Bank Supervision as Information Production: Evidence from U.S. Bank Holding Companies

Mehdi Beyhaghi, Jiyoung Chae, Filippo Curti, and Jeffrey R. Gerlach*

May 2026

Abstract

We examine bank supervision as an information-production process using confidential supervisory ratings for U.S. bank holding companies from 2005 to 2023. We construct a benchmark model using prior ratings, observable financial characteristics, and market data, and analyze deviations from this benchmark as a measure of supervisory innovation. The benchmark explains 70–80 percent of rating variation, indicating that supervision is predominantly systematic and grounded in observable fundamentals. We show that ratings deviate more from the benchmark when examiners have greater access to private information. We also document persistent examiner-level heterogeneity in stringency and procyclical rating patterns that are independent of information access. These supervisory innovations have real effects on bank behavior: unexpected stringency reduces asset growth and lending while increasing capitalization; unexpected leniency is associated with higher future risk-taking.

Keywords: Bank supervision; Information production; Examiner judgment; Regulatory oversight; Bank risk; Financial stability.
JEL Classification: G21; G28; D82; E58.

---

*Beyhaghi is at the Federal Reserve Board (e-mail: mehdi.beyhaghi@frb.gov). Chae and Curti are at the Federal Reserve Bank of Richmond (e-mails: jiyoung.chae@rich.frb.org, filippo.curti@rich.frb.org). Gerlach is at the Federal Reserve Bank of Philadelphia (e-mail: jeffrey.gerlach@phil.frb.org). We thank Yannis Cabossioras, Sarah Dahlgren, Mark Flannery, Priyank Gandhi, Yadav Gopalan, Fiona Hoellmann, Cooper Howes, John Kandrac, David Marques Ibanez, Hamid Mehran, Patricia Mosser, Rahman Sakif, Rimmy Tomy, Lilah Shema-Zlatokrilov, and Greg Weitzner for helpful comments. We also thank seminars and conferences participants at the Federal Reserve Bank of Richmond, the Federal Reserve Board, the 2025 EFA Annual Meeting, the 2025 Behavioral Finance Conference at the University of Chicago, and the 2026 Columbia/BPI Annual Bank Regulation Research Conference at Columbia University, and the 2026 Joint conference by the Research Group of the Basel Committee on Banking Supervision and the Centre for Economic Policy Research for valuable feedback. The views expressed in this article are solely those of the authors and do not necessarily represent the views of the Board of Governors of the Federal Reserve System, the Federal Reserve Bank of Richmond, the Federal Reserve Bank of Philadelphia, or the Federal Reserve System.

1 Introduction

U.S. banking regulators spend billions annually on confidential bank supervision, premised on the belief that examiners produce information beyond what markets can observe.[1] Yet despite this massive investment, we have little direct evidence on whether supervision actually generates incremental information—or whether examiner assessments simply reflect publicly available signals. Following the 2023 failures of Silicon Valley Bank and Signature Bank, this question has urgent policy relevance: if supervision adds no information beyond public disclosures, the rationale for devoting substantial public resources to confidential supervisory oversight is unclear.

Bank opacity poses a fundamental trade-off: opacity enables liquidity creation by maintaining information-insensitive liabilities, that is, deposits and short-term debt that function as money (Dang, Gorton, Holmström, and Ordonez, 2017), but simultaneously weakens market discipline. Confidential supervision addresses this by providing monitoring through regulatory action rather than price mechanisms—but only if supervisors produce incremental information and use it to guide appropriate regulatory actions. Quantifying this information poses significant empirical challenges. Supervisory ratings comprise three distinct components: (1) public information available to all market participants, (2) private supervisory information gathered through examinations and confidential regulatory data,[2] and (3) examiner judgment in synthesizing and weighting this information.[3] Researchers can directly observe only the first component.

In this paper, we provide the first direct evidence that bank supervision produces economically meaningful information beyond public disclosures, and that this information production varies systematically with institutional features that determine examiners' access to soft information. Using confidential RFI composite ratings for 2,593 U.S. bank holding companies from 2005 to 2023, we show that a predictive model using all observable financial and market data explains 70–80 percent of rating variation—but the unexplained 20–30 percent is not noise. These deviations predict

[1]According to a Congressional Research Service report from December 2020, the Federal Reserve reported \$1.84 billion in supervision expenditures in 2019. Similarly, the Federal Deposit Insurance Corporation (FDIC) budgeted \$1.06 billion for supervision in 2020, while the Office of the Comptroller of the Currency (OCC) spent \$965 million on supervisory activities in fiscal year 2019.

[2]Through on-site examinations, direct interactions with bank management, and access to confidential supervisory materials, examiners acquire and synthesize information unavailable to market participants.

[3]In principle, two examiners with identical private information could assign different ratings, reflecting differences in judgment. Similarly, the same examiner might evaluate identical bank conditions differently over time depending on macroeconomic conditions or regulatory priorities.

future bank behavior and vary systematically with examiners' information access: ratings deviate more from observable-information benchmarks when examiners conduct on-site examinations, when they are geographically proximate to banks, and when banks are privately held. A one-grade tougher-than-predicted rating reduces assets by 5.9%, lending by 2.1 percentage points, and increases capital ratios by 0.3–0.4 percentage points. These effects persist even after controlling for persistent examiner stringency, indicating that supervisory information—not merely examiner preferences—drives bank responses.

We document three core findings. *First*, supervision adds value precisely when examiners have greater access to private information. On-site examinations generate deviations 0.027 rating grades larger than off-site exams (12% of the mean deviation), but this effect is concentrated among privately held banks—for publicly traded banks, where market signals are abundant, examination modality has minimal impact. Geographic proximity similarly matters: examiners located 50+ miles from banks assign ratings 0.033 grades closer to model predictions (15% of mean deviation). *Second*, examiner stringency exhibits pronounced variation both across individuals and over time. Some examiners are persistently tougher than others—a one-unit increase in examiner stringency predicts a 0.25-grade increase in rating deviations for other banks. Moreover, average stringency fluctuates dramatically over the cycle, reaching maximum leniency (−0.280) during the Global Financial Crisis and peak stringency (+0.080) in 2014 following Dodd-Frank implementation, with renewed leniency preceding the 2023 banking turmoil. *Third*, these supervisory signals have asymmetric real effects: unexpected stringency triggers immediate deleveraging (5.9% asset reduction) and capital increases (0.3–0.4pp), while unexpected leniency manifests primarily in higher future non-performing loans (0.1pp increase relative to 1.3% mean).

This paper makes three contributions to the literature on bank supervision and external monitoring. First, we provide the first systematic evidence linking supervisory information production to institutional features—examination modality, geographic proximity, bank transparency—that determine examiners' information access. While prior work documents examiner-level variation (Agarwal et al., 2024), our mechanism-based approach explains *when* and *why* these effects arise. We show that examiner-specific deviations are largest when public signals are weakest: on-site examinations generate substantially larger departures from our benchmark for privately held banks than for publicly traded institutions, while off-site examinations significantly compress rating deviations for opaque banks but have little effect on transparent ones. Second, we demonstrate that supervisory information has causal effects on bank behavior using multiple identification strategies including within-bank examiner switches, examiner office fixed effects, and placebo tests that rule out anticipatory assignment. Critically, rating deviations predict bank behavior even after controlling for examiner fixed effects, indicating that time-varying private information—not persistent examiner preferences—drives these responses. Third, we document systematic time-variation in supervisory stringency over the financial cycle. This procyclical pattern raises concerns about regulatory forbearance during booms, when vulnerabilities accumulate, and excessive tightening following crises, when credit provision is most valuable.

Our empirical strategy exploits three sources of variation in examiners' access to private information. First, we use geographic distance between the examiner's office and the bank being examined. Federal Reserve supervision operates through twelve regional districts with substantial geographic variation: examiners may assess banks in their own city or banks hundreds or thousands of miles away.[4] We expect examiners to acquire more granular, context-specific information when examining geographically proximate institutions. Second, we distinguish between on-site and off-site examinations. On-site examinations allow examiners to conduct in-person interviews with management, observe operations directly, and access detailed records in real time, facilitating the collection of soft information that may not be readily quantifiable or available remotely. Third, we distinguish between publicly traded and private banks. Publicly traded banks are subject to extensive disclosure requirements and often covered by equity analysts and financial media, providing a richer public information set. In contrast, private banks operate with less transparency, giving examiners a comparative informational advantage.

We interpret deviations of actual ratings from model-predicted ratings—which we term the *examiner-specific effect* (ESE)—as a revealed measure of incremental supervisory information and judgment. We find that examiners with greater access to private information deviate more from the benchmark. Specifically, examiners located closer to the banks they assess, those conducting on-site examinations, and those evaluating privately held banks are more likely to diverge from model predictions. These patterns suggest that supervision adds value precisely when examiners

[4]For example, San Francisco Fed examiners supervise banks across nine western states (Alaska, Arizona, California, Hawaii, Idaho, Nevada, Oregon, Utah, and Washington), creating considerable variation in examiner-bank proximity.

are better positioned to acquire soft, non-public information.

We implement several identification tests that address potential endogeneity in exam assignment. Exploiting within-bank variation from examiner switches and including examiner office fixed effects, we show that distance effects strengthen rather than weaken when controlling for systematic allocation strategies. Placebo tests confirm that lagged and future exam characteristics do not predict current rating deviations, ruling out anticipatory assignment based on private supervisory concerns. Additional robustness tests confirm that the relationship between information access and rating deviations is stable over time. Notably, while off-site examination effects strengthen during crisis episodes (specifically, COVID-19), distance effects remain remarkably consistent throughout our sample period, supporting our information-production interpretation and ruling out parameter instability as a confounding factor.

Our analysis focuses on U.S. bank holding companies (BHCs) supervised by the Federal Reserve from 2005 to 2023, encompassing 25,245 examinations of 2,593 unique banks. This institutional setting offers several advantages. First, focusing on a single regulator eliminates confounding variation from differences in supervisory practices across agencies or examiner rotation policies that characterize bank-level CAMELS ratings. Second, the Federal Reserve's RFI/C(D) rating framework[5] is explicitly designed to capture holistic supervisory judgment rather than mechanically aggregating component ratings (see Table 1), making it particularly well-suited for studying how supervisory information and judgment enter regulatory assessments. Third, our sample period spans the financial crisis, subsequent regulatory reforms, and recent banking stress episodes, allowing us to examine how the informational content of supervision varies across regulatory regimes and macroeconomic conditions.

Our findings highlight supervision as an information-production technology shaped by both institutional features and human judgment. While regulatory frameworks are uniform, the informational value of supervision depends critically on examiners' access to private information, their ability to acquire soft information through direct interaction, and how they interpret that information when forming assessments. This heterogeneity reveals both when supervision adds the most value and how supervisory outcomes can vary systematically across examiners and over time. Un-

[5]RFI/C(D) evaluates Risk management, Financial condition, Impact of non-depository subsidiaries, Composite condition, and Depository institutions. See Section 3 for details.

derstanding these mechanisms is essential for evaluating the effectiveness of bank supervision and its contribution to financial stability.

The remainder of the paper is organized as follows. Section 2 reviews related literature. Section 3 describes the institutional setting and data. Section 4 develops testable hypotheses. Section 5 outlines our empirical methodology. Section 6 presents the main results on information production, examiner heterogeneity, and real effects on bank behavior. Section 7 concludes. Appendices provide robustness tests and detailed variable definitions.

