Table of Contents

This paper can be publishable if it can be revised to address potential problems and then resubmitted.

1. Outline and Contributions

This paper (Arnould et al., 2023) discusses the bank reaction towards the anticipated regulation change. The paper uses Loan to Value Ratio (LTV) as the measurement for bank reaction to the uncertainty. This uncertainty is raised by policy change that will give supervisors full discretions on the reduction of risk weights for the bank who use standardised approach (SA), which weight is much higher than internal ratings‑based (IRB) modelling approach's risk weights. In the paper, they use Difference in Differences (DID) method to try to prove that banks possible to be eligible for this policy may reduce the LTV levels to get a higher expected reduction for the relative risk weights. This effect is moderated by the interaction between supervisors and banks.

For me, **the key contributions** in the paper are the following:

Most importantly, they use the cost of prepositioning estimated by their research to capture the lower bound cost of the supervision from the regulator. Since most entities used SA are small banks, regulatory capital costs among these banks are invisible as they are usually not listed in the stock market. If their estimates granularity and accuracy can be higher, this paper can give the industry and academia a clearer conceptual framework to figure out those banks' relevance shadow costs that is unrevealed before (Kisin and Manela, 2016).

Second, announcement effect is proposed by the paper. This pre-trend effect before the final implementation can extend the existing discussion of the role of policy change further to different stages. This paper's finding of the regulated bank reaction in the announcement stage can add the current understanding of policy's impact in the whole stages.

Further, it adds to the limited literature on uncertainty effect from regulators. Banks with more interactions with supervisors will adjust the asset composition early. It shows the positive side of the risk drawn by the uncertainty. This uncertainty from the policy change may reach expected outcome set by policymakers.

2. Potential Issues of the Paper

Evaluating the reasoning of the paper, I may **raise those concerns** that will mitigate this paper's importance and potential contributions as discussed above:

For assumptions, this paper assumes that banks can learn whether they can be eligible for the new policy, so the paper set *Eligible* to distinguished treatment group from control group. But it may raise the questions related to these two groups.

One is that it sets "no risk scores above two-third" as the threshold for the eligible group. This criteria for the low risk profile does not have a clear reasoning in the paper. If we select other thresholds, the final estimates will be different.

Another is that the two groups already know they are likely or not likely to be treated, so that's why the paper find the opposite outcome between the groups. In control group, banks are "treated" actually (they reacted toward the change, too), so we cannot infer that the core coefficient of DID setting is the clean effect of announcement. These concerns will impede the accuracy of regulatory capital costs calculation.

For DID setting, it uses two treatment date: (1) the consultation paper (CP) issuing; (2) policy statement (PS) issuing. So, it is two-timing treatments DID—staggered DID. In parallel trends diagnostic test, the paper shows the dynamic effect from the first treatment (CP). The parallel trends assumption may not hold in second treatment (PS). Also, dynamic effect will be problematic in the estimation, especially in time varying dynamic. Different banks in treatment group will have different effects, and the specific bank will have different effect between first and second treatment. For the *Post (PS)* treatment, the effect of the first treatment will impact the estimation of our second treatment effect (PS) since the time gap is short. The paper needs to conduct parallels trends test for the second treatment. We advise the paper to choose first treatment time only to capture total anticipated reaction by using dynamic DID model. Or the paper can use synthetic method to isolate CP and SP effect. Estimation bias will impair the economic explanation of announcement role. We cannot tease out the possibility that it is treatment and control group's different risk aversions that gives them different reaction about LTV since the allocation for treatment group and control group is not random. The announcement causal effect is weakened at this situation.

