 1 and F-test 2 test the linear restriction that the coefficients in the first two and three rows sum *** p<0.01, ** p<0.05, * p<0.1.

Table 2.B2.2 : Urban Headcount Ratio

	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984
	Urban Headcount Ratio														
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Number of Bank Branches per capita in 1961 *:															
	0.47	0.35	0.23	0.08	−0.06	−0.15	−0.21	−0.27	−0.32	−0.36*	−0.38**	−0.39**	−0.40**	−0.40**	−0.39**
(1961 - 2000) trend	(0.36)	(0.34)	(0.33)	(0.32)	(0.31)	(0.29)	(0.26)	(0.24)	(0.21)	(0.19)	(0.17)	(0.16)	(0.15)	(0.14)	(0.14)
	−0.81**	−0.65*	−0.48	−0.30	−0.14	−0.03	0.07	0.15	0.21	0.23	0.23	0.20	0.15	0.05	−0.12
(T - 2000) trend	(0.35)	(0.35)	(0.34)	(0.33)	(0.31)	(0.29)	(0.27)	(0.26)	(0.25)	(0.26)	(0.28)	(0.32)	(0.35)	(0.38)	(0.43)
	−0.09	−0.13	−0.18	−0.22	−0.23	−0.26	−0.29	−0.31	−0.32	−0.30	−0.28	−0.24	−0.18	−0.09	0.08
(1990 - 2000) trend	(0.25)	(0.25)	(0.26)	(0.28)	(0.31)	(0.33)	(0.36)	(0.38)	(0.40)	(0.40)	(0.42)	(0.44)	(0.46)	(0.47)	(0.51)
	−3.85*	−4.47**	−4.87**	−4.67*	−3.90	−3.59	−3.22	−2.76	−2.16	−1.35	−0.65	−0.02	0.57	1.17	1.88
Post-T *D*	(2.12)	(1.84)	(2.00)	(2.26)	(2.56)	(2.43)	(2.36)	(2.29)	(2.16)	(2.03)	(1.96)	(1.84)	(1.57)	(1.08)	(1.27)
	1.64	1.38	1.10	0.91	0.85	0.70	0.58	0.50	0.46	0.52	0.60	0.70	0.84	1.02	1.30
Post-1989 *D*	(1.63)	(1.52)	(1.41)	(1.29)	(1.17)	(1.07)	(1.00)	(0.96)	(0.99)	(1.09)	(1.21)	(1.31)	(1.37)	(1.35)	(1.23)
State & year *D*	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Other controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Adjusted R^2	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91
F-test 1	6.93	6.27	3.70	2.22	1.56	1.00	0.61	0.37	0.28	0.34	0.39	0.47	0.63	1.04	1.47
	[0.02]	[0.02]	[0.07]	[0.16]	[0.23]	[0.33]	[0.45]	[0.55]	[0.60]	[0.57]	[0.54]	[0.51]	[0.44]	[0.33]	[0.25]
F-test 2	3.95	3.95	3.95	3.95	3.95	3.95	3.95	3.95	3.95	3.95	3.95	3.95	3.95	3.95	3.95
	[0.07]	[0.07]	[0.07]	[0.07]	[0.07]	[0.07]	[0.07]	[0.07]	[0.07]	[0.07]	[0.07]	[0.07]	[0.07]	[0.07]	[0.07]
Observations	627	627	627	627	627	627	627	627	627	627	627	627	627	627	627

This table is a replication of Table 2, column 2 in Burgess and Pande (2005), page 788, for different cut-years. The 8th column presents the results from the original regression and is the benchmark for the comparison. For replication, we used data and methodology provided by the authors. F-test 1 and F-test 2 test the linear restriction that the coefficients in the first two and three rows sum to 0. Standard errors clustered by state are in parentheses; p-values are in square brackets. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 2.B2.3 : Aggregate headcount ratio

	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984
	Urban Headcount Ratio														
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Number of Bank Branches per capita in 1961 *:															
(1961 - 2000) trend	−0.59	−0.32	−0.33	−0.42	−0.54	−0.63**	−0.68**	−0.71***	−0.72***	−0.73***	−0.73***	−0.72***	−0.72***	−0.71***	−0.70***
	(0.67)	(0.54)	(0.48)	(0.41)	(0.35)	(0.28)	(0.24)	(0.23)	(0.22)	(0.22)	(0.22)	(0.22)	(0.21)	(0.21)	(0.20)
(T - 2000) trend	0.32	0.17	0.28	0.45	0.63	0.77**	0.89**	0.99***	1.09***	1.18***	1.28***	1.37***	1.43***	1.41***	1.14***
	(0.65)	(0.55)	(0.50)	(0.44)	(0.39)	(0.35)	(0.33)	(0.33)	(0.33)	(0.30)	(0.27)	(0.26)	(0.27)	(0.29)	(0.25)
(1990 - 2000) trend	−0.49**	−0.61**	−0.71***	−0.79***	−0.85***	−0.90***	−0.97***	−1.04***	−1.12***	−1.21***	−1.31***	−1.40***	−1.47***	−1.46***	−1.20**
	(0.22)	(0.23)	(0.24)	(0.25)	(0.27)	(0.28)	(0.30)	(0.31)	(0.32)	(0.33)	(0.36)	(0.39)	(0.43)	(0.46)	(0.43)
Post-T D	−2.43	−5.40**	−6.39**	−6.41**	−5.54**	−4.72**	−4.07**	−3.53*	−3.05*	−2.58	−2.08	−1.45	−0.58	0.78	3.14*
	(2.69)	(2.36)	(2.41)	(2.42)	(2.43)	(2.06)	(1.85)	(1.71)	(1.62)	(1.77)	(2.05)	(2.28)	(2.34)	(2.14)	(1.54)
Post-1989 D	3.47*	2.73	2.17	1.72	1.44	1.18	0.91	0.62	0.32	0.02	−0.26	−0.51	−0.68	−0.65	−0.23
	(1.86)	(1.89)	(1.95)	(1.97)	(1.93)	(1.90)	(1.87)	(1.82)	(1.73)	(1.56)	(1.40)	(1.26)	(1.15)	(1.09)	(1.15)
State & year D	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Other controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Adjusted R^2	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.87
F-test 1	1.41	0.42	0.05	0.02	0.21	0.51	0.99	1.76	3.30	7.99	17.23	22.93	18.48	12.97	6.04
	[0.25]	[0.53]	[0.83]	[0.88]	[0.66]	[0.49]	[0.34]	[0.21]	[0.09]	[0.01]	[0.00]	[0.00]	[0.00]	[0.00]	[0.03]
F-test 2	4.15	4.15	4.15	4.15	4.15	4.15	4.15	4.15	4.15	4.15	4.15	4.15	4.15	4.15	4.15
	[0.06]	[0.06]	[0.06]	[0.06]	[0.06]	[0.06]	[0.06]	[0.06]	[0.06]	[0.06]	[0.06]	[0.06]	[0.06]	[0.06]	[0.06]
Observations	627	627	627	627	627	627	627	627	627	627	627	627	627	627	627

This table is a replication of Table 2, column 4 in Burgess and Pande (2005), page 788, for different cut-years. The 8th column presents the results from the original regression and is the benchmark for the comparison. For replication, we used data and methodology provided by the authors. F-test 1 and F-test 2 test the linear restriction that the coefficients in the first two and three rows sum to 0. Standard errors clustered by state are in parentheses; p-values are in square brackets. *** p<0.01, ** p<0.05, * p<0.1.

2.D Initial financial development and rural branch expansion

Figure 2.D1: Initial financial development and rural branch expansion with cut-years in 1967, 1972, 1977, 1980, 1985, and 1990

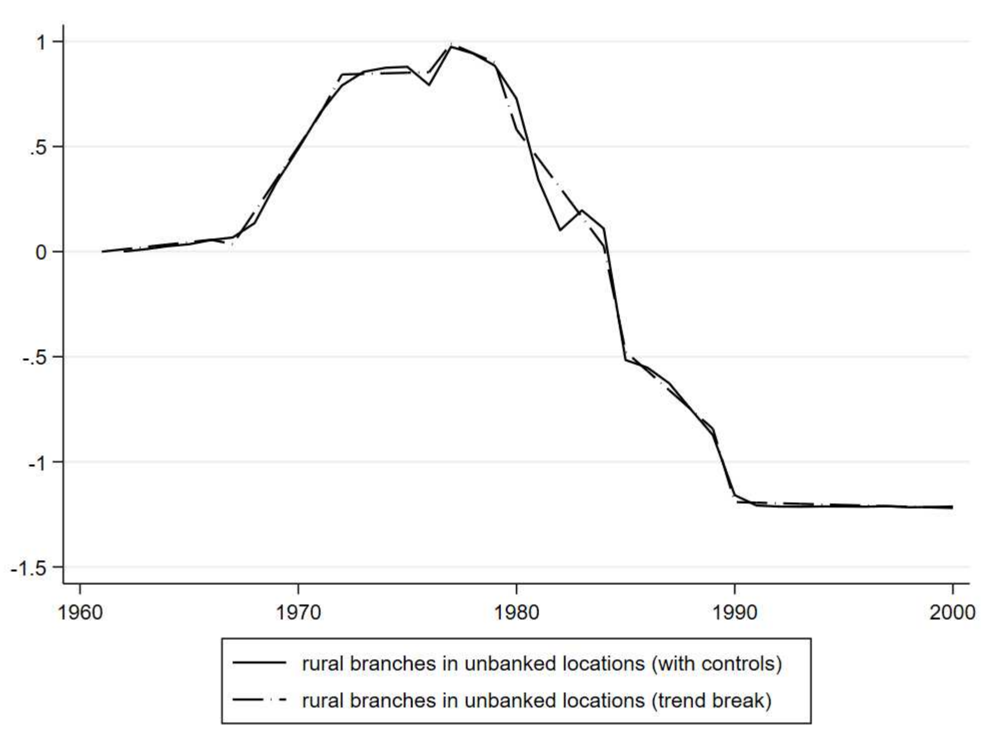


Figure presents the trend reversals obtained from introducing the cut-years in 1967, 1972, 1977, 1980, 1985, and 1990, and their fit to the coefficient of the effect of initial financial development on branch expansion (under a similar exercise to Burgess and Pande, 2005). The cut-years were chosen in correspondance to the historic events This figure corresponds to Figure 1 in Burgess and Pande (2005), p.784. Note: The series "rural branches in unbanked locations (with controls)" shows the annual coefficients of the effect of initial financial development on branch expansion from the equation (1). The series "rural branches unbanked locations (trend break)" graphs the trends obtained from equation (2); the values are reported in Table D1, column 1 and correspond to the results reported in Table (1), column (1).

History: In 1967, Gandhi's Ten-Point Programme was announced, paving the path towards priority sector lending, the nationalization of commercial banks, and other interventions. Starting from 1967, the bank licensing policy requiring a specific ratio of new rural bank branches per every urban branch was applied. These bank licensing rules were updated in February 1970 and September 1971. The year 1972 denotes a major shift towards "socialist" interventionist policies after the 1971 war with Pakistan wherein the Soviet Union sided with India while the United States with Pakistan.

Next, the 1977 policy applies the banking expansion rule highlighted by Burgess and Pande (2005) while in 1980 the Integrated Rural Development Programme was fully extended and the National Rural Employment Programme was launched. The year 1985 marks the start of the Seventh Five-Year Plan with strengthened though not yet systematic liberalization efforts and the year in which the low profitability of rural bank branches started to be addressed. Finally, in 1990, a major liberalization plan was proposed before the balance of payments crisis of 1991.

Table 2.D1: Banking as a function of initial financial development

	Branches in rural unbanked locations	Rural bank: Credit share	Rural bank: Savings share	Branches in banked locations	Credit share: Priority sector	Credit share: Cooperat
	(1)	(2)	(3)	(4)	(5)	(6)
Number of Bank Branches per capita in 1961·						
(1961 - 2000) trend	0.02	−0.13	0.38	0.10***	2.17	0.76
	(0.01)	(0.73)	(0.84)	(0.02)	(1.68)	(1.22)
(1967 - 2000) trend	0.14**	0.55	0.08	0.04	−4.17	0.25
	(0.05)	(1.13)	(1.24)	(0.03)	(2.40)	(2.21)
(1972 - 2000) trend	−0.15***	−0.32	−0.83	0.10**	2.42*	−1.19
	(0.04)	(0.47)	(0.59)	(0.04)	(1.29)	(1.29)
(1977 - 2000) trend	−0.05	−0.47	−0.09	−0.16***	−1.17	−0.42
	(0.04)	(0.38)	(0.18)	(0.04)	(1.16)	(0.37)
(1980 - 2000) trend	−0.09*	−0.59	−0.32	0.02	2.17*	0.92*
	(0.05)	(0.80)	(0.22)	(0.02)	(1.18)	(0.47)
(1985 - 2000) trend	0.05	0.22	0.00	−0.05	−0.79	1.20
	(0.10)	(0.48)	(0.19)	(0.04)	(1.01)	(0.70)
(1990 - 2000) trend	0.09**	0.69	0.37	0.13***	−0.82**	−1.24
	(0.04)	(0.50)	(0.24)	(0.04)	(0.37)	(1.14)
Post-1966 dummy	−0.17**			−0.15***		−3.52*
	(0.06)			(0.05)		(1.88)
Post-1971 dummy	0.19**	−0.01	0.89*	0.01	0.20	1.23
	(0.07)	(0.49)	(0.43)	(0.06)	(2.05)	(2.31)
Post-1976 dummy†	0.18*	−1.60*	−0.29	0.20**	−2.21	−0.37
	(0.10)	(0.77)	(0.47)	(0.08)	(1.64)	(0.91)
Post-1979 dummy	−0.18	0.32	−0.64	0.00	−4.12**	0.73
	(0.14)	(1.36)	(0.67)	(0.03)	(1.88)	(0.44)
Post-1984 dummy	−0.41**	−1.16	−0.40	−0.02	−5.33***	−2.54
	(0.14)	(0.87)	(0.47)	(0.06)	(1.69)	(1.57)
Post-1989 dummy†	−0.34***	1.84	0.36	−0.33***	−0.87	−5.85
	(0.07)	(1.52)	(0.59)	(0.08)	(2.26)	(3.96)
State and year dummies	YES	YES	YES	YES	YES	YES
Other controls	YES	YES	YES	YES	YES	YES
Adjusted R-squared	0.96	0.88	0.86	0.98	0.86	0.81
F-test 1	9.39	0.83	0.59	55.65	4.22	0.93
	[0.01]	[0.38]	[0.46]	[0.00]	[0.06]	[0.35]
F-test 2	0.46	0.09	8.52	25.67	1.69	0.07
	[0.51]	[0.76]	[0.01]	[0.00]	[0.21]	[0.80]
Observations	636	512	512	636	512	494

Source: This table is a replication of Table 1 in Burgess and Pande (2005), page 785. For replication we used data and methodology provided by the authors. †Original paper contains Post-1976 dummy*(1977–2000) trend and Post-1989 dummy*(1990–2000) trend instead, which is not consistent with the text and the stata code. Therefore, we have changed the variable names accordingly. Note: p-values of tests are presented in brackets. Coefficient estimates are from regressions in the form of equation (3). Other controls include state population density, log state income per capita, log rural locations per capita, all measured in 1961. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 2.D2: Bank branch expansion and poverty: reduced form evidence

	Headcount ratio			Wage	
	Rural	Urban	Aggregate	Agricul	Factory
	(1)	(2)	(3)	(4)	(5)
Number of Bank Branches per capita	0.52	0.81	0.49	-0.01	0.12**
in 1961 * (1961 - 2000) trend	(1.182)	(0.677)	(0.995)	(0.014)	(0.055)
Number of Bank Branches per capita	0.48	-0.42	0.36	-0.04	-0.17**
in 1961 * (1967 - 2000) trend	(1.543)	(0.965)	(1.423)	(0.027)	(0.063)
Number of Bank Branches per capita	-2.15	-1.02	-1.81	0.04	0.06**
in 1961 * (1972 - 2000) trend	(1.490)	(1.228)	(1.378)	(0.025)	(0.023)
Number of Bank Branches per capita	0.73	0.27	0.52	0.03	-0.01
in 1961 * (1977 - 2000) trend	(0.811)	(0.611)	(0.661)	(0.037)	(0.026)
Number of Bank Branches per capita	0.32	0.39	0.43	-0.04**	-0.03
in 1961 * (1980 - 2000) trend	(0.542)	(0.317)	(0.369)	(0.017)	(0.020)
Number of Bank Branches per capita	0.52	-0.76	0.10	-0.00	0.02
in 1961 * (1985 - 2000) trend	(0.634)	(0.642)	(0.434)	(0.026)	(0.013)
Number of Bank Branches per capita	-1.19**	0.31	-0.84*	0.05*	-0.02*
in 1961 * (1990 - 2000) trend	(0.532)	(0.761)	(0.418)	(0.027)	(0.010)
Number of Bank Branches per capita	-8.32***	-2.69	-7.31***	0.28***	-0.11
in 1961 * Post-1966 dummy	(2.587)	(3.054)	(2.420)	(0.066)	(0.177)
Number of Bank Branches per capita	-3.98*	-3.16*	-3.90**	0.02	0.03
in 1961 * Post-1971 dummy	(2.200)	(1.646)	(1.629)	(0.059)	(0.048)
Number of Bank Branches per capita	1.05	-0.41	0.46	0.03	0.04
in 1961 * Post-1976 dummy†	(1.584)	(0.718)	(1.180)	(0.063)	(0.031)
Number of Bank Branches per capita	-1.11*	-0.01	-0.82	0.04	0.03
in 1961 * Post-1979 dummy	(0.564)	(0.643)	(0.554)	(0.040)	(0.051)
Number of Bank Branches per capita	3.61***	1.76	3.34**	-0.03	0.03
in 1961 * Post-1984 dummy	(1.033)	(3.216)	(1.279)	(0.043)	(0.031)
Number of Bank Branches per capita	0.17	1.61	0.25	-0.01	0.00
in 1961 * Post-1989 dummy†	(1.619)	(1.296)	(1.275)	(0.060)	(0.024)
State and year dummies	YES	YES	YES	YES	YES
Other controls	YES	YES	YES	YES	YES
Adjusted R-squared	0.841	0.914	0.882	0.905	0.699
F-test 1	1.429	0.223	1.128	9.964	9.859
	[0.250]	[0.643]	[0.305]	[0.00700]	[0.00674]
F-test 2	2.800	3.719	3.906	1.760	5.664
	[0.115]	[0.0729]	[0.0668]	[0.206]	[0.0310]
Observations	627	627	627	545	553

Source: This table is a replication of Table 2 in Burgess and Pande (2005), page 788. For replication we used data and methodology provided by the authors. †Original paper contains Post-1976 dummy*(1977–2000) trend and Post-1989 dummy*(1990–2000) trend instead, which is not consistent with the text and the stata code. Therefore, we have changed the variable names accordingly. Note: p-values of tests in brackets. The first column reports the regression of the annual coefficients at the rural headcount ratio (γ_t, equation (1)) on the annual coefficients on initial financial development (λ_t, equation (2)). The other columns show estimated coefficients from regressions similar to equation (2) but with the respective headcount ratios as dependent variables. For the definition of other control variables see Table 1. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 2.D3: Bank branch expansion and poverty: instrumental variables evidence

	Headcount ratio								Wage	
	Rural			Urban	Aggr	Rural			Agric	Factory
						1961 -1989	1977 -2000	Survey years		
	OLS		IV	IV	IV	IV	IV	IV	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Number branches opened in rural unbanked locations per capita	2.09** (0.785)	1.72 (1.244)	1.66 (2.980)	2.24 (3.025)	1.60 (2.745)	0.12 (0.100)	-0.11* (0.056)	2.95 (3.564)	1.34 (3.120)	12.22 (11.460)
Number of Bank Branches per capita in 1961 * (1961 - 2000) trend		-0.45 (0.312)	-0.45 (0.319)	-0.26* (0.144)	-0.45 (0.270)	-0.01 (0.009)	0.02 (0.014)	0.19 (0.391)	-0.55 (0.346)	-0.78 (0.586)
Number of Bank Branches per capita in 1961 * Post-1966 dummy		-2.14 (3.184)	-2.11 (2.778)	1.20 (2.451)	-1.60 (2.516)	0.11 (0.093)	0.01 (0.165)	-6.08** (2.096)	9.71 (7.346)	-3.80 (3.127)
Number of Bank Branches per capita in 1961 * Post-1971 dummy		-4.06 (2.906)	-4.03 (2.455)	-4.11* (2.196)	-3.66* (1.923)	-0.14 (0.084)	-0.02 (0.042)	-7.86* (4.063)	-3.37 (2.347)	-8.60** (3.781)
Number of Bank Branches per capita in 1961 * Post-1976 dummy†		-0.46 (1.913)	-0.46 (1.790)	-1.59 (1.281)	-0.71 (1.336)	0.06 (0.066)	0.04 (0.053)	-3.12 (2.307)	-0.02 (1.935)	0.86 (5.154)
Number of Bank Branches per capita in 1961 * Post-1979 dummy		1.07 (1.958)	1.02 (2.154)	2.15 (2.164)	1.52 (1.922)	0.09 (0.083)	-0.14** (0.067)	-0.69 (2.025)	1.22 (2.166)	9.92 (12.976)
Number of Bank Branches per capita in 1961 * Post-1984 dummy		8.55*** (1.974)	8.49** (3.616)	3.02 (4.088)	7.35* (3.761)	0.03 (0.092)	-0.19 (0.114)	6.55* (3.516)	8.68** (3.633)	20.46 (13.353)
Number of Bank Branches per capita in 1961 * Post-1989 dummy†		0.70 (2.061)	0.67 (2.454)	0.65 (1.380)	0.21 (2.042)	0.14** (0.057)	-0.20* (0.100)	0.00 (0.000)	1.43 (2.471)	6.76 (6.913)
State and year dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Other controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Adjusted R-squared	0.807	0.847	0.847	0.891	0.883	0.838	0.651	0.871	0.877	0.586
Observations	627	627	627	627	627	545	553	460	535	375

Source: This table is a replication of Table 3 in Burgess and Pande (2005), page 789. For replication we used data and methodology provided by the authors. †Original paper contains Post-1976 dummy*(1977–2000) trend and Post-1989 dummy*(1990–2000) trend instead, which is not consistent with the text and the stata code. Therefore, we have changed the variable names accordingly. Note: The IV estimates correspond to equation (3) for different dependent variables. For the definition of other controls see Table D1. Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 2.D4: Rural credit and savings and poverty: instrumental variables evidence

	Headcount Ratio					
	Rural		Urban		Aggregate	
	(1)	(2)	(3)	(4)	(5)	(6)
Rural bank credit share	-0.94		-0.43		-0.85	
	(0.753)		(0.689)		(0.665)	
Rural bank savings share		-0.85		-0.87		-0.89
		(0.825)		(1.237)		(0.748)
Number of Bank Branches per capita						
in 1961·(1961 - 2000) trend	-0.82	-0.98	-0.53*	-0.80	-0.80*	-0.99**
	(0.505)	(0.559)	(0.287)	(0.525)	(0.429)	(0.462)
in 1961·Post-1966 dummy	7.84	12.84	10.70	15.75*	7.58	12.77
	(9.592)	(10.446)	(6.293)	(8.737)	(8.201)	(8.786)
in 1961·Post-1971 dummy	-2.71	-2.56	-2.92	-1.97	-2.47	-2.12
	(2.080)	(2.649)	(1.736)	(2.374)	(1.763)	(2.370)
in 1961·Post-1976 dummy†	-0.79	0.14	-1.19	-0.95	-0.96	-0.16
	(1.533)	(1.536)	(1.467)	(1.800)	(1.103)	(1.222)
in 1961·Post-1979 dummy	-1.29	-0.86	0.56	-0.14	-0.59	-0.41
	(2.358)	(2.002)	(1.307)	(1.745)	(2.112)	(1.792)
in 1961·Post-1984 dummy	3.80	5.88***	-0.02	-0.19	3.08	4.70**
	(2.418)	(1.716)	(3.227)	(3.297)	(2.369)	(1.721)
in 1961·Post-1989 dummy†	2.58	1.43	1.79	1.24	2.02	0.97
	(2.029)	(2.018)	(2.470)	(1.419)	(1.761)	(1.506)
State and year dummies	YES	YES	YES	YES	YES	YES
Other controls	YES	YES	YES	YES	YES	YES
Adjusted R-squared	0.775	0.804	0.919	0.891	0.828	0.840
Observations	503	503	503	503	503	503

Source: This table is a replication of Table 4 in Burgess and Pande (2005), page 791. For replication we used data and methodology provided by the authors. †Original paper contains Post-1976 dummy*(1977–2000) trend and Post-1989 dummy*(1990–2000) trend instead, which is not consistent with the text and the stata code. Therefore, we have changed the variable names accordingly. Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 2.D5: Bank branch expansion and poverty reduction: robustness checks

	Rural Headcount Ratio		Urban Headcount Ratio	
Number branches opened in rural unbanked locations per capita	3.49 (3.356)	3.97 (3.219)	0.40 (2.285)	0.92 (2.205)
Cumulative land reform	-2.01** (0.784)	-1.97** (0.830)	0.33 (0.392)	0.24 (0.460)
Health and education spending	-5.05 (23.354)	-2.47 (22.425)	24.79 (17.281)	24.38 (17.051)
Other development spending	-14.20 (18.622)	-11.05 (17.528)	11.82 (14.481)	12.38 (12.527)
Fraction legislators from:				
Congress parties		-9.28 (5.505)		0.58 (3.783)
Janata parties		-11.48** (4.321)		1.21 (3.923)
Hindu parties		-0.26 (15.708)		7.37 (13.950)
Hard Left parties		-5.62 (8.046)		3.66 (4.836)
Regional parties		-2.53 (11.366)		0.04 (6.541)
State and year dummies	YES	YES	YES	YES
Other controls	YES	YES	YES	YES
Adjusted R-squared	0.862	0.864	0.910	0.907
Observations	605	603	605	603

Source: This table is a replication of Table 5 in Burgess and Pande (2005), page 792. For replication we used data and methodology provided by the authors. Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

References

Banerjee, A., Cole, S., & Duflo, E. (2004). Banking reform in India. *India Policy Forum*, *1*(1), 277–332.

Burgess, R., & Pande, R. (2005). Do rural banks matter? Evidence from the Indian social banking experiment. *American Economic Review*, *95*(3), 780–795.

Cole, S. (2009). Fixing market failures or fixing elections? agricultural credit in India. *American Economic Journal: Applied Economics*, *1*(1), 219–50.

Datt, G., Ravallion, M., & Murgai, R. (2016). Growth, urbanization and poverty reduction in India. *National Bureau of Economic Research*, (No. w21983).

Kochar, A. (2011). The distributive consequences of social banking: A microempirical analysis of the Indian experience. *Economic Development and Cultural Change*, *59*(2), 251–280.

Mohanty, A., & Acharya, S. (2006). *Operational analysis of regional rural banks.* Gyan Publishing House.

Panagariya, A. (2008). *India: The Emerging Giant.* Oxford University Press.

Chapter 3

Uncertain Trends in Economic Policy Uncertainty

Nino Buliskeria **Jaromir Baxa** **Tomas Sestorad**

Abstract

The news-based Economic Policy Uncertainty indices (EPU) of Germany, France, and the United Kingdom display discernible trends that can be found neither in other European countries nor in other uncertainty indicators. Therefore, we replicate the EPU index of European countries and show that these trends are sensitive to the rather arbitrary choice of normalizing the raw counts of news related to economic policy uncertainty by the count of all newspaper articles. We show that an alternative normalization by news on economic policy leads to different long-term dynamics with less pronounced trends and markedly lower uncertainty during recent periods of uncertainty such as Brexit or the COVID-19 pandemic. Consequently, our results suggest that the effects of uncertainty related to these events on economic activity may have been overestimated.

JEL Codes: D80, E66, E32

Keywords: economic policy uncertainty, trend-cycle decomposition, reproducibility, reliability

Acknowledgment: This work was supported by the Grant Agency of the Czech Republic, Grant No. GACR no. 20-14990S. Baxa and Buliskeria also thank the Charles University Research Center program No.UNCE/HUM/035, and the Cooperation Program at Charles University, research area Economics, for the support. Furthermore, Nino Buliskeria acknowledges the support received from the Charles University Grant Agency under Grant No. 1034519. We want to thank Matěj Míček, Dominik Coupek, Magdalena Báčová, Ivana Krouparová, Ondřej Michalík, and Lukáš Beran for excellent research assistance. Also, we are grateful to Volga Audzei, Jan Babecky, Jozef Baruník, Wojciech Charemza, Zuzana Gric, Davide Furceri, Alejandro Vicondoa and participants of research seminars at Charles University, Czech National Bank and Heidelberg University for their helpful comments and suggestions. The responsibility for all remaining errors and omissions rests solely on us. The views presented in this paper are those of the authors and do not necessarily reflect the official view of the Czech National Bank or any other authors' affiliations. Conflict of interest: None.

3.1 Introduction

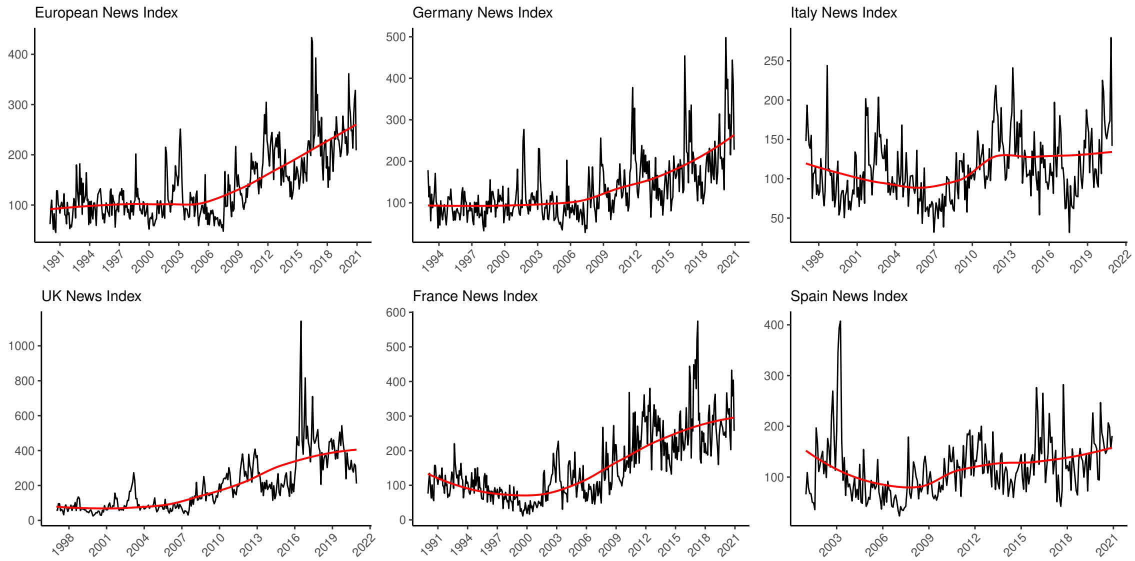
In recent years, there has been widespread interest in estimating the impact of uncertainty on economic performance. Bloom (2009) and Justiniano and Primiceri (2008), among others, presented theoretical models in which uncertainty fluctuations decrease output growth, increase unemployment, and contribute significantly to overall variations in the business cycles. Fernández-Villaverde and Guerrón-Quintana (2020) provide a comprehensive survey of the literature that estimates the effects of uncertainty shocks. However, measuring uncertainty remains challenging, and multiple conceptually different indicators are used in the literature to track uncertainty in financial markets, forecast disagreements, or more general uncertainty in the economy (see Castelnuovo et al. (2017) and Ferrara et al. (2018) for surveys).

One of the most popular proxies for uncertainty in economic models is the Economic Policy Uncertainty Index (EPU) developed by Baker et al. (2016), with more than 9000 citations since its publication.[1] This EPU index is based on the count of newspaper articles containing a set of words associated with uncertainty related to future economic policies. The popularity of this index comes from its success in matching historical data, particularly in the United States, where increases in the EPU index indicate historical periods of economic and political instability. Moreover, the EPU index is available at a monthly frequency and for almost 30 countries, allowing for many different types of analysis.

In this paper, we focus on the properties of the EPU indices of the major European economies: Germany, France, Italy, Spain, and the United Kingdom. We focus in particular on the long-term trends in the EPU indices that are

[1] Google Scholar as of March 2023.

Figure 3.1: The Economic Policy Uncertainty of European Countries



Note: Data from https://www.policyuncertainty.com/europe_monthly.html, retrieved on April 13, 2021. The trends were obtained using the Loess non-parametric method.

apparent in the United Kingdom, Germany, France and the aggregate European EPU index, but not in Italy or Spain (Figure 1).[2] However, the absence of trends in Italy and Spain is counterintuitive, considering the turmoil of the EU debt crisis and the subsequent political instability in both countries. Furthermore, EPU trends are inconsistent with the uncertainty measures derived from the implied volatilities in financial markets (Figure 2) and with the World Uncertainty Index of Ahir et al. (2018) (Figure 3).[3] Other uncertainty

[2]The presence of trends is corroborated by conventional stationarity tests (ADF, ADF-GLS, and KPSS) even for the sample ending before 2020, i.e. before the COVID-19 pandemic. The *KPSS test* rejects the null of stationarity for Europe, Germany, France, and the United Kingdom, but does not reject stationarity in Italy and Spain. The *ADF test* does not reject the unit root for Europe, France, and the United Kingdom, but rejects the unit root for Germany, Italy, and Spain. However, when we account for *heteroskedasticity* using the ADF-GLS, the unit root cannot be rejected for Germany, Italy, and Spain as well (for the results of the stationarity tests, see the Appendix, Table A1).

[3]Note that in contrast to the EPU, the World Uncertainty Index (WUI) has been calculated for almost all countries in the world, which has made this index appealing for cross-country analyses (Ahir et al., 2020 and 2021). However, the WUI also has several drawbacks. First, it is available on a quarterly frequency (the EPU is available on a monthly frequency), and the spikes in the EPU appear (much) earlier than the spikes in the WUI, which is admitted by Ahir et al. (2018). On the other hand, in the case of European countries, the WUI leads

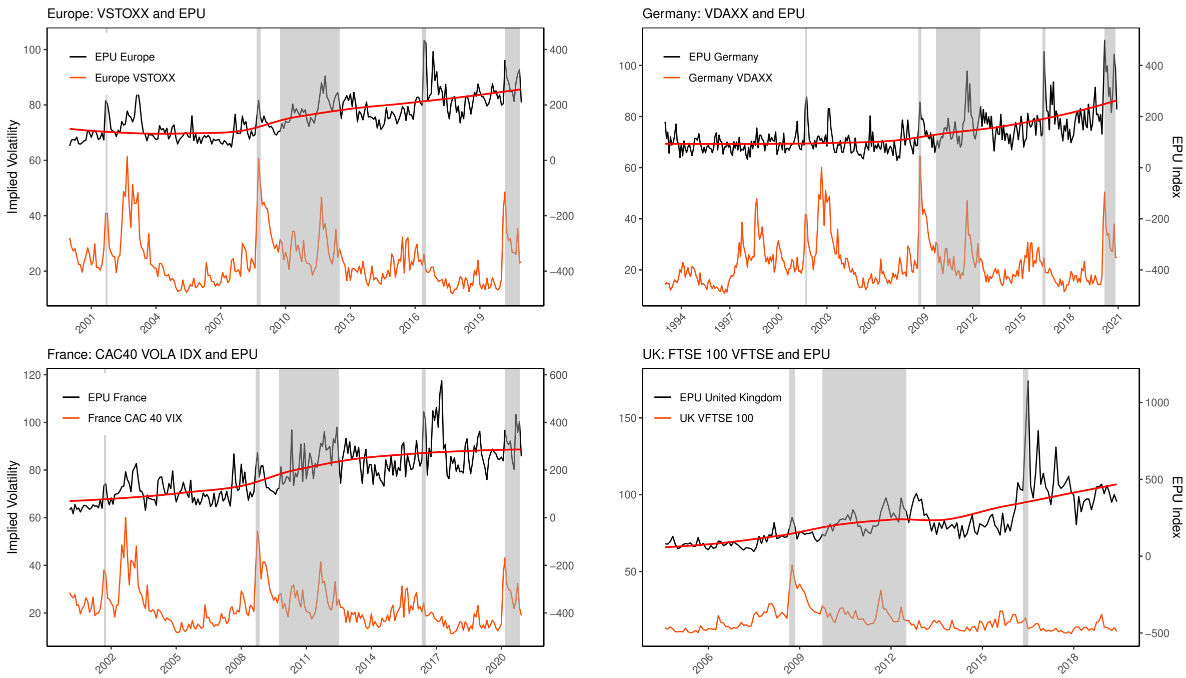
proxies for EU countries also tend to be mean-reversing. Meinen and Röhe (2017) show the developments of the macroeconomic uncertainty indices of Germany, France, Italy, and Spain in two versions of the index, one inspired by Jurado et al. (2015), the second follows Rossi and Sekhposyan (2015). In all cases, macroeconomic uncertainties had returned to pre-2008 levels by 2014. Also, the forecast dispersion in production uncertainties (following Bachmann et al., 2013) returned to pre-crisis levels. The EPU dynamics is also exceptional at the euro area level. The European Central Bank (2016) shows that financial market uncertainty, forecast disagreement, and survey-based proxies for economic uncertainty based on both consumer and business surveys returned to pre-Great Recession levels. The forecast uncertainty from the Survey of Professional Forecasters stabilized at a higher than pre-crisis level, possibly due to a change in forecasters' risk perception after the Great Recession, which was not expected by the majority of forecasters participating in the surveys. Therefore, the comparison of uncertainty measures constructed through various approaches shows discrepancies and raises doubts about the reliability of the increasing trends observed in the EPU indices.