2 Related Literature

This paper contributes to the literature on external monitoring and supervisory oversight, with a particular focus on the role of examiners in bank supervision. A large body of research emphasizes the inherent opacity of banks and the resulting need for supervisory monitoring. Dang, Gorton, Holmström, and Ordonez (2017) describe banks as "secret keepers," highlighting their incentives to withhold loan-level information to prevent costly duplication of private information. Similarly, Morgan (2002) argues that the combination of unique assets, leverage, and systemic risk makes opacity a defining feature of banking. Empirical evidence supports this view: Berger and Davies (1998) show that adverse supervisory information predicts subsequent stock returns, while DeYoung, Flannery, Lang, and Sorescu (2001) document that supervisory ratings contain information not immediately reflected in market prices, indicating that examinations complement market discipline.[6]

This certification role aligns with foundational theories of costly monitoring and delegated verification, in which specialized monitors arise to economize on information production and improve allocation efficiency (Townsend, 1979; Boyd and Prescott, 1986). More recent work emphasizes that the effectiveness of monitoring depends critically on information quality and timely loss recognition. For example, Acharya and Ryan (2016) argue that weak information environments allow risks to remain hidden, undermining both market discipline and regulatory intervention, while Acharya and Yorulmazer (2007) highlight how supervisory forbearance can distort incentives and encourage excessive risk-taking.

[6] See also Hirtle (2016), Flannery, Kwan, and Nimalendran (2004), and Iannotta (2006).

A growing empirical literature demonstrates that supervision mitigates opacity and shapes bank behavior by producing and acting on private information.[7] Gopalan (2022) show that disclosure of supervisory ratings accelerates loan-loss provisioning. Costello, Granja, and Weber (2019) find that stricter regulatory oversight enhances reporting transparency. Angelini, Lotti, Passalacqua, and Soggia (2021) and Bonfim, Cerqueiro, Degryse, and Ongena (2023) document that supervisory inspections reduce risk-taking and curtail lending to weaker firms, and Correia, Luck, and Verner (2025) show that supervisors use private information to anticipate bank failures. Collectively, these studies underscore that supervision complements market discipline by influencing both information revelation and real outcomes. Related work also emphasizes that supervisory interventions can generate real trade-offs between resilience and credit provision (Acharya, Berger, and Roman, 2018).

Our paper builds on this foundation and contributes to the emerging literature on examiner heterogeneity in bank supervision. The study most closely related to ours is the parallel work by Agarwal et al. (2024), who model confidential CAMELS ratings to quantify examiner discretion and show that examiner-specific variation affects bank behavior. While both papers document examiner-level effects, our focus and interpretation differ in important ways.

First, we emphasize supervision as an information-production process rather than framing examiner-specific variation primarily as noise or idiosyncratic discretion. Regulatory guidance explicitly recognizes that composite supervisory ratings are not mechanical functions of component ratings but reflect examiners' holistic assessments of bank condition.[8] Consistent with this design, deviations between model-predicted and assigned ratings may reflect the incorporation of qualitative or forward-looking information—such as governance quality, management responsiveness, or emerging risks—not captured in quantitative data. Our methodology explicitly accommodates this feature by interpreting examiner-specific effects as the joint outcome of private information acquisition and supervisory judgment, rather than attributing all residual variation to taste-based discretion.

Second, our analysis isolates settings in which supervisory information production is more or

[7] In addition to producing soft information through examinations, supervisors have access to confidential regulatory data—such as granular balance-sheet, loan-level, and supervisory reporting—that are substantially richer than publicly available disclosures (Beyhaghi, Fracassi, and Weitzner, 2025).

[8] See the *Uniform Financial Institutions Rating System* (UFIRS), the FDIC *Risk Management Manual*, and the OCC *Large Bank Supervision* handbook.

less constrained. By exploiting variation in examiner proximity, examination modality (on-site versus off-site), and bank transparency, we show that examiner-specific deviations are systematically larger when examiners have greater access to soft, non-public information. This mechanism-based approach complements Agarwal et al. (2024) by explaining when and why examiner effects arise, rather than solely documenting their existence.

Third, we focus on bank holding companies supervised exclusively by the Federal Reserve under a unified supervisory framework. This design avoids confounding cross-agency differences documented in prior work (Agarwal et al., 2014; Beyhaghi and Gerlach, 2024) and allows us to link supervisory assessments to market-based measures, since holding companies are the publicly traded entities within banking organizations.

Finally, we introduce a time-series perspective on supervisory behavior. We show that examiner stringency varies systematically over the business and financial cycle, becoming more conservative following periods of stress and more permissive during expansions. This procyclical pattern relates to a broader literature on time-inconsistent regulation and the challenges of maintaining supervisory discipline over the cycle. Acharya and Yorulmazer (2007) highlight how supervisory forbearance during crises—the ex post reluctance to close distressed institutions—can distort incentives and encourage risk-taking, while Acharya et al. (2017) document that regulatory leniency during expansions contributes to systemic risk accumulation. Our findings complement this work by documenting systematic variation in examiner-level stringency over the financial cycle, with supervisory assessments becoming more lenient during booms and tightening sharply after periods of stress.

Our analysis also relates to research highlighting how information access shapes monitoring outcomes. Gopalan, Kalda, and Manela (2021) show that increases in examiner-bank distance following supervisory office closures weaken oversight and raise bank risk, while Kandrac and Schlusche (2021) document that on-site examinations facilitate deeper information collection and materially affect supervisory ratings. We extend this work by demonstrating that such informational frictions operate at the individual examiner level within a single regulatory regime.

More broadly, our findings connect to evidence of heterogeneity among external monitors in other settings. Studies of auditors and analysts show that monitoring quality varies with individual expertise and information access (Francis and Yu, 2009; Gul, Jaggi, and Krishnan, 2007; Clement, 1999; Ke and Yu, 2006). In banking, research on loan officers and internal governance similarly highlights the importance of monitor identity for risk-taking outcomes (Berger, Kick, and Schaeck, 2014; Srivastav and Hagendorff, 2016; Cole, Kanz, and Klapper, 2015; Karolyi et al., 2024; Ivanov and Wang, 2024). Together, these literatures underscore that even within formalized oversight systems, who monitors—and what they can observe—matters.

3 Institutional Background

U.S. bank supervision relies on several distinct rating systems tailored to institutional scope and regulatory objective. These supervisory ratings summarize examiners' assessments of financial condition, risk management, and governance, and play a central role in determining supervisory actions and regulatory constraints. Supervisory ratings carry significant consequences for banks, ranging from recommended corrective actions and mandated operational changes to restrictions on distributions, mergers, and acquisitions, or even resolution. Examiner assessments can therefore affect banks' intermediation activity and have tangible implications for the real economy.

Supervisory ratings for depository institutions. Commercial banks and thrifts are evaluated using the CAMELS rating system, which assigns composite and component ratings based on Capital adequacy, Asset quality, Management, Earnings, Liquidity, and Sensitivity to market risk. CAMELS ratings are assigned jointly by federal and state regulators, including the Federal Reserve, the Office of the Comptroller of the Currency (OCC), the Federal Deposit Insurance Corporation (FDIC), and state banking authorities. Because many banks alternate between federal and state supervisors over time, CAMELS-based analyses may reflect heterogeneity across both examiners and regulatory agencies.

Supervisory ratings for bank holding companies. Our analysis focuses on bank holding companies, which are supervised exclusively by the Federal Reserve. From 2005 through 2018, the Federal Reserve used the RFI/C(D) rating system to assess the consolidated condition of all BHCs. Beginning in 2019, the largest institutions ($100 billion or more in assets) transitioned to a separate framework (discussed below), while the RFI system continued to apply to all other BHCs. RFI ratings evaluate (i) Risk management, (ii) Financial condition, and (iii) the Impact of non-depository subsidiaries on the organization, culminating in a Composite rating. In addition, the D component summarizes the condition of subsidiary depository institutions based on their CAMELS ratings.

Crucially, RFI composite ratings are not mechanical averages of underlying components. Supervisory guidance emphasizes that the composite rating reflects a holistic and forward-looking assessment of the organization's condition, incorporating examiner judgment about the relative importance of risks, governance quality, and emerging vulnerabilities. This feature makes RFI ratings particularly well suited for studying how supervisory information and examiner judgment enter regulatory outcomes.

Federal Reserve supervision is also shaped by the System's historical district structure. The Federal Reserve comprises 12 districts, each overseen by a Federal Reserve Bank responsible for supervising bank holding companies in its region. These banks operate from headquarters and up to five branches located throughout their districts.[9] Examiners at each location typically supervise geographically proximate banks.

Large Financial Institution (LFI) ratings. Beginning in 2019, the Federal Reserve introduced the Large Financial Institution (LFI) rating framework for BHCs with $100 billion or more in total assets. LFI ratings replace numeric scores with a non-numeric assessment of whether firms broadly meet supervisory expectations in capital, liquidity, and governance and controls. Our baseline sample includes all BHCs with numeric RFI ratings from 2005 to 2023. This naturally includes post-2019 observations for BHCs that remain under the RFI framework (the vast majority) but excludes post-2019 observations for the approximately 10–15 institutions that transitioned to the LFI framework—affecting fewer than 1% of bank-year observations. We verify robustness to alternative sample definitions in Section 6.

Comparison of supervisory rating systems. Table 1 summarizes the key differences between CAMELS, RFI/C(D), and LFI rating systems. The table highlights that RFI ratings apply uni-

[9]The list of Federal Reserve Banks and their branches includes: Boston (no branches); New York (no branches); Philadelphia (no branches); Cleveland (Cincinnati and Pittsburgh); Richmond (Baltimore and Charlotte); Atlanta (Birmingham, Jacksonville, Miami, Nashville, and New Orleans); Chicago (Detroit); St. Louis (Little Rock, Louisville, and Memphis); Minneapolis (Helena); Kansas City (Denver, Oklahoma City, and Omaha); Dallas (El Paso, Houston, and San Antonio); and San Francisco (Los Angeles, Portland, Salt Lake City, and Seattle).

formly to BHCs under a single regulatory authority, rely explicitly on examiner judgment, and maintain a consistent numeric scale over our sample period. These institutional features allow us to cleanly study examiner-specific information and heterogeneity without confounding cross-agency or framework-level variation.

4 Hypothesis Development

Bank supervision relies on individual examiners to translate regulatory standards into assessments of bank condition. Although supervisory guidance and internal review processes promote consistency, examiners differ in their access to information, their interpretation of qualitative signals, and their tolerance for risk. These differences can generate systematic variation in supervisory ratings, even when banks face the same formal regulatory framework.

A central implication of viewing supervision as information production is that examiner assessments should deviate most from observable-information benchmarks when examiners have greater access to private or soft information. On-site examinations facilitate direct interaction with bank management and internal systems, while geographic proximity enhances familiarity with local economic conditions and borrower environments. Conversely, when banks are publicly traded and subject to market scrutiny, examiners' informational advantage relative to public data is reduced.

Hypothesis 1 ***Examiners with greater access to private information are more likely to assign ratings that deviate from those predicted using publicly available data.***

Even conditional on information access, examiners may differ persistently in how they map information into ratings. Some examiners apply more conservative thresholds, while others place greater weight on mitigating factors. Such examiner-specific stringency or leniency may reflect experience, professional judgment, or interpretation of supervisory guidance rather than bank fundamentals.

Hypothesis 2 ***Examiner-specific stringency or leniency contributes to systematic differences in supervisory ratings beyond bank characteristics and information access.***

Supervisory behavior may also vary over time. Following periods of financial stress, heightened regulatory scrutiny and learning may lead examiners to apply more conservative assessments, whereas prolonged expansions may be associated with greater tolerance for risk.

Hypothesis 3 ***Average supervisory stringency varies over time, becoming more conservative following financial stress and more lenient during economic expansions.***

Finally, supervisory ratings communicate information to banks and trigger regulatory responses that can influence subsequent behavior. Unexpectedly stringent ratings may induce banks to contract risk and growth, while lenient assessments may allow vulnerabilities to persist.

Hypothesis 4 ***Banks receiving tougher-than-predicted supervisory ratings reduce risk-taking and growth, while lenient ratings are associated with weaker subsequent risk discipline.***

5 Methodology

Conceptual setup. We conceptualize a bank's assigned rating as the outcome of two general components: (i) observable signals about bank condition—including publicly available financial and market data as well as the bank's most recent supervisory rating—and (ii) incremental information and judgment available to supervisors through the examination process. The first component reflects the observable information set available to supervisors at the beginning of an examination, while the second pertains to both non-public supervisory information revealed during examinations and persistent examiner-level differences in how similar information is translated into ratings.