This paper gives the reasoning that higher communication between supervisor and bank will make banks understand better whether they are eligible. Nevertheless, the research only observes the earlier adjustments in the higher frequency of communication group compared to the low frequency group. We can only conclude that the higher frequency of communication gives banks more information advantage and quicker information dissemination. Those banks who have information advantage take actions early, but it doesn't necessarily mean they know they will be finally in the Supervisory Review and Evaluation Process (SREP) eligible list. Therefore, if we want to check whether interactions can give the hints for eligibility, we need to compare the successful rate of banks enrolled in the policy among all high-interaction eligible banks with that of low-interaction banks. For another, the paper needs to explain why they set 1.5 times higher than the mean interactions as standards to distinguish higher group from lower group. Further, highly supervisory interaction and bank riskiness (LTV) may be jointly determined. And it would better investigate which way of communications (e.g., meetings, calls, and site visits) induce banks to lower the LTV.

Also, section 5.4 in the paper compares the 2019 eligible banks with 2020 eligible banks. This DID-setting can be also problematic. The paper's reasoning is that 2019 cohorts' prepositioning crowded out the following eligible banks' prepositioning by adjusting their portfolios into more low-risk investments later. But the null result of Table 8 in the paper undermines this reasoning as it just means no difference between 2019 and 2020 cohorts. We would better see 2020 cohort vs. clean control's null result to justify this reasoning. Besides, in this DID design, the treatment group at the first stage (2019 eligible) become control group, which can lead to bias. Additional parallel tests may be needed with this new comparison between 2019 and 2020 eligible cohorts. Also, 2020 cohort can see the 2019 cohort's behavior and outcome. Based on that, 2020 cohort will adjust themselves with anticipation, so announcement effect is not clean.

The following section will investigate further about these concerns.

3. Analysis of Potential Issues and Bias

3.1 Varying coefficient based on inclusion different criteria

The paper should give us more justifications why they choose two third threshold as inclusion criteria. We can see in the paper that coefficient of the effect reduced from 0.138 (Main Result-Table 6) to 0.093 (Robustness Result-Table 10). This shows the heterogeneity treatment effect on varying unit across the sample.
If the paper adds the histogram of risk score, we can see the distribution of it and then can have better understanding of it. We set a simulated economy that 50% will benefits t from reducing risks by t%

and 50% will lost t from this action. In Section 3.1, we assume 0-25 risk score can benefit 0.5t, 25-50 risk score can benefit t, 50-75 risk score will lose t, and 75-100 risk score will lose 0.5t. Total four banks are uniformly distributed with risk score. If we set risk score at 50 as the cutoff point, we can generate total 1.5t-(-1.5t) = 3t difference between the groups. If we set risk score at 75 as the cutoff point, we can generate total 0.5t-(-0.5t) = 1t difference between the groups. Sensitive analysis may be applied to check this issue. For LTV, even if the paper conducts the LTV < 50% and LTV < 80%, we are skeptical low LTV will lead to low-risk score, and vice versa. Difference between groups may exist even if they are shown similar in Table 3. If we consider risk preferences are different between groups, we will be confused whether it is announcement effect or the risk preference effect.

3.2 Control group are treated

As Armstrong et al. (2022) discussed, Perfect Compliance is one of the assumptions if we want to isolate the causal effect by using DID-setting. In Arnould et al.'s paper (2023), it shows that control group deviated from its trend in pre-treatment period. It means the coefficient of our interest may not be the Average Treatment Effect (ATE) or the Local Average Treatment Effect (LATE). The announcement will give two group's different direct "treatments", which make the empirical design not well-established. In Table 3.2.1, we show the possible estimation bias on this situation. As following the simulated economy in Section 3.1, we now assume all banks with benefits are allocated in the treatment groups; others are in control groups. Compared to the randomization trials, we find the sign of coefficient will be different. Also, even if we assume treatment group's reaction is the announcement effect, control group's no parallel bias still exists in Figure 3.2.1.