Our goal is to investigate why the trends in the EPU arise in some countries and not in others, whether these trends are reliable, and whether the values of the EPU can be used for policy guidance as an indicator relevant for short-term predictions of economic activity. Thus, we focus mainly on the construction of the EPU index. Baker et al. (2016) first extract the raw counts of articles related to uncertainty from the leading newspapers in the countries, then divide these counts by the counts of all articles published each month.[4] This normalization was used to control for part of the variation in the counts of articles related to

to more intuitive trends than the EPU.

[4] To facilitate interpretation, this ratio is normalized so that the pre-2010 mean equals 100.

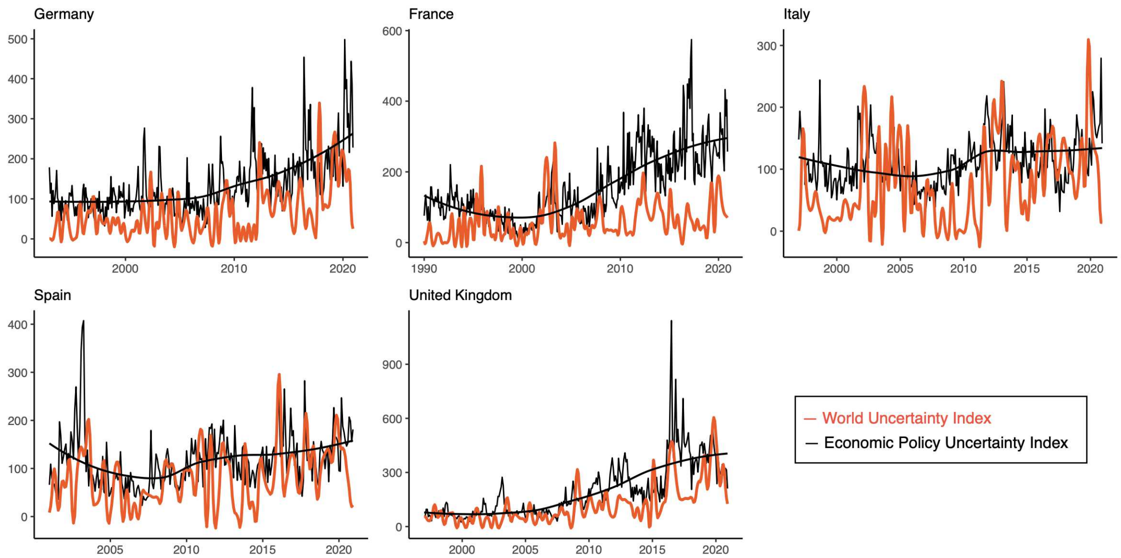
Figure 3.2: Economic Policy Uncertainty and Implied Volatility Indices in Europe



Note: Implied volatility - red (VSTOXX, derived from the Eurozone's EURO STOXX 50, VDAX based on the German DAX index, CAC40 VOLA IDX (VCAC) based on the French CAC40, and the FTSE 100 VIX index (VFTSE) based the British FTSE 100 index, left axes. EPU index - **black**, right axes. Notice the trend in the EPU that is absent in implied volatilities representing risk assessment on financial markets.

uncertainty arising from changes in the count of all articles over time. However, the composition of newspaper articles and their relative counts also evolved markedly for reasons not related to uncertainty. For instance, the counts of German newspaper articles in *Handelsblatt* and *Frankfurter Allgemeine-Zeitung* - which are used to calculate the German EPU – have gradually decreased over the past two decades. There was a shift in the composition of the newspapers toward longer, more in-depth analyses, and the overall article count related to economic policy increased. These changes in composition were driven by changes in readers', journalists', and editors' preferences due to the rising importance of online editions and social media platforms, along with possibly greater interest in economic policy after the Global Financial Crisis. Although these changes are not necessarily caused by uncertainty, they affect the count of all articles and confound the EPU index through the denominator.

Figure 3.3: Economic Policy Uncertainty and World Uncertainty Indices in Europe



Note: This figure shows the World Uncertainty Index (WUI) in red and Economic Policy Uncertainty index in black. The WUI introduced by Ahir et al. (2018), reflects the frequency of the word "uncertainty" in the reports by the Economist Intelligence Unit that are scaled by the total word count in each report. It is available at a quarterly frequency for all countries covered by the EIU reports. The values of the WUI were multiplied by 1000 to make the scale comparable with the EPU.

To disentangle the changes driven by fluctuations in uncertainty-related articles from those in the count of all articles, which are inputs for scaling the EPU index by Baker et al. (2016), we replicate the text mining procedure for the major European economies. Next, we test the sensitivity of the EPU index to alternative normalization. We take the ratio of the raw count of uncertainty-related articles to the count of economic policy-related articles instead of all articles. This alternative scaling removes the effects of structural changes in the publication policies of newspapers, as well as changes in the share of articles not related to economic policy. Although this alternative has pros and cons, the comparison between EPU indices based on two alternative normalizations shows the sensitivity of the EPU to a scaling factor that has not been discussed previously. Furthermore, we discuss the robustness of the index to relatively arbitrary changes in the selection of newspapers and databases used for text mining. To anticipate our results, we find that the trends of the EPU index are relatively sensitive to these changes and that the overall index is difficult to replicate.

Our work is closely related to the literature that discusses the properties of the EPU index and addresses potential biases in the index by using refined algorithms to select articles related to uncertainty[5]. This literature employs language processing methods to select the most appropriate keywords and eliminate irrelevant articles from simple text searches. Azqueta-Gavaldón (2017) uses machine learning to eliminate irrelevant articles from those selected by a simple text search by Baker et al. (2016) for the United States; however, the differences between his uncertainty index and the original EPU are relatively minor. Azqueta-Gavaldón et al. (2023) use machine learning to develop an EPU index for the euro area , Tobback et al. (2018) for Belgium, and Larsen (2021) for Norway. Charemza et al. (2022) construct the EPU index for Russia, where they extend the approach of Azqueta-Gavaldón et al. (2023) for sentiment analysis. They add specific positive and negative weights to the relevant keywords to determine whether articles about uncertainty indicate increasing or decreasing uncertainty. Overall, this literature implies that a more precise extraction of the raw count of uncertainty-related articles leads to EPU dynamics closer to those obtained by human evaluation of the articles. Nevertheless, this literature takes the normalization of the count of uncertainty-

[5]The other biases treated in this literature are related to the pitfalls connected with selecting the appropriate keywords for constructing the EPU index in non-English speaking countries. These biases are related to linguistic differences, differences in journalistic styles, conventions, and the overall social context (Charemza et al., 2022).

related articles by all articles as given, while our analysis shows that the choice of scaling factor matters for some countries and for the overall European EPU.

Finally, we recommend several adjustments to the calculation of the EPU index of European countries to improve its reliability when employed in short-term forecasting, for example, within central banks. The EPU index has the advantage over alternative uncertainty indicators in that it captures different types of uncertainty than uncertainty indices calculated based on the dispersion in predictions or associated with financial instability. First, we recommend scaling the raw count of uncertainty-related articles by economic policy articles rather than all articles. Such adjustment in normalization leads to less pronounced trends and more consistent searches across databases, especially since Factiva does not allow the extraction of the count of *all* articles.[6] We show that using an alternative normalization for the EPU index leads to a quantitatively smaller impact of uncertainty shocks on economic activity and implies lower estimates of Brexit- and COVID-related economic policy uncertainty. Second, we suggest expanding the coverage of newspapers. In the current editions, the EPU indices of European countries are based only on two leading newspapers from the respective country. Considering that adding or changing the selection of newspapers affects the overall index, two newspapers seem insufficient for reasonable, robust, and credible approximations of uncertainty. Third, the same database should preferably be used for text mining of uncertainty-related articles to ensure consistency across countries.

The remainder of the paper is organized as follows. The second section provides more details on the construction of the Economic Policy Uncertainty Index.

[6] Factiva's search engine does not allow "blank" searches to obtain all articles published in a given month. Baker et al. (2016) address this issue by searching for the word "today" instead; however, it appears that when we search for the article "the" ("la" in Italian and French), the count of retrieved articles differs (is larger) from the search results using "today."

Section three presents a replication of the EPU index and examines the impact of alternative normalizations and a different selection of newspapers. Section four studies the implications of alternative EPU normalizations on the quantitative effect of uncertainty shocks on economic activity and Brexit-related uncertainty. Finally, section five concludes with several recommendations for utilizing the EPU indices.

3.2 The Economic Policy Uncertainty Index

Baker et al. (2016) constructed the EPU index for Germany, France, Italy, Spain, and the United Kingdom using the relative frequency of articles reflecting the uncertainty of economic policy in two leading newspapers for each country. Thus, the authors turned to full-text databases of newspaper articles and counted the articles retrieved with the following search query for each month:

> *(economic OR economy OR business OR industry OR commerce OR commercial) AND (spending OR policy OR deficit OR budget OR tax OR regulation OR "Bank of England" OR war OR tariff) AND (uncertain OR uncertainty)*[7]

This query implies that an article is considered an indicator of economic policy uncertainty if it contains at least one word from all three parts of the search query. The first subset of keywords implies that the selected articles are related to economic affairs, the second to policy, and the final one refers to uncertainty.[8] To account for shifts in newspaper composition, the article count

[7]These keywords are used for the calculation of the index for the United Kingdom. The queries for other countries are equivalent, but obviously in the language of the particular newspaper. The exact specification of the keywords is provided in Baker et al. (2016) and this paper's Appendix.

[8]Baker et al. (2016) explain their selection in Section 2 of their paper: *"We aim to capture uncertainty about who will make economic policy decisions, what economic policy actions will be undertaken and when, and the economic effects of policy actions (or inaction) – including uncertainties related to the economic ramifications of "noneconomic" policy matters, e.g., military actions. Our measures capture both near-term concerns (e.g., when will the Fed*

is normalized by the total article count published in a particular newspaper in a given month, that is, $x_{it} = \textit{uncertainty articles}_{it}/\textit{all articles}_{it}$. The ratio is further normalized by the variance of x_{it} until December 2009 to avoid the effect of new observations on the historical values of the index. Next, the ratio x_{it}/σ_i^2 is averaged across newspapers within a given country or, in the case of the European EPU, across all newspapers from European countries. Finally, these averages are rescaled, so the mean until December 2009 is 100. Therefore, the EPU value of 100 reflects the average level of uncertainty in a country of interest until the Great Recession. Since the EPU for European countries usually starts in the 1990s, it mainly reflects the uncertainty during the NICE (non-inflationary, consistently expansionary) decade before 2007.

The construction of the index relies on relatively strong implicit assumptions. First, it is assumed that the same keywords consistently represent the uncertainty of economic policy over time. However, the use of contemporary vocabulary for keyword selection can lead to increasing trends, as some words are used more frequently in more recent times than at the beginning of the sample, for example, because of the evolving perception of the importance of various policies for the economy. Nevertheless, this assumption is not as problematic for samples spanning over a few decades as for historical EPU indices, with samples ranging over many decades.

The second and more important assumption is that the relative frequency of keywords represents changes in uncertainty related to economic policy and does not represent surges in pure interest in macroeconomic factors. There is relatively fresh evidence that readers' preferences evolve and depend on the state of the economy. Jha et al. (2021, 2022) analyze millions of books published

adjust its policy rate) and longer-term concerns (e.g., how to fund entitlement programs), as reflected in newspaper articles." Brandt (2021) has explored the extent to which the selection of keywords affected the index.

in eight countries over more than one hundred years to investigate popular sentiment toward financial institutions. In addition to persistent differences between countries, the authors document systematic movements in sentiments following wars, epidemics, natural disasters, and other major shocks. Sentiment usually improves when insurance coverage mitigates the shock and the financial sector helps to spread the burden of the shock. On the other hand, when a shock is uninsured and contracts cannot be renegotiated, people tend to perceive the financial sector more negatively. In line with these considerations, Duca and Saving (2018) show that the EPU is not exogenous to macroeconomic developments. They argue that macroeconomic fundamentals and political fragmentation cause around 40% of long-run and short-run fluctuations in the EPU indices of the US and Europe.[9] According to these findings, Ludvigson et al. (2021) assert that uncertainty is not only a source, but also a consequence of business cycle fluctuations. They highlight the importance of uncertainty in propagating other macroeconomic shocks in addition to being the primary cause of business cycle fluctuations.

Importantly, Baker et al. (2016) do not control for the structural changes in the newspaper industry that affect the composition and content of newspapers

[9] *"Nevertheless, the Baker et al. (2016) index is often met with skepticism by economists who are concerned that EPU either reflects other economic factors or is so endogenous as to be meaningless. One particular and common shortcoming of studies that analyze EPU is that long-term trends in EPU are ignored or omitted. Accounting for these trends is important because they could shed light on the factors underlying time series, helping social scientists better interpret and gauge short- and long-term movements in economic policy uncertainty."* (Duca and Saving, 2018). Some skepticism toward the EPU index was also expressed in the Deutsche Bank report: *"For the European EPU index, the BBD weights the EU countries equally and does not distinguish between local and international policy uncertainty. Put differently, German newspapers writing about Brexit and associated economic uncertainty in the UK are counted towards an increasing EPU in Germany and contribute to the European index. International news coverage in local newspapers probably inflates index values during major events such as Brexit. Another caveat is the representativeness of the newspapers taken for index construction. BBD uses two major newspapers from each country, which usually have a specific economics and finance focus (and, as such, tend to be more internationally oriented). During episodes of economic uncertainty, this may magnify the surge in EPU."* (Kaya et al. (2018)).

and thus affect the counts of all articles that serve as the denominator of the EPU index. However, the composition of newspapers has evolved dramatically over the past decades, in part because the Internet and social networks have taken up a large part of the market share of the print media. The online space has led to a continuous decrease in the circulation of newspapers and gradual changes in their content. Short articles, such as sports news and news often reprinted from press agencies, have moved to online editions. In addition, there has been a shift by journalists to more detailed and lengthy analyses than before. Consequently, the article count has mostly decreased, although with varying intensity across countries and newspapers. But even a modest drop in the count of sports articles, for example, affects the denominator, the count of all articles, and increases the EPU index independently of the article count related to uncertainty (nominator). This drop in the count of sports articles can be illustrated by the Frankfurter Allgemeine Zeitung. We have compared the article count in different categories in the first week of May in 2000, 2005, 2010, 2015, and 2019. In all years, the highest article count was in the section Economy, but its count decreased from more than 100 in 2000 to below 60 in 2019. The count of sports-related articles also decreased from about 40 in 2000 to about 10 in 2019. Thus, while the article count in both categories decreased, the relative share of economic articles increased.

Overall, a significant shift in the preferences of readers, journalists, and editors materialized in the structural changes in newspapers. These changes in relative counts of articles could affect the development of the EPU independently of fluctuations in fundamental uncertainty. In the following sections, we will show how the counts of uncertainty-related, economic policy-related, and all articles have evolved. The trends in all series will stand out clearly.

3.3 Replication of the EPU index

3.3.1 Text mining exercise

To investigate the reasons behind the differences in EPU trends across countries, we replicate the construction of the EPU index. We performed a full-text search of articles related to uncertainty as specified by Baker et al. (2016) and selected the newspapers and period as close to theirs as possible. However, minor modifications were inevitable, given the availability (or lack thereof) of newspaper archives and databases to us. In particular, the composition of the newspapers used for the EPU index has also evolved over time. The current edition of the European EPU available at the link[10] is based on a different selection of newspapers than in the index in Baker et al. (2016). We follow the EPU available online, which is regularly updated and used in the literature. Therefore, we use the term ”original EPU" for the EPU on the EPU website, despite some small differences from the EPU in the published paper.

The original EPU relies on Le Monde and Le Figaro for France, Handelsblatt and Frankfurter Allgemeine Zeitung for Germany, Corriere Della Sera and La Stampa for Italy[11], El Mundo and El Pais for Spain, and The Times of London and the Financial Times for the United Kingdom. In the case of German newspapers, we use their online archives in line with Baker et al. (2016). We used the Factiva database for other countries. Given the license restrictions, we adjusted the selection of newspapers. Specifically, Le Monde is not available under our Factiva license, so we replaced it with the leading economic newspaper in France, Les Echos.[12] For the same reasons, instead of

[10] https://www.policyuncertainty.com/europe_monthly.html

[11] La Stampa appears in the current EPU published at policyuncertainty.com, the journal version of Baker et al. (2016) uses Corriere Della Sera and La Repubblica.

[12] Baker et al. (2016) used the Lexis Nexis database to access the archive of Le Monde. However, this database was not available to us either. As a sensitivity check, we scraped Le Monde's online archive, and these results show the sensitivity of the EPU index to the

the Financial Times, we opt for The Guardian, which Baker et al. (2016) used in calculating the Historical EPU index for the United Kingdom.[13] Moreover, The Guardian represents a progressive political stance, while The Times of London is more conservative-leaning. Therefore, the replicated EPU index encompasses conservative and progressive perspectives. Additionally, Baker et al. (2016) use the NewsBank Access World News database for the United Kingdom instead of Factiva, which is used in our replication. Consequently, the replicated indices in this paper are not based on exactly the same underlying data as those used by Baker et al. (2016). On the other hand, the changes document the sensitivity of the EPU indices, including differences in the search engine and in the selection of newspapers. The text mining specifications are summarized in Table 1.

In contrast, we used the same search queries as Baker et al. (2016) to obtain uncertainty-related articles:

Query 1: *(economic OR economy OR business OR industry OR commerce OR commercial) AND (spending OR policy OR deficit OR budget OR tax OR regulation OR "central bank" OR war OR tariff) AND (uncertain OR uncertainty)*[14]

The article count obtained from Query 1 was then scaled by the count of all articles or, in the case of Factiva, by articles containing the word "today" because Factiva does not permit searching for all articles in a given period. To address the role of changes in newspaper composition, we propose an alternative index, referred to as the *adjusted EPU* that differs from the original EPU in the denominator. Instead of dividing the count of uncertainty-related articles by

choice of newspaper article database.

[13] The Historical index for the United Kingdom is available at www.policyuncertainty.com.

[14] The country-specific queries can be found in the Appendix.

Table 3.1: Text mining specifications of the EPU

	Newspapers		Database	
	Baker et al. 2016	*Replication*	*Baker et al. 2016*	*Replication*
Germany	Frankfurter Allgemeine Zeitung, Handelsblatt		Newspapers' online archives	
France	Le Monde Le Figaro	Les Echos Le Figaro	Lexis Nexis Factiva	Factiva Factiva
UK	The Times Financial Times	The Times The Guardian	NewsBank Access World News Database	Factiva Factiva
Italy	Corriere Della Sera, La Stampa		Factiva	
Spain	El Mundo, El País		Factiva	
France Sensitivity Check	Le Monde Le Figaro	Le Monde Le Figaro Le Monde Le Figaro Les Echos	Lexis Nexis Factiva	Online Archive Factiva Online Archive Factiva Factiva

Note: This table summarizes the differences in text mining between Baker et al. (2016) and this paper. We performed three alternative searches for France, using the text mining in Les Echos and Le Figaro as a baseline. The other searches in Le Monde's archive were used in the sensitivity analysis presented in Section 3.4.

all articles, we used the count of articles discussing economic policy, obtained using query 2, as the denominator.

Query 2: *(economic OR economy OR business OR industry OR commerce OR commercial) AND (spending OR policy OR deficit OR budget OR tax OR regulation OR "central bank" OR tariff OR war*[15]*)*

Query 2 is equivalent to Query 1 up to the third part, with uncertainty-related keywords excluded from Query 2.

[15]Following Baker et al. (2016), the keyword *war* is used only for the United Kingdom and not for other countries, presumably because military spending is associated mainly with increased military spending and expansionary fiscal policy rather than with uncertainty.

The benefit of our adjusted EPU is that normalization removes articles not related to economic policy from the denominator and infers uncertainty from fluctuations of uncertainty-related articles relative to the writings on economic policy. The underlying assumption of this choice of normalization is that the words in Query 2 are related to economic policy in general and do not necessarily indicate uncertainty. The first set of words *(economic OR economy OR business OR industry OR commerce OR commercial)* is less debatable than the second set of policy-related words *(spending OR policy OR deficit OR budget OR tax OR regulation OR "Bank of England" OR tariff*), and in the case of the United Kingdom, *war*. The context matters, and noise and arbitrariness are inherent in all text-mining exercises. Although there are attempts to tackle this problem with machine learning techniques, see Azqueta-Gavaldón (2017) or Charemza et al. (2022), we stick to straightforward text mining due to its transparency and for a direct comparison of the impact of modified specification with the original index by Baker et al. (2016). Additionally, either the counts of economic policy articles do not increase with major events, such as the European debt crisis, or the pattern is similar to that of *all* articles. The only exception is Germany during the COVID-19 pandemic at the end of the sample.

Our adjusted EPU, with uncertainty-related articles normalized by economic policy articles, is similar to the Monetary Policy Uncertainty Index by Husted et al. (2020) who use text mining to obtain articles relevant to the uncertainty of future monetary policy and scale the count of relevant articles by articles on the Federal Reserve rather than by the count of all articles. Moreover, the scaling by economic policy articles is also similar to Baker et al. (2014). Their work is one of the earlier iterations of papers on the EPU index, where the authors discuss the long-term trends in the US EPU index. As a robustness test, they normalized the EPU index by the frequency of articles with the words *(economic OR economy)* instead of all articles. In the case of this alternative normalization, the rise of the EPU in the United States after 2007 is not as evident as with the baseline EPU index normalized by *all* articles. We use a more restrictive search query for the adjusted EPU to increase the chance that uncertainty fluctuations are driven primarily by articles related to economic policy uncertainty and that the impact of the evolving composition of newspapers is minimized. The comparison between the replicated EPU and the adjusted EPU provided in the following sections of this paper reveals qualitative and quantitative differences and how the policy implications of both indices differ.

3.3.2 Case studies: raw counts for Germany, Italy, and the United Kingdom

Figure 4 presents the raw results of the three search queries to demonstrate the trends in the inputs of the EPU index for Germany, Italy, and the United Kingdom. To keep the discussion concise, we skip the presentation of the raw counts of France and Spain, as their patterns are similar to those of Germany (France) and Italy (Spain). The “uncertainty-related articles" are the denominator of the index ratio, whereas the “all" and “Economic Policy" articles are two alternatives for the denominator.

In Figure 4a, we present the raw counts of the components of the EPU index for Germany. Since the mid-2000s, the count of uncertainty-related articles has increased in line with the EPU index in both German newspapers (Frankfurter Allgemeine Zeitung and Handelsblatt), while the count of all articles has decreased. In the case of Frankfurter Allgemeine Zeitung, the data reveal a dramatic drop in the count of all articles between 2002 and 2003. This drop is related to significant losses at F.A.Z. Group, the parent company of Frankfurter Allgemeine Zeitung following a drop in advertising revenue. The customized sections for Berlin and Munich were scrapped, and an eight-page, English-language edition, published six days a week, was reduced to weekly tabloid format. The Saturday issue shrank from 234 pages at its peak to 40 pages within two years; see Landler (2004). Thus, the ratio of uncertainty-related articles to all articles increased, partly due to a decrease in the denominator and not just because of the count of articles directly pointing towards uncertainty. Conversely, the count of articles related to economic policy (the second candidate for the denominator) increased, moving inversely to the count of all articles. This result indicates that the EPU index calculated using the count of all articles as a denominator is, to some extent, driven by changes in the composition of newspapers.

Italy is an entirely different case, as shown in Figure 4b. Unlike Germany, the dynamics of articles for Italy containing the word "Oggi" (Italian for today) resembles that of economic policy articles.[16] Furthermore, the series act counterintuitively, given that the future of economic policies was rather unpredictable during the 2010s.[17] The Italian economy descended into an

[16] A strong seasonal pattern is particularly evident at Corriere della Sera. There is a periodic, systematic decline in the article count in August, coinciding with Italy's holiday season.

[17] In response to the crisis, the Italian government led by Mario Monti adopted strict austerity policies that were opposed by the general public. This led to a rise in both the left- and right-wing populist parties (Five Star Movement and the League). After the 2013 snap elections following Monti's resignation, Italy experienced a series of political crises. The cabinet was led first by Enrico Letta and then by Matteo Renzi (both members of the Democratic party). The government passed several structural reforms and gradually softened some of its austerity measures. At the same time, Renzi ran a campaign to change the Italian constitution; however, the new constitutional rules were rejected in a referendum in December 2016. Renzi resigned and Paolo Gentilioni was appointed new prime minister. The Democratic party gradually lost public support. Two populist and Eurosceptic movements (the League and the Five Star Movement) won the 2018 elections, forming a government led by an independent prime minister Giuseppe Conte. The government collapsed in 2019; however, Conte continued as prime minister in a new coalition of the Five Star Movement with the Democratic Party until 2021.

unprecedented triple-dip recession, unemployment, and a banking crisis that even threatened its membership of the Eurozone.[18] Despite political turmoil and instability after the Great Recession, the count of uncertainty-related articles remains fluctuating around its mean throughout the sample (particularly in La Stampa). These developments contrast with those in Germany, where we observe a significantly higher increase in the count of economic policy uncertainty-related articles, although its economy evolved relatively smoothly under Chancellor Angela Merkel and recovered quickly from the 2008 Great Recession. A comparison of the German and Italian raw counts shows that the evolution of articles related to uncertainty does not necessarily reflect the intuition behind long-term changes in the uncertainty of economic policy in different countries.

In the case of the United Kingdom (Figure 4c), the count of uncertainty-related articles peaks with the Brexit referendum and subsequent negotiations, while the count of "economic policy" articles compared to articles with the word "today" evolves differently across newspapers. Thus, fluctuations in the denominator contribute to the dynamics of the overall EPU index. Furthermore, the count of economic policy articles is higher than that of articles with the word "today." Therefore, we can conclude that the search result for the word "today" is not equivalent to the count of all articles published in a given month. The approach to scaling the search count of articles related to uncertainty by the count of articles that contain the word today is different from scaling by all articles, contributing to the inconsistencies between countries.[19]

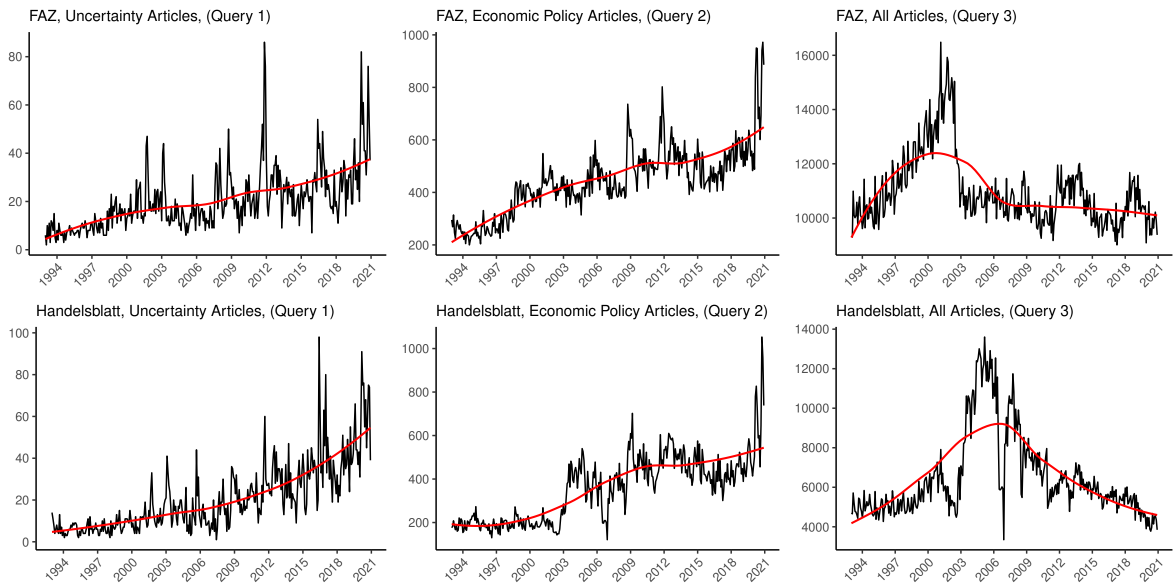
Overall, the raw counts underlying the EPU index reveal non-trivial differences

[18] In particular, Beppe Grillo, the leader of the Five Star Movement, attempted to initiate a referendum over Italy's euro area membership in 2014.

[19] Similarly, searching for articles such as "la" in Italian results in a higher article count than with the equivalent of "today."

Figure 3.4: Counts of Articles Underlying the EPU Index (monthly frequency)

4(a). Germany

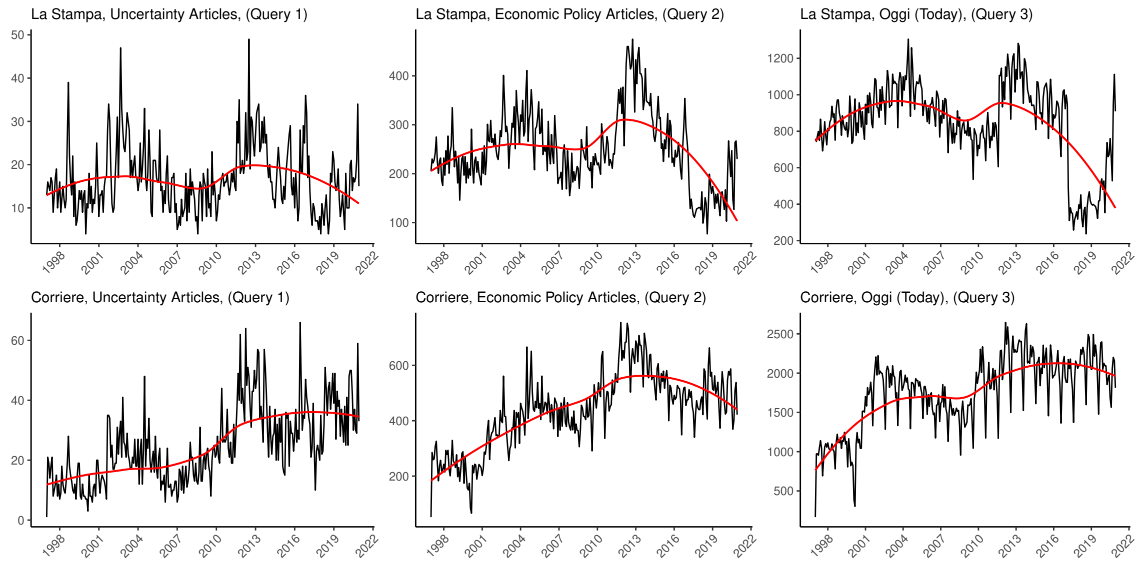


Note: The article counts of the three search queries used to construct the EPU index. Each country is represented by two leading newspapers. The first column shows the results for the first query of uncertainty-related articles. The second column shows the count of all articles (Germany) and articles containing the word “today" (Italy and the UK). The third column depicts the series used for an alternative normalization by the third query (economic policy-related articles). The sample is determined by the availability of articles in the archives. Trends: The Loess non-parametric method.

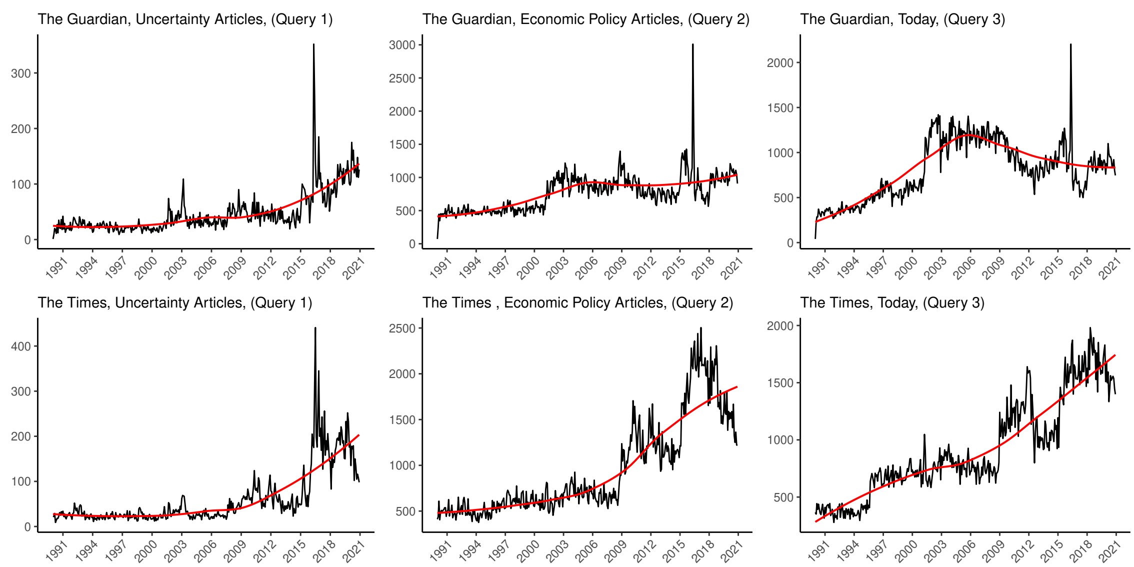
across countries and normalization choices. The cross-country differences are most prominent between Italy and Germany's EPU indices. Moreover, there are significant differences in the developments of the scaling factors (“all," “today," or “economic policy" articles) used in normalizations of the raw count of uncertainty-related articles. These findings cast doubt on the reliability of the EPU index in representing uncertainty since it is not clear how much of the EPU fluctuations are caused by uncertainty per se and how much by the evolving structures of the newspapers.

Figure 3.4: Counts of Articles Underlying the EPU Index (monthly), Cont.

4(b): Italy



4(c): United Kingdom



Note: The counts of articles of the three search queries used for the construction of the EPU index. Each country is represented by two leading newspapers. The first column shows the results for the first query for uncertainty-related articles. The second column shows the count of all articles (for Germany) and articles containing the word 'today' (in the case of Italy and the United Kingdom). The third column shows the series used for alternative normalization by the third query (economic policy-related articles). The sample is determined by the availability of articles in the archives. Trends: The Loess non-parametric method.

3.3.3 Replicated EPU indices

We present our replicated EPU indices in Figure 5, along with the series provided by Baker et al. (2016) on the https://www.policyuncertainty.com/ web page. In the case of the European EPU, our replicated EPU (the gray area) closely matches the original index, although some discrepancies appear in the severity of several uncertainty peaks, including during the EU debt crisis and the Brexit referendum. These discrepancies arise from text mining specifications and time periods different to those used to calculate the European EPU. Although Baker et al. (2016) constructed their index starting in 1987, it is based only on a subset of countries and newspapers up to 2001. In addition, the period 1987-1990 is based solely on the French newspaper Le Monde. Our approach is different. We construct the European EPU only for the period which is based on the complete set of newspapers and countries. The difference in the period before 2009 implies that the standard errors and averages used to scale the index differ.[20]

The alternative *adjusted EPU* relies on normalization by "economic policy" articles. It mimics the original and replicated indices reasonably well until the Global Financial Crisis of the late 2000s. However, the gap between the two differently normalized EPU indices widens after 2009. Our "adjusted EPU" index increases with the 2007 financial crisis and remains at a similar level in the following years despite the turmoil of the European debt crisis. Furthermore, no additional major increases are visible in the original index until all EPU indices peak with the Brexit referendum. However, the adjusted EPU index peaks around 270 points, while the original index reaches 440 points (60% higher).

[20] However, we do not have access to the raw data used by Baker et al. (2016), so we cannot assess how much the difference in time periods contributes to the differences in the resulting EPUs.

More apparent discrepancies arise at the country level between the original, replicated, and adjusted EPU indices. The most striking case is Germany. Despite relying on the same newspapers, archives and queries, we were unable to closely replicate the Germany EPU index by Baker et al. (2016). Although the short-term fluctuations are very similar, the long-term growth in the original EPU that motivated our analysis is even more pronounced in the replicated index. To achieve closer replication to the original EPU, we performed thorough checks of our search exercise, repeated web-scraping, and tried some additional sample restrictions in March 2023. Although such modifications are not mentioned in Baker et al. (2016), we repeated text mining with additional restrictions on search specifications. We selected only printed articles and excluded additional media that are covered by the archives of Handelsblatt and Frankfurter Allgemeine Zeitung, such as the business weekly Wirtschaftswoche and the regional FAZ edition Rhein-Main-Zeitung. However, we were unable to reproduce the original index. Our tentative explanation for these differences is the limited reliability of search results within newspapers' online archives, with possibly unstable article counts over time. The recheck in March 2023 led to different article counts in the Frankfurter Allgemeine Zeitung and slightly changed the trajectory of the German EPU. Baker et al. (2016) also experienced this phenomenon with the New York Times archive. These additional results are provided in Appendix C.[21]

The alternative adjustment of the German EPU removes a large portion of the long-term trends compared to the original and replicated EPU indices. Moreover, unlike the other two indices, in the case of the adjusted EPU,

[21] We performed both automated scraping and manual searching, and the results were identical. We have also experimented with different logical operators to verify whether the search engine interprets them as we expect, but even this exploration did not uncover the cause of the differences.

the implied level of uncertainty in the 2010s appears to be higher than in the previous decade. This is despite a relatively broad consensus on macroeconomic policy and the German economy's successful rebound from the Great Recession. However, the increase in the adjusted EPU with the COVID-19 pandemic is less pronounced than in the original and replicated EPUs. The dynamics of the adjusted EPU is also reasonably robust to additional restrictions to include only print articles and the strict focus on Handelsblatt and Frankfurter Allgemeine Zeitung without additional resources that appear in their online archives (Appendix C).