To operationalize this framework, we start with a two-step decomposition. In the first step, we construct a benchmark forecast of supervisory ratings using the proxies for the first component, as explained above. In the second step, we measure deviations between assigned and predicted ratings—the examiner-specific effect (ESE)—and analyze how these deviations vary with information access, examiner characteristics, and bank behavior.

Step 1: Benchmark model using observable information and prior supervisory assessment. We begin by estimating a benchmark model that predicts supervisory ratings using observable financial and market information together with the bank's most recent supervisory rating. Specifically, we regress the supervisory rating of bank i at time t, denoted $R_{i,t}$, on a comprehensive set of bank characteristics derived from regulatory filings and, where available, market data. These characteristics capture capital adequacy, liquidity, performance, volatility, concentration, business model, and asset and liability composition.

Our baseline specification uses Ordinary Least Squares (OLS):

$$R_{i,t} = \alpha + \beta \text{Bank Characteristics}_{i,t} + \epsilon_{i,t}. \tag{1}$$

Because supervisory ratings are ordinal, we also estimate Ordered Logit specifications:

$$\text{logit}\big(P(R_{i,t} \leq j)\big) = \kappa_j - \beta \text{Bank Characteristics}_{i,t}, \quad j = 1, 2, 3, 4, \tag{2}$$

and a binary Logit specification where $BR_{i,t} = 1$ for unsatisfactory ratings (3–5) and $BR_{i,t} = 0$ for satisfactory ratings (1–2):

$$P(BR_{i,t} = 1) = \frac{1}{1 + \exp\big(-(\alpha + \beta \text{Bank Characteristics}_{i,t})\big)}. \tag{3}$$

While we compare results across these specifications, we use OLS as our primary approach in subsequent analysis for transparency and ease of interpretation. Results are robust to alternative estimation methods.

Step 2: Constructing the examiner-specific effect. We define the examiner-specific effect (ESE) as the deviation between the assigned rating and the model-predicted rating:

$$\text{ESE}_{i,t} = R_{i,t} - \text{Round}(\widehat{R}_{i,t}), \tag{4}$$

where $\widehat{R}_{i,t}$ is the fitted value from the OLS benchmark model. Because ratings are assigned as integers from 1 to 5, we round predicted ratings to the nearest integer to ensure comparability and focus on economically meaningful deviations rather than continuous residual variation. Under this interpretation, ESE captures the supervisory innovation—the component of the assigned rating not explained by observable fundamentals and prior supervisory assessments.

Step 3: Information access and rating deviations. To test whether examiner access to private information affects supervisory assessments (Hypothesis 1), we examine whether ratings deviate more from observable-information benchmarks when examiners have greater access to soft information:

$$|\text{ESE}_{i,t}| = \beta_1 \text{PAIP}_{i,t} + \beta_2 \text{Controls}_{i,t} + \delta_i + \mu_t + \eta_{i,t}, \tag{5}$$

where $\text{PAIP}_{i,t}$ denotes proxies for examiner access to private information, δ_i are bank fixed effects, and μ_t are time fixed effects. Because we cannot observe ex ante whether private information is positive or negative, we focus on the absolute magnitude of rating deviations.

Our key proxies for private information access include geographic distance between the examiner's office and the bank and an indicator for on-site versus off-site examinations. We also distinguish between publicly traded and private banks to evaluate the incremental effect of greater public information availability for traded banks. The identifying assumption is that examiners gain more private information, relative to market participants, when they are geographically closer to the bank, conduct on-site examinations, and evaluate banks for which less public information is available.

Information access and examination modality. To further distinguish information effects from residual discretion, we augment equation (5) by interacting our access proxies with indicators for examination modality and geographic proximity. In particular, we interact distance-based measures with an off-site examination indicator and estimate specifications with discrete distance bins to allow for non-linear proximity effects. Under an information-production interpretation, access to private information should matter most for on-site examinations and at shorter distances, where examiners can more effectively acquire and interpret soft, non-public information.

A key identification challenge is distinguishing information-driven deviations from purely taste-based examiner discretion. Our empirical design addresses this concern in three ways. First, we show that deviations are systematically larger precisely when supervisory information advantages are strongest—during on-site examinations, at shorter distances, and for non-public banks. Second, we demonstrate that these patterns attenuate when examinations are off-site or banks are publicly traded. Third, we show that for the subset of publicly traded banks, examiner access to private information matters most when examinations are conducted on-site, consistent with supervision adding value through the production of non-public information. Fourth, we show that examiner-specific effects predict future bank outcomes even after controlling for examiner fixed effects, indicating that the residual variation reflects information rather than persistent examiner preferences.

Step 4: Examiner stringency. We next examine whether persistent examiner-level differences contribute to rating outcomes by estimating:

$$\text{ESE}_{i,t} = \gamma_1 E_{k,i} + \gamma_2 \text{Controls}_{i,t} + \delta_i + \mu_t + \nu_{i,k,t}, \tag{6}$$

where k indexes the examiner assigned to bank i at time t. The variable $E_{k,i}$ measures examiner stringency using a leave-one-out approach:

$$E_{k,i} = \frac{1}{N_{k,-i}} \sum_{j \neq i} \text{ESE}_{j,t(k,j)}, \tag{7}$$

where $N_{k,-i}$ is the number of examinations conducted by examiner k excluding bank i. This construction mitigates mechanical correlation between the stringency measure and the outcome for the current bank. The coefficient γ_1 captures whether an examiner's average stringency (or leniency) in other examinations predicts their rating deviation for bank i.

Step 5: Time variation in supervisory stringency. To assess whether supervisory stringency varies over time, we compute annual averages of examiner-specific effects across all examinations and plot these averages with confidence intervals. Persistent deviations from zero indicate shifts in aggregate supervisory tone, which we interpret as reflecting changes in regulatory risk tolerance rather than abrupt changes in information availability.

Step 6: Real effects on bank behavior. Finally, we examine whether examiner-specific effects predict subsequent bank behavior:

$$Y_{i,t+1} = \lambda_1 \text{ESE}_{i,t} + \delta_i + \mu_t + \xi_{i,t}, \tag{8}$$

and

$$Y_{i,t+1} = \lambda_2 \text{ESE}_{i,t} + \delta_i + \mu_t + \rho_k + \zeta_{i,k,t}, \tag{9}$$

where $Y_{i,t+1}$ denotes measures of bank growth or risk-taking, including asset growth, loan growth, and capital ratios. The second specification includes examiner fixed effects to control for persistent examiner stringency. Under the information-production interpretation, any remaining predictive power of $\text{ESE}_{i,t}$ after conditioning on examiner fixed effects reflects variation in supervisory information rather than examiner taste.

Robustness: One-step alternative specifications. Our main design serves a specific purpose: the first stage constructs a forecast $\widehat{R}_{i,t}$ using only publicly available information, establishing a benchmark against which we measure examiner-specific deviations. To ensure our findings are not artifacts of this procedure, we implement one-step specifications as robustness checks.

In the one-step approach, we directly estimate:

$$R_{i,t} = \alpha + \beta_1 \text{Bank Characteristics}_{i,t} + \beta_2 \text{Info Access}_{i,t} + \delta_i + \mu_t + \varepsilon_{i,t}, \tag{10}$$

where $\text{Info Access}_{i,t}$ includes distance, off-site examination status, and public trading indicators. We then test whether residual dispersion (absolute or squared residuals) varies systematically with information access, examining the same economic hypothesis without constructing ESE as an intermediate variable.[10] Results from both approaches, presented in Online Appendix A, are quantitatively and statistically consistent, confirming our conclusions are robust to estimation procedure.

6 Empirical analysis

6.1 Data

Our empirical analysis draws from three main data sources, with the Federal Reserve's confidential supervision records serving as the primary source. These records provide examination information and supervisory ratings, specifically the RFI composite ratings of bank holding companies (banks). Our dataset encompasses 25,245 annual examinations of 2,593 unique banks over the period 2005–2023, with 2005 marking the first year of RFI rating enforcement. We include all available ratings

[10] The two-step approach offers a technical advantage: supervisory ratings are integers (1–5) while predictions are continuous. Our procedure rounds predicted ratings before computing ESE and uses absolute values in the second stage—transformations that cannot be implemented in single-step regressions where residuals must remain continuous and signed.

for firms that filed FR Y-9C at least once during the sample period, even for periods when their total assets were below the $3 billion FR Y-9C filing threshold. For the small number of BHCs that transitioned to the Large Financial Institution (LFI) rating framework beginning in 2019 (approximately 10-15 institutions), we include their pre-2019 RFI ratings but exclude their post-2019 LFI ratings, as these use a non-numeric scale not directly comparable to the 1–5 RFI system. This exclusion affects fewer than 1% of bank-year observations in our sample. The RFI composite rating of 1 indicates strong performance, 2 satisfactory, 3 fair, 4 marginal, and 5 unsatisfactory. Table 2 presents the distribution of these ratings across our sample. As can be seen, the majority of banks are rated satisfactory.

Our second data source comprises financial reports from bank holding companies, primarily FR Y-9C filings. Many banks, however, alternate between FR Y-9C and FR Y-9SP filings due to fluctuations around the $3 billion total asset threshold. As explained, to maintain comprehensive coverage, we incorporate measures from both report types. While FR Y-9SP filings are simpler and less frequent, lacking detailed metrics like Tier 1 capital, our examination rating data remains consistent across all filers. To address the varying depth of financial information, we construct two sets of explanatory variables: a baseline set using measures common to both statement types, and an expanded set derived solely from FR Y-9C data. This dual approach maximizes the use of information while utilizing the more detailed information available for larger institutions when possible.

The summary statistics of the variables used in our analysis are provided in Table 3. The first six variables (total assets and its logarithm, as well as ratios of equity to assets, securities to assets, loans to assets, and ROA) constitute our baseline set, which can be constructed for all observations regardless of filing type. These variables were selected based on prior literature and broadly capture bank size, capital, asset distribution, and profitability— factors expected to forecast ratings effectively. We augment this baseline with 23 additional variables derived exclusively from FR Y-9C filings, offering a more detailed financial profile of the larger institutions.[11] These additional

[11]These FR Y-9C-specific variables include the following ratios: trading assets/assets, allowance for loans/assets, leverage ratio, double leverage ratio, tier 1 capital ratio, ratio of residential real estate loans, ratio of commercial real estate loans, ratio of commercial and industrial loans, ratio of consumer loans, ratio of depository institution loans, ratio of foreign institution loans, ratio of agriculture loans, loans Herfindahl-Hirschman Index, assets Herfindahl-Hirschman, risk-weighted assets/assets, NPL/assets, NPL/loans, loan loss reserve/loans, loan loss reserve/NPL, residential real estate NPL, commercial real estate NPL, commercial and industrial NPL, consumer NPL. Definitions are provided in Table A1 in the Appendix.

variables are used as regressors in our expanded forecasting models. The financial data are taken from the most recent statements available prior to the bank examination.

Our third source of data is market data from the Center for Research in Security Prices (CRSP) for publicly traded banks. We obtain data on equity returns, prices, and shares outstanding using the CRSP-FRB Link dataset, which is provided by the Federal Reserve Bank of New York.[12] This dataset establishes a connection between CRSP's permanent company number (PERMCO) and the unique RSSD ID number of the bank holding company. We utilize market information up to the end of the month preceding the end of the examination period, which is before the assignment of a rating. For instance, if a rating is assigned on May 10, we consider all stock market information up until April 30. Consequently, we derive the following metrics: market-to-book ratio and average monthly stock return over the most recent quarter (3-month period). Additionally, we compute average monthly stock return, standard deviation of monthly returns, and average turnover (volume/shares outstanding) over a rolling 12-month period. The summary statistics for these variables are presented at the end of Table 3.