Trial 1	Pre	Post	Difference
Treatment	u	u+t	t
Control	u	u-t	-t
DID			2t

Trial Randomization	Pre	Post	Difference
Treatment	u	u	0
Control	u	u	0
DID			0

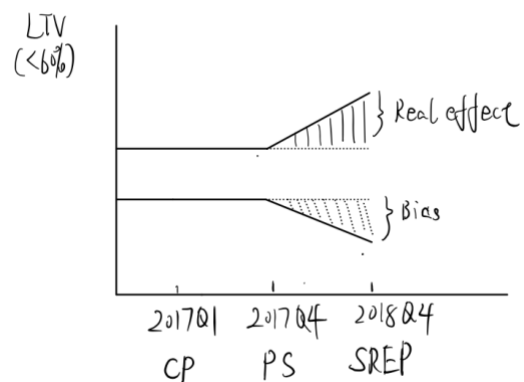Table 3.2.1 Bias from not compliance

Figure 3.2.1 Non-parallel Bias

The issue in the paper is that there may be no announcement effect, but just two groups inherent characteristic like risk aversion cause the outcome (changing in LTV). To make the paper's coefficient more convincible, we advise the paper use Honest DID (Rambachan and Roth, 2022) for the robustness test. It will set the maximum bound for the post-treatment parallel difference compared to the baseline of pre-treatment parallel trend difference. The key takeaway in this method is that it assumes the magnitude of violation for parallel assumption after treatment should be smaller than that of violation before treatment. Honest DID sets the inequality as follows:

$$|\delta_1| \leq M \cdot max|\delta r|$$

M represents the level we can allow for the violation. For example, when M = 2, our result will be invalid only if the violation after treatment is at least twice as large as violation before treatment. Reporting M in the paper can show how confident we can relax the assumption for parallel trends.

3.3 As-if staggered DID issues

The paper use two nodes as treatment timings, which will disturb with each other. Also, there is possible dynamic effect shown in Section 5.2 in that paper. Using Two Way Fixed Effect (TWFE) DID model may be not appropriate in predicting staggered and dynamic treatment effect.

The paper shows there are 8 eligible banks in the 2019 cohorts for main analysis, and this number jumps to 10 at 2020 cohorts. Without more detail information, we cannot know which banks are in these groups respectively. We don't know whether banks eligible in 2019 still be eligible in 2020 or not. So, we cannot make sure that the control group in the additional analysis is the never-treated group. If banks had the preposition in 2019, it's not a surprise to see no significant adjustment when they are also included in 2020 cohorts. So, there is a possibility that the paper compares the mix of 2019 & 2020 eligible and 2019 eligible only groups (2019 cohorts) with the mix of 2019 & 2020 eligible and 2020 eligible only groups (2020 cohorts). We can see the overlapping issue may exist between 2019 and 2020 cohorts. 2020 cohorts may have anticipation effect after seeing 2019 cohort behaviors.
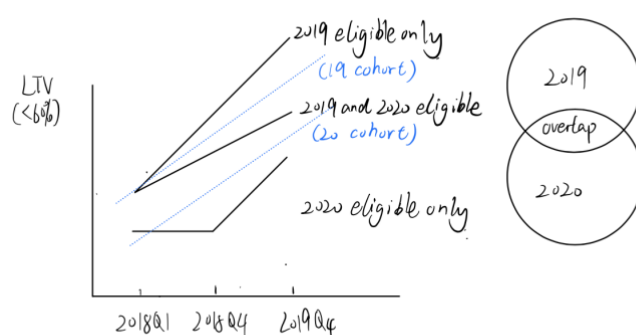
Figure 3.3.1 Overlap between 2019 cohorts and 2020 cohorts

Likewise, two timing nodes in the main analysis—CP and PS—give control group two options. One is never-treated control group, another will be treated at 2017Q4. Although the paper use 2016Q1-2019Q4 as sample period, it's still hard to disentangle and answer at which node eligible banks under the paper's criterion will be treated. Also, the pre-trend when compare 2019 cohorts with 2020

cohorts may violate parallel assumption. It would better also test the parallel assumption in the analysis.

Baker et al. (2022) shows the single treatment timing should be no bias no matter it combines with static effect or dynamic effect. So, we advise the paper might choose only first treatment date because this can capture the whole accumulated announcement effect and alleviate the anticipation effect's concern without staggered issue.