The French case illustrates the sensitivity of the EPU index to newspaper selection. Our sample is shorter than that of Baker et al. (2016), and we replaced the leading newspaper Le Monde with the leading economic newspaper Les Echos. These two largely arbitrary changes were enough to produce very different long-term trends in the French EPU. Even our replicated EPU lacks most of the trend increase of the original EPU, whereas the normalization of uncertainty-related articles by economic policy articles removes an additional portion of trend growth (adjusted EPU). In the next section, we will show that this result remains robust even after the inclusion of Le Monde in the sample.

The results are similar for the United Kingdom to those of France. Changes in newspaper selection - note that we have replaced the Financial Times with the Guardian - and in the database lead to a persistently lower EPU index in the latter part of the sample. We do not observe prominent peaks during the European debt crisis, and the uncertainty related to Brexit decreased compared to the original EPU. On the other hand, the difference between the replicated series and the alternative normalization is relatively small.

Finally, in the cases of the two southern countries, Italy and Spain, both the replicated and adjusted EPU indices produce similar dynamics to the original series without apparent increases in the long-run trend. We attribute this similarity to the fact that our search specifications match perfectly with those used by Baker et al. (2016), who also rely on the Factiva database. However, even in this case, our replicated EPU index does not align perfectly with the original EPU index.

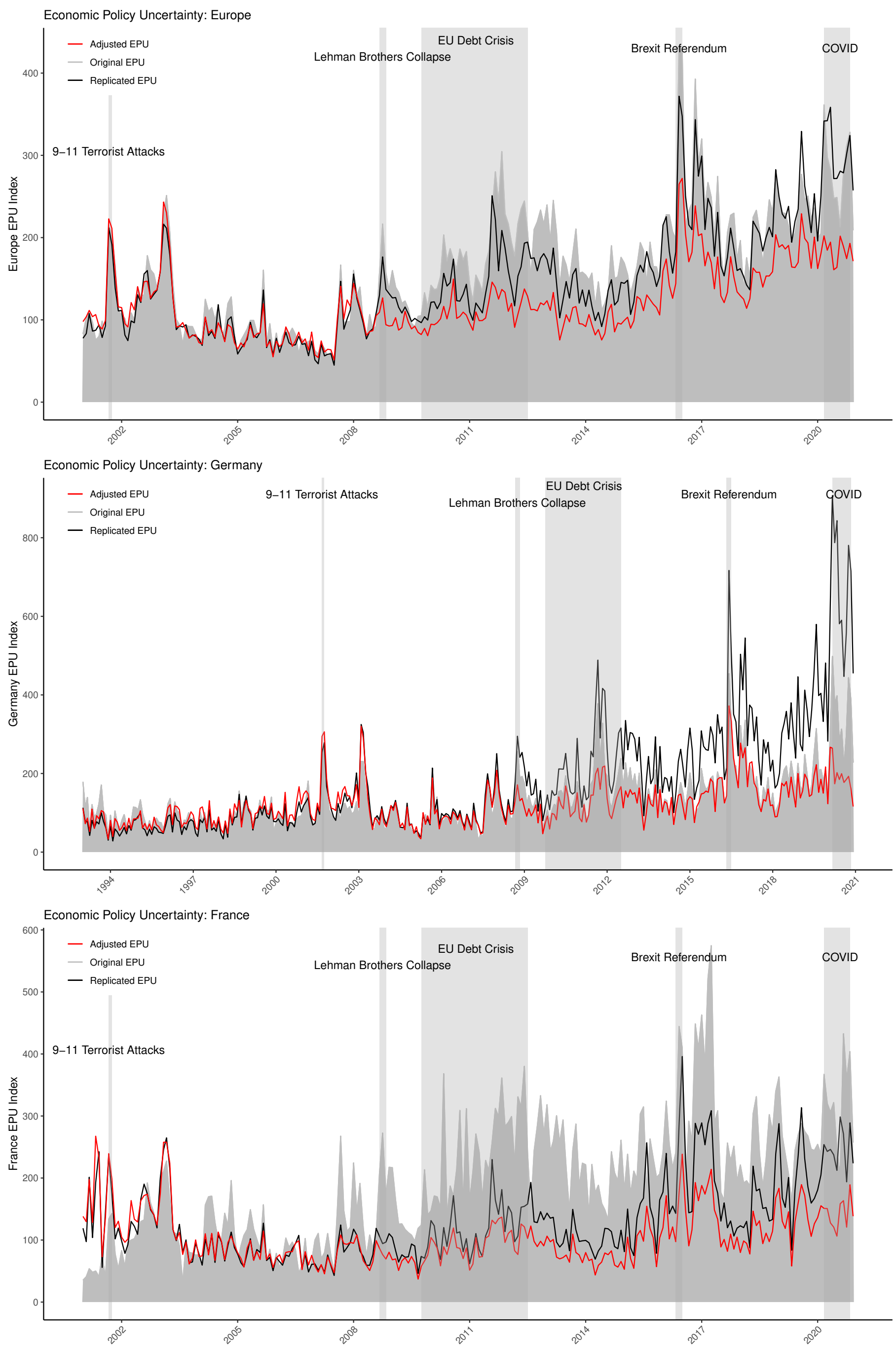
3.3.4 Sensitivity analysis: French EPU based on alternative newspapers.

In our replication of the French EPU, we rely on the newspapers Le Figaro and Les Echos, since both are available in the Factiva database. As a sensitivity check, we regenerate the index for France constructed using Le Monde, retrieved through the online scraping of its archives. This option allows to explore the impact of different sets of newspapers and databases on the EPU index.[22] Interestingly, replacing Les Echos with Le Monde results in an even greater difference between the replicated and original EPU indices (Figure 6 (*a*)). Despite the same composition of the newspapers as in Baker et al. (2016), the long-term trend is much less apparent in the replicated index. Furthermore, contrary to the original EPU index by Baker et al. (2016), the European debt crisis does not stand out as a major period of uncertainty.

We also calculated the French EPU from all three newspapers, Le Monde, Les Echos, and Le Figaro, to present the impact of extended newspaper coverage. This index constructed using three newspapers is between the indices based on two newspapers. Therefore, the EPU index is sensitive to newspaper selection,

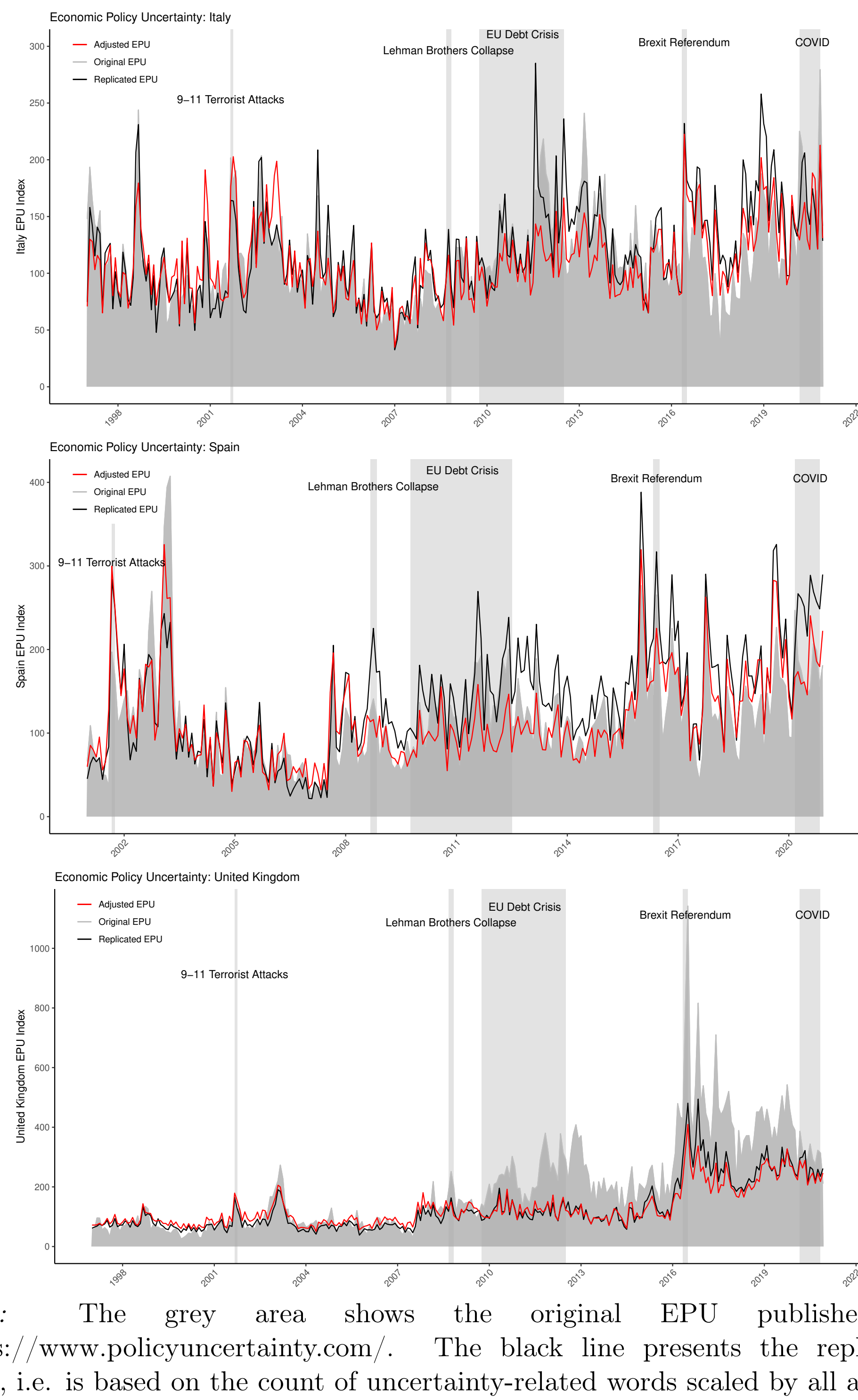
[22] Since we have higher confidence in the completeness of the Factiva database than in the scraped archive, we prefer to use the same database for both newspapers in one country to avoid the bias caused by different scopes of articles, such as whether sport-related news are included or not. Also, as mentioned in Section 3.1, we do not have the opportunity to access the Lexis Nexis database used by Baker et al. (2016) as a source of Le Monde articles.

Figure 3.5: Comparison of the Original, Replicated, and Adjusted EPU Indices



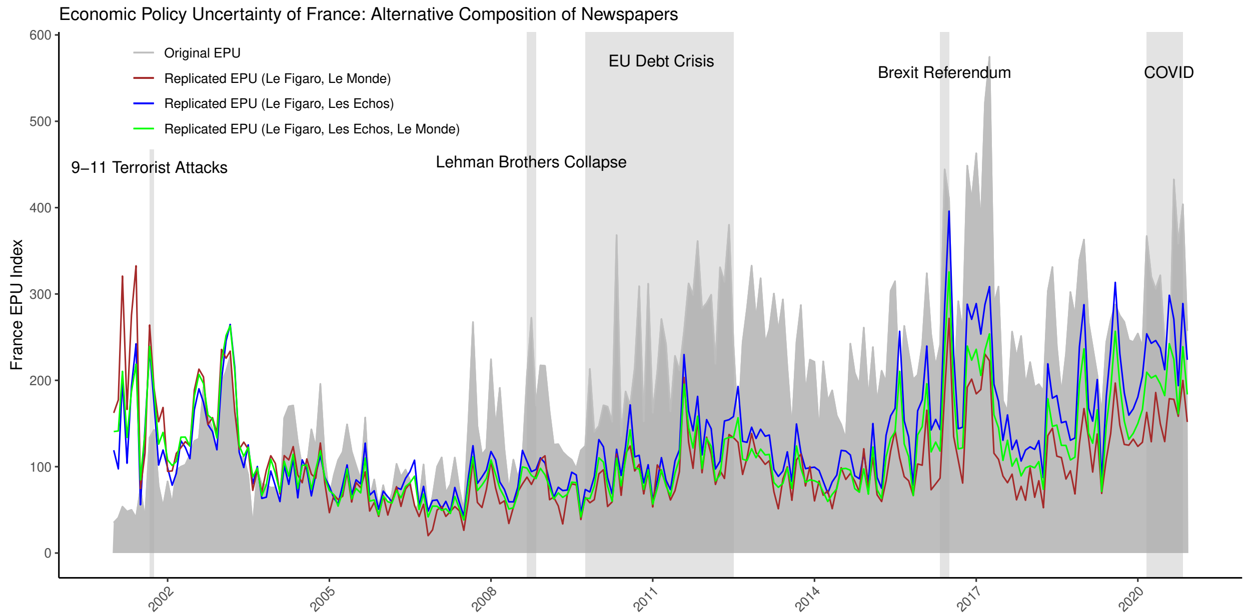
Note: The grey area shows the original EPU published at https://www.policyuncertainty.com/. The black line presents the replicated EPU, i.e. is based on the count of uncertainty-related words scaled by all articles. The red line - adjusted EPU - shows the EPU index based on the count of uncertainty-related articles scaled by economic-policy-related articles.

Figure 3.5: Comparison of the Original, Replicated, and Adjusted EPU Indices, Cont.



Note: The grey area shows the original EPU published at https://www.policyuncertainty.com/. The black line presents the replicated EPU, i.e. is based on the count of uncertainty-related words scaled by all articles. The red line – adjusted EPU - shows the EPU index based on the count of uncertainty-related articles scaled by economic-policy-related articles.

Figure 3.6: (*a*) French EPU Indices Based on Alternative Newspaper Selections



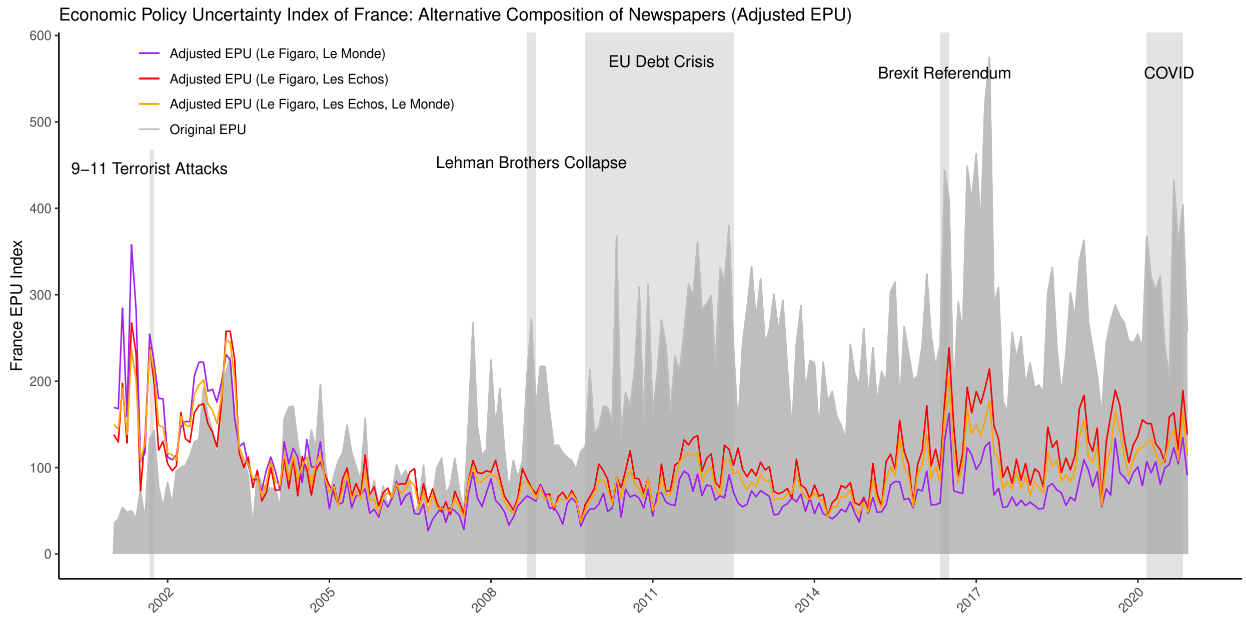
Note: The grey area shows the original EPU published at https://www.policyuncertainty.com/ (both panels). Top panel: Blue = replicated EPU based on the Factiva database, with Le Figaro and Les Echos. Green = replicated EPU, Le Figaro (Factiva), Le Monde (own archive scraped using Google). Brown = all three newspapers. Bottom panel: Original EPU and replicated EPUs normalized by economic-policy articles. Red = Le Figaro and Les Echos; orange = Le Figaro and Le Monde; purple = all three newspapers.

but the difference is not as pronounced as the utilization of different archives and search engines. Along with the original EPU index, the bottom panel of Figure 6(*b*) shows adjusted EPU indices normalized by "economic policy" articles for different combinations of French newspapers. The rising trend of the original EPU index is almost completely absent from the replicated indices.

3.3.5 Implications

Our calculations show that it is not easy to replicate the EPU index by Baker et al. (2016). We obtained the closest, albeit imperfect, match between our replication and the original index in the cases of Italy and Spain, where we used identical newspapers and databases. Additionally, although we used the exact specification of text mining and sources for Germany, our replicated

Figure 3.6: (*b*) French Adjusted EPU Indices Based on Alternative Newspaper Selections



Note: The grey area shows the original EPU published at https://www.policyuncertainty.com/ (both panels). Top panel: Blue = replicated EPU based on Factiva database, with Le Figaro and Les Echos. Green = replicated EPU, Le Figaro (Factiva), Le Monde (own archive scraped using Google). Brown = all three newspapers. Bottom panel: Original EPU and replicated EPUs normalized by economic-policy articles. Red = Le Figaro and Les Echos; orange = Le Figaro and Le Monde; purple = all three newspapers.

index is very different. This suggests that the choice of database plays a more important role than the selection of newspapers in replicating the French EPU. However, a change in the selection of newspapers also impacts the EPU.

Lastly, we show that the alternative normalization of the count of “uncertainty" articles by “economic policy" articles leads to less pronounced trends in the EPU indices of Germany and France while remaining relatively inconsequential to indices in the other countries. Furthermore, the most prominent differences between the original and adjusted EPUs appear in recent years, around the Brexit referendum and at the beginning of the COVID-19 pandemic. In the cases of Germany, France, and the United Kingdom, the peaks of the adjusted

EPU are about one-half of the peaks of the original EPU.[23] On the positive side, all variants are highly correlated, notably at higher frequencies. The correlation coefficients are provided in the Appendix, Table A.3.

3.4 Applications

3.4.1 EPU and economic activity in Europe

Next, we estimate panel VAR models to investigate how the alternative normalization of the EPU translates into estimates of the impact of uncertainty shocks on economic activity. Our benchmark specification contains the EPU index, the log of stock prices, the 10-year government bond yield, the short-term interest rate, the unemployment rate, and the log of industrial production in manufacturing. The choice of variables and their ordering is inspired by the specification in Baker et al. (2016), who conducted a similar exercise for a wider range of countries.[24]

The panel VAR model was estimated on monthly data beginning in January 2001 and ending in December 2019, prior to the COVID-19 pandemic. We adopted a pooled estimator, which is more suitable for panel VAR models with larger T than the GMM estimator (Canova and Ciccarelli, 2013), and the model was estimated using Bayesian techniques based on the Normal-Wishart prior (Dieppe et al., 2016).[25] We used three lags for estimation. Regarding the identification of orthogonal shocks, we rely on the Cholesky identification, although timing restrictions are always debatable. However, our main goal

[23] Table A.1 in the Appendix shows the differences in peaks for different EPU indices.

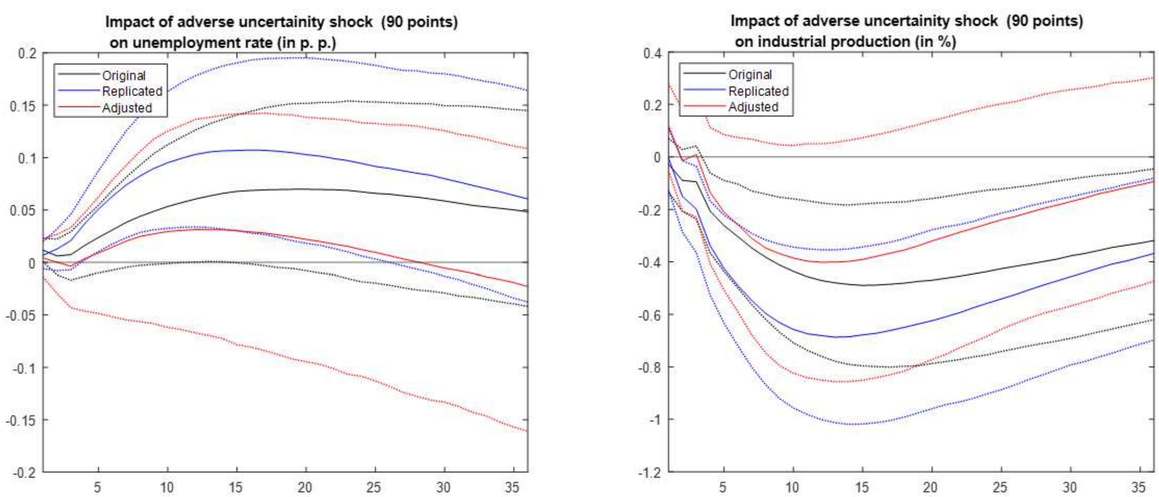
[24] Baker et al. (2016) include several VAR specifications in their paper. Their baseline for the United States includes the EPU index, the log of the S&P 500 index, the federal funds rate, log employment, and log industrial production. The panel VAR model for all countries for which the EPU has been developed uses the same specification but with the unemployment rate instead of employment. We added bond yields to track EU financial market conditions with more precision.

[25] Baker et al. (2016) rely on the pooled estimator as well. However, they use a standard OLS estimator.

is to compare the impulse responses implied by alternative EPU indices and not causal inference, and for this purpose, this simple identification scheme is sufficient.

The resulting impulse responses to the effects of 90-point increases in alternative EPU indices on the unemployment rate and industrial production are shown in Figure 7.[26] Qualitatively, the results are similar: no matter whether the original, replicated or adjusted EPU is used, it is predicted that the unemployment rate will increase and industrial production will decrease. However, the adjusted index leads to weaker implications because, in this case, none of the responses to the uncertainty shock are statistically significant.

Figure 3.7: Impact of an Uncertainty Shock on Unemployment and Industrial Production for Different EPU Indices



Note: Panel VAR model, Germany, France, Spain, Italy, and the United Kingdom, monthly data 2001M01 - 2019M12. 90% credible intervals obtained from 1,000 bootstrap simulations. Responses correspond to an increase in the EPU index by 90 points. The 'Difference' column shows the distribution of the difference between impulse responses under the original and replicated index of uncertainty. A positive value means that the variable in question has a stronger response to an innovation in uncertainty when the original EPU index is considered.

Regarding the impact on the unemployment rate, the predicted effect of

[26]The 90-point increase of the EPU follows Baker et al., 2016 for comparability.

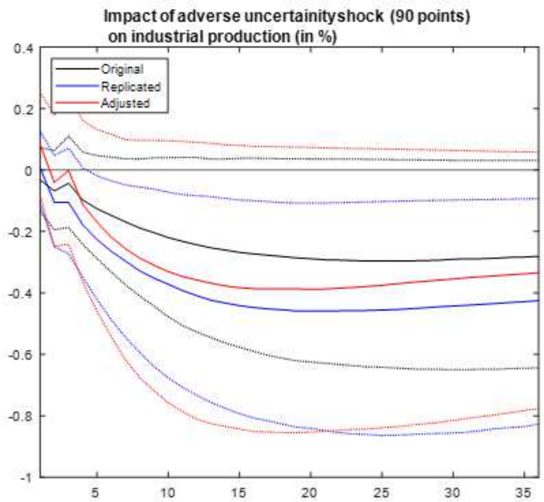
uncertainty is lowest with the adjusted EPU, about half of the effect of the original EPU, and one-third of the effect of the replicated EPU. Furthermore, the effects lose their statistical significance and a large part of the 90% credible interval is below zero. The effect on industrial production at the one-year horizon is again lowest with the adjusted index and highest with the replicated EPU, but the difference between the adjusted and the original EPU is not that pronounced. Quantitatively, a 90-point increase in adjusted and original EPU implies a decrease in industrial production of about 0.4%; nevertheless, the impact increases to more than 0.6% with the replicated EPU.

As a sensitivity check, we also estimate a bivariate model with industrial production and the EPU index. In this case, the estimated impulse responses of industrial production to alternative EPU indices are very similar, with the original EPU having the quantitatively smallest effect (Figure 8). However, the adjusted EPU again implies lower effects than the replicated EPU, and the responses are statistically insignificant with 90% credible intervals.

As we have shown in previous sections, the largest discrepancies between the adjusted EPU and the original or replicated EPU arise in the latter part of the sample and are driven by different trends in the alternative EPU indices, whereas the timing of spikes in EPUs remains similar. To investigate whether differences in trends drive differences in estimated impulse responses, we re-estimated our baseline VAR panel VAR model with EPUs detrended by the Hodrick-Prescott filter, with the smoothing parameter λ set to 10,000. The resulting impulse responses (Figure 9) reveal that the responses of the unemployment rate and industrial production are remarkably similar, implying that the effects of the high-frequency component of all EPU indices are similar.

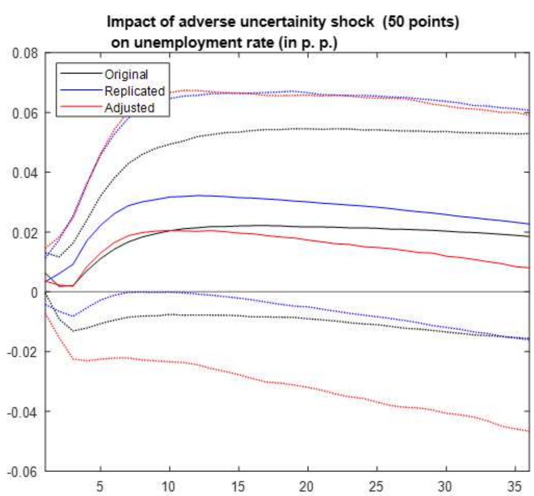
In general, these exercises confirm the negative effects of uncertainty on eco-

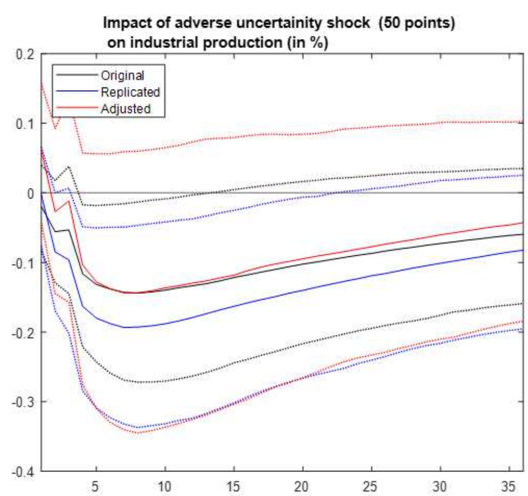
Figure 3.8: Impact of an Uncertainty Shock on Industrial Production for Different EPU Indices: Bivariate Model



Note: Panel VAR model, Germany, France, Spain, Italy, and the United Kingdom, monthly data 2001M01 - 2019M12. 90% credible intervals obtained from 1,000 bootstrap simulations. Responses correspond to an increase in the EPU index by 90 points.

Figure 3.9: Impact of an Uncertainty Shock on Unemployment and Industrial Production for Different EPU Indices: Cyclical Components of EPUs.





Note: Panel VAR model, Germany, France, Spain, Italy, and the United Kingdom, monthly data 2001M01 - 2019M12. 90% credible intervals obtained from 1,000 bootstrap simulations. The responses correspond to a 50-point increase in the cyclical component of the EPU obtained using the Hodrick-Prescott filter by 50 points.

nomic activity reported by Baker et al. (2016) and many others since then. Therefore, from a policy perspective, the difference in the predicted effects of uncertainty on economic activity is in the assessment of the size of the shock implied by the adjusted EPU scaled by economic policy versus the original EPU, and not in the different responses of economic activity to uncertainty.

3.4.2 Brexit-related uncertainty

One of the most prominent European EPU peaks corresponds to the Brexit referendum of June 23, 2016. Uncertainty around future trade relations with the United Kingdom was expected to have a detrimental impact on the British economy. Bloom (2016) had already predicted a negative impact of uncertainty before the referendum. Based on the findings in Baker et al. (2016), Bloom (2016) asserts that a stop in investment and hiring until clarification of the UK's status would amplify economic uncertainty and its negative impact on the economy.[27] International Monetary Fund (2016) and HM Treasury (2016) also highlight that the effect of the Brexit referendum was transmitted to the British economy through uncertainty. In these publications, the uncertainty was represented by a common factor of several uncertainty indicators. Therefore, the impact of measurement problems on the EPU was lower compared to the analysis by Bloom (2016). [28]

[27] *"In that analysis, we estimated that a 90-point upward innovation in the US EPU Index led to short-term declines of 1.2% in US industrial production, about 0.6% in its gross investment, and about 0.5% in its level of employment. Since the Brexit-related increase in the UK EPU index appears to be even greater, we believe that Brexit-related uncertainty has a material negative effect on UK economic performance"* (Bloom, 2016).

[28] The HM Treasury constructed the uncertainty factor by averaging the following normalized measures: the EPU; FTSE 100 implied volatility; sterling implied volatility; CBI Industrial Trends uncertainty measure; European Commission consumer uncertainty measure, and GfK unemployment expectations. All series were normalized by subtracting their means and dividing by the respective standard deviations in 2000-2015. The IMF used the average standard deviation of the current and future years of the consensus forecast, the EPU, the survey-based indicator of the uncertainty effect on industrial demand, GfK unemployment expectations, and the volatility of the stock market and the exchange rate.

Table 3.2: Brexit-related uncertainty, the United Kingdom

	original EPU	*replicated EPU*	*adjusted EPU*
January 2016	249.4	153.0	146.0
February 2016	312.6	208.7	177.7
March 2016	479.3	229.4	177.4
April 2016	434.6	194.1	159.8
May 2016	428.5	296.6	226.1
June 2016	799.9	384.2	321.9
July 2016	1141.8	480.8	409.4
August 2016	458.7	336.9	273.2
September 2016	379.0	266.3	225.7
October 2016	545.1	344.0	285.5
November 2016	816.2	494.4	337.3
December 2016	468.0	321.7	252.0

Note: The United Kingdom European Union membership referendum took place on June 23, 2016. The original EPU is taken from https://www.policyuncertainty.com/. The replicated EPU is based on the count of uncertainty-related words scaled by all articles. The adjusted EPU is the EPU index with alternative normalization, that is, the counts of articles related to uncertainty scaled by articles related to economic policy.

In Table 2, we present the evolution of the British EPU around the period of Brexit referendum. Our replicated and adjusted EPU indices reveal a markedly lower increase in uncertainty. Although the original EPU in July is 4.5 times higher than the January level, our replicated indices are about three times higher than their January levels. Therefore, the increase in uncertainty is corroborated by our replication, but its magnitude is lower. Our results have quantitative implications for the predictions of the short-term impact of the Brexit referendum. According to Bloom's assessment of the Baker et al. (2016) US model, every 90-point upward innovation in the EPU implies a 0.5% decrease in employment. Therefore, an increase in the EPU of 890 points between January and July 2016 decreases employment by 5%. On the other hand, the replicated and adjusted EPUs imply a moderate decrease in employment of 1.4 to 1.8%. Thus, a relatively arbitrary change in the composition of newspapers, along with a change in the database used for scraping the newspaper articles, lead to a significantly lower predicted impact of the Brexit referendum on employment. A similar exercise for industrial production leads to a prediction of a 12% decline in industrial production with the original EPU. In comparison, the replicated EPU again suggests a more moderate 4% decline.[29]

3.4.3 COVID-related uncertainty

The uncertainty indices spiked again with the start of the COVID-19 pandemic. Altig et al. (2020) compare the evolution of a wide range of indicators that map uncertainty in the United States. The authors found that while all indices imply huge jumps in uncertainty in reaction to the pandemic, their amplitudes and time paths differ greatly. We document similar variations solely among the original, replicated, and adjusted EPU indices in the case of European countries. Table 3 presents the European EPU indices and shows that the original and replicated EPU peaked at values greater than 80% above the January 2020 levels, while the adjusted EPU peak was only 25% higher.

Again, a substantially larger variation appears in the country-level data. We observe the largest increases in the EPU indices in Germany, where the original EPU almost quadrupled and the adjusted index doubled. Large discrepancies between the original and adjusted EPU indices appear in all other countries, but the sizes of the differences between the peaks and the values in January 2020 are smaller. Still, the EPU index scaled by all articles leads to roughly twice as large an uncertainty increase as the adjusted EPU normalized by

[29] Note that following the result of the Brexit referendum, the Bank of England cut the policy rate, launched quantitative easing, and allowed the pound to depreciate. All these measures contributed to mitigating the short-term impact of the referendum on economic performance, and industrial production accelerated after those monetary policy interventions. Employment also continued to rise.

economic policy articles in all countries. In addition, the timings of the peaks are different in France, Spain, and the United Kingdom.

To conclude, we confirm that the COVID-19 pandemic caused unprecedented jumps in the uncertainty index of Europe, but the timing of the peaks and the implied changes in uncertainty differ markedly across EPU indices.

3.5 Conclusion

In this paper, we explore the properties of the EPU indices of the major European economies (Germany, France, Italy, Spain, and the United Kingdom) and their sensitivity to minor modifications to estimation procedures.

First, we discuss the long-term increasing trends in the EPU indices in the United Kingdom, Germany, France, and the aggregate European EPU index, and the lack thereof in Italy and Spain, despite the turmoil during the EU debt crisis and the subsequent political instability in both countries. Moreover, the trends in the EPU are inconsistent with the financial instability measured by implied volatility and with the World Uncertainty Index. These uncertainty indices tend to be mean-reverting and do not contain any clear upward-sloping trends observed in the EPU.

Next, we replicate the text mining and construction of the EPU indices to reveal the nature of the trends and their sensitivity to minor changes in their construction. We found that our replicated EPU indices differed from the original indices, despite using the closest possible text mining specification. The differences appear mainly in the long-run behavior of the indices of Germany, France, and, to some extent, the United Kingdom as well. These differences arise not only from the utilization of alternative databases and/or newspapers (France, United Kingdom), but also appear in countries where our text-mining exercise matches the one conducted by Baker et al. (2016), the authors of the original EPU.

Finally, we show that some dynamics of the original EPU indices are driven by non-trivial movements in the count of all articles used for the normalization of the raw count of uncertainty-related articles rather than by the frequency of articles related to economic policy uncertainty that are supposed to provide signals about genuine uncertainty. To eliminate the impact of changing newspaper composition on the EPU, we normalize the counts of uncertainty-related articles by the counts of articles related to economic policy, instead of all articles. We found that this normalization leads to less pronounced trends in the EPU, more consistent results using different newspapers and/or databases, and that it possibly increases the reproducibility and reliability of the EPU index, as seen in our experiments with French and German newspapers.

Our findings have several implications for empirical research on the effects of uncertainty shocks. Most importantly, the EPU index could have overestimated the uncertainty in European countries during recent uncertainty episodes by 20 - 50%. Furthermore, our panel VAR experiments revealed that the impulse responses of industrial production and the unemployment rate to the shock of uncertainty are relatively similar across alternative EPU indices, so the negative effects of uncertainty shocks on economic activity are also confirmed when articles related to economic policy are used instead of all articles for the normalization of the EPU. Therefore, the differences in the sizes of alternative EPUs lead to a quantitatively different assessment of the importance of uncertainty shocks, which is particularly relevant for policymakers aiming to offset their negative effects through policy interventions. Our results also imply that policymakers must consider the limited reliability of uncertainty indicators to track uncertainty. Finally, alternative scaling factors for the count of uncertainty-related articles should be considered when developing new uncertainty indicators based on the methodology of the EPU index.

Table 3.3: Economic policy uncertainty during COVID-19

	Peak value	Increase in the peak since January (%)	Peak date
		Europe	
original EPU	361.4	81.4	March
replicated EPU	358.5	83.0	May
adjusted EPU	202.1	24.5	March
		Germany	
original EPU	498.1	278.7	March
replicated EPU	907.0	221.2	March
adjusted EPU	267.0	117.1	March
		France	
original EPU	432.7	70.2	September
replicated EPU	298.5	65.2	August
adjusted EPU	189.2	38.9	November
		Italy	
original EPU	279.4	118.5	November
replicated EPU	209.1	49.0	November
adjusted EPU	212.9	46.8	November
		Spain	
original EPU	246.8	158.9	March
replicated EPU	289.7	91.0	December
adjusted EPU	240.4	68.0	August
		*United Kingdom**	
original EPU	386.6	62.4	March
replicated EPU	322.0	36.7	May
adjusted EPU	289.0	29.2	May

Note: Our sample ends in December 2020. The original EPU is taken from https://www.policyuncertainty.com/. The replicated EPU is based on the count of uncertainty-related words scaled by all articles. The adjusted EPU is the EPU index with the alternative normalization, i.e., the counts of uncertainty-related articles scaled by articles related to economic policy. *In the case of the United Kingdom, the COVID-related peak is compared to the EPU levels in February. In January, the EPU indices were determined by the final phase of negotiation of the Withdrawal Agreement Bill that was finally passed through the House of Commons on 22 January, and the United Kingdom officially left the EU on 31 January 2020. The original EPU peaks in January 2020, before the spread of the COVID-19 epidemic in Europe.