These variables are defined in A1 in the Appendix. All ratios are winsorized at the 3rd and 97th percentiles to mitigate the effect of outliers.

6.2 Estimating Examiner-Specific Effect

We initiate our analysis with a forecasting model that predicts bank ratings using observable financial and market information together with the bank's prior supervisory rating, as outlined in Step 1 of our methodology. Starting with a parsimonious set of bank characteristics derived from FR Y-9C or FR Y-9SP statements, we progressively expand our model before proceeding with subsequent steps. Our findings are presented sequentially, aligned with our hypotheses.

Our baseline forecasting model is an Ordinary Least Squares (OLS) regression using six bank-year level characteristics to predict ratings across all data. Notably we use last/starting composite rating as the first variable, which is also an input to the examiner's rating. The other five are the variables we extract from financial statements. Given the discrete nature of bank ratings (1-5), we also employ alternative discrete choice models as per step 1 of our methodology. These include an ordered logit model and a logistic model. The dependent variable for the latter is binary (1 for

[12] `https://www.newyorkfed.org/research/banking_research/datasets.html`

ratings 3-5, 0 for ratings 1-2). We compare these models to assess their relative performance in forecasting ratings. Results are presented in Table 4.

Table 4 shows consistent coefficient signs and significance across all three specifications. The OLS model's adjusted R-squared indicates that it explains 72.5% of the variation in bank ratings. We use the residuals from this model as a measure of examiner-specific effects in subsequent analyses.

It's important to note that this is a forecasting model, and the coefficients should be interpreted with caution due to potential correlations among variables. In forecasting models, we aim to incorporate as many relevant variables as possible to provide the most accurate prediction. The current model uses bank characteristics derived from both FR Y-9C and FR Y-9SP filings, which we later augment with additional variables in an expanded version when we focus on FR Y-9C banks.

In Figure 1 we compare the performances of the OLS and the Ordered Logit model fits in predicting bank ratings. The figure illustrates the distribution of examiner-specific effects computed using both models. In Figure 1, we observe that both models show a similar concentration of examiner-specific effects around zero. This suggests that both models have a similar level of accuracy in predicting the actual ratings assigned by examiners. Given this, we use the examiner-specific effects derived from the OLS model in all our subsequent analyses.

To further explore the model's predictive power, we focus on the subsample of publicly traded banks in Table 5. This subset allows us to access a broader range of regressors. Column 1 presents our baseline model with six explanatory variables, achieving an adjusted R-squared of 0.676. In Column 2, we expand to 29 explanatory variables by adding 23 FR Y-9C-specific ratios to the baseline set (starting composite rating and five financial ratios available in both Y-9C and Y-9SP statements). Including these additional regressors increases the adjusted R-squared to 0.702. Column 3 further augments the model by adding five market-based metrics (stock returns, volatility, and turnover), bringing the total to 34 variables and marginally improving the adjusted R-squared to 0.710. The results from Table 5 demonstrate that additional variables offer incremental improvements to the model's forecast. They also show that our parsimonious baseline model performs relatively well in predicting bank ratings. The modest gains in predictive power from the expanded models underscore the effectiveness of our initial variable selection.

Following this step, we assign measures of the examiner-specific effect to each bank-year observation based on the maximum amount of available information available for each observation. That means, for publicly traded banks, we utilize the saturated model incorporating both financial statement data and additional market variables. FR Y-9C filers that are not publicly traded are assessed using the saturated model with FR Y-9C variables, while the remaining banks are evaluated using our baseline model. This stratified approach maximizes the use of available information for each bank type.

A potential concern with this stratified approach is that using different prediction models across bank types may mechanically affect the magnitude of ESE. In particular, because publicly traded banks are evaluated against a more powerful model (Table 5, Column 3), their residual variance may be mechanically smaller than that of other banks, independent of true differences in transparency. To address this concern, we verify in Online Appendix B that our main findings are robust to using a single common baseline model (the 6-variable specification from Table 4) for all banks. Tables OAB1 and OAB2 demonstrate that public banks continue to exhibit significantly smaller examiner-specific deviations even when evaluated using the same benchmark as all other institutions, confirming that the transparency effect is not a mechanical artifact of differential model power.

6.3 Testing H1 - Access to Private Information and Rating Divergence

Interpreting supervision as information production, we expect that when supervisory contact is weaker (e.g., greater distance or off-site work), assigned ratings should be more tightly spanned by public observables, producing smaller deviations from the benchmark prediction.

To test Hypothesis 1, we analyze how examiners' access to private information affects the deviation of their ratings from model predictions based on observable fundamentals and prior supervisory assessments. Results for this analysis are presented in Tables 6 and 7.

As outlined in the Methodology section, we employ three proxies for access to private information: examiner-bank distance, examination type (off-site vs. on-site), and bank public trading status.[13] Table 6 presents results for distance and examination type, while Table 7 focuses on

[13]A fundamental difference exists between our distance and exam type measures versus public trading status, though both affect the private-to-public information ratio in the same direction. Distance and off-site exams reduce examiners' access to private information. In contrast, public trading status increases the availability of public information. Despite these distinct mechanisms, both reduce the relative share of private information available to examiners, leading to the same directional effect on rating deviations.

publicly traded banks. The dependent variable in these Tables is the absolute examiner-specific deviation.

In Table 6, we consider two different ways to capture access to private information based on distance. The first is a simple measure with a dummy that takes a value of one if the distance between bank and examiner's office is more than the median in the sample (Models 1, 3, and 7). The second uses three dummies based on whether the distance between bank and examiner office is less than 25 miles, between 25 miles and 50 miles, and more than 50 miles, following DeYoung, Glennon, and Nigro (2008) (Models 2, 4, and 8, with the omitted variable being the less than 25 miles dummy).

Consistent with Hypothesis 1, we find that examiners' ratings for banks further from their offices are more likely to align with model-predicted ratings based on public information. Specifically, being above median distance reduces the absolute examiner-specific deviation by 0.021 rating grades (Model 1), representing approximately 10% of the average deviation of 0.219. Further, being 50+ miles away—effectively in a different town—reduces the absolute deviation by 0.033 rating grades (Model 2), representing approximately 15% of the average deviation in the sample. These results are robust to controlling for time fixed effects (Model 4) and including bank-level controls (Model 8). Overall the results show that examiners that are examining a bank further away from their office are more likely to produce ratings that are closer to predicted ratings based on publicly available information. This finding aligns with our hypothesis that publicly traded banks are subject to greater scrutiny from various market participants, including stock analysts, investors, and regulatory bodies such as the SEC. This increased information production and transparency effectively diminishes the examiner's private information advantage, resulting in ratings that more closely align with model predictions based on publicly available information. As shown in Online Appendix B (Table OAB1), this result is robust to using a uniform baseline model for all banks, indicating that the transparency effect reflects genuine information differences rather than mechanical variation from model stratification.

Furthermore, our other measure of access to private information is a dummy variable that takes a value of one if the exam is conducted off-site. Importantly, while there is a positive correlation between a bank's distance from the examiner's office and the likelihood of an off-site exam, we find that the off-site dummy provides additional explanatory power even after controlling for distance variables. In Model 7, the coefficient on the off-site dummy is -0.027, significant at the 1% level, indicating that off-site examinations are associated with absolute deviations that are 0.027 rating grades smaller on average. Relative to the mean deviation of 0.219, this represents a 12% reduction. Model 8 demonstrates that when a bank is at least 50 miles away from the examiner's office and the exam is conducted off-site, the combined effect reduces the absolute examiner-specific deviation by approximately 0.07 rating grades compared to close-proximity, on-site examinations. This reduction is equivalent to roughly one-third of the average deviation in the sample.

Table 7 provides additional support for Hypothesis 1 by demonstrating that examiner estimates for publicly traded banks exhibit smaller deviations from model predictions compared to other banks. The table presents four different models, each varying in their choice of year fixed effects and controls. Notably, all models consistently show a significant negative relationship between the publicly traded dummy variable and the absolute examiner-specific deviation.

In the most restrictive specification (Model 4), the coefficient on the publicly traded dummy is -0.042, statistically significant at the 1% level. This indicates that public trading status reduces the discrepancy between examiner ratings and model predictions by 0.042 rating grades. Considering the average deviation of 0.219 rating grades, this reduction is equivalent to approximately 20% of the average deviation.

This finding aligns with our hypothesis that publicly traded banks are subject to greater scrutiny from various market participants, including stock analysts, investors, and regulatory bodies such as the SEC. This increased information production and transparency effectively diminishes the examiner's private information advantage, resulting in ratings that more closely align with model predictions based on publicly available information.

Overall, the results in Tables 6 and 7 are consistent with Hypothesis 1, demonstrating that parts of heterogeneity in examiners' ratings can be attributed to the examiners' access to private information.

The robustness of these findings to alternative estimation methods is demonstrated in Online Appendix A, which presents one-step specifications that directly test whether rating residuals vary with information access without constructing ESE as an intermediate variable. Across all specifications— one-step absolute deviation tests, variance tests, and our baseline two-step approach—coefficient estimates are quantitatively similar and statistical significance remains unchanged, confirming that our results are not driven by the two-step estimation procedure.

Next, we examine how the information effects of bank examinations vary with bank opacity. If examiner-specific deviations reflect the production of private supervisory information, then access to such information should matter most when public signals about bank condition are weak.

Table 8 reports the results. Columns 1–4 interact the publicly traded indicator with the off-site examination dummy, controlling for their individual effects. The baseline category represents on-site examinations of privately held banks. Both the off-site dummy and the publicly traded indicator enter negatively, indicating that each is associated with smaller rating deviations. Critically, the positive interaction term reveals that off-site examinations reduce rating deviations more for private banks than for public banks. This pattern suggests that off-site examinations compress information differences primarily for opaque institutions, while they provide little additional reduction in informational gaps for banks with readily available market-based information.

Columns 5–6 augment these specifications by including distance-bin fixed effects that capture non-linear proximity effects. These results reveal that on-site examinations generate substantially larger deviations from model-predicted ratings when examiners are geographically close to the bank, with proximity effects concentrated at shorter distances and attenuating sharply beyond 50 miles. These patterns are difficult to reconcile with examiner preferences or arbitrary discretion, but align naturally with a framework in which examiners' ability to gather soft information depends on physical proximity and direct interaction with bank management.

Taken together, these findings indicate that examiner access to private information has significantly stronger effects on rating deviations for opaque banks. Online Appendix B (Table OAB2) confirms that these interaction effects persist when all banks are evaluated using the same baseline model. In the following subsection, we address whether these patterns reflect causal information acquisition or systematic selection in exam assignment.

6.4 Addressing Endogeneity Concerns

A key identification challenge is whether exam assignment or bank location is endogenous to unobserved bank conditions. Four substantive concerns arise. First, although banks do not control examiner assignments, bank owners might theoretically select locations to increase the likelihood of working with certain examiners. Second, exam type and distance may reflect supervisory triage based on private information about bank risk. Third, time-varying bank characteristics may predict the assignment of examiners. Fourth, matching could occur around examiner reassignments. We address these concerns systematically.

Within-bank identification and examiner office fixed effects. Table 9 presents our core identification tests, which examine whether the relationship between information access and rating deviations reflects causal information acquisition or endogenous assignment. The dependent variable is the absolute examiner-specific effect, consistent with our baseline specification. Column 1 replicates the baseline specification from Table 6, Column 7. Column 2 demonstrates that adding examiner office fixed effects—which absorb all location-specific allocation strategies—*strengthens* rather than weakens the distance effect: the coefficient increases in magnitude from -0.022 to -0.036 (statistically significant at the 1% level). The same pattern emerges for the off-site examination dummy, for which the coefficient increases from -0.026 (t-statistic = 2.564) to -0.030 (t-statistic = 2.900). This pattern is inconsistent with office-level confounds driving our results and instead suggests that controlling for systematic office-level factors reveals stronger effects of information access based on proximity..