As the paper states the persistent effect of announcement, we infer that it means potential dynamic effect as shown in Section 5.2 Figure 1a and 1b in the paper. However, Goodman-Bacon (2021) uses decomposition to show the key problem in the staggered TWFE DID model with dynamic effect. In this situation, treatment group at previous period will transit into control group at later period, and then contaminate the whole estimates. In the paper's analysis, the control group can be divided into later-treated group and never-treated as discussed above. This may partly explain why control group are reacted with announcement as we said in Section 3.2. In Bacon Decomposition Method, later treated compared to early treated will incur negative weighted issue. The heterogeneity across the time and unit also causes this issue (Borusyak and Jaravel, 2017; De Chaisemartin and D'Haultfœuille, 2020). Above all, dynamic (time-varing) effect and different magnitude of treatment effect in different treatment timing will be the source of the bias.

$$\beta_{kU}^{2\times2} = ATT_k\left(POST(k)\right) + \left[\Delta Y_k^0\left(POST(k), PRE(k)\right) - \Delta Y_U^0\left(POST\left(k\right), PRE\left(k\right)\right)\right] \tag{14a}$$

$$\beta_{k\ell}^{2\times2,k} = ATT_k\left(MID\left(k,\ell\right)\right) + \left[\Delta Y_k^0\left(MID\left(k,\ell\right), PRE(k)\right) - \Delta Y_\ell^0\left(MID\left(k,\ell\right), PRE\left(k\right)\right)\right] \tag{14b}$$

$$\beta_{k\ell}^{2\times2,\ell} = ATT_\ell\left(POST(\ell)\right) + \left[\Delta Y_\ell^0\left(POST(\ell), MID\left(k,\ell\right)\right) - \Delta Y_k^0\left(POST\left(\ell\right), MID\left(k,\ell\right)\right)\right]$$
$$- \left[ATT_k\left(POST\left(\ell\right)\right) - ATT_k\left(MID\left(k,\ell\right)\right)\right] \quad \text{sources of the negative weight} \tag{14c}$$
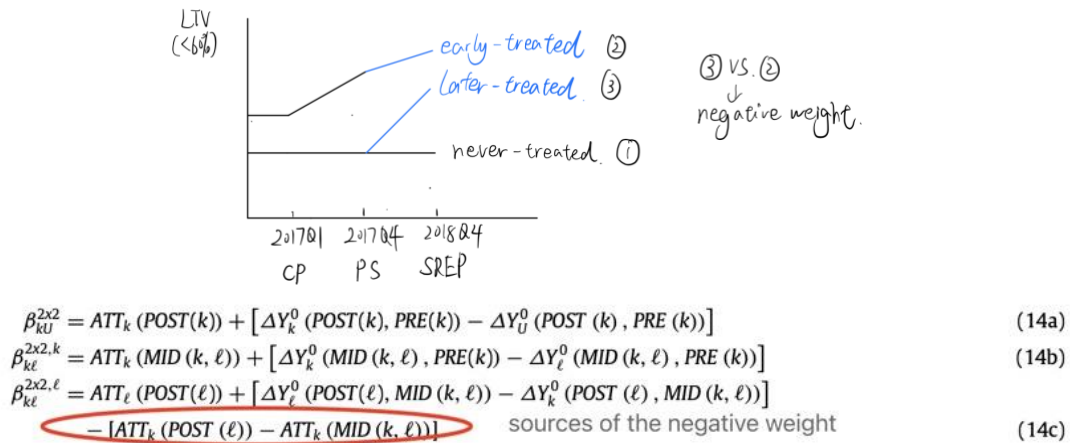
Figure 3.3.2 sub-DID: later treated vs. early treated (Goodman-Bacon, 2021)

We advise the paper to find never-treated group as the control if it is available. Following Roth et al. (2022)'s recommendation, the paper can also try Callaway and Sant' Anna(CS)'s（2021）or Sun and Abraham(SA)'s (2021) group-time estimators to address the possible negative weight issue. Both two methods use never-treated group as control group if applicable. While CS use the last treated group as the control groups if the sample doesn't include never treated, SA use all yet untreated to compare. The paper may firstly figure out the status of never treated in its setting before choosing from these two methods. Or they may concentrate at one treatment node—CP at 2017Q1—to discover the whole announcement effect and to avoid anticipation and staggered issues.