References

Ahir, H., Bloom, N., & Furceri, D. (2018). The world uncertainty index. *Mimeo, Available at SSRN 3275033*.

Ahir, H., Bloom, N., & Furceri, D. (2020). 60 years of uncertainty. *Finance & Development*, *57*(001).

Ahir, H., Bloom, N., & Furceri, D. (2021). What the continued global uncertainty means for you. *IMF Blog, January 19, 2021*.

Altig, D., Baker, S., Barrero, J. M., Bloom, N., Bunn, P., Chen, S., Davis, S. J., Leather, J., Meyer, B., Mihaylov, E., et al. (2020). Economic uncertainty before and during the COVID-19 pandemic. *Journal of Public Economics*, *191*, 104274.

Azqueta-Gavaldón, A. (2017). Developing news-based economic policy uncertainty index with unsupervised machine learning. *Economics Letters*, *158*, 47–50.

Azqueta-Gavaldón, A., Hirschbühl, D., Onorante, L., & Saiz, L. (2023). Sources of economic policy uncertainty in the euro area. *European Economic Review*, 104373.

Bachmann, R., Elstner, S., & Sims, E. R. (2013). Uncertainty and economic activity: Evidence from business survey data. *American Economic Journal: Macroeconomics*, *5*(2), 217–249.

Baker, S., Bloom, N., Canes-Wrone, B., Davis, S. J., & Rodden, J. (2014). Why has US policy uncertainty risen since 1960? *American Economic Review*, *104*(5), 56–60.

Baker, S., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, *131*(4), 1593–1636.

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica*, *77*(3), 623–685.

Bloom, N. (2016). What is Brexit-related uncertainty doing to UK growth? *British Politics and Policy at LSE*.

Brandt, R. (2021). Economic policy uncertainty index: Extension and optimization of Scott R. Baker, Nicholas Bloom and Steven J. Davis's search term. *DoCMA Working Paper*, (5).

Canova, F., & Ciccarelli, M. (2013). Panel vector autoregressive models: A survey. In *Var Models in Macroeconomics – New Developments and Applications: Essays in Honor of Christopher A. Sims* (pp. 205–246). Emerald Group Publishing Limited.

Castelnuovo, E., Lim, G., & Pellegrino, G. (2017). A short review of the recent literature on uncertainty. *Australian Economic Review*, *50*(1), 68–78.

Charemza, W., Makarova, S., & Rybinski, K. (2022). Economic uncertainty and natural language processing; the case of Russia. *Economic Analysis and Policy*, *73*, 546–562.

Dieppe, A., Legrand, R., & Van Roye, B. (2016). The Bayesian Estimation, Analysis and Regression (bear) Toolbox. *ECB working paper*.

Duca, J. V., & Saving, J. L. (2018). What drives economic policy uncertainty in the long and short runs: European and US evidence over several decades. *Journal of Macroeconomics*, *55*, 128–145.

European Central Bank. (2016). The impact of uncertainty on activity in the euro area. *ECB Economic Bulletin No. 2016/8*.

Fernández-Villaverde, J., & Guerrón-Quintana, P. A. (2020). Uncertainty shocks and business cycle research. *Review of Economic Dynamics*, *37*, S118–S146.

Ferrara, L., Lhuissier, S., & Tripier, F. (2018). Uncertainty fluctuations: Measures, effects and macroeconomic policy challenges. In *International Macroeconomics in the Wake of the Global Financial Crisis* (pp. 159–181). Springer.

HM Treasury. (2016). HM Treasury analysis: The immediate economic impact of leaving the EU. *Report presented to Parliament by the Chancellor of the Exchequer by Command of Her Majesty*.

Husted, L., Rogers, J., & Sun, B. (2020). Monetary policy uncertainty. *Journal of Monetary Economics*, *115*, 20–36.

International Monetary Fund. (2016). United Kingdom: Selected issues. *IMF Country Report No. 16/169*.

Jha, M., Liu, H., & Manela, A. (2021). Natural disaster effects on popular sentiment toward finance. *Journal of Financial and Quantitative Analysis*, *56*(7), 2584–2604.

Jha, M., Liu, H., & Manela, A. (2022). Does finance benefit society? a language embedding approach. *SSRN Scholarly Paper ID, 3655263*.

Jurado, K., Ludvigson, S. C., & Ng, S. (2015). Measuring uncertainty. *American Economic Review*, *105*(3), 1177–1216.

Justiniano, A., & Primiceri, G. E. (2008). The time-varying volatility of macroeconomic fluctuations. *American Economic Review*, *98*(3), 604–41.

Kaya, O., Schildbach, J., AG, D. B., & Schneider, S. (2018). Economic policy uncertainty in Europe. *Deutsche Bank Research*.

Landler, M. (2004). Media; Woes at two pillars of German journalism (National edition, Section C). *The New York Times*, 8–8.

Larsen, V. H. (2021). Components of uncertainty. *International Economic Review*, *62*(2), 769–788.

Ludvigson, S. C., Ma, S., & Ng, S. (2021). Uncertainty and business cycles: Exogenous impulse or endogenous response? *American Economic Journal: Macroeconomics*, *13*(4), 369–410.

Meinen, P., & Röhe, O. (2017). On measuring uncertainty and its impact on investment: Cross-country evidence from the euro area. *European Economic Review*, *92*, 161–179.

Rossi, B., & Sekhposyan, T. (2015). Macroeconomic uncertainty indices based on nowcast and forecast error distributions. *American Economic Review*, *105*(5), 650–655.

Tobback, E., Naudts, H., Daelemans, W., de Fortuny, E. J., & Martens, D. (2018). Belgian economic policy uncertainty index: Improvement through text mining. *International Journal of Forecasting*, *34*(2), 355–365.

3.A Additional tables

Table 3.A1: Stationarity tests

I. Index by Baker, Bloom, and Davis (2016)

	ADF		ADF-GLS		KPSS	
	tau	p-value	tau	p-value	t-stat	p-value
Europe	-2.403	0.141	-0.577	0.527	2.724	$<$0.01
Germany	-5.409	0.000	-0.859	0.392	2.058	$<$0.01
France	-2.453	0.127	-0.502	0.559	3.141	$<$0.01
Italy	-3.704	0.004	-1.023	0.314	0.494	0.045
Spain	-3.609	0.006	-1.999	0.045	0.435	0.062
UK	-0.803	0.818	0.192	0.795	3.141	$<$0.01

II. Replicated EPU indices

	ADF		ADF-GLS		KPSS	
	tau	p-value	tau	p-value	t-stat	p-value
Europe	-2.051	0.265	-0.667	0.487	2.719	$<$.01
Germany	-2.024	0.277	0.048	0.753	3.029	$<$.01
France	-2.624	0.088	-2.308	0.020	1.476	$<$.01
Italy	-4.279	0.001	-1.899	0.058	1.488	$<$.01
Spain	-3.414	0.011	-0.639	0.501	1.805	$<$.01
UK	-1.404	0.582	-0.718	0.461	2.701	$<$.01

Note: Augmented Dickey-Fuller test with lag length based on AIC, test down from 12 lags. ADF-GLS test, demeaned and detrended data, lag length based on AIC, test down from 12 lags. ADF and ADF-GLS null hypothesis: unit root. KPSS test, automatic window size, null hypothesis: stationarity. Sample: 2001:1 - 2019:10.

Table 3.A2: Summary statistics

Variable	Since	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max	AR(1)
Original EPU Index									
Europe	1990-01	372	139	67	45	89	179	433	0.8277
Germany	1993-01	336	132	73	28	81	167	498	0.7095
Italy	1997-01	288	112	41	32	81	135	279	0.6049
UK	1997-01	288	194	156	25	75	274	1142	0.8712
France	1990-01	372	155	98	11	81	218	575	0.8190
Spain	2001-01	240	117	58	23	75	146	407	0.6051
Replicated EPU Index									
Europe	2001-01	240	147	68	45	96	183	372	0.8632
Germany	1993-01	336	180	145	28	81	251	907	0.8469
Italy	1997-01	288	119	43	33	87	143	285	0.5863
UK	1990-01	372	130	77	38	73	153	494	0.8810
France	2001-01	240	131	64	43	83	159	396	0.7249
Spain	2001-01	240	139	68	22	88	181	388	0.6828
Adjusted EPU Index									
Europe	2001-01	240	121	43	51	91	143	272	0.8386
Germany	1993-01	336	120	53	33	83	148	372	0.6665
Italy	1997-01	288	110	34	35	85	128	223	0.5461
UK	1990-01	372	125	60	51	86	145	409	0.8623
France	2001-01	240	105	43	37	73	126	267	0.6866
Spain	2001-01	240	117	55	30	78	147	326	0.6787

Summary Statistics for Original, Replicated and Adjusted EPU indeces, for Europe, Germany, Italy, UK, France and Spain.

Table 3.A3: Uncertainty indicators: correlation coefficients

			Europe				
	EPU_{Orig}	EPU_{Rep}	EPU_{Adj}	WUI	VSTOXX		
EPU_{Orig}	1	0.938	0.851	0.696	0.154		
EPU_{Rep}		1	0.917	0.665	0.16		
EPU_{Adj}			1	0.627	0.189		
WUI				1	-0.105		
			Germany				
	EPU_{Orig}	EPU_{Rep}	EPU_{Adj}	WUI	Forecast	Macro	VDAXX
EPU_{Orig}	1	0.946	0.850	0.374	0.325	-0.003	0.281
EPU_{Rep}		1	0.794	0.471	0.394	-0.110	0.226
EPU_{Adj}			1	0.321	0.117	0.033	0.302
WUI				1	0.318	-0.131	-0.053
Forecast					1	0.239	0.224
Macro						1	0.446
			France				
	EPU_{Orig}	EPU_{Rep}	EPU_{Adj}	WUI	Forecast	Macro	VCAC
EPU_{Orig}	1	0.665	0.387	0.312	0.252	-0.035	0.035
EPU_{Rep}		1	0.864	0.199	0.005	0.128	0.175
EPU_{Adj}			1	0.220	0.001	0.178	0.277
WUI				1	0.238	-0.085	0.217
Forecast					1	0.168	0.231
Macro						1	0.618
			Italy				
	EPU_{Orig}	EPU_{Rep}	EPU_{Adj}	WUI	Forecast	Macro	
EPU_{Orig}	1	0.777	0.730	0.263	-0.004	0.021	
EPU_{Rep}		1	0.865	0.339	-0.128	-0.009	
EPU_{Adj}			1	0.275	0.046	0.067	
WUI				1	-0.223	-0.274	
Forecast					1	0.371	
			Spain				
	EPU_{Orig}	EPU_{Rep}	EPU_{Adj}	WUI	Forecast	Macro	
EPU_{Orig}	1	0.801	0.800	0.498	0.022	0.162	
EPU_{Rep}		1	0.875	0.507	0.179	0.356	
EPU_{Adj}			1	0.500	0.109	0.192	
WUI				1	-0.024	0.200	
Forecast					1	0.335	
			United Kingdom				
	EPU_{Orig}	EPU_{Rep}	EPU_{Adj}	WUI	VFTSE		
EPU_{Orig}	1	0.907	0.867	0.748	-0.124		
EPU_{Rep}		1	0.951	0.695	-0.136		
EPU_{Adj}			1	0.766	-0.112		
WUI				1	-0.057		

Note: The correlations between the original EPU by Baker et al. (2016), the replicated EPU and the adjusted EPU scaled by economic policy articles and other uncertainty indicators. Those include (i) the respective world uncertainty indices (Ahir et al. (2018)), (ii) the implied volatility of the stock market, (iii) the uncertainty indices derived from the forecast dispersion and (iv) macroeconomic uncertainty. The last two indicators were calculated by Meinen and Röhe (2017)). Because the WUI is available on a quarterly basis, its correlations are based on quarterly data.

3.B Replication of the EPU - search queries

Germany

Q1 (Wirtschaft OR wirtschaftlich) AND (steuer OR wirtschaftspolitik OR regulierung OR regulierungs OR ausgaben OR bundesbank OR EZB OR zentralbank OR haushalt OR defizit OR haushaltsdefizit) AND (unsicher OR Unsicherheit)

Q2 (Wirtschaft OR wirtschaftlich) AND (steuer OR wirtschaftspolitik OR regulierung OR regulierungs OR ausgaben OR bundesbank OR EZB OR zentralbank OR haushalt OR defizit OR haushaltsdefizit)

Q3 (empty)

Sources: Own archives of Frankfurter Allgemeine Zeitung and Handelsblatt available at https://fazarchiv.faz.net/?dosearch=new and https://archiv.handelsblatt.com/

France

Q1 (economie OR economique OR economiques) AND (taxe OR taxes OR impot OR impots OR politique OR politiques OR regulation OR regulations OR reglementation OR loi OR “lois reglementations” OR depense OR depenses OR deficit OR deficits OR “banque centrale" OR “BCE" OR “Reserve Federale" OR budget OR budgetaire) AND (incertitude OR incertain OR incertitudes OR incertains)

Q2 (economie OR economique OR economiques) AND (taxe OR taxes OR impot OR impots OR politique OR politiques OR regulation OR regulations OR reglementation OR loi OR “lois reglementations” OR depense OR depenses OR deficit OR deficits OR “banque centrale" OR “BCE" OR “Reserve Federale" OR budget OR budgetaire)

Q3 aujourd'hui

Source: Factiva. Searched for Le Figaro (France, French Language) and Les Echos (France, French Language)

Search specifications:

Language: French

Exclude: Republished news; Recurring pricing and market data; Obituaries, sports, calendars, etc.

Starting date: January 2001

Note - Original sources: Figaro - Factiva, Le Monde - Lexis Nexis

Italy

Q1 (economia OR economico OR economica OR economici OR economiche) AND (tassa OR tasse OR politica OR regolamento OR regolamenti OR spesa OR spese OR spesa OR deficit OR "Banca Centrale" OR "Banca d'Italia" OR budget OR bilancio) AND (incerto OR incerta OR incerti OR incerte OR incertezza)

Q2 (economia OR economico OR economica OR economici OR economiche) AND (tassa OR tasse OR politica OR regolamento OR regolamenti OR spesa OR spese OR spesa OR deficit OR "Banca Centrale" OR "Banca d'Italia" OR budget OR bilancio)

Q3 oggi

Source: Factiva. Newspapers: Corriere della Sera (Italy, Italian Language) and La Stampa (Italy, Italian Language)

Search specifications:

Language: Italian

Exclude: Starting date: January 1997

Republished news; Recurring pricing and market data; Obituaries, sports, calendars...

Spain

Q1 (económica OR economía) AND (impuesto OR tarifa OR regulacion OR politica OR gastar OR gasta OR gasto OR presupuesto OR deficit OR "banco central") AND (incierto OR incertidumbre)

Q2 (económica OR economica) AND (impuesto OR tarifa OR regulacion OR politica OR gastar OR gasta OR gasto OR presupuesto OR deficit OR "banco central")

Q3 hoy

Source: Factiva. Newspapers: El Mundo (Spain, Spanish Language) El País - Nacional (Spain, Spanish Language)

Search specifications:

Language: Spanish

Exclude: Republished news; Recurring pricing and market data; Obituaries, sports, calendars, etc.

Starting date: January 2001 (limited by El País).

United Kingdom

Q1 (economic OR economy OR business OR industry OR commerce OR commercial) AND (spending OR policy OR deficit OR budget OR tax OR regulation OR "Bank of England" OR war OR tariff) AND (uncertain OR uncertainty)

Q2 (economic OR economy OR business OR industry OR commerce OR commercial) AND (spending OR policy OR deficit OR budget OR tax

OR regulation OR "Bank of England" OR war OR tariff)

Q3 today

Source: Factiva. Newspapers: The Times (UK), The Guardian (UK).

Search specifications:

Language: English

Exclude: Republished news; Recurring pricing and market data; Obituaries, sports, calendars, etc.

Starting date: January 1990 (limited by The Guardian).

The original EPU was retrieved from
https://www.policyuncertainty.com/europe_monthly.html, on April 13, 2021
Factiva searches were performed in March 2021.

3.C Additional results for Germany

This appendix tests the robustness of the differences between the original EPU index by Baker et al. (2016) and our replicated EPU. We have considered the following alterations to the baseline text mining exercise.

First, we manually retrieved the underlying EPU data at a quarterly frequency from the Frankfurter Allgemeine Zeitung archive. The results matched our monthly data.

Second, we reran the text mining in March 2023, two years after our data collection, with and without additional restrictions.

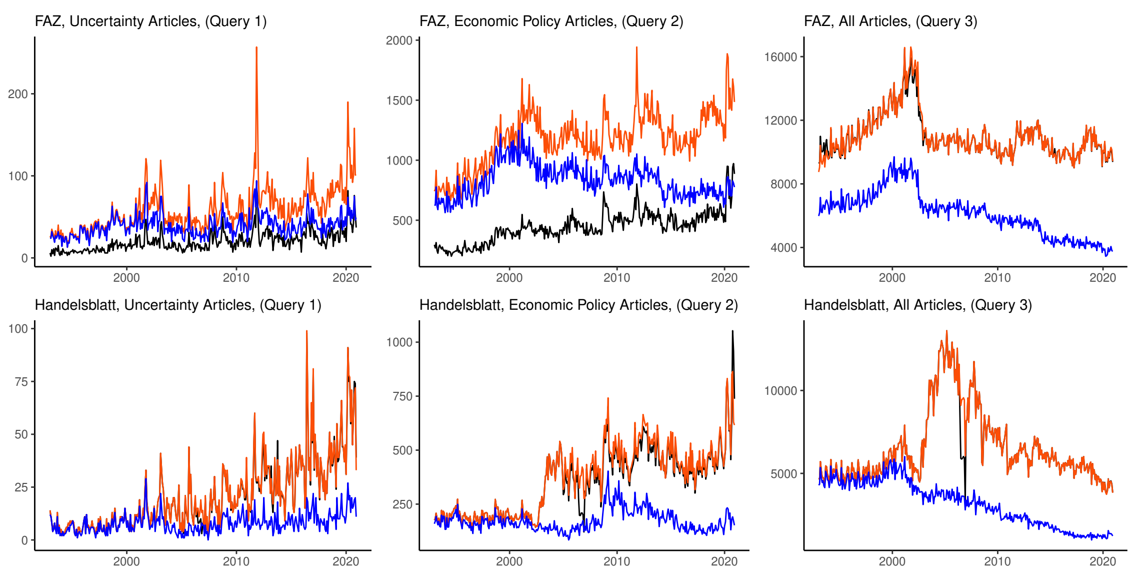
- Handelsblatt online archive: Only Handelsblatt print edition was used (previously, all resources (*Alle Quellen*) were used).

- Frankfurter Allgemeine Zeitung: Instead of *Alle Quellen*, all resources, only "Frankfurter Allgemeine Zeitung" was used.

Differences from these experiments are depicted in Figures 2.C1 and 2.C2.

Most strikingly, even with the restrictions, the counts of uncertainty and economic policy articles in the Frankfurter Allgemeine Zeitung are *higher* than in our original text mining exercise in 2021, although their short-run pattern remains very similar. The count of all articles obtained from our search differs relatively little. On the other hand, we do not observe such large discrepancies in the case of Handelsblatt, where the data from the 2023 data collection matched those collected in 2021 (Figure 2.C1). This result shows that the collection of data from the FAZ online archive does not produce consistent results over time.

Figure 3.C1: Raw Counts for Germany Retrieved in 2021 and 2023, With and Without Restrictions

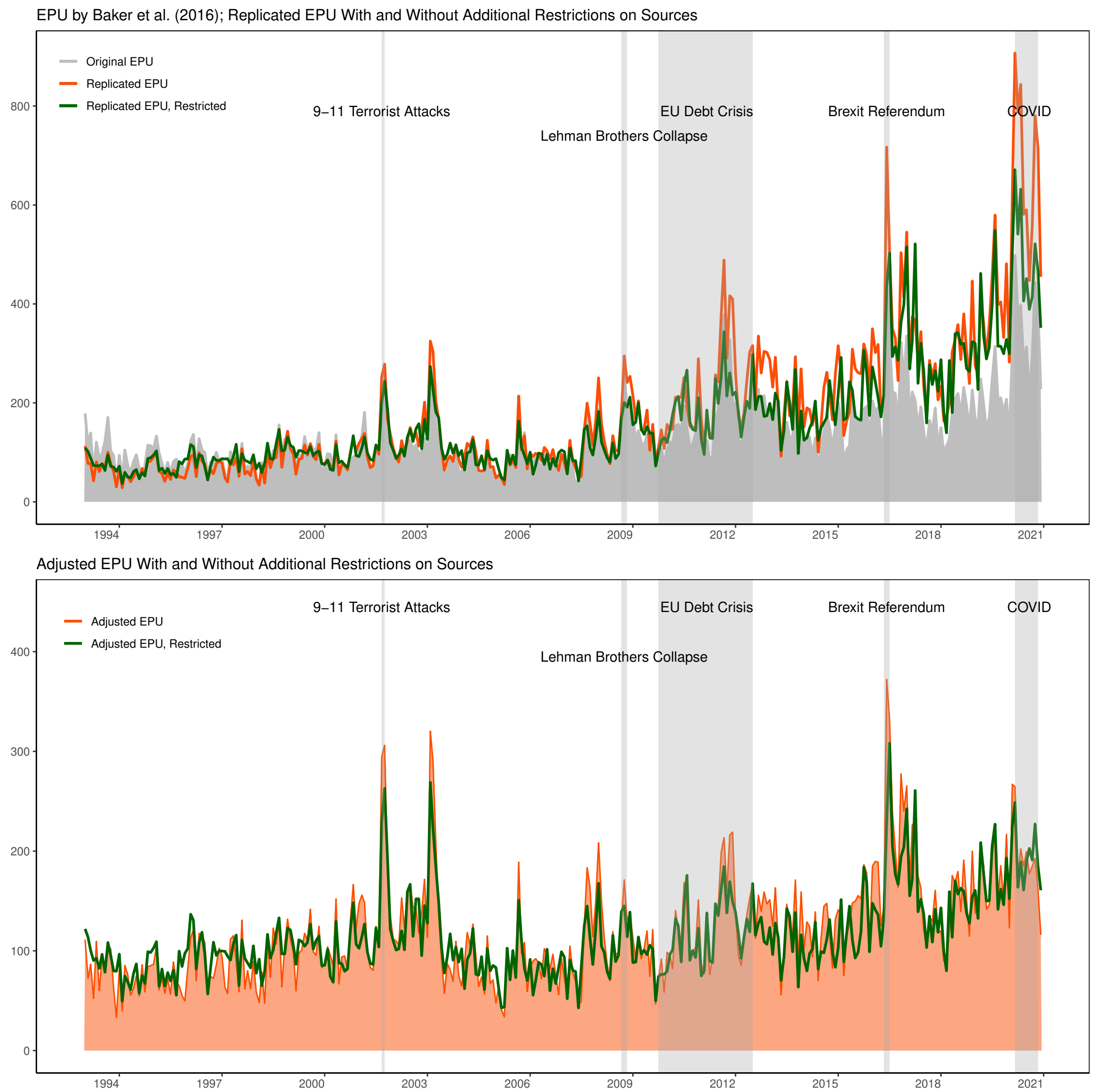


Note: Search in 2021 (baseline) - black, Search in 2023 - red, Search in 2023 with restrictions (Handelsblatt, print edition, Frankfurter Allgemeine Zeitung - articles appearing only in FAZ and not in other resources included in the archive) - blue. Query 1: economic + policy + uncertainty-related articles. Query 2: economic + policy articles. Query 3: All articles.

Despite these differences in FAZ counts, the replicated EPU index did not change with the data collected in 2023. This was due to the normalization of the overall index to the pre-2009 means and standard deviations. The restrictions (the print edition of Handelsblatt and articles published in the Frankfurter Allgemeine Zeitung) leads to an index that resembles the original EPU by Baker et al. (2016) more than our replication without those additional restrictions (Figure C2). However, the replicated index has remained higher in recent years than the index by Baker et al. (2016). On the other hand, the dynamics of the adjusted EPU index normalized by economic-policy articles is relatively robust to additional restrictions on resources (Figure 2.C2, lower panel).

Overall, this experiment confirms that the adjusted EPU is more robust to various changes in search specifications than the baseline EPU.

Figure 3.C2: Germany - EPU Indices, Sensitivity Analysis



Note: A comparison of the EPU by Baker et al. (2016) (gray) with the replicated EPU (orange) and replicated EPU with additional restrictions (Handelsblatt, print edition, Frankfurter Allgemeine Zeitung - articles appearing only in FAZ and not in other resources included in the archive (green)). The bottom plot shows a comparison of the adjusted EPU (orange) and the adjusted EPU with these restrictions (green).

References

Ahir, H., Bloom, N., & Furceri, D. (2018). The world uncertainty index. *Mimeo, Available at SSRN 3275033*.

Baker, S., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, *131*(4), 1593–1636.

Meinen, P., & Röhe, O. (2017). On measuring uncertainty and its impact on investment: Cross-country evidence from the euro area. *European Economic Review*, *92*, 161–179.

Chapter 4

Disentangling p-Hacking and Publication Bias

Nino Buliskeria

Abstract

This study differentiates p-hacking from publication bias by examining biases resulting from selective reporting within studies versus selective publication of entire studies. Analyzing a dataset of 400 meta-studies, which covers nearly 200,000 estimates from approximately 19,000 individual studies in economics and related social sciences, I observe a notably higher incidence of p-hacking compared to selective publication. Using various meta-regression methods, I find that selective reporting within studies is more prevalent than publication bias arising from selection among studies. This finding underscores the considerable influence of practices such as p-hacking and method-searching, suggesting that they contribute significantly to selection bias in the economic literature and could affect the perceived reliability of published findings.

JEL Codes: A11, C13, C40

Keywords: selective reporting, publication bias, p-hacking

Acknowledgment: This work was supported by the Charles University Research Center program No. UNCE/HUM/035. This manuscript is part of a project that has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 870245. I thank Jaromir Baxa, Pedro Bom, Abel Brodeur, Chishio Furukawa, Ali Elminejad, Alessandro Ferrari, Tomas Havranek, Bob Reed, Tom Stanley, Oliko Vardishvili and Ia Vardishvili for their helpful comments and suggestions. I also thank the participants of the MAER-Net Colloquium at the Universitat de Les Illes Balears and the research seminar at the University of California, Irvine, for their helpful suggestions and comments. The responsibility for all remaining errors and omissions rests solely with me.

4.1 Introduction

Selective reporting of empirical results can distort our understanding of how robust documented regularities are and give a false impression of their generalizability. Since the early 1980s, the critical examination of empirical research, initiated by Edward Leamer, has catalyzed what is now known as the credibility revolution in economics. This movement has strongly emphasized the importance of meta-research and the replicability of published work.[1] The credibility of empirical research is the cornerstone of scientific progress, yet it remains vulnerable to the influences of p-hacking and publication biases.

Publication bias arises when editorial teams and reviewers prefer studies that demonstrate statistically significant results. Meanwhile, the perception that publication bias is prevalent can lead researchers to abandon studies with unexpected or unpromising results, exacerbating publication bias. On the other hand, p-hacking involves various tactics researchers use, sometimes unintentionally, to achieve more favorable p-values, including "specification search," "p-hacking," or "data dredging" (Brodeur et al., 2020, 2023; Lang, 2023; Mathur, 2022). These tactics can include collecting data until the results appear significant, adjusting econometric models, or setting specific sample criteria to reach desired levels of statistical significance. The urge to engage in p-hacking can come from the perceived importance of statistical significance for the probability of publication (Andrews and Kasy, 2019).

[1] This wave of change has influenced research beyond economics to address what is commonly referred to as the "replication crisis" (Camerer et al., 2018), affecting fields such as medicine and epidemiology with John P. A. Ioannidis at the forefront (Begley & Ioannidis, 2015; Ioannidis, 2005; Ioannidis et al., 2017), as well as psychology and social sciences. An expanding body of work explores the issues of potential publication biases within economics and various other fields (Andrews & Kasy, 2019; Ashenfelter et al., 1999; Bruns et al., 2019; De Long & Lang, 1992; Doucouliagos & Stanley, 2013; Ferraro & Shukla, 2020; Furukawa, 2019; Havránek, 2015; Ioannidis, 2005; Ioannidis et al., 2017; Leamer, 1983; Miguel et al., 2014; Stanley, 2005, 2008).

Meta-regression analyzes are widely used to assess the extent of selection bias and to estimate the true population mean, often referred to as "mean-beyond bias" in the literature.[2] These methods generally conceptualize publication bias as a filtering mechanism that impacts a collection of point estimates, which are presumed to be unbiased estimators of the true population effects.[3] However, this foundational assumption is notably vulnerable to selection bias caused by p-hacking, as noted by Irsova, Bom, et al. (2023). The practice of p-hacking, which involves actively seeking specifications that yield significant results, significantly undermines this crucial assumption. p-Hacking can potentially modify both the effect size and the standard error, resulting in spurious precision (Irsova, Doucouliagos, et al., 2023). Although theoretically the difference between publication bias and p-hacking is distinct, they are observationally equivalent. This observational equivalence challenges the classical metaregression analysis, since it cannot differentiate between the two. The key presumption underpinning the metaregression analysis is the statistical unbiasedness of point estimates and standard errors. The literature acknowledges the consequences of published p-hacked coefficients, but the extent and measurement of p-hacking remain ambiguous. While Brodeur et al. (2023) argue for the dominant role of p-hacking in publication bias, Lang (2023) finds

[2] There are two primary categories of statistical techniques for detecting and adjusting for publication bias. The first encompasses traditional methods, such as funnel plot analysis and the "incidental" truncation theorem outlined in Greene (1990), which are based on the assumption that results that are statistically significant and align with the desired hypotheses are more likely to be published (Bom & Rachinger, 2019; Duval & Tweedie, 2000; Egger et al., 1997; Furukawa, 2019; Ioannidis et al., 2017; Stanley, 2008; Stanley & Doucouliagos, 2012, 2014). The second category involves modeling the relationship between a study's likelihood of being published and its p-value, thereby defining a parametric structure for the distribution of population effects before selection. Models in this category, such as two-parameter selection models, often show a bias toward the publication of positive results (Andrews & Kasy, 2019; Hedges, 1984, 1992; Iyengar & Greenhouse, 1988; van Aert & Van Assen, 2021; van Assen et al., 2015; Vevea & Hedges, 1995).

[3] Publication bias is traditionally viewed as a sieve influencing the research submission and publication process, involving decisions made by researchers, journal editors, and peer reviewers. This bias, resulting from study-level selection, is termed "selection across studies" (SAS) by Mathur (2022).

limited evidence for this phenomenon.

The selective publication of significant and large results causes a truncation in the distribution of observed coefficient estimates. As shown in Greene (1990) and elaborated in more detail in Section 2, this truncation leads to a correlation between the observed coefficients ($coef_i$) and their standard errors (SE_i). Through meta-regression analysis, the strength of this correlation (β) is estimated, serving as an indicator of the extent of selection bias [4]:

$$coef_{ij} = \alpha + \beta \cdot SE_{ij} + [\epsilon_i + u_{ij}]$$

Meanwhile, the estimated intercept (α) from this analysis measures the *true mean beyond bias*, adjusted to account for selection bias.

I define p-hacking as the biased selection of the reported point estimate and the standard error pairs within the study, usually by the authors. By controlling for study-specific characteristics, I isolate the bias arising from p-hacking:

$$\text{FE:} \quad coef_{ij} - \overline{coef}_j = \beta^{FE}(SE_{ij} - \overline{SE}_j) + u_{ij}$$

Employing fixed-effects analysis enables the comparison of estimates while canceling the impact of study heterogeneity. By doing so, it becomes possible to identify variations in selection bias that are specifically attributable to variations in within-study coefficient selection, known as p-hacking.

Next, to identify the selection bias between studies, I apply the between-effect estimation on means of coefficient and standard error pairs for each study.

[4] Equations in this section are presented for clarity. Please refer to the section 4.2 and 4.4 for further details on theory and application

$$\text{BE:} \quad \overline{coef}_j = \alpha + \beta^{BE}\overline{SE}_j + u_j$$

This approach measures the magnitude of selection across studies, the selection type that does not introduce bias in point estimates.

The focus is on five key bias correction estimators: the Egger equation, quantile regression, the Precision-Effect Estimate with Standard Errors (PEESE), the combined PET-PEESE approach, and the endogenous kink model (EK). My objective is to evaluate the extent of selection bias arising from within-study manipulations versus across-study biases. To control for the impressions in meta-regressions coming from the potential presence of the p-hacking, I adopt the instrumental variable approach detailed by Irsova, Bom, et al. (2023) for each estimation technique.

This study also stands out due to its extensive and unique data, encompassing 400 meta-studies that include nearly 200,000 estimates derived from about 19,000 distinct studies. The data for these 400 meta-studies was obtained from the authors when not available in online journal directories (see the Appendix for the list of meta-studies). Next, I combined 412 distinct data sets, synchronizing meta-study and study-level journal titles, and identified the status (working or published article) of the study at the time of meta-study publication (in the journal of online series). Finally, I merged it with a dataset of the SCImago Science Journal Rank on the journal research areas classification to identify the field of meta-study. I base my analysis on this unique and comprehensive data set, which provides a robust platform to examine how biases manifest in published research.

In my analysis of 412 meta-studies, I implement two sets of five key bias correction estimators, each employing an instrumental variable approach. I perform a fixed effect analysis to estimate the extent of bias attributable to p-hacking. Whereas I use a between-effect approach to assess the degree of selection bias arising from selection across studies. This dual approach results in 412 bias estimates for each between- and fixed-effect estimation, which is 4120 regressions in total. To analyze these findings further, I employ a ratio to compare the between- and fixed-effect estimates. Theoretically, as suggested by (Angrist & Pischke, 2009), this ratio, in absolute terms, should be less than one due to the attenuation bias inherent in fixed-effect estimation. However, the median ratio consistently exceeds 1 in all the methodological specifications in my study. My analysis reveals that p-hacking is more prevalent compared to selection between studies, aligned with Brodeur et al. (2023). The results consistently show a higher level of bias in fixed-effect analyzes, indicating a substantial contribution of practices such as p-hacking to selection bias in the economic literature. This outcome indicates a substantial contribution of practices such as p-hacking and method searching to selection bias in the economic literature, leading to a potentially inaccurate perception of robustness in published findings.

The paper is structured as follows. Section 2 discusses the theoretical foundations of bias detection techniques. Section 3 examines the data. Section 4 introduces the empirical techniques and discusses the results. The final section summarizes the findings and implications.

4.2 Theoretical foundation

According to the traditional definition of publication bias, the research results are selected for publication according to their direction and statistical significance. Although this selective publication process partially truncates the overall distribution of reported results in the literature, in most meta-literature, it is assumed that the chosen results are unbiased estimations of the true underlying effect relative to their respective population. Therefore, most publication bias detection and correction techniques rely on this assumption.

However, Brodeur et al. (2016, 2023), Irsova, Bom, et al. (2023), and Mathur (2022) point to the possible manipulation of design choices that influence standard errors and coefficients to increase the probability of publication. In observational research, the derivation of the standard error is subject to various complicated design choices and with different choices of model specification, both effect size and standard error change. Since both jointly contribute to statistical significance, design choices aiming at increased significance can cause spurious precision and violate the core assumption of unbiased estimates. Violation of this assumption renders meta-regression analysis incapable of correcting for publication bias. Irsova, Bom, et al. (2023) state that in this case *"the simple unweighted mean is often the best, but still no good"*. Although the literature agrees on the potential consequences of published p-hacked coefficients, the significance of the matter or the way to measure it is ambiguous.