Columns 3 and 4 exploit within-bank variation by restricting the sample to bank-year observations with examiner switches. This approach compares the same bank examined by different examiners. The distance effect remains economically and statistically significant (-0.019, t-statistic = 2.226) and strengthens further when combined with office fixed effects (-0.035, t-statistic = 4.082). These within-bank tests provide our most demanding identification strategy, as they control for all time-invariant bank characteristics and isolate variation arising solely from examiner reassignments. The off-site examination effect, while significant in cross-sectional specifications, attenuates in within-bank tests. This likely reflects limited within-bank variation in off-site status compared to examiner identity, reducing statistical power. Importantly, the persistence of distance effects across all specifications—including the most restrictive combination of within-bank variation and office fixed effects—supports a causal interpretation: geographic proximity enables examiners to acquire private information that manifests in rating deviations.

Placebo tests: Ruling out anticipatory assignment. If supervisors assigned exam types or examiners based on private concerns about future bank deterioration, we would expect lagged or future exam characteristics to predict current examiner-specific effects. Table 10 tests this directly.

Column 1 shows that whether the prior exam was off-site have no predictive power for current rating deviations. The joint test yields p=0.290, strongly rejecting anticipatory assignment based on prior supervisory information. Column 2 tests whether future assignment changes (examiner switches or off-site status changes in period $t + 1$) predict current ratings, which would indicate forward-looking triage. Again, we find no relationship (joint p=0.236).

These placebo tests directly address the matching on time-varying private concerns. The absence of predictive power from lagged and future exam characteristics indicates that a potential relationship between assignments and unobservable information about bank trajectory cannot explain our results. Combined with the within-bank identification in Table 9, this evidence supports a causal interpretation: examiner-specific effects reflect information acquisition during examinations rather than selection into assignment.

Interpretation and remaining concerns. We acknowledge that exam assignment correlates with observable bank characteristics, as documented in unreported balance tests. However, this correlation likely reflects legitimate risk-based supervision—where larger, more complex banks receive different supervisory intensity—rather than confounding selection. Three findings support this interpretation:

First, our *within-bank* tests isolate variation where the same bank experiences different examiners, holding constant all permanent bank characteristics. The persistence of distance effects in these specifications (Table 9, Columns 3-4) indicates that distance-based information access effects operate even after absorbing level differences across banks.

Second, our *placebo tests* show no evidence that assignment anticipates future problems based on private supervisory information. If systematic selection were driving our results, we would expect lagged and future exam characteristics to predict current ratings. The absence of such relationships (Table 10) rules out the most concerning form of endogeneity.

Third, the *strengthening* of distance effects when adding office fixed effects (Table 9, Column 2) is inconsistent with omitted office-level factors confounding our estimates. If distance merely proxied for systematic differences in how regional offices allocate examiners, controlling for office fixed effects would attenuate rather than strengthen coefficients.

Taken together, these tests indicate that examiner-specific effects reflect supervisory information production rather than endogenous assignment based on unobserved bank conditions. While we cannot rule out all forms of selection, our evidence supports a causal interpretation of the core information access mechanisms documented in Section 6.

6.5 Testing H2 - Examiner Stringency and Rating Consistency

Having established that information access effects reflect causal mechanisms rather than selection, we now examine the role of individual examiner characteristics, particularly, examiner's relative stringency on rating assignments.[14] Our analysis of Hypothesis 2, presented in Table 11, reveals that individual examiner characteristics influence bank ratings, even after controlling for bank-specific factors and economic conditions. In the first five models of Table 11, we use a variety of controls to check the robustness of our findings. Our key explanatory variable is the measure of examiner stringency that is calculated as the average discrepancy between the examiner's ratings and model-predicted ratings for other banks the examiner examine. The coefficients for this variable are significant at the 1% level across all specifications, including those with time and bank fixed effects.

In the most restrictive model (Model 5), the coefficient for examiner stringency is 0.246. This indicates that an examiner who has, on average, given ratings one grade worse than the model-predicted rating in examinations of other banks assigns ratings that are, on average, 0.246 grades more stringent than other examiners for the current bank examination. Put differently, a one-unit increase in examiner stringency (measured as the leave-one-out average ESE across other banks) is associated with approximately a quarter-grade increase in the examiner-specific effect for the current examination. This persistent relationship demonstrates that examiners who have been tough in previous examinations tend to remain tough in current examinations as well. In other words, examiners who have been tough in previous examinations are more likely to be tough in current examinations as well. These results are robust when accounting for firm- and time-fixed

[14] While ESE measures the rating deviation for a specific examination, examiner stringency (equation (7)) captures an examiner's persistent tendency toward toughness or leniency across all their examinations. We use these terms to distinguish between examination-specific effects and examiner-level characteristics.

effects, as well as the choice of bank controls.

A potential concern regarding our results is the possibility of selection bias in examiner assignment, specifically, if tougher examiners are systematically assigned to weaker banks—those that are in a more precarious financial position than their public data or financial statements would indicate. If this were the case, it could lead to a spurious correlation between examiner stringency and bank outcomes, potentially biasing our results. To address this concern, we implement a falsification test in the final specification of Table 11. This test involves regressing the starting rating of the bank on examiner stringency. The rationale behind this approach is that if tougher examiners were indeed systematically assigned to banks with weaker initial conditions, we would expect to see a significant positive relationship between examiner stringency and the starting bank rating.

Our results, however, do not support this alternative explanation. The coefficient on examiner stringency in this regression is not statistically different from zero, which suggests that there is no systematic relationship between examiner stringency and the initial condition of the banks they are assigned to examine. This finding aligns with our understanding of the examiner assignment process, which is primarily based on examiner availability rather than an intentional matching of examiner characteristics to bank conditions.

Overall, the results in Table 11 support Hypothesis 2 in that the variations we observe in bank outcomes following examinations are partially attributable to differences in examiner stringency. In other words, examiners' intrinsic characteristics can influence their ratings.

6.6 Testing H3 - Temporal Trends in Examiner Stringency

Next, we move to test Hypothesis 3, which posits that the average stringency of bank examinations varies over time, with examiners tending to be more stringent following periods of financial stress or economic downturns, and relatively more lenient during periods of economic stability and growth. To investigate this hypothesis, we plot the annual average examiner-specific effects in Figure 2.

Figure 2 shows the temporal variation in examiner stringency. The figure plots the mean examiner-specific effects with 90% confidence intervals for each year, where a negative value indicates that ratings in that year were on average more lenient than expected, and a positive value indicates that ratings are on average more stringent than expected. The shaded areas on the figure represent the Global Financial Crisis (2007Q4-2009Q2) and the banking turmoil of 2023Q1.

Our results reveal clear patterns of leniency during periods leading up to financial crises and increased stringency following crises. For instance, during the Global Financial Crisis, the average examiner-specific effect reached its lowest level (-0.280), indicating a period of relative leniency, and subsequently rebounded to its highest level (0.080) in 2014, following the crisis, as examiners responded to increased public scrutiny and attention from regulators, as well as the enactment of the Dodd-Frank Act, by adopting a more stringent approach to bank examinations. We also observe that bank ratings tend to be relatively lenient in the years leading up to the banking turmoil of 2023 that led to the failure of Silicon Valley Bank, Signature Bank and First Republic Bank following a period of economic expansion. Overall, the results in this section support our hypothesis that the average examiner is possibly influenced by the macroeconomic environment.

A potential concern with interpreting time-series variation in average examiner-specific effects is that these patterns may reflect parameter instability in our forecasting model rather than genuine shifts in supervisory stringency. Because our benchmark model is estimated over the full 2005–2023 sample, it imposes a constant mapping from financial ratios to ratings across distinct macroeconomic regimes. If the informational value of bank fundamentals varies systematically with economic conditions, this could mechanically generate time-varying deviations even absent changes in examiner behavior.

To address this concern, we test whether the relationships between information access measures (examiner-bank distance and examination modality) and rating deviations remain stable across time periods. We estimate interaction models that allow coefficients to vary across four distinct periods: Pre-Crisis (2006–2008), Crisis/Recovery (2009–2015), Post-Regulation (2016–2019), and COVID Era (2020–2023). Results, reported in Online Appendix C, show that distance effects remain stable across all periods (joint test p-value = 0.692), while offsite examination effects exhibit meaningful time variation (joint test p-value = 0.029). Notably, offsite examinations are associated with significantly larger reductions in rating stringency during crisis periods (GFC and COVID) compared to normal times. This pattern is consistent with our interpretation of Figure 2 as reflecting shifts in supervisory risk tolerance rather than model misspecification, while also revealing that information access effects vary systematically with the supervisory environment.

The results so far support our Hypotheses 1-3, demonstrating that factors such as access to private information, examiner intrinsic stringency, and changing economic conditions contribute to heterogeneity in examiner practices and affect bank ratings. In the next set of tests, we examine how these examiner-specific effects are related to future bank risk-taking and growth, addressing Hypothesis 4.

6.7 Testing H4 - Impact of Examiner-Specific Effects on Bank Behavior

To analyze Hypothesis 4, we investigate the relationship between examiner-specific effects and future bank behavior. We use our main measure of examiner-specific effect (ESE_t),[15] as well as two decomposed measures to explore potential asymmetric effects: a positive component (ESE_t^+), capturing tougher-than-predicted ratings, and a negative component (ESE_t^-), capturing more lenient ratings. Formally, $ESE_t^+ = \max(ESE_t, 0)$ and $ESE_t^- = \max(-ESE_t, 0)$.

To ensure interpretive consistency, we address mechanical censoring at the bounds of the 1–5 rating scale. We exclude from the construction of ESE_t^+ those bank-year observations where the actual rating is 1, as such observations cannot reflect stricter-than-predicted outcomes.[16] Rating 1 cannot be stricter-than-predicted, and rating 5 cannot be more lenient-than-predicted, so the corresponding boundary observations are excluded only from the mechanically constrained side of the decomposition.

We estimate panel regressions with bank and year fixed effects to evaluate how these measures relate to bank behavior in the subsequent year. Table 12 presents results using our baseline specification for Log(Assets) in Model 1. Due to evidence of asymmetric responses to examiner-specific effects across all dependent variables, we primarily report results for ESE_t^+ and ESE_t^- in Models 2–11 to conserve space. Our findings generally indicate that banks respond more strongly to tougher-than-predicted ratings (ESE_t^+). This is evidenced by statistically and economically significant effects on key metrics such as bank size, lending activity, and capital ratios. These results suggest that examiner stringency has a more pronounced impact on future bank behavior compared to examiner leniency.

Notably, a one-grade increase above the model-predicted rating is associated with approximately

[15] ESE_t represents $ESE_{i,t}$ for brevity. All regressions are conducted at the bank(i)-year(t) level, consistent with our previous analyses.

[16] A rating of 1 is already the most favorable supervisory rating and cannot be stricter than expected. Similarly, for the construction of ESE_t^-, we exclude observations with an actual rating of 5, which is the harshest possible rating. These constraints ensure that the decomposed measures capture meaningful upward and downward rating deviations.

a 5.9% reduction in total assets (Model 2) and a 2.1 percentage point decrease in the loan-to-asset ratio (compared to a sample mean of 34.5%) (Model 4). Moreover, we observe a positive relationship with bank capital. Specifically, a one-grade higher than predicted rating correlates with a 0.3 percentage point increase in the equity-to-assets ratio (Model 6) and a 0.4 percentage point increase in the Tier 1 ratio (Model 8). These effects are substantial when compared to the sample averages of 9.7% for equity-to-assets and 13.1% for the Tier 1 ratio. These findings strongly support Hypothesis 4, demonstrating that banks receiving tougher-than-predicted ratings subsequently reduce both risk-taking and growth. The observed changes in asset size, lending activity, and capital ratios collectively indicate a more conservative approach to banking operations following stricter supervisory assessments.