3.4 High supervisory interaction reasoning

As discussed in Section 2, the argument for the high supervisory interaction needs to be polished. For the measurement, the author might state in detail how they measure the level of communication. Instead of using number of meetings, we advise to check the content of Meeting Minute like using words number to proxy the degree of interaction. Besides, we recommend use NLP-based dictionaries (Garcia et al., 2023) to predict the positive or negative communication between supervisors and banks, this would help the author to explain the different signs in the Table 7 better.

Even if we believe this part of analysis, we will intuitively ask how to distinguish the communication effect from announcement effect. We are curious whether it is the banks interaction characteristics that affect the LTV change, or it is the announcement that affect bank's behavior. When use the term "interaction", we assume this is two-way communication, which will impact both two sides. In accounting, analysts will adjust their forecasts so many times throughout the year in the anticipation of management upward bias to beat the targets (Brown, 2022). So, what if every year's eligible standards will be changed based on bank's anticipated behavior?

4. Discussion

This paper may contribute to the literature of policy effect on bank's investment preference and to the cost of regulatory capital if they can disentangle the mess of comparison in the DID setting and explain the accuracy of their estimation.

The paper shows potential dynamic effect with two timing treatment. It raises concerns about the violation for the parallel assumption and heterogenous effect issue. These can lead to estimation bias. They can use Honest DID and Bacon Decomposition to check the robustness of their results. Also, finding a "good" control group—never treated—is as important as limiting the "bad" control group—always treated and early treated. In comparison between 2019 and 2020 cohorts, the author should be aware of anticipation effect and sample overlapping issue between two cohorts. If it is hard to find "good" control groups, synthetic control method might be considered.

Due to no randomization, the coefficient is sensitive to the eligible and non-eligible bank's allocation. To estimate more accurate regulatory capital cost, the author should justify why they set the two-third standard and conduct more standards to check if the outcomes are consistent.

Overall, I am inspired by this paper and learn more about the banks' behavior changes. In accounting, financial intermediaries also play important roles on firm's risk management and information disseminations. It's also a good learning journey for me especially for analysts' behaviors field. Their potential contributions to the industry and literature are worthy of publication.

Reference

Armstrong, C. S., Glaeser, S. A., & Huang, S. (2022). Contracting with controllable risk. The Accounting Review, 97(4), 27-50.

Arnould, G., Guin, B., Ongena, S., & Siciliani, P. (2020). (When) do banks react to anticipated capital reliefs?.

Borusyak, K., & Jaravel, X. (2017). Consistency and inference in bartik research designs. Technical Report, Working paper 2017. 187 and,"Revisiting event study designs.

Brown, A. B., Lin, G., & Zhou, A. (2022). Analysts' forecast optimism: The effects of managers' incentives on analysts' forecasts. Journal of Behavioral and Experimental Finance, 35, 100708.

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. Journal of Econometrics, 225(2), 200-230.

De Chaisemartin, C., & d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. American Economic Review, 110(9), 2964-2996.

Garcia, D., Hu, X., & Rohrer, M. (2023). The colour of finance words. Journal of Financial Economics, 147(3), 525-549.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. Journal of Econometrics, 225(2), 254-277.

Baker, A. C., Larcker, D. F., & Wang, C. C. (2022). How much should we trust staggered difference-in-differences estimates?. Journal of Financial Economics, 144(2), 370-395.

Kisin, R., and A. Manela (2016), "The shadow cost of bank capital requirements", Review of Financial Studies 29 (7).

Rambachan, A., & Roth, J. (2020). Design-based uncertainty for quasi-experiments. arXiv preprint arXiv:2008.00602.

Roth, J., Sant'Anna, P. H., Bilinski, A., & Poe, J. (2022). What's trending in difference-in-differences? A synthesis of the recent econometrics literature. arXiv preprint arXiv:2201.01194.