In this section, I discuss the theoretical foundation of metaregression analysis (MRA) and the importance of the underlying assumption of unbiasedness of the point estimate. First, I present the theory behind identifying the true mean beyond bias, then I discuss estimation techniques when the assumption of unbiasdness holds and when it does not. Finally, I show my identification strategy to measure the magnitude of p-hacking compared to selection across studies. For simplicity, I consider a strict rule of selection bias where coefficient estimates that do not satisfy the significance requirement do not get published.[5]

[5] Andrews and Kasy (2019) conclude that studies with a 5% significance level have 30 times higher chances of being published than insignificant results. They estimate the publication probabilities based on replication and meta-analysis approach and provide strong evidence

Consider a series of studies that estimate the effect size of a specific research question[6]. Each study uses different sample specifications and techniques to achieve unbiased estimates. In this scenario, the study i estimates an unbiased effect $\hat{\alpha}_i$ expected to be close to the actual true effect, denoted as α_i. The discrepancies between these estimated and true effect sizes result from sampling errors and measurement inaccuracies; therefore $\hat{\alpha}_i$ can be expressed as true effect α_i plus error.

$$\hat{\alpha}_i = \alpha_i + u_i \tag{4.1}$$

Following the Central Limit Theorem[7], the distribution of the estimated effect size is:

$$\hat{\alpha}_i \sim N(\alpha_i, \sigma_i^2) \tag{4.2}$$

Furthermore, I follow the conventional assumption that the true effect size follows a normal distribution with a Θ mean and $\aleph^2$ variance[8]:

$$\alpha_i \sim N(\Theta, \aleph^2) \tag{4.3}$$

This assumption is widely assumed in the meta-research and implies that as the number of studies increases, the distribution of their estimated effects, even with sampling and measurement errors, tends to follow a normal distribution

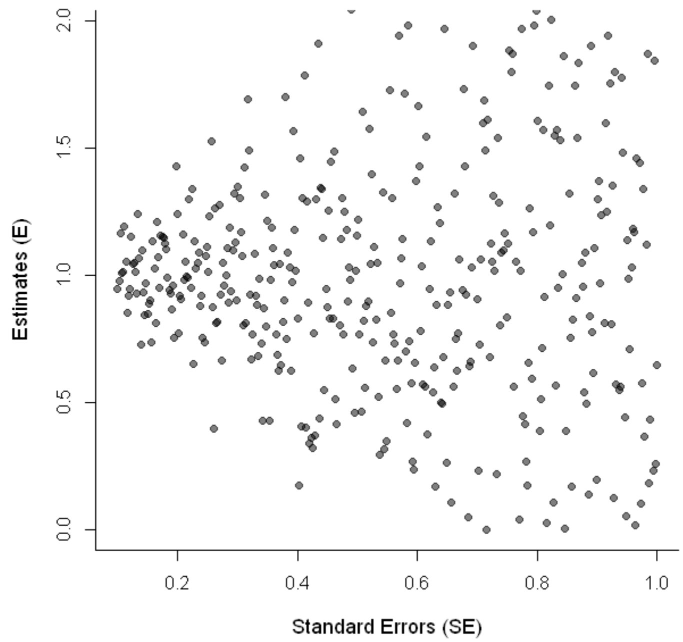
of selectivity based on significance.

[6]Similarly to Jackson and Mackevicius (2023), I start by building the discussion from the point estimates in each study.

[7]The central limit theorem (CLT) states that the average from a random sample for any population (with finite variance), when standardized, has an asymptotic standard normal distribution (Wooldridge, 2002). Here, estimates have not been standardized; therefore, they are normally distributed with mean and variance.

[8]Normality assumption is not essential, here I rather adopt it for ease of demonstration. Most popular meta-analysis techniques assume that the true coefficient estimate, α_i, is statistically independent of its standard error, $sigma_i$, in the population, this easily follows if one assumes that both α_i and $\hat{\alpha}_i$ have the same constant mean Θ across the published studies within a research area. One of the straightforward and most frequently assumed distributions that satisfies the aforementioned requirements in normal distribution

Figure 4.1: A normally distributed population



centered around the true effect:

$$\hat{\alpha}_i \sim N(\Theta, \sigma_i^2 + \aleph^2) \tag{4.4}$$

Therefore:

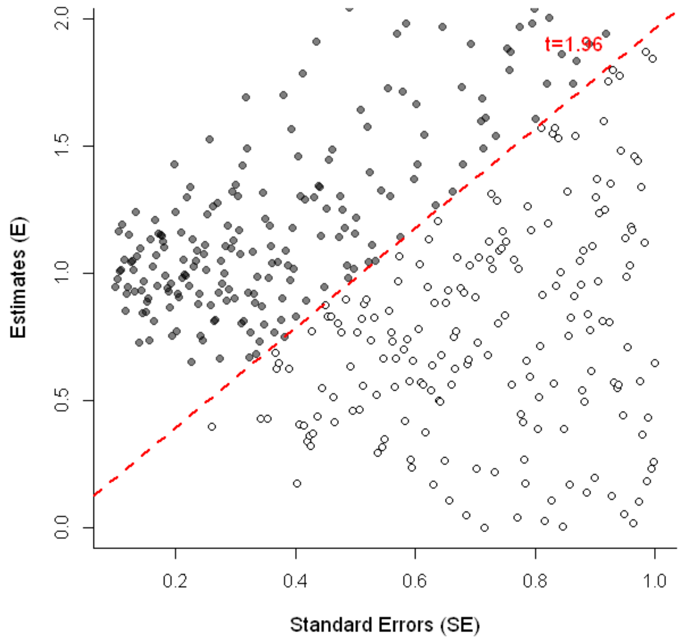
$$\hat{\alpha}_i = \Theta + u_i \tag{4.5}$$

where $u_i \sim iid\ N(0, \sigma_u)$ is noise due to the sampling or measurement error, as shown in figure 4.1.

Let us now consider the classical definition of publication bias. Articles are selected for publication on the basis of their coefficient estimate and significance. This selection criterion leads to missing observations, conditional on coefficient size $\hat{\alpha}_i|\hat{\alpha}_i > a$, and significance level $\hat{\alpha}_i|t_{\hat{\alpha}_i} > c$, where a and c are some constant thresholds. This truncation then creates publication bias (see Figure 4.2.

The preferences for the coefficient estimate can be in its direction, magnitude,

Figure 4.2: Distribution truncated based on significance, no evidence of p-hacking



or proximity to conventional beliefs. Let me assume that coefficients larger than some constant a are preferred for simplicity. In the case of truncation based on the coefficient value, only $\hat{\alpha} > a$ are observed; therefore, Equation (4) becomes $\hat{\alpha}_i|\hat{\alpha}_i > a = \hat{\alpha}_i + u|\alpha_i > a$, where $E[u|\alpha_i > a] \neq 0$, and based on (3), to deduct the population mean of true effect Θ bias introduced by truncation needs to be studied:

$$\begin{aligned} E[\hat{\alpha}_i|\hat{\alpha}_i > a] &= \Theta + E[u_i|\hat{\alpha}_i > a] \\ &= \Theta + E[u_i|u_i > a - \Theta] \end{aligned} \tag{4.6}$$

where σ_i is estimated standard error from study i, $E[u_i|u_i > a - \Theta] = \sigma_i\phi(\kappa)/[1 - \Phi(\kappa)]$ and $\kappa = (a - \hat{\alpha}_i)/\sigma_i$ (see Greene, 1990, Theorem 2.2; Wooldridge, 2002; Johnson et al., 1995). Therefore, the conditional expectation of the error term u_i is the product of the estimated standard error and the inverse Mill ratio, which is the ratio of the probability density function to the complementary cumulative distribution function.

$$E[\hat{\alpha}_i|\hat{\alpha}_i > a] = \Theta + \sigma_i \frac{\phi(\kappa)}{[1-\Phi(\kappa)]}$$

Therefore, the meta-regression is as follows:

$$E[\hat{\alpha}_i|\hat{\alpha}_i > a] = \Theta + \sigma_i \lambda(\kappa) \tag{4.7}$$

Thus, $\lambda(\kappa)$ represents the inverse Mills ratio. If the truncation of the estimated coefficient is above $\alpha_i|\alpha_i < a$, then $\lambda(\kappa) = -\phi(\kappa)/\Phi(\kappa)$.

The truncation of the significance is similar to the truncation of the coefficient estimate, also referred to as incidental truncation [9]. Now, I look at $E[\hat{\alpha}_i|\hat{\alpha}_i/\sigma_i > c]$, where c is the critical value at which the coefficient estimate becomes significant (frequently taken at $c = 1.96$ for the significance level of 5%. To apply the same logic here, it is important to look at the distribution of $\hat{\alpha}_i$ and $\hat{\alpha}_i/\sigma_i$. As discussed above, using CLT, $\alpha_i \sim N(\alpha_i, \sigma_i)$, therefore,

$$\hat{\alpha}_i/\sigma_i \sim N(\alpha_i/\sigma_i, 1) \tag{4.8}$$

with bivariate normal joint distribution. Therefore, following Theorem 2.5 in Greene (1990)[10]

$$E[\hat{\alpha}_i|\hat{t} > c] = \Theta + \sigma_i \rho \frac{\phi(\kappa_{it})}{1-\Phi(\kappa_{it})} \tag{4.9}$$

where $\hat{t} = \hat{\alpha}_i/\sigma_i$, $\kappa_{\hat{t}} = (c - \hat{t})/\sigma_{i\hat{t}}$, and $\rho = corr(\alpha_i, \hat{t}) = 1$. However, considering Equation (7), $\rho = 1$ and $\kappa_{\hat{t}} = (c - \hat{\alpha}_i/\sigma_i)$ result in the same form of meta-

[9] see in Greene (1990), Theorem 2.5; see Heckman (1979)

[10] first moment of incidental truncation is $\alpha + \rho\sigma\lambda(\kappa_t)$, where ρ is correlation coefficient. However, here $corr(\alpha, \alpha/se) = 1$

regression as shown in Equation (4.7):

$$E[\hat{\alpha}_i|\hat{t} > c] = \Theta + \sigma_i\lambda(\kappa) \tag{4.10}$$

To estimate Θ, often referred to as mean beyond bias in the meta-literature, one needs to consistently estimate $\lambda(\kappa)$ first. However, in both cases, the conditional mean is a complex non-linear function of the truncation value σ, α, and λ, while the second term of the equation, $\lambda(\kappa)$, is not constant with respect to α and σ_i. To express the complexity of this term, I take the derivative of $E[\hat{\alpha}|truncation]$ with respect to σ, I drop i for simplicity, however, it is assumed as before:

$$\begin{aligned}\partial E[\hat{\alpha}|truncation]/\partial\sigma &= \lambda(\kappa) + \sigma\partial\lambda(\kappa)/\partial\sigma \\ &= \lambda(\kappa) + \sigma\partial\lambda(\kappa)/\partial\kappa \cdot (\partial\kappa/\partial\sigma)\end{aligned}$$

where:

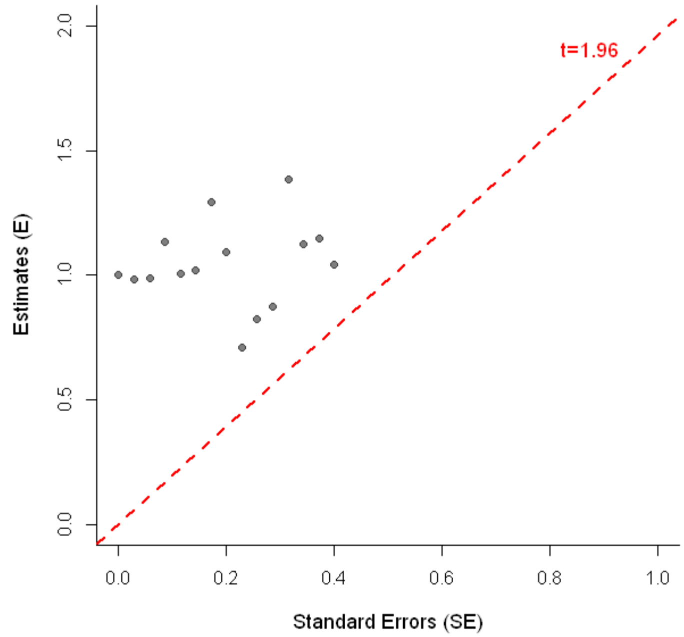
$$\begin{aligned}\partial\lambda(\kappa)/\partial\kappa &= \frac{\phi'(\kappa)[1-\Phi(\kappa)] + \phi(\kappa)\Phi'(\kappa)}{[1-\Phi(\kappa)]^2} \\ &= \frac{\phi'(\kappa)[1-\Phi(\kappa)] + \phi(\kappa)^2}{[1-\Phi(\kappa)]^2} \\ &= -\frac{\phi(\kappa)\cdot\kappa}{[1-\Phi(\kappa)]} + \frac{\phi(\kappa)^2}{[1-\Phi(\kappa)]^2} \\ &= \lambda^2(\kappa) - \kappa\cdot\lambda(\kappa)\end{aligned} \tag{4.11}$$

as also shown in Heckman (1979). Therefore, after plugging in this derivative and derivative of κ with respect to σ, I have:

$$\partial E[\hat{\alpha}|truncation]/\partial\sigma = \lambda(\kappa) + \frac{\alpha}{\sigma}[\lambda^2(\kappa) - \kappa\cdot\lambda(\kappa)]$$

Equations (4.7) and (4.10) is the statistical foundation of the meta-regression model for bias detection, and Equation (4.2) shows the relation between the expected mean of the truncated estimates and their standard error.

Figure 4.3: Study A, no evidence of p-hacking, simulation



In this figure, I present the example of Study A, where there is no evidence of p-hacking since the $t = 1.96$ is not a binding constraint and all results naturally fell on the left side of the line. Hypothetically speaking, study with all naturally significant results would suffer from no selection within study.

A common approach in the literature to detect bias is to employ a truncated regression model (see Equation 4.7), also known as the Egger's equation.[11]

$$\hat{\alpha}_i = \alpha + \lambda\sigma_i + \epsilon_i \tag{4.12}$$

This model aims to determine the presence of bias and to deduce the mean of the target coefficient adjusted for bias from the observed truncated distribution. To alleviate heteroskedasticity, this equation is estimated using weighted least

[11] Frequently written as $coef_i = \alpha + \beta SE_i + u_i$ in the literature, where $coef$ is a coefficient estimate, and SE stands for the standard error. However, here I opted to follow the initial notation.

squares, weighted by precision, where t_i is the reported t statistics.

$$t_i = \lambda + \alpha(1/\sigma_i) + u_i \tag{4.13}$$

The test $H_0 : \alpha = 0$ is known as the *Precision Effect Test* (PET) in the literature and provides a valid test to determine whether there is a nonzero empirical effect after correcting for publication bias (Stanley, 2008). However, Egger's equation struggles to correctly identify the true mean α in cases of nonzero effect size. This is intuitive after comparing Equation (4.12) with (4.7), since Egger's regression estimates λ as a constant, while it is a complex function $\lambda(\kappa_i)$ of $\hat{\alpha}$, σ, and the truncation value c, see Equations 4.11 & 4.2. Therefore, Egger's equation can correctly measure the extent of bias and identify the mean beyond bias if the underlying empirical effect is zero ($\alpha = 0$), granting the second quadratic term of Equation 4.2 obsolete - $\partial E[\hat{\alpha}|truncation]/\partial\sigma = \lambda(\kappa)$ and leading to a linear relation between the expected effect and the standard error. However, nonzero cases remain challenging for PET approach.

The literature strand successfully addresses this issue, using different weighting and Taylor approximation techniques to appeal to the second-order structure of the equation 4.2 (Bom & Rachinger, 2019; Havránek, 2010; Ioannidis et al., 2017; Stanley, Doucouliagos, et al., 2007; Stanley & Doucouliagos, 2012, 2014). Stanley and Doucouliagos (2014) recommends adopting a quadratic approximation approach, using the weighted least squares (WLS) estimate of the mean beyond bias α.

$$\hat{\alpha}_i = \alpha + \lambda\sigma_i^2 + \epsilon_i \qquad \text{or} \tag{4.14}$$

$$t_i = \lambda\sigma_i + \alpha(1/\sigma_i) + u_i \tag{4.15}$$

where meta-regression (6) is using $1/\sigma_i$ or $1/\sigma_i^2$ as the weights for the weighted least squared estimation. In the literature, the estimated α is called the *precision effect estimate with standard error* (PEESE) (Havránek, 2010; Stanley, Doucouliagos, et al., 2007; Stanley & Doucouliagos, 2012). Stanley and Doucouliagos (2014) suggest employing the PEESE estimator, Equation 4.15 only when there is evidence of a nonzero effect (i.e., rejecting $H_0 : \alpha = 0$), and the PET estimator, Equation (4.12) when accepting $H_0 : \alpha = 0$, which results in the PET-PEESE estimator.

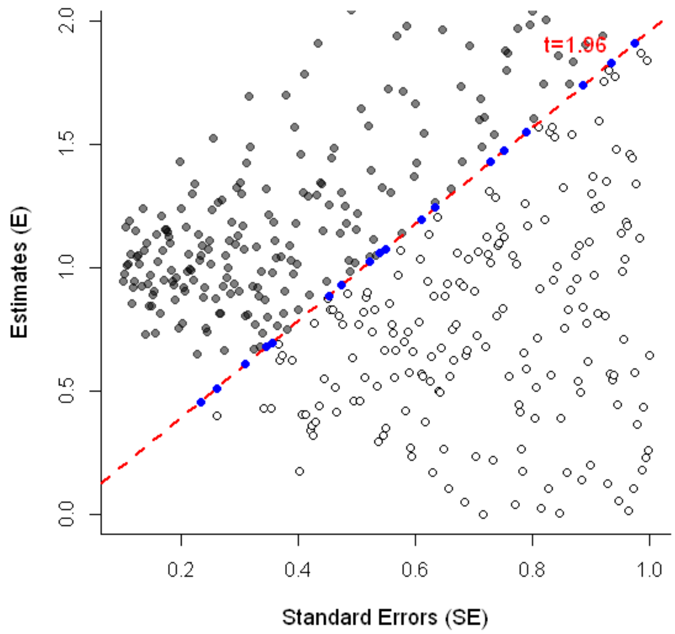
Bom and Rachinger (2019) improve PET-PEESE by proposing the endogenous kink (EK) metaregression model, offering a novel approach to correct for publication bias. A distinctive feature of the EK model is the presence of a 'kink' at a specific cut-off value of the standard error. Below this cutoff point, publication selection is deemed unlikely. Therefore, the EK model approximates $\lambda(\kappa)$ using a piecewise linear metaregression:

$$\hat{\alpha}_i = \alpha + \delta[\sigma_i - a]I_{\sigma_i \geq a} + \epsilon_i \tag{4.16}$$

where, $\mathrm{I}_{\sigma_i \geq a}$ is an indicator function that takes the value of one if σ_i is greater than or equal to a, and zero otherwise. Similarly to PET, PET-PEESE, the EK model addresses the heteroskedasticity of $\hat{\alpha}_i$ by dividing each term by $1/\sigma_i$. The EK model endogenously determines the cutoff value based on a preliminary estimate of the true effect and a predefined threshold of statistical significance.

However, the literature is silent on bias detection and correction techniques in the case of spurious precision. All of these methods are based on the implicit belief that the reported nominal precision accurately reflects the true underlying precision. Irsova, Bom, et al. (2023) show that the simple unweighted mean can often outperform complex estimators even when the share of reported spurious

Figure 4.4: Distribution truncated based on significance, with the evidence of p-hacking

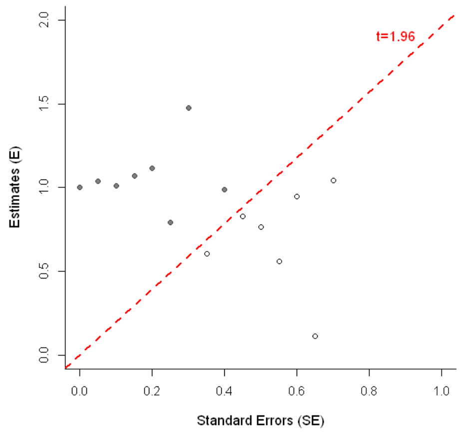


precision is very low in the meta-sample. Thus, they argue that when reported standard errors are manipulated conventional solutions, designed to address publication bias, lead further away from true mean. In observational studies, calculating the standard error is often a crucial part of the research process. The process is complex, and varying the computation of confidence intervals will lead the researcher to report different levels of precision for the same estimated effect size, potentially leading to misleading results and spurious precision.

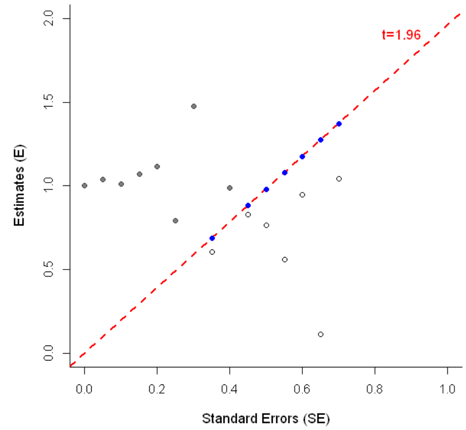
Figure 4.4 illustrates the distributional consequences of various actions such as cheating, clustering, correcting for heteroskedasticity, and addressing non-stationarity, all undertaken to obtain statistically significant results without a solid theoretical or reasonable basis.

The action of p-hacking can take place in the cases in which researchers increase their selection efforts towards larger estimates in response to noise

Figure 4.5: Study B, evidence of p-hacking, simulation



(a) Reporting only significant coefficients



(b) Manipulating model specifications to achieve significance

(larger standard errors) in their data or methods leading to imprecision and insignificance. With these manipulations, the most precise estimates stay close to the true effect. Therefore, inverse-variance weighting plays a role in reducing bias and improving the efficiency of the aggregated estimate. In contrast, researchers may also achieve statistical significance by reducing the standard error. However, in this case, there is no bias in the reported effect sizes; both the filled and hollow circles would represent identical effect sizes, with the only difference being in precision. The straightforward unweighted average of these estimates is unbiased, but applying inverse-variance weighting would introduce an additional downward bias.

Figure 4.5 presents the two scenarios of p-hacking, in (a) the author, after conducting a number of estimations and robustness checks, reports only significant results; while (b) shows the case where the author adjusts the specifications of the exercise to achieve significance at the 5% level. The presence of p-hacking introduces the spurious relation between coefficient estimate and standard error, undermining the effectiveness of techniques for detecting and correcting bias.

To control for the spurious relation between estimated coefficients and their standard errors, I use the Meta-analysis Instrumental Variable Estimator (MAIVE) model, where I instrument standard error with the inverse of the sample size[12] , i.e., replace the reported standard error with the portion of the error that can be explained by the sample size. Since in most contexts, the sample size is more difficult to increase than the standard error, the adjusted measure potentially captures the underlying precision better.

$$\sigma_i^2 = \phi_0 + \phi_1(1/n_i) + \nu_i \tag{4.17}$$

$$\sigma_i = \sqrt{\phi_0 + \phi_1(1/n_i) + \nu_i} \tag{4.18}$$

where Equation 4.17 is the first stage regression for the PEESE and Equation 4.18 for the PET estimation techniques; σ_i is the standard error of the effect size as reported in a primary study; ψ_o is the constant term, n_i denotes the sample size of the primary study, and ν_i is an error term. The error term of the first stage regression, ν_i, absorbs the spurious components of the reported standard error that are attributable to p-hacking. Irsova, Bom, et al. (2023) simulate a realistic p-hacking scenario, suggesting that the MAIVE version of PET-PEESE, without additional inverse variance weights, is more resistant to spurious precision than other existing methods.

The primary objective of the paper is to assess the degree of selection bias resulting from selection within studies (p-hacking) compared to selection across studies (publication bias, file drawer effect). To this end, I plan to conduct my analysis using the instrumental approach as outlined by Irsova, Bom, et al. (2023). My focus is on the five bias correction estimators mentioned above: linear meta-regression, quantile regression, precision effect estimate with

[12]here I follow Irsova, Bom, et al. (2023), who offer the MAIVE technique to control for the spurious relation

standard errors (PEESE), PET-PEESE, and the Endogenous Kink (EK) model. I begin with the linear Egger equation. This is in line with the consensus in the literature that Egger's method is a reliable tool for detecting the presence of selection bias.

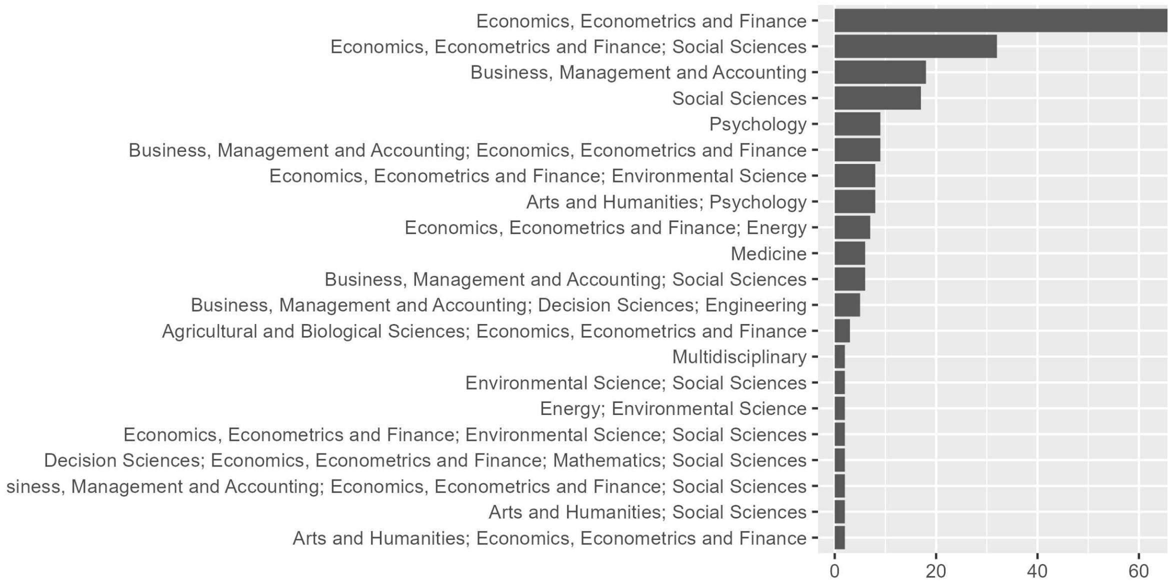
4.3 Data description

This thesis investigates the sources of selective reporting by examining within-study selection and across-study selection in 400 meta-analyzes, encompassing more than 20,000 studies and 200,000 coefficient estimates from various fields of social sciences, mainly economics. The meta-data set is a collection of data from previous and newly published meta-studies. It contains meta-study and study-level information on authors, titles, publication years, and journals. In addition, the metadata contain coefficient estimates, their respective standard errors, and the sample size of each estimation technique from each study.

Many meta-studies examine closely related questions, often analyzing multiple coefficients of interest corresponding to different true means. In such cases, data from these meta-studies are classified into separate categories and included in the analysis as distinct entities at the meta-level. For example, Balima et al. (2020) analyze the impact of publication selection bias on the macroeconomic effects of inflation targeting. They consider a variety of macroeconomic indicators, including the effects of inflation targeting on inflation, GDP, interest rate volatility, inflation volatility, growth volatility, exchange rate volatility, and deficit. I retain the categorization of Balima et al. (2020)'s data, assigning a unique meta-ID to each category and treating them as independent meta-studies.

An analysis of the journals where these meta-studies have been published reveals a concentration in various economic disciplines. Figure 4.6 presents

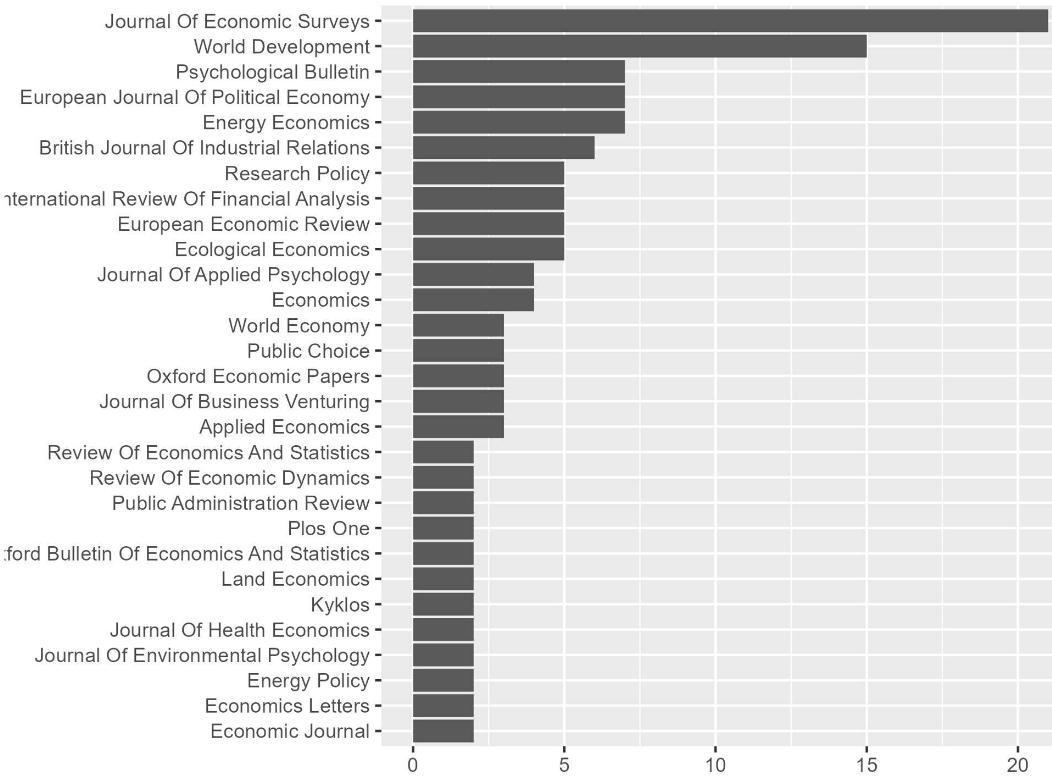
Figure 4.6: The meta-analyses published in journals areas



Note: Journal research areas classification according to the SCImago Science Journal Rank (SJR), https://www.scimagojr.com/journalrank.php?area=2000

this distribution, categorizing research areas according to the SCImago Journal Rank (SJR). It also shows the frequency of publications within each research area. In particular, the fields of *Economics, Econometrics, and Finance*, with more than 100 meta-analyses, are also mentioned as part of the majority of other area classifications. The repeated appearance of the *Economics, Econometrics, and Finance* classification throughout Figure 4.6 indicates that our data set mainly comprises estimates drawn from economic research.

Figure 4.7: Meta-analyses per journal

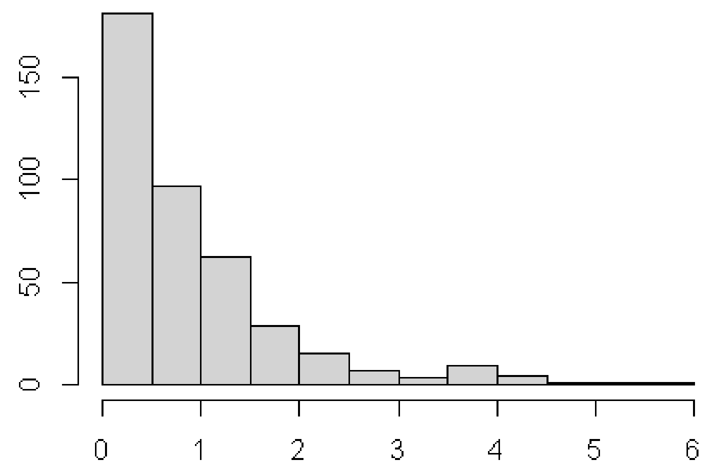


Note: a list of journals that are the most frequent publishers of meta-studies included in the dataset.

Figure 4.6 shows the journals that most frequently publish meta-analyses in the data. Not surprisingly, it reflects the picture that can be seen in Figure 4.6, where the most frequent research area is economics. In Figure 4.7, it is apparent that these meta-studies are published more frequently in economic outlets, sometimes psychology, or in interdisciplinary journals such as *Journal of Health Economics*. I present only those journals that have published meta-study in the sample at least twice; however, similarly to Figure 4.6, the economic journals are the majority of the journals, and social science and interdisciplinary journals are the second most frequent and rarely medicine.

To understand the extent of bias in the literature, I use Egger's regression $coef_{ij} = \alpha + \beta SE_{ij} + \epsilon_{ij}$, where $coef_{ij}$ & SE_{ij} is the estimated coefficient and

Figure 4.8: Distribution of Selectivity in Empirical Economics.



Note: Bias estimated from Egger's regression, $coef_i = \alpha + \beta SE_i + \epsilon_i$. The bias is considered *small to modest* if $|\beta < 1|$, *substantial* if $1 \leq |\beta| \leq 2$, and *severe* for $|\beta > 2|$. I find *substantial* selectivity across 91 different topics and *severe* in 44 topics in economics & social sciences. For 278 areas, bias falls in the little to modest category.

standard error pair j of study i, α is the mean beyond bias, β estimates the extent and existence of bias. I run this regression analysis separately on data from k meta-studies, obtaining the k number of β coefficients for each topic. Figure 4.8 shows the distribution of β_k on different topics. Doucouliagos and Stanley (2013) categorizes the biases in *little to modest* category if $|\beta < 1|$, *substantial* if $1 \leq |\beta| \leq 2$ and *severe* for $|\beta > 2|$. I find *substantial* selectivity across 91 different topics and *severe* in 44 topics in economics & social sciences. For 278 areas, bias falls into the little to modest category.

Finally, in Figure 4.9, I look at the distribution of t-statistics in published articles and show evidence of potential p-hacking, as discussed in Brodeur et al. (2023). I use the de-rounding technique and weight the z-statistics (measured as $coef_{ij}/SE_{ij}$) with the inverse of the number of tests present in each article and superimpose an Epanechnikov kernel density curve on the histogram. De-rounding does not change the shape of the distribution; it only

Figure 4.9: De-rounded & weighted distribution of z-statistics of published papers.

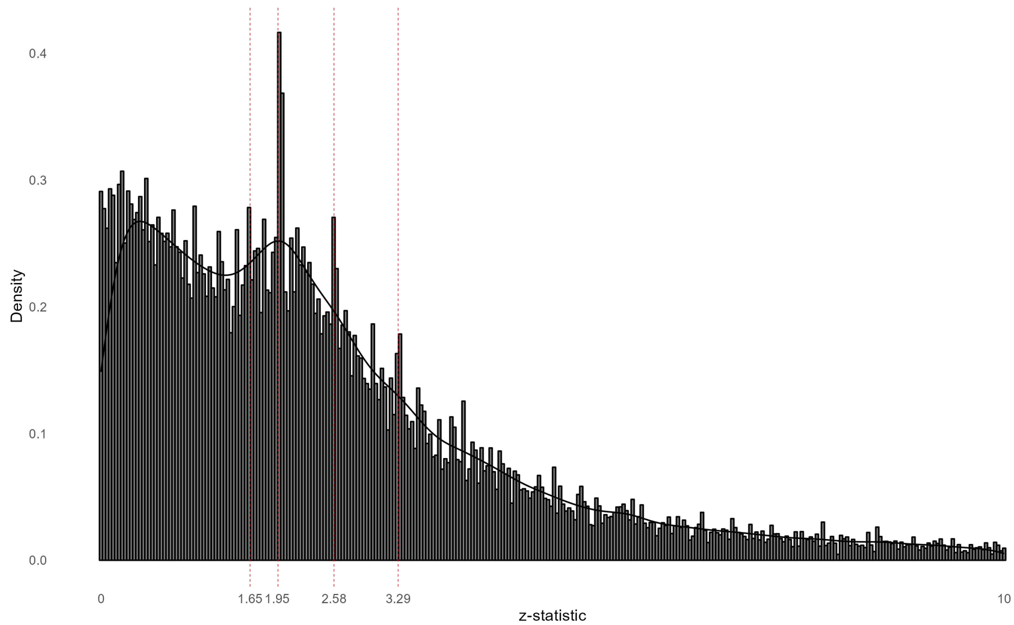


Figure is showing the distribution of z-statistics of coefficient estimates in the published papers. The distribution is de-rounded to control for the Note: The two-humped camel-shaped pattern, similar to Brodeur et al. (2020, 2023), is evident. I superimpose an Epanechnikov kernel density curve.

smooths potential discontinuities in histograms. Figure 4.9 presents the two-humped camel-shaped pattern, bunching at $z = 1.96$, indicating the existence of p-hacking. However, as pointed out in Kranz and Pütz (2021), this approach cannot explain the excess share of observed z-statistics near zero.

The observed distribution of z-statistics, even adjusted for rounding, consistently shows two distinct peaks, one at zero and one around $z = 2$, Figure 4.9. However, Kranz and Pütz (2021) point out that this second peak does not necessarily indicate p-hacking or publication bias. It could also be explained by a latent mixed distribution resulting from varying research objectives. For example, some studies could refine previous findings with significant effects, while others could be more exploratory, lacking a solid prior assumption of the actual effects being present. To demonstrate this numerically, Kranz and Pütz

(2021) consider 5,000 random samples from a combination of three Cauchy distributions, each with a scale parameter of 0.8: one distribution has a center at 0, representing exploratory research, while the other two, centered at -2 and 2, represent more focused research. They show that the resulting distribution of absolute z-statistics is very similar to the empirical distribution in the pooled data in Figure 4.9. This paper contributes to this discussion by analyzing similar questions based on metaregression analysis.

4.4 Estimation and results

There should be no correlation between estimates and standard errors if there is no publication bias, that is, selection within (SWS) or across studies (SAS). Therefore, for now I assume that any correlation between the coefficient $coef_{ij}$ and its standard error SE_{ij} indicates the existence of bias. Therefore, the correlation between $coef_{ij}$ and SE_{ij} within the study indicates bias from SWS, and the correlation between the mean study estimates indicates bias due to SAS[13]. I run 800 regressions to estimate bias coefficients for each research question and separately evaluate the extent of the selection of the results coming from the within-study and between-study variation.