In contrast to the pronounced effects of stricter ratings, our analysis reveals that responses to lenient ratings (ESE_t^-) are generally muted or statistically insignificant across most metrics. However, there is one notable exception: the non-performing loans (NPL) ratio. We find that lenient ratings are associated with a statistically significant increase in NPLs, suggesting that supervisory leniency may allow risk to accumulate within banks. Quantitatively, a rating that is one grade more lenient than the model-predicted rating corresponds to a 0.1 percentage point increase in the NPL-to-asset ratio. This effect is meaningful when compared to the sample average NPL-to-assets ratio of 1.3%. These findings provide nuanced support for the second part of Hypothesis 4, indicating that while banks may not significantly alter their overall behavior in response to lenient ratings, there is evidence of increased risk accumulation in their loan portfolios.[17]

Table 13 replicates the analysis from Table 12 while incorporating examiner fixed effects. This approach aims to isolate the component of the examiner-specific effect related to private information, net of persistent examiner-level toughness or leniency. Notably, our results remain robust under this specification. For instance, the effect on asset growth is -0.056, nearly identical to the baseline model. Similarly, the lenient-rating effect on NPLs remains positive and statistically significant (0.001, significant at the 5% level), indicating that both tough and lenient supervisory

[17] One potential concern is that our findings could be influenced by changes in supervisory frameworks over time. While our baseline sample already excludes post-2019 LFI observations (affecting fewer than 1% of bank-year observations), institutions that later transitioned to LFI may have been systematically different even during their RFI period. To address this concern, we verify that our main results are robust to (i) excluding these institutions from the entire sample period (2005–2023) and (ii) restricting the full sample to the pre-2019 period only. In both cases, the magnitude and statistical significance of examiner-specific effects, their relationship with information access, and their impact on subsequent bank behavior remain unchanged.

information—beyond persistent examiner stringency—affect bank behavior. This consistency suggests that a significant portion of the effect is driven by private, non-public information uncovered during the examination process.

Our findings continue to support an asymmetric response pattern. Tougher-than-expected ratings consistently drive contraction and de-risking across various metrics. In contrast, lenient ratings show less systematic effects, primarily manifesting in higher NPLs and, in some instances, a modest increase in total assets. Importantly, both patterns persist after controlling for examiner identity, reinforcing the information-channel interpretation. These results not only corroborate Hypothesis 4 but also carry substantial regulatory implications.

Overall, our evidence demonstrates that examiner-specific effects are not mere statistical artifacts but meaningful supervisory signals. Ratings that deviate from model predictions based on observable fundamentals and prior supervisory assessments have tangible consequences for banks' balance sheet strategies and risk postures in the subsequent year. These findings underscore the importance of considering examiner effects in the broader context of banking supervision and regulation.

7 Conclusion

This paper studies bank supervision as an information-production process and examines how individual examiners shape supervisory outcomes and subsequent bank behavior. Using confidential supervisory ratings for U.S. bank holding companies from 2005 to 2023, we develop a novel measure—the examiner-specific effect (ESE)—defined as the deviation between assigned ratings and ratings predicted using observable financial and market information together with prior supervisory ratings. This framework separates the systematic component of supervision grounded in observable fundamentals from the incremental information and judgment introduced through the examination process.

Our first set of findings demonstrates that supervision adds value through information production. Examiner-specific effects are closely linked to information access: examiners who are geographically closer to the banks they assess, who conduct on-site examinations, or who evaluate privately held institutions deviate more from observable-information benchmarks. By contrast, ratings for publicly traded banks—where market signals are richer—exhibit significantly smaller deviations. This pattern is strongest for on-site examinations of privately held banks and attenuates sharply when examinations are conducted remotely or when public market signals are abundant, establishing that supervision generates incremental information beyond what is available through public disclosures and market discipline.

Second, we document persistent heterogeneity across examiners. Even conditional on information access, some examiners are systematically more stringent or lenient than others, and these differences persist across banks and over time. This evidence indicates that human judgment plays a complementary role in supervisory outcomes. Third, supervisory stringency varies systematically over the financial cycle, with ratings relatively lenient during economic expansions and more stringent following periods of financial stress, adding a dynamic dimension to supervision's information-production role.

Most importantly, examiner-specific effects have economically meaningful real consequences. Banks receiving tougher-than-predicted ratings subsequently reduce asset growth, curtail lending, and increase capital ratios. These responses are asymmetric: unexpectedly stringent ratings trigger sharp balance-sheet adjustments, while lenient ratings have muted short-run effects but are associated with greater future risk accumulation, as reflected in higher nonperforming loans.

Our findings highlight a fundamental tension in supervisory oversight. Intensive supervision that facilitates deep information acquisition promotes safety and soundness but can constrain bank growth and lending, particularly during downturns. Understanding both the informational advantages of supervision and the sources of examiner-level variation is essential for evaluating how bank oversight contributes to financial stability.

More broadly, this paper contributes to a growing literature emphasizing the role of information production in regulatory oversight. Our findings demonstrate that the value of supervision depends critically on examiners' ability to access and process non-public information, with human judgment playing a complementary role in interpreting that information.

References

Acharya, Viral V, Allen N Berger, and Raluca A Roman, 2018, Lending implications of us bank stress tests: Costs or benefits?, *Journal of Financial Intermediation* 34, 58–90.

Acharya, Viral V, Lasse H Pedersen, Thomas Philippon, and Matthew Richardson, 2017, Measuring systemic risk, *The review of financial studies* 30, 2–47.

Acharya, Viral V and Stephen G Ryan, 2016, Banks' financial reporting and financial system stability, *Journal of Accounting Research* 54, 277–340.

Acharya, Viral V and Tanju Yorulmazer, 2007, Too many to fail—an analysis of time-inconsistency in bank closure policies, *Journal of financial intermediation* 16, 1–31.

Agarwal, Sumit, David Lucca, Amit Seru, and Francesco Trebbi, 2014, Inconsistent regulators: Evidence from banking, *The Quarterly Journal of Economics* 129, 889–938.

Agarwal, Sumit, Bernardo C Morais, Amit Seru, and Kelly Shue, 2024, Noisy experts? discretion in regulation, Working paper, National Bureau of Economic Research.

Angelini, Paolo, Francesca Lotti, Andrea Passalacqua, and Giovanni Soggia, 2021, The real effects of bank supervision: evidence from on-site bank inspections, *SSRN Work. Pap* 3705558.

Berger, Allen N and Sally M Davies, 1998, The information content of bank examinations, *Journal of Financial Services Research* 14, 117–144.

Berger, Allen N, Thomas Kick, and Klaus Schaeck, 2014, Executive board composition and bank risk taking, *Journal of Corporate Finance* 28, 48–65.

Beyhaghi, Mehdi, Cesare Fracassi, and Gregory Weitzner, 2025, Adverse selection in corporate loan markets, *The Journal of Finance* .

Beyhaghi, Mehdi and Jeffrey R Gerlach, 2024, Discount window borrowing 2003-2019, *Available at SSRN 4919012* .

Bonfim, Diana, Geraldo Cerqueiro, Hans Degryse, and Steven Ongena, 2023, On-site inspecting zombie lending, *Management Science* 69, 2547–2567.

Boyd, John H and Edward C Prescott, 1986, Financial intermediary-coalitions, *Journal of Economic theory* 38, 211–232.

Clement, Michael B, 1999, Analyst forecast accuracy: Do ability, resources, and portfolio complexity matter?, *Journal of Accounting and Economics* 27, 285–303.

Cole, Shawn, Martin Kanz, and Leora Klapper, 2015, Incentivizing calculated risk-taking: Evidence from an experiment with commercial bank loan officers, *Journal of Finance* 70, 537–575.

Correia, Sergio, Stephan Luck, and Emil Verner, 2025, Supervising failing banks, *SSRN Working Paper* .

Costello, Anna M, Joao Granja, and Joseph Weber, 2019, Do strict regulators increase the transparency of banks?, *Journal of Accounting Research* 57, 603–637.

Dang, Tri Vi, Gary Gorton, Bengt Holmström, and Guillermo Ordonez, 2017, Banks as secret keepers, *American Economic Review* 107, 1005–1029.

DeYoung, Robert, Mark J Flannery, William W Lang, and Sorin M Sorescu, 2001, The information content of bank exam ratings and subordinated debt prices, *Journal of Money, Credit and Banking* 900–925.

DeYoung, Robert, Dennis Glennon, and Peter Nigro, 2008, Borrower–lender distance, credit scoring, and loan performance: Evidence from informational-opaque small business borrowers, *Journal of financial intermediation* 17, 113–143.

Flannery, Mark J, Simon H Kwan, and Mahendrarajah Nimalendran, 2004, Market evidence on the opaqueness of banking firms' assets, *Journal of Financial Economics* 71, 419–460.

Francis, Jere R and Dechun Yu, 2009, Big 4 office size and audit quality, *The Accounting Review* 84, 1521–1552.

Gopalan, Yadav, 2022, The effects of ratings disclosure by bank regulators, *Journal of Accounting and Economics* 73, 101438.

Gopalan, Yadav, Ankit Kalda, and Asaf Manela, 2021, Hub-and-spoke regulation and bank leverage, *Review of Finance* 25, 1499–1545.

Gul, Ferdinand A, Bikki L Jaggi, and Gopal V Krishnan, 2007, Auditor independence: Evidence on the joint effects of auditor tenure and nonaudit fees, *Auditing: a journal of practice & theory* 26, 117–142.

Hirtle, Beverly, 2016, Public disclosure and risk-adjusted performance at bank holding companies, *Economic Policy Review, Issue Aug* 151–173.

Iannotta, Giuliano, 2006, Testing for opaqueness in the european banking industry: evidence from bond credit ratings, *Journal of Financial Services Research* 30, 287–309.

Ivanov, Ivan T and James Z Wang, 2024, Bank supervision and corporate credit supply, *Management Science* 70, 3338–3361.

Kandrac, John and Bernd Schlusche, 2021, The effect of bank supervision and examination on risk taking: Evidence from a natural experiment, *The Review of Financial Studies* 34, 3181–3212.

Karolyi, Stephen A, Lewis Gaul, Jonathan Jones, and Pinar Uysal, 2024, Internal loan ratings, supervision, and procyclical leverage, *SRRN Working Paper* .

Ke, Bin and Yong Yu, 2006, The effect of issuing biased earnings forecasts on analysts' access to management and survival, *Journal of Accounting Research* 44, 965–999.

Morgan, Donald P, 2002, Rating banks: Risk and uncertainty in an opaque industry, *American Economic Review* 92, 874–888.

Srivastav, Abhishek and Jens Hagendorff, 2016, Corporate governance and bank risk-taking, *Corporate Governance: An International Review* 24, 334–345.

Townsend, Robert M, 1979, Optimal contracts and competitive markets with costly state verification, *Journal of Economic theory* 21, 265–293.

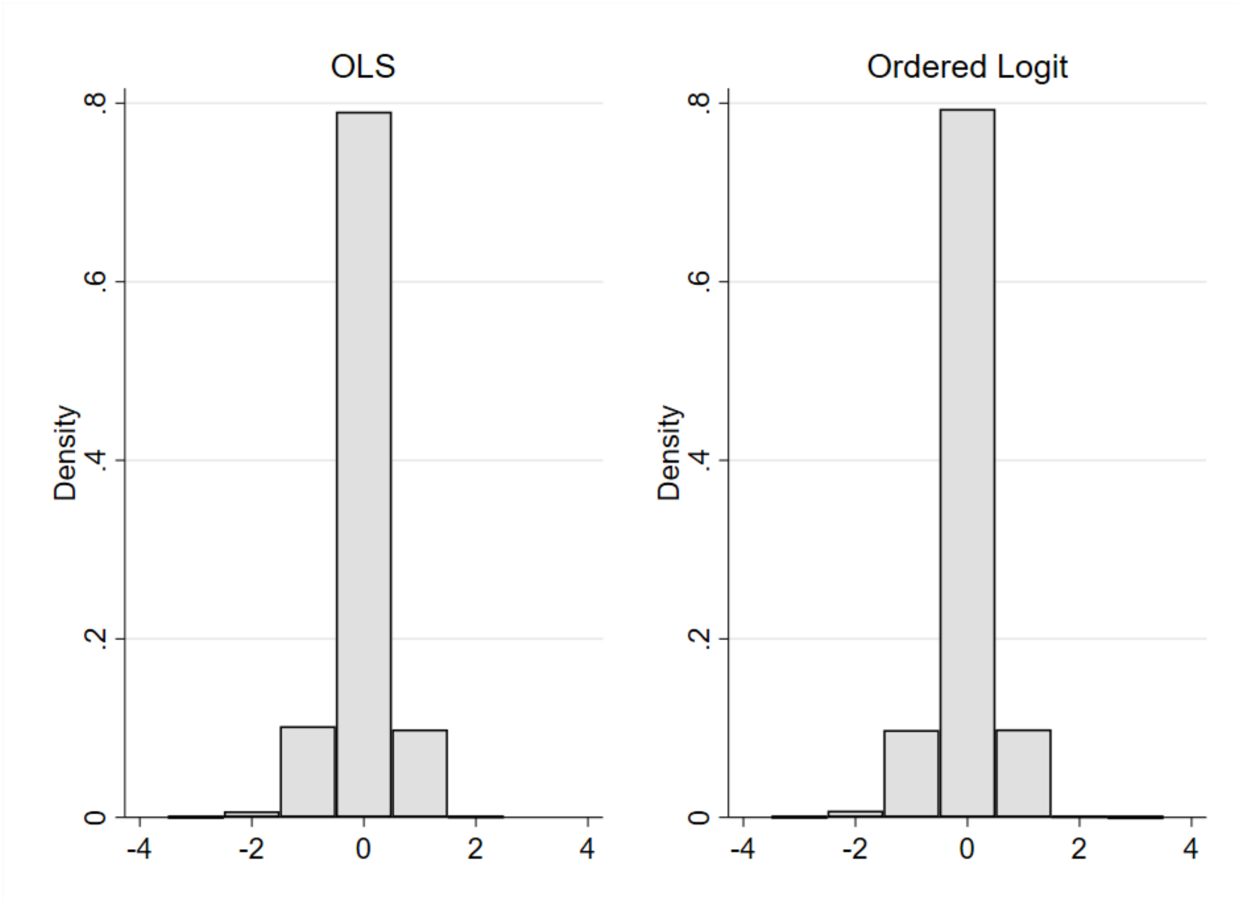

Figure 1: **Distribution of Examiner-Specific Effects.** This figure compares the distribution of estimated examiner-specific effects from two models: an ordinary least squares (OLS) model and an ordered logit (OLOGIT) model. The x-axis represents the difference between the predicted model rating and the actual examiner rating, with a value of zero indicating perfect agreement between the two. Positive values indicate that the examiner's rating was higher than the predicted model rating, while negative values indicate that the examiner's rating was lower.

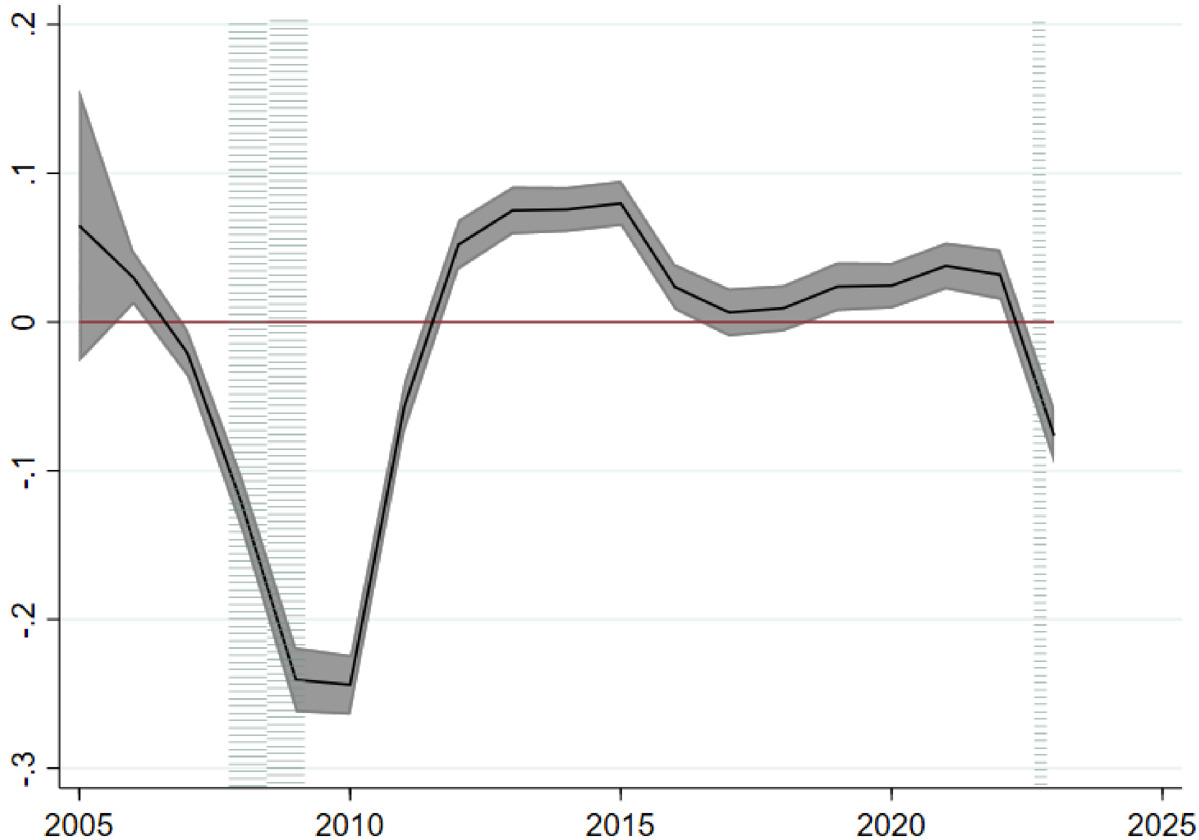

Figure 2: **Average Examiner-Specific Effect on Rating.** This figure presents the average examiner-specific effect on bank ratings, calculated as the difference between actual ratings and model-predicted ratings. The predictions use models that incorporate observable financial and market information together with prior supervisory ratings to forecast examination ratings. The figure plots the average examiner effect over time from 2005 to 2023, along with a 90% confidence band (5th-95th percentile). Shaded areas represent the Global Financial Crisis (2007Q4–2009Q2) and the banking turmoil of 2023Q1.

Table 1: **Comparison of Major U.S. Bank Supervisory Rating Systems**

This table summarizes key institutional differences across U.S. supervisory rating systems. The RFI/C(D) framework applies uniformly to bank holding companies under Federal Reserve supervision and emphasizes holistic examiner judgment rather than mechanical aggregation of components, making it particularly suitable for studying examiner-specific information and heterogeneity.

	CAMELS	**RFI/C(D)**	**LFI**
Applicability	Depository institutions	Bank holding companies	Large BHCs ($100B+, since 2019)
Components	Capital adequacy; Asset quality; Management; Earnings; Liquidity; Sensitivity to risk	Risk management; Financial condition; Impact of non-depository entities; Composite condition; Depository CAMELS	Capital planning and positions; Liquidity risk management; Governance and controls
Composite rating	Yes	Yes	No
Rating scale	Numeric: 1 (best) to 5 (worst)	Numeric: 1 (best) to 5 (worst)	Non-numeric: Broadly Meets Expectations; Conditionally Meets Expectations; Deficient-1; Deficient-2
Supervisory agencies	Federal Reserve OCC, FDIC, State regulators	Federal Reserve only	Federal Reserve only

Table 2: **Distribution of RFI Composite Ratings**

This table presents the frequency distribution of RFI composite ratings. The sample include all bank holding companies that have filed FR Y9-C forms at least once during the sample during 2005–2023. Ratings range from 1 (strong) to 5 (unsatisfactory).

RFI Composite Rating	Freq.	Percent
1	5,957	23.60
2	14,474	57.33
3	3,070	12.16
4	1,095	4.34
5	649	2.57
Total	25,245	100.00

Table 3: **Summary Statistics**

This table presents summary statistics for a panel of 2,593 unique bank holding companies that filed Form Y-9C at least once between 2005 and 2023. The data are observed at the bank-year level. The variables used in this analysis are defined in the Appendix.

	N	Mean	SD	10%	Median	90%
Total Assets (thousand USD)	25,190	11,241,954	95,984,616	222,739	583,463	6,755,556
Log(Assets)	25,190	13.672	1.511	12.314	13.277	15.726
Equity/Assets	25,190	0.097	0.033	0.058	0.094	0.137
Securities/Assets	25,190	0.098	0.115	0.000	0.038	0.293
Loans/Assets	25,189	0.345	0.342	0.000	0.390	0.779
ROA	24,996	0.006	0.007	-0.001	0.007	0.015
Trading Assets/Assets	13,237	0.002	0.006	0.000	0.000	0.002
Allowance for Loans/Assets	13,237	0.010	0.005	0.005	0.009	0.018
Leverage Ratio	13,213	0.091	0.030	0.062	0.092	0.124
Double Leverage Ratio	13,237	0.005	0.018	0.000	0.000	0.005
Tier 1 Capital Ratio	12,707	0.131	0.042	0.087	0.123	0.186
Ratio of Residential Real Estate Loans	13,216	0.256	0.147	0.076	0.241	0.463
Ratio of Commercial Real Estate Loans	13,216	0.437	0.180	0.187	0.443	0.672
Ratio of Commercial and Industrial Loans	13,216	0.147	0.096	0.039	0.132	0.281
Ratio of Consumer Loans	13,216	0.053	0.068	0.003	0.027	0.146
Ratio of Depository Institution Loans	13,216	0.000	0.002	0.000	0.000	0.000
Ratio of Foreign Institution Loans	13,212	0.000	0.000	0.000	0.000	0.000
Ratio of Agriculture Loans	13,216	0.018	0.039	0.000	0.001	0.062
Loans Herfindahl-Hirschman Index	13,216	0.370	0.128	0.216	0.355	0.541
Assets Herfindahl-Hirschman Index	13,216	0.587	0.405	0.297	0.470	0.947
Risk-Weighted Assets/Assets	12,729	0.728	0.111	0.572	0.738	0.867
NPL/Assets	13,237	0.013	0.016	0.001	0.007	0.034
NPL/Loans	13,216	0.020	0.024	0.002	0.010	0.052
Loan Loss Reserve/Loans	13,216	0.016	0.008	0.008	0.013	0.027
Loan Loss Reserve/NPL	13,036	2.533	3.638	0.388	1.224	5.820
NPL-Residential Real Estate	13,117	0.018	0.022	0.001	0.009	0.046
NPL-Commercial Real Esate	13,119	0.024	0.033	0.000	0.010	0.071
NPL-Commercial and Industrial	12,789	0.015	0.020	0.000	0.007	0.040
NPL-Consumer	13,070	0.006	0.009	0.000	0.003	0.016
Market-to-Book Ratio	3,090	1.248	0.588	0.585	1.155	2.052
Monthly Stock Return (1 quarter) (%)	3,542	0.208	0.650	-0.483	0.098	1.000
Monthly Stock Return (1 year) (%)	3,500	0.072	0.292	-0.296	0.060	0.449
Stock Return Volatility	3,500	0.082	0.044	0.039	0.071	0.143
Average Stock Turnover	3,542	0.097	0.107	0.011	0.064	0.206