I estimate the extent of selection for each meta-analysis k, study j, and estimate i, using the following meta-regression:

$$coef_{ij} = \alpha + \beta SE_{ij} + e_j + u_{ij} \tag{4.19}$$

Where $coef_{ij}$ is the coefficient estimate i of the study j; SE_{ij} is the corresponding standard error; e_j indicates characteristics specific to the study and u_{ij} is the error term. This regression cannot differentiate between the selection within-

[13] The caveat here is that coefficients within study are less likely to be independent, however when controlling for the fixed effects, in case of SWS, and taking mean estimates, in case of SAS, this issue should resolve.

and between-studies, however, it can serve as a benchmark for the comparison. Meta-regression of this type is most frequently used in the literature; however, there can be two issues that present the problem of identifying the estimated β as a measure of selection bias as a whole. First, it is implausible that the pairs of ($coeff_{ij}$; SE_{ij}) and ($coeff_{kj}$ and SE_{kj}) are independent. This assumption can be relaxed if one assumes that the authors and editors select each coefficient estimate independently and separately.[14] However, if the researcher is involved in p-hacking, then the assumption that each coefficient estimate was selected on its own merit is implausible. The second problem arises when one considers the existence of p-hacking, since the necessary assumption that estimated standard errors are unbiased SE_{ij} is also unlikely, therefore, equation 4.19 suffers from the spurious correlation and cannot accurately estimate the extent of selection bias β in the literature. To address this issue, I use the Meta-analysis Instrumental Variable Estimator (MAIVE) and instrument standard errors using the respective sample size in the first stage to replace the reported standard error, SE_{ij}, with the portion of the error that can be explained by the sample size. Irsova, Doucouliagos, et al. (2023) argue in favor of using the sample size as an instrument for reported standard errors. The reported variance (SE^2) is a linear function of the inverse of the sample size used in the primary study by definition. The sample size is not estimated, so it is free from measurement error. Changes in methodology generally have no effect on the sample size and neither do the choice of control variables. The sample size appears to be more resistant to selection bias, as gathering additional data is more challenging than manipulating the standard error to reach significance. Endogeneity might still persist if researchers, anticipating smaller effects, opt for larger experiments. However, in the context of observational studies, researchers generally use all

[14] see Andrews and Kasy (2019) for more detailed discussion.

available data.

To isolate the bias coming from within-study selection, I need to control the study-specific characteristics. I do this by applying fixed effects estimation, demeaning the estimates by the study mean effect and mean standard error:

$$\text{FE:} \quad coef_{ij} - \overline{coef}_j = \beta^{FE}(SE_{ij} - \overline{SE}_j) + u_{ij} \tag{4.20}$$

The fixed effect estimator takes care of the fixed effect of e_j for the unobserved study by subtracting the mean estimates of the study. This approach allows me to estimate the measure of bias, $\hat{\beta}^{FE}$, coming from the within-study variation.

Next, to study the extent of publication bias, I look at the extent of selection between studies. Here, I need to proxy a selection criterion for each study - ideally, it would be a main result or a set of results based on which the paper was selected for publication. Unfortunately, I do not have information on which of the estimates is more important in the pool of reported estimates. Therefore, I revert to taking mean estimates as the average story told in the manuscript and the average criteria based on which the publication decision is made.

$$\text{BE:} \quad \overline{coef}_j = \alpha + \beta^{BE}\overline{SE}_j + u_j \tag{4.21}$$

Therefore, I study the variations between studies using the averages of the estimates for each study.

Finally, with similar rationality, I employ the PEESE, PET-PEESE, and EK model approaches to consistently estimate the extent of selection bias. As above, I run these regressions on demeaned reported estimates first and mean estimates second, to analyze the extent of selection bias that arises from selection within the study and between the studies, respectively.

Figure 4.10: Different types of selection biases influencing published work

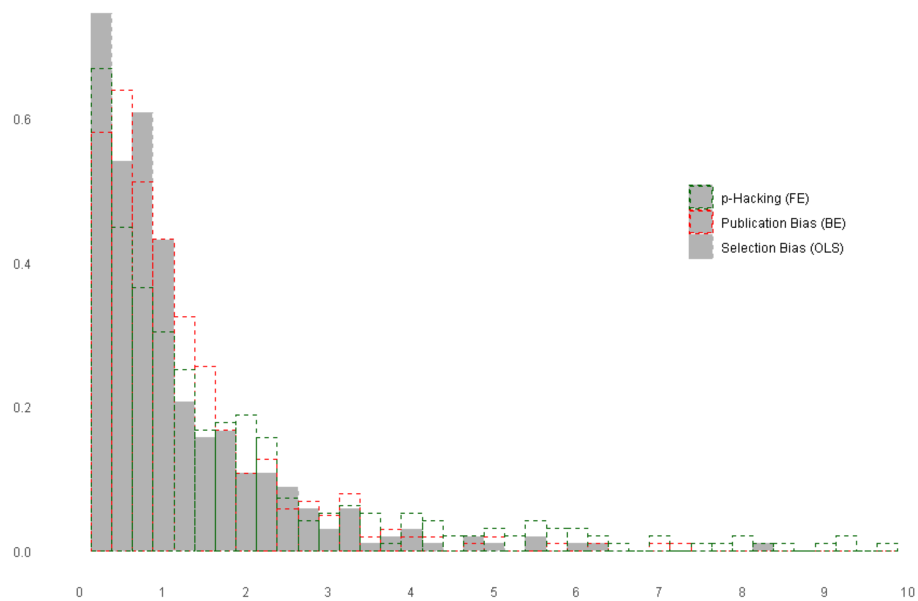


Figure presents the distribution of estimated $\hat{\beta}$ from fixed effect, between effect and OLS estimations, where β^{FE} is extent of within study selection - measure of p-hacking, β^{BE} measures the extent of publication bias defined as selection across study, β^{OLS} estimates the average selectivity in the literature and is the most common version of the meta-regression. Note that these results are retrieved from analysis of Published Paper sub-sample.

4.4.1 Selection within vs. across study

The Figure 4.10 shows the distribution of β coefficient from the Fixed-effect (4.20), between-effect (4.21), and OLS (4.19) estimated for 400 subsamples separately. The distribution of the coefficient β estimated from the OLS regressions, presented as the gray shadow in the figure, is the average effect of selection in the published literature. The measure of bias from the within-study variation indicates the extent of p-hacking (in green); and the measure of bias coming from the between-study variation indicates the extent of publication bias (in red). In Figure 4.10, when looking at part of the distribution that shows little or no bias $|\beta| < 1$, as well as the moderate level of bias $1 < |\beta| < 2$, the selection between studies seems to be more relevant. But as the severity of the selection bias increases, p-hacking plays a larger role in the selection bias.

Figure 4.11: Distribution of $\Psi_k = |\beta_k^{FE}/\beta_k^{BE}|$

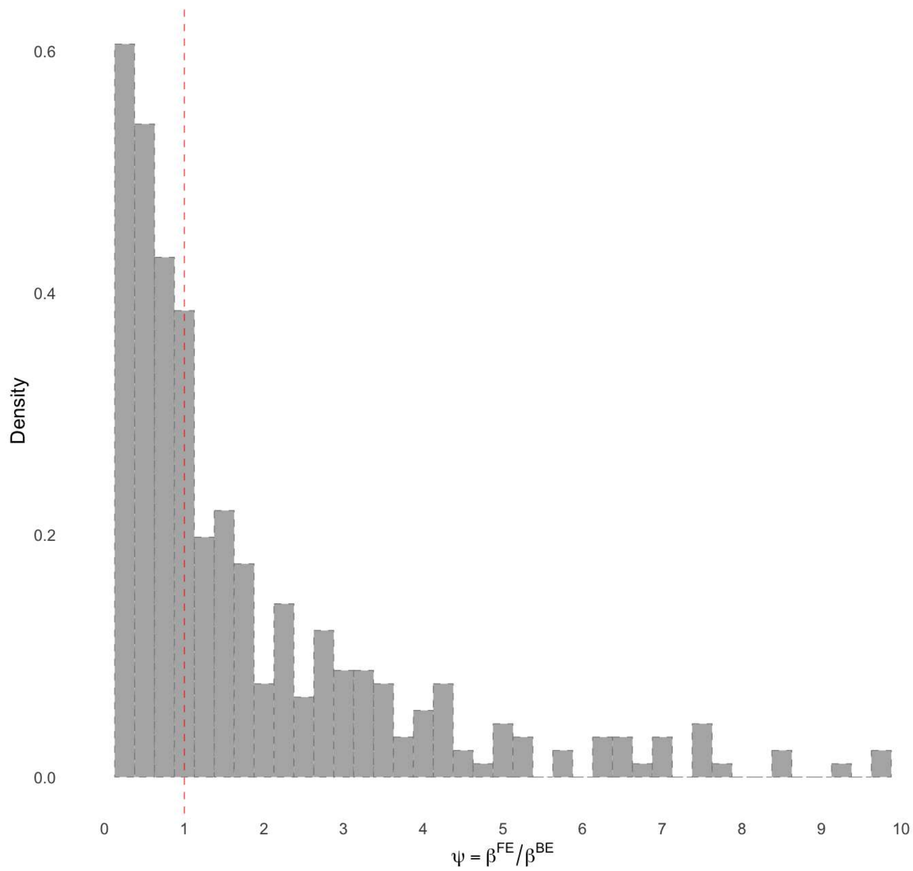


Figure shows comparison of within $|\beta_k^{FE}|$ and between $|\beta_k^{BE}|$ selection using ratio.

Finally, I calculate β_k^{FE} and β_k^{BE} and derive $\psi_k = \beta_k^{FE}/\beta_k^{BE}$ for each meta-study k based on the subsample of published results. Figure 4.11 shows the distribution of ψ_k with a significant part of the distribution on the right side of red line indicating threshold where $\beta_k^{FE} > \beta_k^{BE}$ has a long tail.

I estimate the ψ_k ratio from the fixed effect and between the effect models[15] and I present the median and mean values of ψ_k with the 95% confidence interval (CI) constructed using t statistics for mean and bootstrapping with a sample with multiple repetitions for the median. Next, to alleviate the effect of

[15]winsorized on 1, 2.5, and 5%. Table 4.1 shows the results of the most liberal 1% winsorization. However, 2.4% and 5% winsorization showed very similar results.

outliers, I apply median regression on the original data without winsorization. The both results are consistent in that, they both predict over 10% larger effect of p-hacking compared to the publication bias in the bias caused by selection of the results for publication. Next, in Table 4.2, I show the analysis based on

Table 4.1: Selection within vs. across study, published papers

	Linear Regression	**Quantile Regression**
Median	1.18	1.11
	[1.03; 1.48]	[0.96; 1.28]
Mean	7.78	9.52
	[5.13; 10.44]	[4.31; 14.73]
Number of Meta-Studies	409	407

In the table, the median and mean values of ψ_k are detailed, each accompanied by a 95% confidence interval (CI). These intervals are calculated using the t-statistics for the mean and using bootstrapping with multiple repetitions for the median. Additionally, the data set has been winsorized at the 1st and 99th percentiles to enhance its statistical robustness. The data set comprises estimates exclusively from published papers.

PEESE, PET-PEESE, and EK regressions. To control for possible p-hacking and more accurately estimate the extent of biased selection, I instrument the reported standard errors, SE_i, in the first stage [16] with the inverse of the sample size to the instrument for the standard errors. In Table 4.2 I report the median and means of estimates that show strong correlation on the first stage as evidence of instrument's relevance.

In all five approaches (Tables 1 & 2), I find that the bias arising from the variation within the study is greater than the selection between studies. Although the mean value is greater than 5 in all cases, this is probably due to the long tails of selection bias and ration ψ_k, see the figures 4.10 and 4.11. Therefore, looking at the median value of ψ_k is essential. Together, the median and mean

[16] suggestions Irsova, Bom, et al. (2023)

Table 4.2: Selection within vs. across study, published papers

	PEESE	**PET-PEESE**	**EK**
Median	1.33	1.29	1.22
	[1.15; 1.51]	[1.05; 1.76]	[1.07; 1.44]
Mean	7.44	7.58	4.41
	[1.66; 13.22]	[1.91; 13.25]	[2.66; 6.17]
Number of Meta-Studies	191	191	191

In this table, the median and mean values of ψ_k are presented, derived from the Instrumental Variable (IV) regressions of the PEESE, PET-PEESE and EK models. These values are accompanied by 95% confidence intervals (CIs), which are constructed using t-statistics for the mean and bootstrapping with multiple repetitions for the median. The dataset has been winsorized at the 1st and 99th percentiles. The number of meta-studies included in this analysis has been reduced to 206, as ψ_k values from regressions with first-stage F statistics less than 10 have been excluded. The data set comprises estimates exclusively from published papers.

values of the ratio suggest that selection within studies is consistently larger compared to selection across studies, pointing to the prevalent evidence of practices like method searching and p hacking in the published literature.

Table 4.3: Selection within vs. across study, all papers

	Linear Regression	**Quantile Regression**
Median	1.16	1.12
Median CI	[1.06; 1.46]	[0.97; 1.38]
Mean	7.85	8.84
Mean CI	[4.84; 10.87]	[1.63; 16.06]
Number of Meta-Studies	412	368

In the table, the median and mean values of ψ_k are detailed, each accompanied by a 95% confidence interval (CI). These intervals are calculated using t-statistics for the mean and bootstrapping with multiple repetitions for the median. Additionally, the data set has been winsorized at the 1st and 99th percentiles to enhance its statistical robustness.

These conclusions are drawn from looking at the published results. Next, I look at a complete dataset that contains results from published papers and working papers to evaluate the comparison of selection within and across studies in general.

Table 4.4: Selection within vs. across study all papers

	PEESE	**PET-PEESE**	**EK**
Median	1.21	1.28	1.28
Median CI	[1.12; 1.44]	[1.10; 1.82]	[1.08; 1.51]
Mean	8.33	7.02	4.45
Mean CI	[2.21; 14.44]	[1.73; 12.31]	[1.93; 6.96]
Number of Meta-Studies	206	206	206

In this table, the median and mean values of ψ_k are presented, derived from the Instrumental Variable (IV) regressions of the PEESE, PET-PEESE, and EK models. These values are accompanied by 95% confidence intervals (CIs), which are constructed using t statistics for the mean and bootstrapping with multiple repetitions for the median. The data set has been winsorized at the 1st and 99th percentiles. The number of meta-studies included in this analysis has been reduced to 206, as the ψ_k values of regressions with first-stage F-statistics less than 10 have been excluded.

However, Tables 4 and 5 demonstrate that the findings derived exclusively from the published literature are consistent with those obtained from the entire data set. The Selection Within Studies (SWS) is consistently found to be more pronounced than Selection Across Studies (SAS). This pattern reinforces the notion that significant selection occurs at the research stage, indicating a tendency to report certain results while omitting others, potentially to strengthen the researcher's argument or narrative.

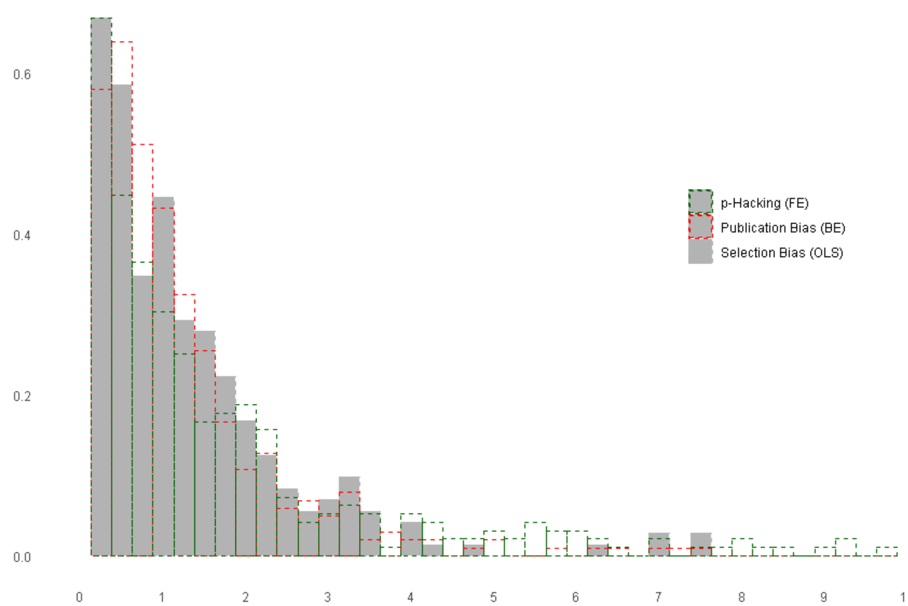
The patterns of selection across and within studies are repeated when analyzing the whole dataset consisting of over 15000 published and 3500 working papers. Next, I look at the selection bias in working papers in comparison to published papers.

4.4.2 Working papers vs published papers

To understand how to correct and potentially prevent selection bias within and between studies, it is important to explore the stages at which selection occurs. Selection across studies may occur at the submission and revision stage, or much earlier, when the researcher decides whether or not to write the paper. Moreover, while previous results have suggested the existence of a significant level of within-study selection, understanding the effect of the publication process on p-hacking is crucial. To this end, in this section, I first investigate the extent of within and between study selection in a working paper subsample, comparing these two types of biases. Subsequently, I compare the within and between study selection in working papers with that in published papers.

Figure 4.12 shows the distribution of selection bias, p-hacking and publication bias in working papers. In the realm of working papers, "publication bias" should be viewed as the decision by researchers to write the paper after receiving initial results or not. The phenomenon in which the research chooses to write the research paper according to the obtained results is frequently referred to as a "file-drawer problem" in the literature. Here, also, selection across studies

Figure 4.12: Selection bias in working paper $|\beta_k^{WP}|$ subset.



dominates for the low selectivity in reported results, and as the selection bias becomes more severe in different fields of research, the effect of selection within study becomes more prominent. To compare the effect of the publication process on bias, I perform a similar analysis as before and compare the extent of these selection biases in the results reported in the working and published articles, see Table 4.5.

In Table 4.5. I have reported results from linear, quantile, PEESE, PET-PEESE, and endogenous kick model estimations. As before, the last three use the instrumental variable approach to control for the spurious relation caused by the existence of p-hacking. The first section of the table shows the medians of the $\Psi_k = \left|\beta_{WP;k}/\beta_{P;k}\right|$ ratio comparing the average selection bias in the results of the working and published papers. Although linear estimations show larger selectivity in the results reported in the working papers, non-linear estimation models do not show such a large difference.

Next, to explore the question of whether the publication process accelerates or reduces selection, I look at the within- and between-study selection comparison separately. Comparison of p-hacking in the working and published papers shows that within-study selection is significantly larger in the results reported in the working papers. In contrast, there are no significant differences in the selection between studies in published papers compared to working papers.

The results in Tables 4.1, 4.2 and 4.5, show that the p hacking dominates compared to the publication bias in published research; however, published results suffer from less within-study selection compared to working papers. Table 4.5 shows on average greater evidence of p-hacking in working compared to published papers. Therefore, I conclude that the publication process filters out a significant portion of p-hacked results.[17]

These results highlight the widespread nature of selection biases in academic research. The upper section of Table 4.5 shows that the decision to write a research paper suffers from a selection bias similar to the journal's decision to publish. In essence, the biases affecting what gets written are strongly mirrored in what gets published. However, the primary driver of this phenomenon remains unclear, whether it is shaped more by the anticipations and decisions of journals and editors, or by researchers' beliefs about what is likely to be accepted. On the one hand, researchers could potentially correctly foresee the publication potential of their work and choose not to draft a manuscript that has a lower chance of acceptance. On the other hand, they might only submit manuscripts that they *believe* to likely be published, thereby limiting the array of choices available to journals, creating a self-fulfilling prophecy: even if journals exhibit no selection bias, they end up publishing only a partial narrative because they receive a non-representative sample of research outcomes. However, these results also point to the mitigating role of the publication process in the selection of estimates *within* the study. Table 4.5, middle section shows that selection within study dominates in working paper sub-sample, leading me to believe that significant portion of p-hacking is filtered before the studies are published.

[17]This conclusion is inline with the findings in Brodeur et al. (2023).

Table 4.5: Comparison of biased selection in working and published papers

	Linear	**Quantile**	**PEESE**	**PET-PEESE**	**EK**
Selective Reporting $\Psi_k = \lvert \beta_{WP;k}/\beta_{P;k} \rvert$					
Median	1.23	1.27	1.02	1.13	1.08
	[1.05; 1.55]	[1.06; 1.61]	[0.86; 1.22]	[1.00; 1.44]	[0.88; 1.21]
Meta-Studies	269	284	187	186	152
p-Hacking, Selective Reporting *within* study, $\Psi_k^{FE} = \lvert \beta_{WP;k}^{FE}/\beta_{P;k}^{FE} \rvert$					
Median	1.16	1.76	1.31	1.67	1.12
	[0.86; 1.28]	[1.36; 2.11]	[0.90; 1.74]	[1.12; 2.32]	[0.99; 1.68]
Meta-Studies	194	282	169	169	169
Publication Bias, Selective Reporting *between* studies, $\Psi_k^{BE} = \lvert \beta_{WP;k}^{BE}/\beta_{P;k}^{BE} \rvert$					
Median	1.16	1.34	0.93	1.05	0.97
	[0.86; 1.29]	[1.13; 1.66]	[0.74; 1.07]	[0.85; 1.24]	[0.86; 1.07]
Meta-Studies	195	288	134	134	134

This table shows the comparison of biased selection in working papers and published papers. For this, I show the median values of $\Psi_k = \lvert \beta_{WP;k}/\beta_{P;k} \rvert$; while, Ψ_k^{FE} compares the extent of p-hacking and Ψ_k^{BE} compares the extent of publication bias in working and published papers. In the columns (1) & (2), the median and mean values of ψ_k are detailed, each accompanied by a 95% confidence interval (CI). These intervals are calculated using the t-statistics for the mean and using bootstrapping with multiple repetitions for the median. Additionally, the data set has been winorized at the 1st and 99th percentiles to enhance its statistical robustness. In columns (3) to (5), the median and mean values of ψ_k are presented, derived from the Instrumental Variable (IV) regressions of the PEESE, PET-PEESE, and EK models. These values are accompanied by 95% confidence intervals (CIs), which are constructed using t-statistics for the mean and bootstrapping with multiple repetitions for the median. The data set has been winsorized at the 1st and 99th percentiles. The number of meta-studies included in this analysis has been reduced to 206, as psi_k values of regressions with first-stage F-statistics less than 10 have been excluded. The data set comprises estimates exclusively from published papers.

4.5 Conclusion

In this study, I have conducted an analysis of a comprehensive meta-dataset comprising more than 200,000 estimates from more than 19,000 studies across 400 different fields. Utilizing key meta-regression methodologies, I present substantial evidence of selective reporting of coefficient estimates within studies that also find their way into the published literature.

This paper highlights the importance of p-hacking in the academic literature, contributing to the emerging body of work such as Brodeur et al. (2023), Lang (2023), Irsova, Doucouliagos, et al. (2023). It supports the issues raised by Irsova, Bom, et al. (2023), underscoring the critical need for meta-analytical methodologies that address the biases of p-hacking in conjunction with selection biases across studies. Furthermore, the paper underscores the risks posed by practices such as p-hacking and method searching to the robustness of established academic beliefs. It provides evidence challenging the notion that these practices are merely concerns for unpublished research, indicating their broader implications in the field.

References

Andrews, I., & Kasy, M. (2019). Identification of and correction for publication bias. *American Economic Review*, *109*(8), 2766–94.

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Ashenfelter, O., Harmon, C., & Oosterbeek, H. (1999). A review of estimates of the schooling/earnings relationship, with tests for publication bias. *Labour Economics*, *6*(4), 453–470.

Balima, H. W., Kilama, E. G., & Tapsoba, R. (2020). Inflation targeting: Genuine effects or publication selection bias? *European Economic Review*, *128*, 103520.

Begley, C. G., & Ioannidis, J. P. (2015). Reproducibility in science: Improving the standard for basic and preclinical research. *Circulation Research*, *116*(1), 116–126.

Bom, P. R., & Rachinger, H. (2019). A kinked meta-regression model for publication bias correction. *Research Synthesis Methods*, *10*(4), 497–514.

Brodeur, A., Carrell, S., Figlio, D., & Lusher, L. (2023). Unpacking p-hacking and publication bias. *American Economic Review*, *113*(11), 2974–3002.

Brodeur, A., Cook, N., & Heyes, A. (2020). Methods matter: P-hacking and publication bias in causal analysis in economics. *American Economic Review*, *110*(11), 3634–3660.

Brodeur, A., Lé, M., Sangnier, M., & Zylberberg, Y. (2016). Star wars: The empirics strike back. *American Economic Journal: Applied Economics*, *8*(1), 1–32.

Bruns, S. B., Asanov, I., Bode, R., Dunger, M., Funk, C., Hassan, S. M., Hauschildt, J., Heinisch, D., Kempa, K., König, J., et al. (2019). Reporting errors and biases in published empirical findings: Evidence from innovation research. *Research Policy*, *48*(9), 103796.

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., et al. (2018). Evaluating the replicability of social science experiments in nature and science between 2010 and 2015. *Nature Human Behaviour*, *2*(9), 637–644.

De Long, J. B., & Lang, K. (1992). Are all economic hypotheses false? *Journal of Political Economy*, *100*(6), 1257–1272.

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

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455–463.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *Bmj*, *315*(7109), 629–634.

Ferraro, P. J., & Shukla, P. (2020). Feature—is a replicability crisis on the horizon for environmental and resource economics? *Review of Environmental Economics and Policy*.

Furukawa, C. (2019). Publication bias under aggregation frictions: From communication model to new correction method. *Unpublished Paper, Massachusetts Institute of Technology*.

Greene, W. H. (1990). *Econometric analysis*. Pearson.

Havránek, T. (2010). Rose effect and the euro: Is the magic gone? *Review of World Economics*, *146*(2), 241–261.

Havránek, T. (2015). Measuring intertemporal substitution: The importance of method choices and selective reporting. *Journal of the European Economic Association*, *13*(6), 1180–1204.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the Econometric Society*, 153–161.

Hedges, L. V. (1984). Estimation of effect size under nonrandom sampling: The effects of censoring studies yielding statistically insignificant mean differences. *Journal of Educational Statistics*, *9*(1), 61–85.

Hedges, L. V. (1992). Modeling publication selection effects in meta-analysis. *Statistical Science*, *7*(2), 246–255.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLOS Medicine*, *2*(8), e124.

Ioannidis, J. P. A., Stanley, T. D., & Doucouliagos, H. (2017). The Power of Bias in Economics Research. *The Economic Journal*, *127*(605), F236–F265.

Irsova, Z., Bom, P. R., Havranek, T., & Rachinger, H. (2023). Spurious precision in meta-analysis.

Irsova, Z., Doucouliagos, H., Havranek, T., & Stanley, T. (2023). Meta-analysis of social science research: A practitioner's guide. *Journal of Economic Surveys*.

Iyengar, S., & Greenhouse, J. B. (1988). Selection models and the file drawer problem. *Statistical Science*, 109–117.

Jackson, C. K., & Mackevicius, C. L. (2023). What impacts can we expect from school spending policy? evidence from evaluations in the us. *American Economic Journal: Applied Economics*.

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1995). *Continuous univariate distributions* (Vol. 289). John Siley & Sons.

Kranz, S., & Pütz, P. (2021). Rounding and other pitfalls in meta-studies on p-hacking and publication bias: A comment on brodeur et al.(2020). *Available at SSRN 3848786*.

Lang, K. (2023). *How credible is the credibility revolution?* (Tech. rep.). National Bureau of Economic Research.

Leamer, E. E. (1983). Let's take the con out of econometrics. *The American Economic Review*, *73*(1), 31–43.

Mathur, M. (2022). Sensitivity analysis for p-hacking in meta-analyses. *OSF Preprints*.

Miguel, E., Camerer, C., Casey, K., Cohen, J., Esterling, K. M., Gerber, A., Glennerster, R., Green, D. P., Humphreys, M., Imbens, G., et al. (2014). Promoting transparency in social science research. *Science*, *343*(6166), 30–31.

Stanley, T. D. (2005). Beyond publication bias. *Journal of Economic Surveys*, *19*(3), 309–345.

Stanley, T. D. (2008). Meta-regression methods for detecting and estimating empirical effects in the presence of publication selection. *Oxford Bulletin of Economics and Statistics*, *70*(1), 103–127.

Stanley, T. D., Doucouliagos, H., et al. (2007). Identifying and correcting publication selection bias in the efficiency-wage literature: Heckman meta-regression. *Economics Series*, *11*, 2007.

Stanley, T. D., & Doucouliagos, H. (2012). *Meta-regression analysis in economics and business*. routledge.

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, *5*(1), 60–78.

van Aert, R. C., & Van Assen, M. (2021). Correcting for publication bias in a meta-analysis with the p-uniform* method. *Manuscript submitted for publication Retrieved from: https://osfio/preprints/bitss/zqjr92018*.

van Assen, M. A., van Aert, R., & Wicherts, J. M. (2015). Meta-analysis using effect size distributions of only statistically significant studies. *Psychological Methods*, *20*(3), 293.

Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika*, *60*, 419–435.

Wooldridge, J. M. (2002). Econometric analysis of crosssection and panel data.

Appendix A

Response to Comments from Reviewers

I extend my gratitude to the reviewers and the committee for their insightful comments on the pre-defense version of my dissertation. The comments are typed in *in italics*, and my responses are in roman.

4.1 Professor Brodeur's Report on Dissertation Thesis

Address the following questions in your report, please:

a) Can you recognize an original contribution of the author?
b) Is the thesis based on relevant references?
c) Is the thesis defendable at your home institution or another respected institution where you gave lectures?
d) Do the results of the thesis allow their publication in a respected economic journal?
e) Are there any additional major comments on what should be improved?
f) What is your overall assessment of the thesis? (a) I recommend the thesis for defense without substantial changes, (b) the thesis can be defended after revision indicated in my comments, (c) not-defendable in this form.

a) This thesis makes a significant contribution to the ongoing discussions surrounding the credibility revolution in economics. It addresses critical issues related to meta-research, the replicability of studies, and the widespread occurrence of publication biases and p-hacking across scholarly publications through a set of comprehensive examinations. The initial study reevaluates the influential work of Burgess and Pande on the impact of bank expansion led by the Indian state on poverty reduction. It highlights that the significant effects initially *reported become less pronounced when broader policy contexts are taken into account, stressing the importance of considering external policy impacts for the reliability of empirical findings.*

The subsequent study scrutinizes the Economic Policy Uncertainty (EPU) indices. It uncovers that the observed trends in these indices are more a reflection of the methodologies employed than of genuine economic uncertainties.

Furthermore, an extensive examination of 20,000 studies investigates the prevalence of publication bias and p-hacking in the field of economics. The analysis indicates that p hacking is far more widespread than previously believed, contributing substantially to selection bias in the academic literature. The results suggest major selection bias in economic research, potentially compromising the trustworthiness of published results.

b) Yes

c) Yes, the thesis is well above the bar for a thesis at the University of Ottawa.

d) Yes, particularly the last chapter.

e) No major changes necessary for the defense.

f) I recommend the thesis for defense without substantial changes.

Chapter 3: *Page 104: Remove $$; Page 106: Remove parentheses after However; Page 107: Finally, I show; Page 119: Balima et al. (2020) analyze.*

Thank you for noticing these errors. I corrected them in the manuscript.

Page 102: Maybe add that it can lead researchers to p-hack more, in addition to abandoning studies.

I agree! Now, the text reads: “Meanwhile, the perception that publication bias is prevalent can lead researchers to abandon studies with unexpected or unpromising results or to p-hack in search for such results, exacerbating publication bias.”

Page 104: Your method is somewhat unclear at this point. Make it clearer

To clarify the method, I have included the equations for linear regressions in the introduction. I hope this change makes it easy to see what I mean." I have added:

$coef_{ij} = \alpha + \beta \cdot SE_{ij} + [\epsilon_i + u_{ij}]$;

FE: $coef_{ij} - \overline{coef}_j = \beta^{FE}(SE_{ij} - \overline{SE}_j) + u_{ij}$;

BE: $\overline{coef}_j = \alpha + \beta^{BE}\overline{SE}_j + u_j$

to the paragraphs in the Introduction where they are discussed.

Pages 118-9: Maybe explain why you are not using Andrews and Kasy’s method.

For my next major revision, I plan to approach the research question using other methodologies, including WAAP, Andre & Kasy, p-uniform, and Mathur (2024).

Page 125: "The rationale behind this instrument is that it is more costly to increase the sample size than to adjust for standard errors.” I don’t think this is the rationale for using this IV. You may want to expand and explain why it is a valid IV. Why should we expect a strong first stage and that Z affects Y only through X.

Thank you for this comment. Now, the text reads:

Irsova et al. (2023) argue in favor of using the sample size as an instrument for reported standard errors. The reported variance (SE^2) is a linear function of the inverse of the sample size used in the primary study by definition. The sample size is not estimated, so it is free from measurement error. Changes in methodology generally have no effect on the sample size and neither do the choice of control variables. The sample size appears to be more resistant to selection bias, as gathering additional data is more challenging than manipulating the standard error to reach significance. Endogeneity might still persist if researchers, anticipating smaller effects, opt for larger experiments. However, in the context of observational studies, researchers generally use all available data.

Page 126: "Unfortunately, I do not have information on which of the estimates is more important in the pool of reported estimates." Do you know which table is the result from?

Although I do not have information on which table the result is from, the order of the coefficients is according to how they were reported in the original paper. However, it would still not allow me to draw conclusions about which coefficient is more important than others since, frequently, first-reported coefficients are not from the preferred methodology of the authors.

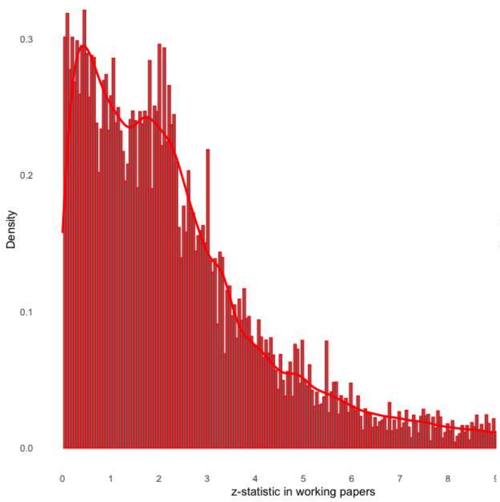
Page 131: by researchers Thank you for noticing, corrected.

For the WP vs. published research, could you show the distribution of z-statistics? A direct comparison would be very interesting. I am expecting that the two distributions are almost on top of each other.

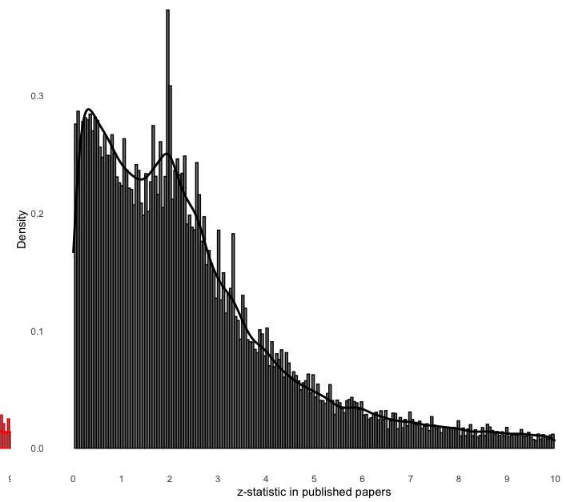
I agree that the comparison of z-statistics' distributions from working papers and published papers is indeed interesting. I will incorporate this discussion into the manuscript during the next revision. However, let me first present the comparison of the distributions in Figure 4.1:

Figure 4.1: Distribution of z-statistics

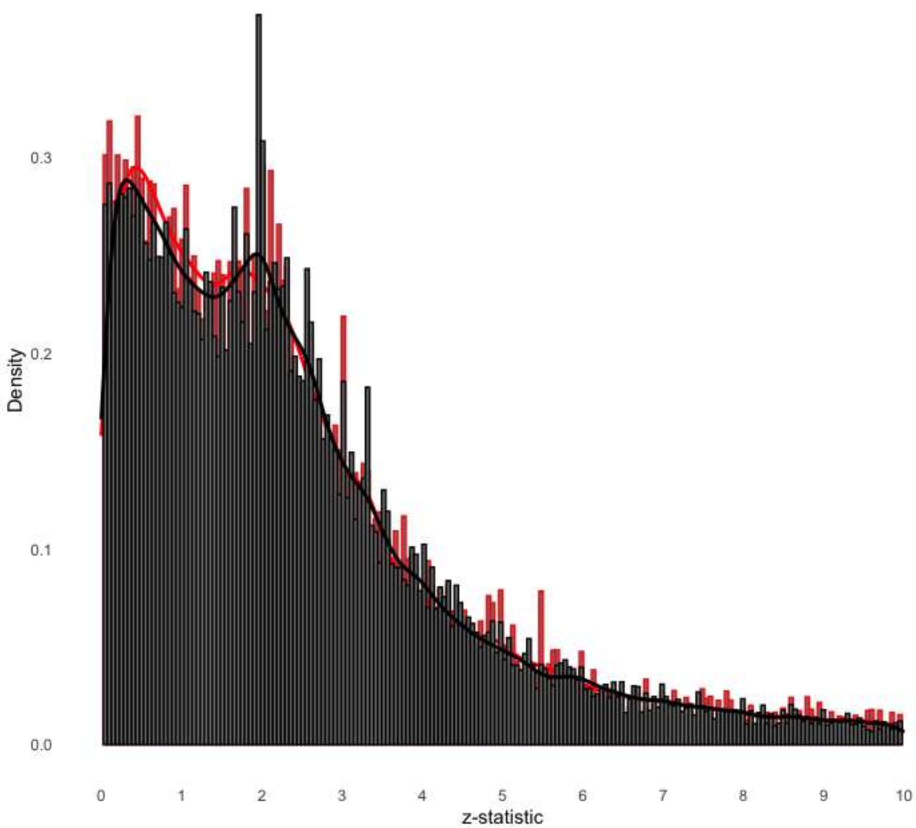
(a) Working Papers sub-sample



(b) Published Papers sub-sample



(c) Comparison of Working and Published papers



This figure presents distribution of z-statistics expressed by ratio of coefficient and its standard error as reported in the studies, de-meaned and de-weighted. Figure (a) presents the distribution in working paper sub-sample and figure (b) shows the distribution derived from published coefficients and standard errors. Finally, figure (c) presents the overlay of z-distributions from working papers (red) and published papers (**black**).

Page 133: It would be useful to conclude this section with explanations for why this result makes sense. Perhaps this should even come at the beginning of the section 3.4.2. Why are you doing such an exercise? Why do we care? How can this section helps us better understand your previous results, etc.

This comment will serve as guidance when reshaping the chapter and specifically the concluding part of Section 4.2. However, to incorporate the comment to the best of my ability at the moment, I have edited the section as follows.

The section begins: "To understand how to correct and potentially prevent selection bias within and between studies, it is important to explore the stages at which selection occurs. Selection across studies may occur at the submission and revision stage, or much earlier, when the researcher decides whether or not to write the paper. Moreover, while previous results have suggested the existence of a significant level of within-study selection, understanding the effect of the publication process on p-hacking is crucial. To this end, in this section, I first investigate the extent of within- and between-study selection in a working paper subsample, comparing these two types of biases. Subsequently, I compare the within and between study selection in working papers with that in published papers."

The section ends: "These results highlight the widespread nature of selection biases in academic research. The upper section of Table 4.5 shows that the decision to write a research paper suffers from a selection bias similar to the journal's decision to publish. In essence, the biases affecting what gets written are strongly mirrored in what gets published. However, the primary driver of this phenomenon remains unclear, whether it is shaped more by the anticipations and decisions of journals and editors, or by researchers' beliefs about what is likely to be accepted. On the one hand, researchers could potentially correctly foresee the publication potential of their work and choose not to draft a manuscript that has a lower chance of acceptance. On the other hand, they might only submit manuscripts that they *believe* to likely be published, thereby limiting the array of choices available to journals, creating a self-fulfilling prophecy: even if journals exhibit no selection bias, they end up publishing only a partial narrative because they receive a non-representative sample of research outcomes. However, these results also point to the mitigating role of the publication process in the selection of estimates *within* the study. Table 4.5, middle section shows that selection within study dominates in working paper sub-sample, leading me to believe that significant portion of p-hacking is filtered before the studies are published. "

Chapter 1: *In the introduction, I would like a more detailed discussion of why you are "After switching 1977 with other hypothetical trend break years, we find significant trend reversals almost every year between 1970 and 1984." And why you "analyze the impact of inclusion of additional trend reversal along with the one in 1977."? Readers who do not know the original study will be confused about the importance of these trend reversals. Adding 2-3 sentences here would be helpful. (e.g., the trend reversal in bank expansion should be driven solely by the ... The significance of those cut-years increases suspicion of the importance of different poverty-targeting policies for the dynamics of the banking network...)*

Thank you for your comment. Much of the discussion was lost due to page limitation; however, I completely agree with this point. Although I cannot edit the published paper, I have addressed the above comment within the dissertation in the following manner (edited parts are in **bold**):

"In this paper, we successfully **reproduce** the empirical results of Burgess and Pande (2005) and then analyze their identification strategy. **Since the authors use the 1977 trend break in rural bank branch openings as the instrument in the original identification strategy, this trend reversal should be significantly driven by the trend break in 1977 (imposition and removal of the 1977 bank branch licensing rule).** After switching 1977 with other hypothetical trend break years, we find significant trend reversals almost every year between 1970 and 1984. **The significance of those cut-years increases suspicion of the importance of different poverty-targeting policies for the dynamics of the banking network.** Some of these years coincide with the introduction of other programs targeting poverty reduction rather than bank expansion.

Next, we analyze the impact of the inclusion of additional trend reversal along with the one in 1977 **to control for the introduction of other amendments of bank licensing policy and other policy interventions.** These results show that with an additional cut-year around 1985 when the government reduced the pace of rural expansion, the effect of bank expansion on poverty decreases and becomes statistically insignificant.“

Your work is closer to a robustness reproduction than a replication as you are using the same data as the original authors. It's a detail, but I just thought I'd mention it.

I agree, I have changed “replicate” to “reproduce” in the dissertation version of this manuscript.

Chapter 2: *I do not have much to say on this chapter. It's outside of my area of expertise.*

4.2 Professor Reed's Report on Dissertation Thesis

Address the following questions in your report, please:

a) Can you recognize an original contribution of the author?

Yes.

b) Is the thesis based on relevant references?

Yes.

c) Is the thesis defendable at your home institution or another respected institution where you gave lectures?

Yes.

d) Do the results of the thesis allow their publication in a respected economic journal?

Yes.

e) Are there any additional major comments on what should be improved?

See below.

f) What is your overall assessment of the thesis? (a) I recommend the thesis for defense without substantial changes, (b) the thesis can be defended after revision indicated in my comments, (c) not-defendable in this form.

My recommendation is that the thesis progress to defense without substantial changes.

I choose to provide my comments chapter by chapter.

Chapter 1:

Do Rural Banks Matter That Much?

This chapter investigates the validity of Burgess and Plante (2005)'s claim in their 2005 AER paper that the expansion of branch banking into rural areas of India was responsible for a significant decrease in poverty. B&D's paper has been very influential. Google Scholar shows it currently having almost 2100 *citations.*

There is an obvious problem with endogeneity: banks are likely to expand in those areas that are experiencing economic growth. Thus, a positive correlation between expansion of the banking system and a reduction in poverty could be due to the former causing the latter, or the latter causing the former.

To address this problem, B&D use two instruments, a modification in the law requiring banks to open branches in "unbanked" regions in India that occurred in 1977, and the elimination of the branch banking requirement in 1990.

Buliskeria challenges the selection of these instruments by arguing that they ignore other government policies that were implemented at the same time as the changes in the branch banking requirements. To demonstrate that there is nothing special about the year 1977, she substitutes different cut-years for 1977 and gets the same basic result. She also adds a cut-year in addition to 1977 and the estimated causal effect of banking expansion becomes insignificant.

These kind of placebo tests are fairly common now and constitute a powerful refutation of B&D's causal interpretation. See below for my chapter assessment and comments:

a) Can you recognize an original contribution of the author?

   Yes

b) Is the thesis based on relevant references?

   Yes

c) Is the thesis defendable at your home institution or another respected institution where you gave lectures?

   Yes

d) Do the results of the thesis allow their publication in a respected economic journal?

Yes, it has already been published in the Journal of Applied Econometrics: Buliskeria, N., & Baxa, J. (2022). Do rural banks matter that much? Burgess and Pande (2005) reconsidered. Journal of Applied Econometrics, 37(6), 1266-1274.

e) Are there any additional major comments on what should be improved?

See below.

Given that the empirical case is so strong and it has already been published in an outstanding journal, I don't think anything else needs to be done.

Minor comments:

1) There are no tests for weak instruments. Should the candidate have found that the instruments were weak, it would have been useful to provide Anderson-Rubin confidence sets to re-evaluate the inference claims of both B&D and the results from this chapter.

Our analysis confirmed that the instruments used are indeed strong. Consequently, we decided not to discuss F-tests. However, the F-tests for our main regressions are presented in Table 1.A1 [now 2A.1]. Additional F-tests related to trend breaks can be found in Figure 1.1 (b) [now 2.1] and Table 1.B1 [now 2B.2]. I have also added further diagnostic tests to the Appendix of the Chapter 2. These were initially shown in Figure 8 (here Figure 2B1.1) of the working paper version of our study. Unfortunately, due to the journal's page limitation, we could only include the main results in the published version. However, thanks to your comment, I have now included F-tests for varying cut years as well as multiple cut-year exercises. Please refer to Figure 2B1.1 (Figure 8 in the working paper version) and Tables 2B1.1 – 2B1.4 and 2B2.1 – 2B2.3.

2) It would have been helpful if Buliskeria had included their code in the Appendix *to this chapter. This would have helped the reader to figure things out when the text wasn't sufficiently clear.*

I am afraid the code is too long to be provided as part of the dissertation; however, I have provided the link to the code and data in the footnote on the title page of Chapter 1 [now 2], and also a printout of the main code in Appendix B, at the end of this document.

3) In the interests of open science, Buliskeria might think of making the data and code publicly available so that the results of the chapter are push-button replicable.

The data and code are publicly available at *Journal of Applied Econometrics Data Archive* [http://qed.econ.queensu.ca/jae/datasets/buliskeria001/]

I also provide the link to the replication package in the footnote on the title page of Chapter 1 [now 2].

Chapter 2:

Uncertain Trends in Economic Policy Uncertainty

This chapter investigates the reliability of the Economic Policy Uncertainty (EPU) index, developed by Baker et al. (2016). The Baker et al. (2016) paper has been extremely influential. It currently has over 11000 Google Scholar citations.

The headline findings from Baker et al. (2016), as stated in their abstract are: "Using firm-level data, we find that policy uncertainty is associated with greater stock price volatility and reduced investment and employment in policy-sensitive sectors like defense, health care, finance, and infrastructure construction. At the macro level, innovations in policy uncertainty foreshadow declines in investment, output, and employment in the United States and, in a panel vector *autoregressive setting, for 12 major economies."*

The main analysis of this chapter investigates two robustness checks. First, it attempts to replicate the EPU index using the same text mining strategy and construction methods employed by Baker et al. (2016). It also proposes an alternative index that scales the number of articles by the number of economic policy articles, rather than the number of total articles.

Buliskeria provides a compelling example from the Frankfurter Allgemeine Zeitung where these two numbers trend in opposite directions. She demonstrates that the use of these alternative indices are associated with the same qualitative impacts of economic policy uncertainty that Baker et al. (2016) found, but the quantitative effects are generally smaller.

a) Can you recognize an original contribution of the author?

   Yes

b) Is the thesis based on relevant references?

   Yes

c) Is the thesis defendable at your home institution or another respected institution where you gave lectures?

   Yes

d) Do the results of the thesis allow their publication in a respected economic journal?

   Yes, given the prominence of the Baker et al. (2016) paper, it should be published in a good journal.

e) Are there any additional major comments on what should be improved?

   See below.

Despite being outside of my area of expertise, I found this chapter very interesting. There are many directions Buliskeria could have taken this chapter. While I do not think it is necessary to do this for the thesis, I have a suggestion plus *some minor comments if Buliskeria wishes to do further work on this.*

1) Our team discussed the potential application of uncertainty within an IV framework several times, but we have not arrived at any meaningful specification yet. The main problem is that it is hard to argue that one indicator of uncertainty is exogenous while the others are endogenous because all of those indicators seem to be, to some extent, causing shifts in economic activity and being affected by economic activity at the same time. One of the possibilities would be to define some events that could be considered as exogenous shocks to uncertainty, unrelated to developments of economic activity and unexpected, and use a binary variable set at one of those events as an instrument in the spirit of narrative identification. However, the candidates for such shocks are relatively scarce, perhaps the Brexit referendum and few elections with unexpected outcomes. Also, the shock identification could help with a more precise estimation of the effects of uncertainty shocks on economic activity, but we have not arrived at any meaningful way for the IV approach to compare the alternative indices of uncertainty. Therefore, any idea on this matter is more than welcome.

Our team discussed the potential application of uncertainty within an IV framework several times, but we have not yet reached any meaningful specification. The main problem is that it is hard to argue that one indicator of uncertainty is exogenous while the others are endogenous because all of those indicators seem to be, to some extent, causing shifts in economic activity and being affected by economic activity at the same time. One of the possibilities would be to define some events that could be considered as exogenous shocks to uncertainty, unrelated to developments of economic activity and unexpected, and use a binary variable set at one of those events as an instrument in the spirit of narrative identification. However, the candidates for such shocks are relatively scarce, perhaps the Brexit referendum and few elections with unexpected outcomes. In addition, shock identification could help with a more precise estimation of the effects of uncertainty shocks on economic activity, but we have not found any meaningful way for the IV approach to compare alternative indices of uncertainty. Therefore, any idea on this matter is more than welcome.

Minor comments:

1) I know Buliskeria is simply replicating the work of Baker et al. (2016), but I didn't understand why the word "today" was used in the keyword searching. Some explanation of this would have been helpful.

Baker et al. (2016) chose the word "today" as the word that has no relation to economics, politics & uncertainty. Therefore it would serve as a reasonable proxy for all articles. Please also see footnote 6, chapter 3, on page 63: "Factiva's search engine does not allow "blank" searches to obtain all articles published in a given month. Baker et al. (2016) address this issue by searching for the word "today" instead; however, it appears that when we search for the article "the" ("la" in Italian and French), the count of retrieved articles differs (is larger) from the search results using "today." Given the sensitivity of the series to the "generic" word, we find the use of "today" to be problematic and agree with your concern. "

2) Figure 2.3 should either have a note or a legend that identifies the black line as the EPU and the red line as the WUI.

Thank you for noticing. I have added the legend to Figure 2.3 [now Figure 3.3]

3) It also would have been interesting to pursue the use of AI (such as ChatGPT) to produce key words. On a lark, I asked ChatGPT for keywords and it gave

> **ChatGPT**
>
> For researching articles on economic uncertainty, consider using a combination of the following keywords:
>
> 1. "Economic uncertainty" - The primary topic of interest.
> 2. "Macroeconomic indicators" - To find discussions on how economic uncertainty affects or is affected by broader economic conditions.
> 3. "Business cycle volatility" - For insights into how economic uncertainty fluctuates with economic cycles.
> 4. "Policy uncertainty" - To explore how government policies contribute to economic uncertainty.
> 5. "Market uncertainty" - For a focus on how financial markets respond to or are impacted by economic uncertainty.
> 6. "Risk assessment" - Useful for finding articles on the evaluation and management of risks associated with economic uncertainty.
> 7. "Investor sentiment" - To understand how economic uncertainty influences investor behavior.
> 8. "Global economic outlook" - For articles that contextualize economic uncertainty within global economic trends.
>
> Combining these keywords with specific countries or regions, time periods, or particular aspects of the economy (like "unemployment," "inflation," or "foreign investment") can help refine your search results.

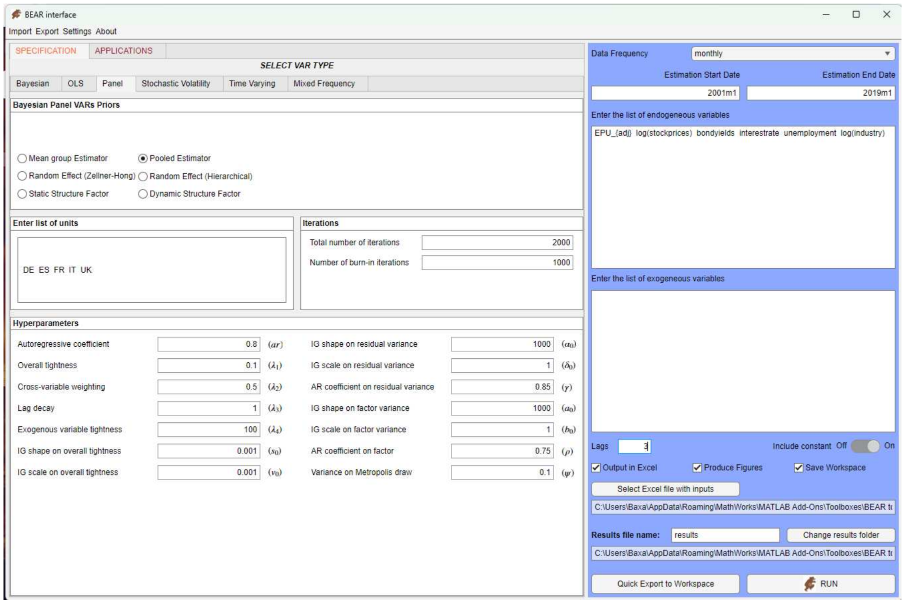
me the list below.

I agree that this could indeed be an interesting pursuit and thank you for the suggestion. Scraping databases was a very time-consuming endeavor therefore, while we consider his suggestion interesting, I am not sure if it is feasible to dive into collecting data for all countries again. The problem is that FACTIVA does not allow quick text mining analysis, so we had to perform all the searching for most of the newspapers manually.

4) It would have been helpful to provide a table(s) in the appendix reporting some of the key VAR results, along with the code used to produce those results (and the associated impulse response functions).

We have used the Matlab package BEAR with a neat Graphical User Interface that, as can be seen in the figure below, does not require any coding at all:

We have provided following specifications to the bear toolbox:



- Panel VAR: pooled estimator;
- structural decomposition: choleski factorisation;
- units: DE; ES; FR; IT; UK;
- endogenous variables: EPU_{adj}; log(stockprices); bondy ields; interest rate; unemployment; log(industry);
- exogenous variables: constant;
- estimation sample: 2001m1-2019m12;
- sample size (omitting initial conditions): 225;
- number of lags included in regression: 3;

→ hyperparameters:

- autoregressive coefficient (ar): 0.8;
- overall tightness (lambda1): 0.1;
- lag decay (lambda3): 1;
- exogenous variable tightness (lambda4): 100;

However, we will prepare the workspace file for publication in the future to allow for the replicability of our research.

5) In the interests of open science, Buliskeria might think of making the data and code publicly available so that the results of the chapter are push-button replicable.

Absolutely! Currently, the data set is available on my webpage ninobuliskeria.com; the replication package is also being prepared for the publication stage and will be available soon.

CHAPTER THREE:

Disentangling p-Hacking and Publication Bias

This chapter proposes a clever procedure for disentangling selection within studies (SWS) and selection across studies (SAS). SWS occurs when researchers p-hack their data and decide which estimates to include in the paper. SAS occurs when journals (and researchers) select which papers to submit and publish in journals. Notice that one type of selection occurs within studies, the other occurs across studies.

Buliskeria proposes to disentangle the two sources of selection via panel fixed effects and panel between effects. In both cases, a standard Egger regression is used to estimate selection bias as a function of the SE variable. Buliskeria uses panel fixed effects to identify SWS and panel between effects to identify SAS. She forms the ratio of the two estimates and finds that SWS bias is frequently multiple times the size of SAS bias. Also noteworthy is that she replaces SE in the Egger regressions with its predicted value using MAIVE estimation, with sample size is used as a instrumental variable.

a) Can you recognize an original contribution of the author?

   Yes

b) Is the thesis based on relevant references?

   Yes

c) Is the thesis defendable at your home institution or another respected institution where you gave lectures?

   Yes

d) Do the results of the thesis allow their publication in a respected economic journal?

Yes, given the prominence of the Baker et al. (2016) paper, it should be published in a good journal.

e) Are there any additional major comments on what should be improved?

See below.

Minor comments:

1) I'm not sure whether this affects the results of the analysis, and I am not 100% certain this is correct, but I think the MAIVE estimator produces a biased estimate of the SE. This arises because it relates average sample size to average SE values. But average SE values are biased downwards because researchers always "cheat" in one direction. They never increase the size of their SEs. They always decrease them. Thus the first stage regression of the MAIVE will underestimate the true value of the SE.

Thank you for your comment. I believe that it can indeed bias the fitted SE values. This concern is valid since a downward measurement error in SE would decrease the strength of the relation between the instrument and SE, thus threatening the validity of the procedure.

I have discussed this point with Tomas Havranek, who with his coauthors of the MAIVE paper plans to develop the extension that can count for the downward bias in the first stage of estimation.

On the other hand, I also want to point to the rigorous simulations that show the efficiency of the MAIVE technique in the case of the existence of p-hacking. In their simulations, the authors compare the MAIVE version of conventional methods to their baseline models (simple average, FE/SLW, PET-PEESE, EK, WAAP, Andre & Kasy, p-uniform) in the case of p-hacking.

Irsova et al. (2024) show three stylized facts in their simulations. First, spurious precision can plausibly arise in observational research. Second, a small portion of spuriousness creates serious problems for current meta-analysis models. Third, and importantly, the new meta-analysis instrumental variable estimator (MAIVE) substantially limits the resulting bias in meta-analysis.

Therefore, while I believe that your concern is very important and should be addressed in the future, MAIVE still performs better than baseline models in case of p-hacking.

Although perfecting the MAIVE approach is beyond the scope of this paper, rather relates more to the original paper by Irsova et al. (2024), I greatly appreciate your comments and I plan to think more on this issue in the next revisions and mention this point in the final version of the paper. In my next major revision, I intend to extend the paper by including methods by Andrews & Kasy (2019), Mathur (2022), RoBMA and others that will allow further analysis of the robustness of the results.

2) It looks like the β_{FE} and β_{BE} coefficients are estimated in separate regressions. I don't know whether you can do this with IV estimation, but without IV it is possible to estimate both in the same equation use REWB (Random Effects Within Between) estimation. See: Bell, A., Fairbrother, M., & Jones, K. (2019). Fixed and random effects models: making an informed choice. Quality & Quantity, 53, 1051-1074. This is what Wooldridge calls "correlated random effects" and I believe it also called the Mundlak model.

Thank you for pointing me to this study and method; I will explore it more during my next major revision, where I also plan to approach the research question using other methodologies, including WAAP, Andre & Kasy, p-uniform, and Mathur (2024).

3) Figure 3.11. Please label axes. Thank you for noticing, corrected.

4) Page 128, 7 lines from the bottom. "I imply median regression" should be "I apply median regression". Thank you for noticing, corrected.

5) Page 108, sentence immediately above Equation (3.3): It should be footnote "7", not "27". Thank you for noticing, corrected.

6) In the interests of open science, Buliskeria might think of making the data and code publicly available so that the results of the chapter are push-button replicable.

I intend to build a replication package and make it available in the future.

4.3 Professor Furukawa's Report on Dissertation Thesis

The first two papers, "Do Rural Banks Matter That Much?" and "Uncertain Trends in Economic Policy Uncertainty," are thorough replications of existing and important papers. The first paper examines the robustness of the results that an apparently exogenous change in rural bank expansion policy has alleviated poverty. In particular, the paper finds that similar estimates can be obtained even when we examine the placebo estimates where the bank expansion policy took place several years before and after the policy change.

Chapter 2:

The second paper examines the economic policy uncertainty index data by replicating the entire process of index computation. This effort leads to the finding that the original claim of events such as Brexit and COVID-19 leading to the larger uncertainty may have been overstated. The finding seems particularly important given the large influence of original studies. In terms of exposition, however, the main results come only in Figures 5 and 6, making it difficult to visually understand the main results. Even though the paper seems very thorough, and while I understand that these expositional judgments are always subject to the authors' tastes, I thought that the draft may benefit from having the most important figures upfront.

Thank you for your comment. The paper has now been submitted and we are waiting for the additional feedback of the reviewers. I believe it would be most efficient to revise the manuscript after also other reviewers' feedback at which point I will also work on refining the structure are readability of the paper. Our intention was first to show the trends themselves in the index and raw numbers, illustrate that they are driven not only by the number of uncertainty-related articles, but also by the normalizing constant, and finally to present the results.

I am deeply grateful for your patience that you read the paper despite having to waited for the most important results until the final figures!

Chapter 3

In this report, I wish to concentrate on the third paper, "Disentangling p-hacking and publication bias," for two reasons: first, my expertise is more in the area of publication bias, and second, there remains more work for the paper so that there is potential for an additional gain from my comments.

I believe the paper asks a central question in the meta-analysis literature – the relative contribution of selection within or across studies for publication bias. The paper also employs an impressively large number of studies and estimates. That being said, I have several questions and concerns regarding the paper:

Main Concern 1: *I was unsure whether the main exposition in the abstract "the selective reporting within studies is about 20% more prevalent than publication bias arising from selection among studies" accurately describes the main results from the estimates. My first reading of the paper was either of the following [i] or [ii]:*

i *when a researcher has a "null" result by chance, he engages in either (A) p-hacking - selection within studies, or (B) publication bias – not publishing. The chance of (A) is x%, and the chance of (B) is y%, and x/y=1.2.*

ii *the overall estimate from the literature has magnitude x of the bias, and the ratio of the contribution from (A) or (B) is 1.2 to 1.*

Thank you for the thorough analysis, and I apologize for any vagueness in the text. To enhance the clarity of the Introduction, I have included the equations for linear regressions. I hope that this adjustment makes the text more easily understandable. I have added:

$coef_{ij} = \alpha + \beta \cdot SE_{ij} + [\epsilon_i + u_{ij}]$;

FE: $coef_{ij} - \overline{coef}_j = \beta^{FE}(SE_{ij} - \overline{SE}_j) + u_{ij}$;

BE: $\overline{coef}_j = \alpha + \beta^{BE}\overline{SE}_j + u_j$

to the paragraphs in the Introduction where they are discussed.

However, upon reading this paper, I realized that the paper is not about these results, but instead about the relative coefficient size of Egger regression, where the standard error is instrumented by the sample size (i.e. MAIVE regression). I think I understand that the relative coefficient size could be indicative of the relative importance, I am not sure if the quantity can be interpreted in the way claimed in the abstract. I am sympathetic to the desire to quantify the magnitude but I think it requires some more effort to establish an interpretation of the ratio (i.e., for example, in the Mills ratio formula described in the paper, how does the first-order approximation - the slope - change with respect to the probability of omission/ selective publication?)

I agree with your concerns. I will continue to explore ways to quantify the difference (thank you for suggesting to explore the slope, I will certainly go in this direction). However, as an immediate remedy, I chose to avoid using specific measures in abstract and in text.

Abstract: [...] Using various meta-regression methods, I find that selective reporting within studies is ~~about 20%~~ more prevalent than publication bias arising from selection among studies. [...]

Main Concern 2: *in the fixed effects regression, the author interprets the positive correlation between the standard error and the coefficients among various estimates within each study. The author then interprets this as the* *evidence for p-hacking. I agree that the correlation may potentially reflect p-hacking, but I have two other possible explanations:*

- *if the author of the original papers estimates among various sub-groups, and puts one group with statistically significant coefficients as one, and puts the remaining groups as "all other groups" together as one group, then there could still be the positive correlation. Even though there is more emphasis put on the statistically significant coefficient, I was not entirely certain if this would constitute a form of p-hacking (even though it is claimed that way in terms of bunching at the threshold). The author still reports all results, but is merely putting more emphasis on the significant results.*

- Thank you for this example. This issue is likely common across most bias detection techniques; in cases of frequent occurrences, the majority of observed coefficients would be those that are significant, while non-significant results would be grouped together, and hence appear under-reported (even if original author reported all the relevant details for the subgroup). However, this grouping can also impact precision and may be mistaken to p-hacking, although grouping can considered an acceptable way of presenting secondary results. While I acknowledge your concern, I currently lack the tools to address this scenario, as I have no information on which coefficients is reported as the representative (probably weighted mean) for "all other groups."

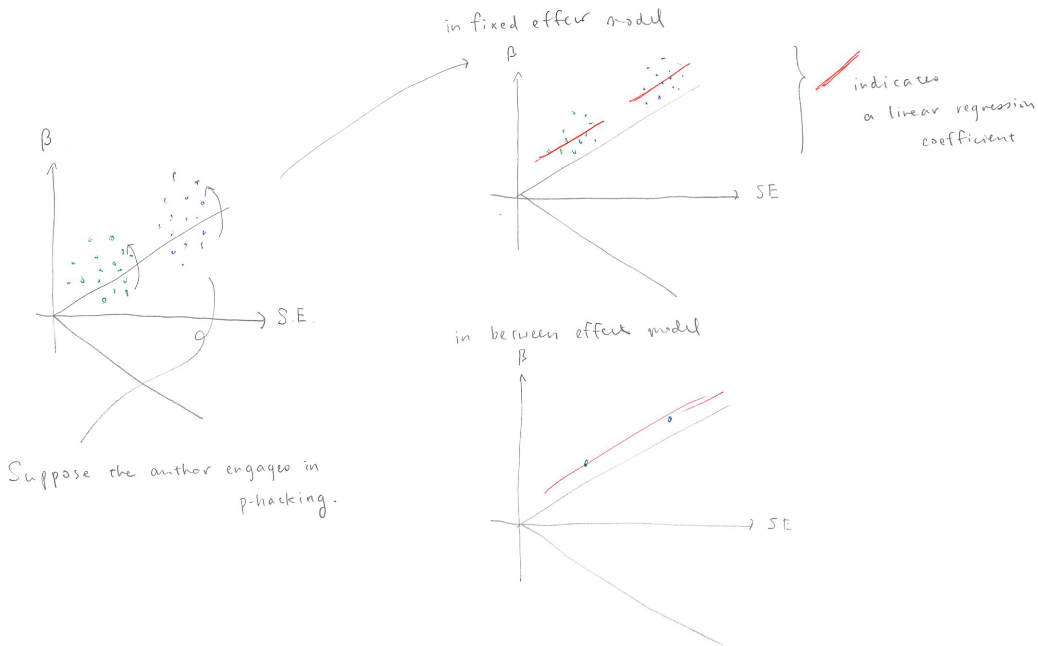
- *Egger regression depends on the assumption that the standard errors and the underlying coefficient estimates are independent in their underlying distributions. However, in some examples, there could be "small study effects" where less precise effects (estimates that arise small samples)* *have larger effects because, say in an experiment, one can have actually more effective treatments when there is less precise effects.*

- I believe this concern would also apply on the broader literature and generally Egger's regression. One way to address this issue would be to introduce precision (inverse of standard error) as the weight, this way less precise coefficients would receive less weight. I also plan to incorporate other methods aside from Egger's regression and broaden the scope of the manuscript.

Perhaps I am being overly worried since these two critiques would probably apply to the existing methods, too.

Main claim 3: *I was not fully convinced that the fixed effects vs. between effects estimates actually separately identify between vs. within selection patterns. Please let me raise one example. Suppose there is dispersion in the precision across estimates of two studies. Both studies originally have various estimates, and the authors of both studies both engage in p-hacking. In this case, as illustrated in the figure below, there can be a positive coefficient for between effect estimates, even though there is only p-hacking. I thought this pattern might be worth discussing.*

I agree that there would be a positive correlation between mean coefficients and mean standard errors in this example. Such a correlation would inflate the estimate of between-study bias, here publication bias, which in turn would mean that the ratio of p-hacking to publication bias would be underestimated. If this is the case, it would strengthen my argument that p-hacking is prevalent in the literature and represents a significant issue that warrants greater attention. I think obtaining a precise measure of selection within and across studies is beyond the capabilities of currently existing techniques. The approach used in this chapter is intended to serve as an indicator of the importance of considering p-hacking when developing bias correction techniques.



Main Comment 4:

- *I am thinking that the ratio between the two coefficients is not easy to interpret because it is possible that, by noise, one of the coefficients could be negative. Then, would it be possible to interpret the ratio that is also negative? Given that there could be such pathological cases, would it be more reasonable to compare the magnitude by taking the difference (rather than taking the ratio)? This approach avoids this concern.*

- *On a related point, how do I think about the publication selection that happens on the negative side? In some studies, such as the effect of labor union on firm productivity, the effect is ambiguous, and so, the selection could occur on both sides.*

- I use the absolute term for selection bias estimate, since I am interested in the absolute magnitude of bias. But I will incorporate this suggestion and do the additional analysis on the difference as well in the next revisions of the paper. Thank you for this comment.

Exposition and Organization

Please let me now share two comments related to the exposition and organization of the paper.

1. *the main specification of this paper is written in page 21. As the reviewer, it was very difficult to understand the paper's substantive content before reading the fixed effects and between effects specifications. The introduction does discuss and describe these methods, but without seeing the formula, it was difficult to interpret.*

    - *I suggest the author to put the Section 2 Theoretical Foundation and Section 4 Selection within vs across studies in the same section.*
    - *I think it could be more reader friendly if the main specification comes at the very beginning, and then explain the supporting models later.*

    I agree that restructuring in this manner could enhance readability, and I plan to revise the structure of the third paper based on this feedback in the future. In the short term, for the sake of clarity, I have addressed this comment by including an equation together with the discussion in the Introduction.

2. *there are some terminologies that seemed somewhat confusing to me.*

    - *if I am reading correctly, the paper refers to "within study" selection as p-hacking, and "between study" selection as publication bias.*

However, in some other parts of the literature, publication bias is regarded as any kind of biases in overall estimates due to selective publication. In this sense, p-hacking is a part of publication bias. I do not wish to push too much for this, but just wished to raise this point.

- *I was not sure if "fixed effects" vs "between effects" is the best term to describe the current method. Usually "fixed effects" is combined with "random effects," and I was not sure of the term "between effects" since this is really "between studies." I just wished to suggest these even though I do not have a good alternative.*

Thank you for raising these points. I agree that in the literature the definition of p-hacking and publication bias varies. I intend to devote a discussion to the various definitions in the literature. In this paper, I follow the definition of Mathur (2024), where publication bias is defined as selection across studies (SAS), and p-hacking is defined as selection within studies (SWS).

Please do not take these comments as the requirements for the Ph.D. thesis, but rather please take these points as possible considerations as the paper becomes revised and submitted. Once again, thank you for giving me a chance to learn about the new approach proposed and thinking through the assumptions and the implications.

Thank you for such thorough comments!

Code for Chapter 2

Do rural banks matter that much?

```
*==========================================================
* Date : October 2021
* Paper: Do Rural Banks Matter That Much?
*        Burgess and Pande (2005) Reconsidered
* The original code is from:
* Burgess, R., \& Pande, R. (2005).
* "Do rural banks matter?
* evidence from the Indian social banking experiment."
* American Economic Review, 95(3)
* This code replicates Burgess and Pande (2005)
*        with different cut-years
* substituted for 1977.
*==========================================================

clear all
*set more off
use "burgesspandeaerfinal.dta"
*------------------------------
*==========================================================
*==========================================================
* SECTION 1: replicate FIGURE 1  for different cut-years.
*==========================================================
*==========================================================

* generate the coefficients for the Figures
drop if year>2000

*Figure 1: rural branches in unbanked locations (with controls)
quietly areg runbfor unbk62-unbk00 rshare62-rshare00
popn62-popn00 stin62-stin00 stdum*, absorb(year) cluster(state)

*Save coeff of ubks
quietly mata: b=st_matrix("e(b)")'
quietly getmata b, force
quietly gen temp=b[_n-1]
quietly replace b=temp
```

```
quietly replace b=. if _n==_N
quietly replace b = . if state > 1
quietly label variable b ///
    "rural branches in unbanked locations (with controls)"

local graphs ""

forvalues i = 1961(1)1996 {
gen T_‘i’ = ‘i’

gen trend_‘i’ = year - (T_‘i’-1)
replace trend_‘i’ =0 if year < T_‘i’
gen dum_‘i’ =0
replace dum_‘i’ =1 if year>(T_‘i’-1)
gen ubt_‘i’ =unb61*trend_‘i’ *dum_‘i’
gen ubdum_‘i’ =unb61*dum_‘i’
gen popt_‘i’ =((pop61)/larea)*trend_‘i’ *dum_‘i’
gen popd_‘i’ =((pop61)/larea)*dum_‘i’
gen stint_‘i’ =stinc61*trend_‘i’
gen stind_‘i’ =stinc61*dum_‘i’
gen rursht_‘i’ =rursh61*trend_‘i’ *dum_‘i’
gen rurshd_‘i’ =rursh61*dum_‘i’


macro define controls "popt61 popt_‘i’  popd_‘i’  popt90 popdum90
rursht61 rursht_‘i’ rurshd_‘i’ rursht90 rurshd90  stint61 stint_‘i’
stind_‘i’ stint90 stindum90"

quietly reg runbfor ubt61 ubt_‘i’ ubt90  ubdum_‘i’ ubdum90
$controls stdum* yrdum*,cluster(state)

quietly mata: coef_‘i’=st_matrix("e(b)")’
quietly getmata coef_‘i’, force
quietly replace coef_‘i’=. if _n>5
quietly gen gamma1_‘i’ = coef_‘i’[1]
quietly gen gamma2_‘i’ = coef_‘i’[2]
quietly gen gamma3_‘i’ = coef_‘i’[3]
quietly gen gamma4_‘i’ = coef_‘i’[4]
quietly gen gamma5_‘i’ = coef_‘i’[5]

quietly gen tr_‘i’ =  gamma1_‘i’*ubt61 + gamma2_‘i’*ubt_‘i’ +
gamma3_‘i’*ubt90 + gamma4_‘i’*ubdum_‘i’ + gamma5_‘i’*ubdum90

quietly egen trend_m_‘i’ = mean(tr_‘i’), by (year)
quietly gen mtrend_‘i’ = trend_m_‘i’ if state == 1
```

```
quietly gen mytrend_‘i’ = mtrend_‘i’ - mtrend_‘i’[1]

quietly label variable mytrend_‘i’ "rural branches in
    unbanked locations (trend break)"

quietly line b mytrend_‘i’ year, ///
legend(size(small) cols(1) symxsize(7)) ///
lpattern(solid longdash_dot) lcolor(black black) ///
ylabel(, labsize(small) angle(horizontal)) ///
xlabel(, labsize(small)) graphregion(color(white))  ///
xtitle(Policy introduced in ‘i’, size(small)) ///
saving(Figure1_‘i’, replace) nodraw

 }

local all "Figure1_1970.gph Figure1_1971.gph Figure1_1972.gph
Figure1_1973.gph Figure1_1974.gph Figure1_1975.gph Figure1_1976.gph
Figure1_1977.gph Figure1_1978.gph Figure1_1979.gph Figure1_1980.gph
Figure1_1981.gph"

grc1leg2 ‘all’, col(3) ///
l1("Initial financial development X year - coefficient
{it}{&gamma}{subscript:t}", size(small))
*title(FIGURE 1. 1961 Initial Financial Development and
Rural Branch Expansion, size(small))  ///

graph export all.eps, replace




*==================================================
*==================================================
* SECTION 2: TABLES
*==================================================
*==================================================



forvalues i = 1961(1)1990 {
*------------------------------
*------------------------------
* generate the coefficients for the Figures
drop if year > 2000
*For the regressions we drop Haryana before 1965
```

```
drop if state == 5 & year < 1965


*=================================================================
*Table 1: Banking as a Function of Initial financial development
*=================================================================


macro define controls "popt61 popt_‘i’ popd_‘i’ popt90 popdum90
rursht61 rursht_‘i’ rurshd_‘i’ rursht90 rurshd90  stint61 stint_‘i’
stind_‘i’ stint90 stindum90"

local table1outcomes "rurcrsh rursavsh bfor priorsh pcoopsh"

*Column 1: Branches in rural unbanked locations
*------------------------------------------------------------
quietly reg runbfor ubt61 ubt_‘i’ ubt90 ubdum_‘i’ ubdum90
$controls stdum* yrdum*,cluster(state)

quietly test ubt61+ubt_‘i’=0
local F1 = r(F)
local p1 = r(p)
quietly test ubt61+ubt_‘i’+ubt90=0

outreg2 using "table_1_‘i’.txt", se ///
addtext(State and year dummies, YES, Other controls, YES) ///
keep(ubt61 ubt_‘i’ ubt90 ubdum_‘i’ ubdum90) ///
addstat(F-test 1,‘F1’, P1 , ‘p1’, F-test 2,  ‘r(F)’, P2, ‘r(p)’) ///
coefastr paren bdec(2) adjr2 symbol(***, **, *) nocons
    replace


*Columns 2 & 3 - Rural bank: Rural bank Credit and Savings share
*& Column 4 - Branches in banked locations
*&& Columns 5 & 6 - Credit share: Priority sector and Cooperative
*----------------------------------------------------------------

foreach v of local table1outcomes {
quietly reg ‘v’ ubt61 ubt_‘i’ ubt90 ubdum_‘i’ ubdum90 $controls
stdum* yrdum*, cluster(state)

quietly test ubt61+ubt_‘i’=0
local F1 = r(F)
local p1 = r(p)
quietly test ubt61+ubt_‘i’+ubt90=0
```

```
outreg2 using "table_1_‘i’.txt", se ///
addtext(State and year dummies, YES, Other controls, YES) ///
keep(ubt61 ubt_‘i’ ubt90 ubdum_‘i’ ubdum90) ///
addstat(F-test 1,‘F1’, P1 , ‘p1’, F-test 2,‘r(F)’, P2,‘r(p)’) ///
coefastr paren bdec(2) adjr2 symbol(***, **, *) nocons append

}


*============================================================
*Table 2:  Bank Branch Expansion and Poverty: Reduced Form
*============================================================

local table2outcomes " h2 head lrwagenew lsalpw"

*Column 2 - Rural Headcount Ratio
*----------------------------------------------------------
quietly reg h1 ubt61 ubt_‘i’ ubt90 ubdum_‘i’ ubdum90
$controls stdum* yrdum*, cluster(state)

quietly test ubt61+ubt_‘i’=0
local F12 = r(F)
local p12 = r(p)
quietly test ubt61+ubt_‘i’+ubt90=0

outreg2 using "table_2_‘i’.txt", se ///
addtext(State and year dummies, YES, Other controls, YES) ///
keep(ubt61 ubt_‘i’ ubt90 ubdum_‘i’ ubdum90) ///
addstat(F-test 1,‘F12’, P1 , ‘p12’, F-test 2,  ‘r(F)’, P2, ‘r(p)’)  ///
coefastr  paren bdec(2) adjr2 symbol(***, **, *)  nolabel nocons replace




*Columns 3 & 4 - Urban & Aggregate Headcount Ratio
* & Columns 5 & 6 - Agricultural & Factory wages
*------------------------------------------------------------
foreach v of local table2outcomes {

quietly reg ‘v’ ubt61 ubt_‘i’ ubt90   ubdum_‘i’ ubdum90
$controls stdum* yrdum*, cluster(state)

quietly test ubt61+ubt_‘i’ = 0
quietly test ubt61+ubt_‘i’ = 0
```

```
local F12 = r(F)
local p12 = r(p)
quietly test ubt61+ubt_‘i’+ubt90=0

outreg2 using "table_2_‘i’.txt", se ///
addtext(State and year dummies, YES, Other controls, YES) ///
keep(ubt61 ubt_‘i’ ubt90 ubdum_‘i’ ubdum90) ///
addstat(F-test 1,‘F12’, P1 , ‘p12’, F-test 2,  ‘r(F)’, P2, ‘r(p)’) ///
coefastr  paren bdec(2) adjr2 symbol(***, **, *)  nolabel nocons append
}




*==================================================
*Table 3: Bank Branch Expansion and Poverty:
* Instrumental Variables Evidence
*==================================================
local table3 "h1 h2 head"


*Column 1: OLS with one regressor:
*Number branches opened in rural unbanked locations per capita
*------------------------------------------------------------
quietly reg h1 runbfor stdum* yrdum*, cluster(state)
outreg2 using "table_3_‘i’.txt", se ///
keep(runbfor) ///
coefastr paren bdec(2) adjr2 symbol(***, **, *) nolabel
    nocons ///
addtext(State and year dummies, YES, Other controls, YES)
    replace


*Column 2: OLS with additional controls
*------------------------------------------
quietly reg h1 runbfor ubt61 ubdum_‘i’ ubdum90
$controls  stdum* yrdum*, cluster(state)

outreg2 using "table_3_‘i’.txt", se   ///
keep(runbfor ubt61 ubdum_‘i’ ubdum90)   ///
coefastr paren bdec(2) adjr2 symbol(***, **, *) nolabel nocons  ///
addtext(State and year dummies, YES, Other controls, YES) append


*Columns 3, 4 & 5  - IV: Rural, Urban & Aggraget Headcount Ratio
```

```
*----------------------------------------------------------------
foreach v of local table3 {
quietly reg ‘v’  runbfor ubt61 ubdum_‘i’ ubdum90
$controls stdum* yrdum* ///
(ubt_‘i’ ubt61 ubdum_‘i’ ubt90 ubdum90
$controls stdum* yrdum* ), cluster(state)

outreg2 using "table_3_‘i’.txt", se ///
keep(runbfor ubt61 ubdum_‘i’ ubdum90)  ///
coefastr paren bdec(2)  symbol(***, **, *) nolabel nocons ///
addtext(State and year dummies, YES, Other controls, YES) append


capture drop resinc
predict resinc, res
quietly reg resinc  ubt_‘i’ ubt90, noc robust
capture drop sargan
ge sargan=_result(7)*_result(1)
capture drop prob
ge prob=chiprob(3,sargan)
disp sargan prob

}


*Columns 6, 7, 8 -- restricted sample for h1
*-----------------------------------------------------

*column 6, earlier then 1990
*-----------------------------------------------------
quietly reg h1 runbfor ubt61 ubdum_‘i’  popt61 popt_‘i’ popd_‘i’
rursht61 rursht_‘i’ rurshd_‘i’ stint61 stint_‘i’ stind_‘i’ stdum*
yrdum* ///
(ubt_‘i’ ubt61 ubdum_‘i’ popt61 popt_‘i’ popd_‘i’ rursht61 rursht_‘i’
rurshd_‘i’  stint61 stint_‘i’ stind_‘i’  stdum* yrdum* )
if year < 1990, cluster(state)

outreg2 using "table_3_‘i’.txt", se ///
keep(runbfor ubt61 ubdum_‘i’) ///
coefastr paren bdec(2) symbol(***, **, *) nolabel nocons ///
addtext(State and year dummies, YES, Other controls, YES) append



*column 7, earlier then treatment year (BP05 1976)
```

```
*----------------------------------------------------
quietly reg h1 runbfor ubt61 ubdum90 popt_`i' popt90 popdum90
rursht_`i' rursht90 rurshd90  stint_`i' stint90 stindum90
stdum* yrdum* ///
(ubt90 ubt_`i' ubdum90 popt_`i' popt90 popdum90  rursht_`i'
rursht90 rurshd90  stint_`i' stint90 stindum90 stdum* yrdum* )
if year > 1976, cluster(state)

outreg2 using "table_3_`i'.txt", se ///
keep(runbfor ubt61 ubdum90)  ///
coefastr  paren bdec(2) symbol(***, **, *) nolabel nocons ///
addtext(State and year dummies, YES, Other controls, YES) append



*column 8, survey years
*------------------------------------------------------
quietly reg h1 runbfor ubt61 ubdum_`i'  ubdum90
$controls stdum* yrdum* (ubt_`i' ubt61 ubdum_`i' ubt90 ubdum90
$controls stdum* yrdum* ) if round~=.,cluster(state)

outreg2 using "table_3_`i'.txt", se ///
keep(runbfor ubt61 ubdum_`i'  ubdum90) ///
coefastr paren bdec(2) symbol(***, **, *) nolabel nocons ///
addtext(State and year dummies, YES, Other controls, YES) append

capture drop resinc
predict resinc, res

quietly reg resinc  ubt_`i' ubt90, noc robust
capture drop sargan

ge sargan=_result(7)*_result(1)
capture drop prob

ge prob=chiprob(3,sargan)
disp sargan prob



*Columns 9 & 10  - IV: Agricultural & Factory wages
*-----------------------------------------------------
local table31 "lrwagenew lsalpw"
*-----------------------------------------------------
foreach v of local table31 {
```

```
quietly reg `v'  runbfor ubt61 ubdum_`i' ubdum90
$controls stdum* yrdum* ///
(ubt_`i' ubt61 ubdum_`i' ubt90 ubdum90
$controls stdum* yrdum* ), cluster(state)

outreg2 using "table_3_`i'.txt", se ///
keep(runbfor ubt61 ubdum_`i' ubdum90) ///
coefastr paren bdec(2) symbol(***, **, *) nolabel nocons ///
addtext(State and year dummies, YES, Other controls, YES) append

capture drop resinc
predict resinc, res

quietly reg resinc  ubt_`i' ubt90, noc robust
capture drop sargan

ge sargan=_result(7)*_result(1)
capture drop prob

ge prob=chiprob(3,sargan)
disp sargan prob

}



*==================================================
*Table 4: Rural credit and Savings and Poverty:
* Instrumental Variables Evidence
*==================================================

local table4outcomes "h2 head"

*Column 1. Rural Headcount Ratio --
* with Rural bank credit share
*--------------------------------------------

quietly reg h1 rurcrsh ubt61 ubdum_`i' ubdum90
$controls stdum* yrdum* ///
(ubt_`i' ubt61 ubdum_`i' ubt90 ubdum90
$controls stdum* yrdum* ), cluster(state)

outreg2 using "table_4_`i'.txt", se  ///
keep(rurcrsh ubt61 ubdum_`i' ubdum90)  ///
coefastr paren bdec(2) adjr2 symbol(***, **, *) nolabel nocons ///
```

```
addtext(State and year dummies, YES, Other controls, YES) replace

capture drop resinc
predict resinc, res

quietly reg resinc  ubt_`i' ubt90, noc robust
capture drop sargan

ge sargan=_result(7)*_result(1)
capture drop prob

ge prob=chiprob(3,sargan)
disp sargan prob


*Column 2. Rural Headcount Ratio --
* with Rural bank savings share
*---------------------------------------

quietly reg h1 rursavsh ubt61 ubdum_`i' ubdum90
$controls stdum* yrdum* ///
(ubt_`i' ubt61 ubdum_`i' ubt90 ubdum90
$controls stdum* yrdum* ), cluster(state)

outreg2 using "table_4_`i'.txt", se  ///
keep(rursavsh ubt61 ubdum_`i' ubdum90)  ///
coefastr  paren bdec(2) symbol(***, **, *) nolabel nocons ///
addtext(State and year dummies, YES, Other controls, YES) append

capture drop resinc
predict resinc, res

quietly reg resinc  ubt_`i' ubt90, noc robust
capture drop sargan

ge sargan=_result(7)*_result(1)
capture drop prob

ge prob=chiprob(3,sargan)
disp sargan prob

*Columns 3 & 5. Urban and Aggregate Headcount Ratio --
* with Rural bank credit share
*---------------------------------------------------------
foreach v of local table4outcomes {
```

```
quietly reg ‘v’ rurcrsh ubt61 ubdum_‘i’ ubdum90
$controls stdum* yrdum* ///
(ubt_‘i’ ubt61 ubdum_‘i’ ubt90 ubdum90
$controls stdum* yrdum* ), cluster(state)

outreg2 using "table_4_‘i’.txt", se  ///
keep(rurcrsh ubt61 ubdum_‘i’ ubdum90)  ///
coefastr  paren bdec(2) adjr2 symbol(***, **, *) nolabel nocons ///
addtext(State and year dummies, YES, Other controls, YES) append

capture drop resinc
predict resinc, res

quietly reg resinc  ubt_‘i’ ubt90, noc robust
capture drop sargan

ge sargan=_result(7)*_result(1)
capture drop prob

ge prob=chiprob(3,sargan)
disp sargan prob

*Columns 4 & 6. Urban and Aggregate Headcount Ratio --
*with Rural bank savings share
*------------------------------------------------------

quietly reg ‘v’ rursavsh ubt61 ubdum_‘i’ ubdum90
$controls stdum* yrdum*  ///
(ubt_‘i’ ubt61 ubdum_‘i’ ubt90 ubdum90
$controls stdum* yrdum* ), cluster(state)


outreg2 using "table_4_‘i’.txt", se ///
keep(rursavsh ubt61 ubdum_‘i’ ubdum90)  ///
coefastr paren bdec(2) adjr2 symbol(***, **, *) nolabel nocons ///
addtext(State and year dummies, YES, Other controls, YES) append

capture drop resinc
predict resinc, res

quietly reg resinc  ubt_‘i’ ubt90, noc robust
capture drop sargan

ge sargan=_result(7)*_result(1)
```

```
capture drop prob

ge prob=chiprob(3,sargan)
disp sargan prob

}




*=======================================================
*Table 5: Bank Branch Expansion And Poverty Reduction:
* Robustness Checks
*=======================================================

macro define policy "clre hcapsh2 odevsh2"
macro define politics "pcona pjan phdlft preg phind"


*Column 1. Rural Headcount Ratio --
* with policy variables
*-----------------------------------------------
quietly reg h1  runbfor $policy  ubt61 ubdum_`i'
ubdum90  $controls stdum* yrdum* ///
($policy ubt_`i' ubt61 ubdum_`i' ubt90 ubdum90
$controls stdum* yrdum* ), cluster(state)

outreg2 using "table_5_`i'.txt", se  ///
keep(runbfor $policy)   ///
coefastr  paren bdec(2) adjr2 symbol(***, **, *) nolabel nocons ///
addtext(State and year dummies, YES, Other controls, YES) replace

capture drop resinc
predict resinc, res

quietly reg resinc  ubt_`i' ubt90, noc robust
capture drop sargan

ge sargan=_result(7)*_result(1)
capture drop prob

ge prob=chiprob(3,sargan)
disp sargan prob
```

```
*Column 2. Rural Headcount Ratio --
* with policy and political variables
*--------------------------------------------
quietly reg h1 runbfor $policy  $politics ubt61
ubdum_`i' ubdum90  $controls stdum* yrdum* ///
($policy $politics ubt_`i' ubt61 ubdum_`i'
ubt90 ubdum90 $controls stdum* yrdum* ), cluster(state)

outreg2 using "table_5_`i'.txt", se  ///
keep(runbfor $policy $politics )    ///
coefastr paren bdec(2) adjr2 symbol(***, **, *) nolabel nocons ///
addtext(State and year dummies, YES, Other controls, YES) append

capture drop resinc
predict resinc, res
quietly reg resinc  ubt_`i' ubt90, noc robust
capture drop sargan
ge sargan=_result(7)*_result(1)
capture drop prob
ge prob=chiprob(3,sargan)
disp sargan prob


*Column 3. Urban Headcount Ratio --
* with policy variables
*-----------------------------------------------
quietly reg h2  runbfor $policy  ubt61 ubdum_`i'
ubdum90 $controls stdum* yrdum* ///
($policy ubt_`i' ubt61 ubdum_`i' ubt90 ubdum90
$controls stdum* yrdum* ), cluster(state)

outreg2 using "table_5_`i'.txt", se  ///
keep(runbfor $policy)    ///
coefastr paren bdec(2) adjr2 symbol(***, **, *) nolabel nocons ///
addtext(State and year dummies, YES, Other controls, YES) append

capture drop resinc
predict resinc, res

quietly reg resinc  ubt_`i' ubt90, noc robust
capture drop sargan

ge sargan=_result(7)*_result(1)
capture drop prob
```

```
ge prob=chiprob(3,sargan)
disp sargan prob


*Column 4. Urban Headcount Ratio --
* with policy and political variables
*--------------------------------------------------
quietly reg h2 runbfor $policy $politics ubt61 ubdum_`i'
ubdum90 $controls stdum* yrdum* ///
($policy $politics ubt_`i' ubt61 ubdum_`i' ubt90 ubdum90
$controls stdum* yrdum* ), cluster(state)

outreg2 using "table_5_`i'.txt", se  ///
keep(runbfor $policy $politics)    ///
coefastr paren bdec(2) adjr2 symbol(***, **, *) nolabel nocons ///
addtext(State and year dummies, YES, Other controls, YES) append

capture drop resinc
predict resinc, res

quietly reg resinc  ubt_`i' ubt90, noc robust
capture drop sargan

ge sargan=_result(7)*_result(1)
capture drop prob

ge prob=chiprob(3,sargan)
disp sargan prob
}
*------------------------------------------
```

```
*============================================
* Date : October 2021
* Author: Nino Buliskeria
* Paper: Do Rural Banks Matter That Much?
* Burgess and Pande (2005) Reconsidered
* by Nino Buliskeria and Jaromir Baxa.

* This code allows to replicates:
* Burgess, R., & Pande, R. (2005).
* "Do rural banks matter?
* evidence from the Indian social banking experiment."
* American Economic Review, 95(3)
*
* This code also offers possibility to extend
* instrumental variables
* exercise presented in Burgess and Pande (2005)
* by introducing additional
* cut-years when summarizing the rural
* bank branch opening trend reversal.
*------------------------------------------
clear all
macro drop _all
*------------------------------------------
*load data
*use "burgesspandeaerfinal.dta"
use "buliskeriabaxafinal.dta"

/*
*Additional Characteristics
infmor "State-wise infant mortality rate for 1961"
litr "State-wise literacy rates for 1961 in
population aged five years and above."
prod "Crop production per farm worker"
rwork "Percent laborers of rural farm workers"

var in 1961 * (T - 2000) trend = vart
var in 1961 * Post-T dummy = vard
*/

drop if year>2000

// The cut-years are named alphabethically,
first being A, up to J.
global manes A B C D E F G H I J
```

```
*___EDIT HERE: ____________________
* Burgess and Pande (2005):
global years 77 90
*txbreak: 68 74 79 84 89
* Historic years:
* global years 67 72 77 80 85 90
*___________________________________

*Generate Control Dummy Variables x*trend and x*dummy

local num : word count $years
scalar n = ‘num’

forvalues i = 1(1)10{
local l‘i’ : word ‘i’ of $manes
di "‘l‘i’’"
}

forvalues i = 1(1)10{
local y‘i’ : word ‘i’ of $years
di "‘y‘i’’"
}

forvalues i = 1(1)‘=n’{

gen trend‘l‘i’’= year - (19‘y‘i’’-1)
replace trend‘l‘i’’=0 if year <19‘y‘i’’

gen dum‘l‘i’’=0
replace dum‘l‘i’’=1 if year>(19‘y‘i’’-1)

gen ubt‘l‘i’’=unb61*trend‘l‘i’’*dum‘l‘i’’
gen ubdum‘l‘i’’=unb61*dum‘l‘i’’

gen popt‘l‘i’’=((pop61)/larea)*trend‘l‘i’’*dum‘l‘i’’
gen popd‘l‘i’’=((pop61)/larea)*dum‘l‘i’’

gen stint‘l‘i’’=stinc61*trend‘l‘i’’
gen stind‘l‘i’’=stinc61*dum‘l‘i’’

gen rursht‘l‘i’’=rursh61*trend‘l‘i’’*dum‘l‘i’’
gen rurshd‘l‘i’’=rursh61*dum‘l‘i’’

gen infmort‘l‘i’’=infmor*trend‘l‘i’’*dum‘l‘i’’
gen infmord‘l‘i’’=infmor*dum‘l‘i’’
```

```
gen litrt`l`i''=litr*trend`l`i''*dum`l`i''
gen litrd`l`i''=litr*dum`l`i''

gen prodt`l`i''=prod*trend`l`i''*dum`l`i''
gen prodd`l`i''=prod*dum`l`i''

gen rworkt`l`i''=rwork*trend`l`i''*dum`l`i''
gen rworkd`l`i''=rwork*dum`l`i''
}
*--------------------------------------------
*Define Global Variables
global In "popt61 rursht61 stint61"
global ubt61 "ubt61"

forvalues i = 1(1)`=n'{
global `l`i'' "popt`l`i'' popd`l`i'' rursht`l`i'' rurshd`l`i''
stint`l`i'' stind`l`i'' infmort`l`i'' infmord`l`i'' litrt`l`i''
litrd`l`i'' prodt`l`i'' prodd`l`i'' rworkt`l`i'' rworkd`l`i'' "
}

forvalues i = 1(1)`=n'{
global ubt`l`i''"ubt`l`i''"
}

forvalues i = 1(1)`=n'{
global ubdum`l`i''"ubdum`l`i''"
}
*--------------------------------------------------
*Define Control Variables:

global controls $In $A $B $C $D $E $F $G $H $I $J
di "$controls"

global ubt $ubtA $ubtB $ubtC $ubtD $ubtE $ubtF $ubtG $ubtH $ubtI $ubtJ
global ubdum $ubdumA $ubdumB $ubdumC $ubdumD $ubdumE $ubdumF
$ubdumG $ubdumH $ubdumI $ubdumJ
global main $ubt61 $ubt $ubdum
di "$main"

global iv $ubt61 $ubdumA $ubdumB $ubdumC $ubdumD $ubdumE
$ubdumF $ubdumG $ubdumH $ubdumI $ubdumJ
di "$iv"

*---------------------------------------------
```

```
*___EDIT HERE: _____________
* Burgess and Pande (2005):
global test "ubtA + ubtB"
* Historic years:
* global test "ubtA + ubtB  + ubtC + ubtD + ubtE + ubtF"
*__________________________

*==========================
// STEP 2: Figrue 1
*==========================
* generate the coefficients for the Figure 1
drop if year>2000
*Figure 1: rural branches in unbanked locations (with controls)
areg runbfor unbk62-unbk00 rshare62-rshare00 popn62-popn00 stin62-stin00 infmor62-
    prod62-prod00 rwork62-rwork00 stdum*,
    absorb(year) cluster(state)
*Start NB
mata: b=st_matrix("e(b)")'
getmata b, force
gen temp=b[_n-1]
replace b=temp
replace b=. if _n==_N
replace b = . if state > 1
label variable b "rural branches in unbanked locations
    (with controls)"

*summarize these trend reversals by a linear trend break model:
reg runbfor $main $controls stdum* yrdum*,cluster(state)
mata: coef=st_matrix("e(b)")'
getmata coef, force
replace coef=. if _n> 2*n+1

mkmat $main, matrix(X)
mkmat coef if coef < . , matrix(B)
matrix tr1 = X*B
svmat tr1, names(trd)

egen trend_m = mean(trd), by (year)
gen mtrend = trend_m if state == 1
gen trend = mtrend - mtrend[1]

label variable trend "rural branches in unbanked
    locations (trend break)"
#delimit ;
line b trend year,
```

```
legend(size(small) cols(1) symxsize(7))

lpattern(solid longdash_dot)
lcolor(black black)
graphregion(color(white))

xlabel(, labsize(small))
ylabel(, labsize(small) angle(horizontal))

xtitle("")
ytitle("Initial financial development X year - coefficient
{it}{&gamma}{subscript:t}",
size(small));
#delimit cr

*======================
//STEP 3: Regressions:
*======================
*For the regressions Burgess and Pande (2005) drop
*   Haryana before 1965:
drop if state==5 & year<1965

*=========================================
*Table 1: Banking as a Function of Initial financial development
local table1outcomes "rurcrsh rursavsh bfor priorsh pcoopsh"
*---------------------------------------------------------
*Column 1: Branches in rural unbanked locations

reg runbfor $main $controls stdum* yrdum*,cluster(state)

test ubt61+ubtA=0
local F1 = r(F)
local p1 = r(p)
*test ubt61+ubtA+ubtB+ubtC+ubtD+ubtE=0
test ubt61 + $test  = 0

outreg2 using "table_1.tex", se ///
addtext(State and year dummies, YES, Other controls, YES) ///
keep($main) ///
addstat(F-test 1,‘F1’, P1 , ‘p1’, F-test 2,  ‘r(F)’, P2, ‘r(p)’) ///
coefastr paren bdec(2) adjr2 symbol(***, **, *) nocons   replace

*---------------------------------------------------------
*Columns 2 & 3 - Rural bank: Rural bank Credit and Savings share
*& Column 4 - Branches in banked locations
```

```
*&& Columns 5 & 6 - Credit share: Priority sector and Cooperative

foreach v of local table1outcomes {
reg ‘v’ $main $controls stdum* yrdum*, cluster(state)

test ubt61+ubtA=0
local F1 = r(F)
local p1 = r(p)
test ubt61 + $test = 0

outreg2 using "table_1.tex", se ///
addtext(State and year dummies, YES, Other controls, YES) ///
keep($main) ///
addstat(F-test 1,‘F1’, P1 , ‘p1’, F-test 2,‘r(F)’, P2,‘r(p)’) ///
coefastr paren bdec(2) adjr2 symbol(***, **, *) nocons append
}

*seeout using "table_1.txt"

*============================================================
*Table 2:  Bank Branch Expansion and Poverty: Reduced Form
local table2outcomes "h2 head lrwagenew lsalpw"
*---------------------------------------------------------
*Column 2 - Rural Headcount Ratio
reg h1 $main $controls stdum* yrdum*, cluster(state)

test ubt61+ubtA=0
local F12 = r(F)
local p12 = r(p)
*test ubt61+ubtA+ubtB+ubtC+ubtD+ubtE=0
test ubt61 + $test = 0

outreg2 using "table_2.tex", se ///
addtext(State and year dummies, YES, Other controls, YES) ///
keep($main) ///
addstat(F-test 1,‘F12’, P1 , ‘p12’, F-test 2,  ‘r(F)’, P2, ‘r(p)’)  ///
coefastr  paren bdec(2) adjr2 symbol(***, **, *)  nolabel nocons replace


*---------------------------------------------------------
*Columns 3 & 4 - Urban & Aggregate Headcount Ratio
* & Columns 5 & 6 - Agricultural & Factory wages

foreach v of local table2outcomes {
 reg ‘v’ $main $controls stdum* yrdum*, cluster(state)
```

```
test ubt61+ubtA=0
test ubt61+ubtA=0
local F12 = r(F)
local p12 = r(p)
test ubt61 + $test = 0

outreg2 using "table_2.tex", se ///
addtext(State and year dummies, YES, Other controls, YES) ///
keep($main) ///
addstat(F-test 1,‘F12’, P1 , ‘p12’, F-test 2,  ‘r(F)’, P2, ‘r(p)’) ///
coefastr  paren bdec(2) adjr2 symbol(***, **, *)  nolabel nocons append
}
*seeout using "table_2.txt"

*=============================================
*Table3: BANK BRANCH EXPANSION AND POVERTY:
*INSTRUMENTAL VARIABLES EVIDENCE
*===========================================
*___EDIT HERE: _________________
*choose value for "Tr" - sample restriction for Table 3, column 7.
*(after treatment period T - 1 )
*Burgess and Pande (2005):
 scalar Tr = 1976
*Historic years:
*scalar Tr = 1966
*______________________________

local table3 "h1 h2 head lrwagenew lsalpw"
*---------------------------------------------
*Column 1: OLS with one regressor:
*Number branches opened in rural unbanked locations per capita

reg h1 runbfor stdum* yrdum*, cluster(state)

outreg2 using "table_3.tex", se keep(runbfor) coefastr  paren ///
bdec(2) adjr2 symbol(***, **, *) nolabel nocons ///
addtext(State and year dummies, YES, Other controls, YES) replace


*------------------------------------------------
*Column 2: OLS with additional controls

reg h1 runbfor $iv $controls stdum* yrdum*, cluster(state)
```

```
outreg2 using "table_3.tex", se  ///
keep(runbfor $iv)    ///
addtext(State and year dummies, YES, Other controls, YES) ///
coefastr paren bdec(2) adjr2 symbol(***, **, *) nolabel nocons append


*---------------------------------------------------
*Columns 3, 4, 5  - IV: Rural, Urban & Aggraget Headcount Ratio
*& Columns 9, 10  - IV: Agricultural & Factory wages

foreach v of local table3{
reg ‘v’  runbfor $iv $controls stdum* yrdum* ///
($main $controls stdum* yrdum*), cluster(state)

outreg2 using "table_3.tex", se ///
keep(runbfor $iv) ///
addtext(State and year dummies, YES, Other controls, YES) ///
coefastr paren bdec(2) adjr2 symbol(***, **, *) nolabel nocons append

capture drop resinc
predict resinc, res
reg resinc $ubt, noc robust
capture drop sargan
ge sargan=_result(7)*_result(1)
capture drop prob
ge prob=chiprob(3,sargan)
disp sargan prob
}

*-----------------------------------------------
*Columns 6, 7, 8 -- *restricted sample for h1
*-----------------------------------------------
*column 6, earlier then 1990

reg h1 runbfor  $iv $controls stdum* yrdum* ///
($main $controls stdum* yrdum*) if year<1990, cluster(state)

outreg2 using "table_3.tex", se ///
keep(runbfor $iv)  ///
addtext(State and year dummies, YES, Other controls, YES) ///
coefastr paren bdec(2) symbol(***, **, *) nolabel nocons append
*coefastr paren bdec(2) adjr2 symbol(***, **, *) nolabel nocons append
*----------------------------------------------
*column 7, earlier then treatment year (BP05 1976)
```

```
*Tr set in file 0_master.do
reg h1 runbfor  $iv $controls stdum* yrdum* ///
($main $controls stdum* yrdum*) if year>Tr, cluster(state)

outreg2 using "table_3.tex", se ///
keep(runbfor $iv)  ///
addtext(State and year dummies, YES, Other controls, YES) ///
coefastr paren bdec(2) adjr2 symbol(***, **, *) nolabel nocons append

*---------------------------------------------
*column 8, survey years

reg h1 runbfor $iv $controls stdum* yrdum* ///
($main $controls stdum* yrdum*) if round~=.,cluster(state)

outreg2 using "table_3.tex", se ///
keep(runbfor $iv) ///
addtext(State and year dummies, YES, Other controls, YES) ///
coefastr paren bdec(2) adjr2 symbol(***, **, *) nolabel nocons append

capture drop resinc
predict resinc, res
reg resinc $ubt, noc robust
capture drop sargan
ge sargan=_result(7)*_result(1)
capture drop prob
ge prob=chiprob(3,sargan)
disp sargan prob

*seeout using "table_3.txt"

*============================================
*Table4: RURAL CREDIT AND SAVINGS AND POVERTY:
* INSTRUMENTAL VARIABLES EVIDENCE
*============================================
*Table 4: rural credit and savings
local table4outcomes " h2 head "
*-------------------------------------
*Column 1
reg h1  rurcrsh $iv $controls stdum* yrdum* ///
($main $controls stdum* yrdum* ), cluster(state)

outreg2 using "table_4.tex", se  ///
keep(rurcrsh $iv)  ///
coefastr  paren bdec(2) adjr2 symbol(***, **, *) nolabel nocons ///
```

```
addtext(State and year dummies, YES, Other controls, YES) replace

capture drop resinc
predict resinc, res
reg resinc $ubt, noc robust
capture drop sargan
ge sargan=_result(7)*_result(1)
capture drop prob
ge prob=chiprob(3,sargan)
disp sargan prob
*------------------------------------------
*Column 2
reg h1  rursavsh $iv $controls stdum* yrdum* ///
($main $controls stdum* yrdum* ), cluster(state)

outreg2 using "table_4.tex", se  ///
keep(rursavsh $iv)  ///
coefastr  paren bdec(2) adjr2 symbol(***, **, *) nolabel nocons ///
addtext(State and year dummies, YES, Other controls, YES) append

capture drop resinc
predict resinc, res
reg resinc $ubt, noc robust
capture drop sargan
ge sargan=_result(7)*_result(1)
capture drop prob
ge prob=chiprob(3,sargan)

*-------------------------------------------
*Columns 3 and 5
disp sargan prob
foreach v of local table4outcomes {
reg ‘v’  rurcrsh $iv $controls stdum* yrdum* ///
($main  $controls stdum* yrdum* ), cluster(state)

outreg2 using "table_4.tex", se  ///
keep(rurcrsh $iv)  ///
coefastr  paren bdec(2) adjr2 symbol(***, **, *) nolabel nocons ///
addtext(State and year dummies, YES, Other controls, YES) append

capture drop resinc
predict resinc, res
reg resinc $ubt, noc robust
capture drop sargan
ge sargan=_result(7)*_result(1)
```

```
capture drop prob
ge prob=chiprob(3,sargan)
disp sargan prob

*---------------------------------------------
*Columns 4 and 6
reg `v' rursavsh $iv $controls stdum* yrdum* ///
($main $controls stdum* yrdum* ), cluster(state)

outreg2 using "table_4.tex", se ///
keep(rursavsh $iv)  ///
coefastr  paren bdec(2) adjr2 symbol(***, **, *) nolabel nocons ///
addtext(State and year dummies, YES, Other controls, YES) append

capture drop resinc
predict resinc, res
reg resinc $ubt, noc robust
capture drop sargan
ge sargan=_result(7)*_result(1)
capture drop prob
ge prob=chiprob(3,sargan)
disp sargan prob

}

*seeout using "table_4.txt"

*====================================
*Table5: BANK BRANCH EXPANSION AND
* POVERTY REDUCTION: ROBUSTNESS CHECKS
*====================================
*Table 5: robustness
macro define policy  "clre hcapsh2 odevsh2"
macro define politics  "pcona pjan phdlft preg phind"
*---------------------------------------------
*Column 1
reg h1  runbfor $policy  $iv  $controls stdum* yrdum* ///
($policy $main $controls stdum* yrdum* ), cluster(state)

outreg2 using "table_5.tex", se  ///
keep(runbfor $policy)   ///
coefastr  paren bdec(2) adjr2 symbol(***, **, *) nolabel nocons ///
addtext(State and year dummies, YES, Other controls, YES) replace

*coefastr  paren bdec(2) symbol(***, **, *) nolabel nocons ///
```

```
capture drop resinc
predict resinc, res
reg resinc $ubt, noc robust
capture drop sargan
ge sargan=_result(7)*_result(1)
capture drop prob
ge prob=chiprob(3,sargan)
disp sargan prob
*-----------------------------------------
*Column 2
reg h1  runbfor $policy  $politics $iv  $controls stdum* yrdum* ///
($policy $politics $main $controls stdum* yrdum* ), cluster(state)

outreg2 using "table_5.tex", se     ///
keep(runbfor $policy $politics ) ///
coefastr paren bdec(2) adjr2 symbol(***, **, *) nolabel nocons ///
addtext(State and year dummies, YES, Other controls, YES) append

capture drop resinc
predict resinc, res
reg resinc $ubt, noc robust
capture drop sargan
ge sargan=_result(7)*_result(1)
capture drop prob
ge prob=chiprob(3,sargan)
disp sargan prob
*-------------------------------------------
*Column 3
reg h2  runbfor $policy  $iv $controls stdum* yrdum* ///
($policy $main $controls stdum* yrdum* ), cluster(state)

outreg2 using "table_5.tex", se  ///
keep(runbfor $policy)   ///
coefastr paren bdec(2) adjr2 symbol(***, **, *) nolabel nocons ///
addtext(State and year dummies, YES, Other controls, YES) append
*coefastr paren bdec(2) symbol(***, **, *) nolabel nocons ///

capture drop resinc
predict resinc, res
reg resinc $ubt, noc robust
capture drop sargan
ge sargan=_result(7)*_result(1)
capture drop prob
ge prob=chiprob(3,sargan)
```

```
disp sargan prob
*--------------------------------------------
*Column 4
reg h2  runbfor $policy $politics $iv $controls stdum* yrdum* ///
($policy $politics $main $controls stdum* yrdum* ), cluster(state)

outreg2 using "table_5.tex", se  ///
keep(runbfor $policy $politics)  ///
coefastr paren bdec(2) adjr2 symbol(***, **, *) nolabel nocons ///
addtext(State and year dummies, YES, Other controls, YES) append

capture drop resinc
predict resinc, res
reg resinc $ubt, noc robust
capture drop sargan
ge sargan=_result(7)*_result(1)
capture drop prob
ge prob=chiprob(3,sargan)
disp sargan prob
*seeout using "table_5.txt"
*===================================
```