Table 4: **Benchmark Forecast Model**

This table presents the results of a benchmark forecast model that predicts examiner ratings based on bank fundamentals, using a set of ratios that are available for all banks. We estimate three alternative models with the same set of variables: an ordered logit model, a logistic model, and an ordinary least squares (OLS) model. T-statistics are shown below the parameter estimates in parentheses and are calculated using robust standard errors clustered by bank. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Composite Rating		
	(1)	(2)	(3)
	Ordered Logit	Logit	OLS
Last Composite Rating	3.021***	2.769***	0.684***
	(78.053)	(50.452)	(103.869)
Log(Assets)	0.105***	0.121***	0.015***
	(6.976)	(4.724)	(4.932)
Equity/Assets	−8.322***	−9.303***	−1.884***
	(10.187)	(7.241)	(11.685)
Securities/Assets	−1.440***	−1.635***	−0.300***
	(5.700)	(4.251)	(5.969)
Loans/Assets	0.863***	0.859***	0.184***
	(9.605)	(6.760)	(9.937)
ROA	−136.416***	−181.396***	−32.211***
	(36.627)	(26.678)	(39.714)
Constant		−7.908***	0.820***
		(21.599)	(18.119)
Threshold for category 2 vs. 1	3.456***		
	(15.645)		
Threshold for category 3 vs. 1-2	8.635***		
	(36.992)		
Threshold for category 4 vs. 1-3	12.073***		
	(47.517)		
Threshold for category 5 vs. 1-4	15.233***		
	(54.585)		
Observations	22,754	22,754	22,754
Adj. R-squared			0.725
Pseudo R-squared	0.481	0.570	

Table 5: **Forecast Models Using Observable Financial Information and Prior Supervisory Ratings**

This table presents the results of three forecast models that predict the composite rating of a bank based on publicly available information. We report the adjusted R-squared for each model, which measures model fit and allows for comparison across the three models. The models differ in their complexity and the types of variables used: Model 1 (Baseline): Uses basic ratios obtained from financial statements, available for all banks, as described in Table 2. Model 2 (Saturated): Incorporates a wide range of ratios built from Y-9C filings. Model 3 (Market-Enhanced): Adds market-related information to the variables used in Model 2, including average monthly stock returns, market return volatility, turnover, and market-to-book ratio in the month preceding the examination. T-statistics are shown below the parameter estimates in parentheses and are calculated using robust standard errors clustered by bank. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Composite Rating		
	(1)	(2)	(3)
	Baseline Model	Saturated Model (from Statements)	With Additional Market Variables
Number of Variables	6	29	34
Observations	3,419	2,707	2,684
Adj. R-squared	0.676	0.702	0.710

Table 6: **Relationship between Examiner-Specific Effect and Measures of Access to Private Information (1)**

This table examines the relationship between the absolute Examiner-Specific Effect on Rating, which measures the absolute difference between the model-predicted and actual rating, and various measures of distance between the bank and the examiner's office. Additionally, the table investigates the relationship between the absolute Examiner-Specific Effect on Rating and whether the exam was conducted online. A larger value of the dependent variable indicates that the model was less predictive of the actual rating based on available public information. T-statistics are shown below the parameter estimates in parentheses and are calculated using robust standard errors clustered by bank. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Absolute Examiner-Specific Deviation							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Distance Above Median	-0.021***		-0.020***				-0.022***	
	(2.773)		(2.676)				(3.150)	
Distance 25 to 50 Miles		-0.003		-0.005				-0.019
		(0.257)		(0.406)				(1.496)
Distance > 50 Miles		-0.033***		-0.031***				-0.044***
		(3.548)		(3.438)				(4.887)
Offsite Exam					-0.040***	-0.026**	-0.027***	-0.026**
					(3.713)	(2.360)	(2.599)	(2.525)
Examination Year FE			YES	YES		YES	YES	YES
Bank FE					YES	YES		
Controls							YES	YES
Observations	22,754	22,754	22,754	22,754	22,625	22,625	22,754	22,754
Adj. R-squared	0.001	0.001	0.027	0.027	0.077	0.097	0.040	0.041

Table 7: **Relationship between Examiner-Specific Effect and Measures of Access to Private Information (2)**

This table examines the relationship between the absolute Examiner-Specific Effect on Rating, which measures the absolute difference between the model-predicted and actual rating, and whether the bank is publicly traded. A larger value of the dependent variable indicates that the model was less predictive of the actual rating based on available public information. T-statistics are shown below the parameter estimates in parentheses and are calculated using robust standard errors clustered by bank. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Absolute Examiner-Specific Deviation			
	(1)	(2)	(3)	(4)
Publicly Traded	−0.085***	−0.074***	−0.034***	−0.042***
	(10.508)	(9.173)	(3.673)	(3.817)
Examination Year FE		YES	YES	YES
Controls (Baseline)			YES	
Controls (Saturated)				YES
Observations	22,754	22,754	22,754	10,041
Adj. R-squared	0.005	0.030	0.039	0.052

Table 8: **Information Access, Examination Modality, and Rating Deviations**

This table examines how examiner access to private information affects deviations between actual supervisory ratings and model-predicted ratings. The dependent variable is the absolute examiner-specific effect, defined as the absolute difference between the assigned rating and the rounded model-predicted rating. Columns (1)–(4) replicate the baseline specifications from Table 6 and introduce interactions with an indicator for off-site examinations. Columns (5)–(6) augment the specification by including distance-bin indicators capturing examiner proximity to the bank. All specifications include bank fixed effects and year fixed effects, and standard errors are clustered at the bank level. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels.

	Absolute Examiner-Specific Deviation					
	(1)	(2)	(3)	(4)	(5)	(6)
Publicly Traded	-0.131^{***}	-0.113^{***}	-0.068^{***}	-0.072^{***}	-0.065^{***}	-0.071^{***}
	(9.929)	(8.660)	(4.973)	(5.001)	(4.768)	(4.917)
Offsite Exam	-0.038^{***}	-0.021^{**}	-0.043^{***}	-0.027^{**}	-0.040^{***}	-0.027^{**}
	(4.348)	(2.430)	(4.512)	(2.353)	(4.257)	(2.352)
Publicly Traded × Offsite Exam	0.066^{***}	0.065^{***}	0.054^{***}	0.061^{***}	0.051^{***}	0.061^{***}
	(3.821)	(3.800)	(3.163)	(3.165)	(2.985)	(3.174)
Examination Year FE		YES	YES	YES	YES	YES
Controls			YES		YES	
Controls (Saturated)				YES		YES
Distance Bin FE					YES	YES
Observations	22,754	22,754	22,754	10,041	22,754	10,041
Adj. R-squared	0.006	0.030	0.040	0.053	0.041	0.053

Table 9: **Information Access and Within-Bank Identification**

This table tests whether the relationship between information access and rating deviations reflects causal information acquisition or endogenous assignment. The dependent variable is the absolute examiner-specific effect. Column 1 presents the baseline specification. Column 2 adds examiner office fixed effects to control for location-specific allocation strategies. Columns 3 and 4 restrict the sample to examiner switches within banks, exploiting variation where the same bank is examined by different examiners. Column 4 combines within-bank identification with office fixed effects, providing the most demanding specification. All specifications include year fixed effects and baseline controls (log assets, equity/assets, securities/assets, loans/assets, ROA). Standard errors are clustered by bank. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Absolute Examiner-Specific Deviation			
	(1)	(2)	(3)	(4)
Distance Above Median	−0.022***	−0.036***	−0.019**	−0.035***
	(3.150)	(5.050)	(2.226)	(4.082)
Offsite Exam	−0.026**	−0.030***	−0.003	−0.004
	(2.564)	(2.900)	(0.289)	(0.294)
Examination Year FE	YES	YES	YES	YES
Examiner Office FE	No	Yes	No	Yes
Sample	All	All	Examiner Switches	Examiner Switches
Identification	Cross-Sectional	Cross-Sectional	Within-Bank	Within-Bank
Observations	22,754	22,754	14,230	14,230
Adj. R-squared	0.040	0.045	0.050	0.059

Table 10: **Placebo Tests: Ruling Out Anticipatory Assignment**

This table tests whether exam assignment is based on private supervisory information about future bank deterioration. If supervisors assigned exam types or examiners anticipating problems, lagged or future exam characteristics should predict current rating deviations. Column 1 tests whether lagged exam characteristics (whether the prior exam was off-site and whether distance was above median) predict current examiner-specific effects. Column 2 tests whether future assignment changes (examiner switches or off-site changes in period $t + 1$) predict current ratings. Column 3 reproduces baseline results for comparison. The joint test p-values test whether lagged/future characteristics are jointly significant. High p-values indicate no anticipatory assignment. All specifications include year and bank fixed effects and baseline controls. Standard errors are clustered by bank. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Absolute Examiner-Specific Deviation	
	(1)	(2)
Lagged Offsite Status	0.012	
	(1.059)	
Change in Examiner Assignment		0.006
		(0.818)
Change in Offsite Status		0.017
		(1.433)
Joint Test p-value	0.290	0.236
Examination Year FE	YES	YES
Bank FE	YES	YES
Controls	YES	YES
Observations	22,495	20,206
Adj. R-squared	0.098	0.094

Table 11: **Relationship between Examiner-Specific Effect and Examiner Stringency**

This table examines the relationship between the examiner-specific effect on a bank's rating and the examiner's stringency, as measured by their behavior in exams of other banks. We estimate several specifications that control for examination year and bank fixed effects, as well as baseline or saturated bank-level controls (excluding starting rating). The first five columns report the results of these regressions, with the examiner-specific effect as the dependent variable. In the last column, we use the starting rating as the dependent variable, to test whether stringent examiners are systematically assigned to banks with weaker initial ratings, which would suggest selection bias in examiner assignment. T-statistics are shown below the parameter estimates in parentheses and are calculated using robust standard errors clustered by bank. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Examiner Specific Effect on Rating					Starting Rating
	(1)	(2)	(3)	(4)	(5)	(6)
Examiner Stringency	0.558***	0.380***	0.384***	0.376***	0.246***	-0.049
	(16.388)	(11.361)	(10.370)	(10.149)	(5.073)	(0.943)
Examination Year FE		YES	YES	YES	YES	YES
Bank FE			YES	YES	YES	YES
Controls (Baseline)				YES		
Controls (Saturated)					YES	YES
Observations	22,237	22,237	22,101	22,101	9,340	9,340
Adj. R-squared	0.019	0.048	0.049	0.054	0.048	0.686

Table 12: **Examiner-Specific Effects on Bank Ratings and Their Influence on Future Risk-Taking and Investments**

This table analyzes how examiner-specific effects on bank ratings correlate with future bank risk-taking behavior and investment decisions. All specifications include controls for examination year and bank fixed effects. We employ three measures of examiner-specific effects. The original measure is the difference between the actual rating and the rating predicted by a model using all available public information about the bank. Additionally, we introduce two decomposed measures: the positive component of the examiner-specific effect when the actual rating is harsher than the model-predicted rating (ESE_t^+), and the negative component when the actual rating is more lenient than the model-predicted rating (ESE_t^-). T-statistics are shown below the parameter estimates in parentheses and are calculated using robust standard errors clustered by bank. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Year $t+1$										
	Log(Assets)			Loan/Assets		Equity/Assets		Tier1 Ratio		NPL/Assets	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
ESE_t	-0.018***										
	(4.179)										
ESE_t^+		-0.059***		-0.021***		0.003***		0.004***		-0.000	
		(6.597)		(3.450)		(4.629)		(3.519)		(0.502)	
ESE_t^-			0.010		0.005		-0.001		0.001		0.001***
			(1.608)		(1.398)		(1.522)		(1.044)		(3.454)
Examination Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Bank FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	20,206	15,603	19,751	15,602	19,750	15,603	19,751	7,503	8,828	7,686	9,067
Adj. R-squared	0.972	0.975	0.975	0.734	0.723	0.769	0.766	0.756	0.752	0.694	0.669

Table 13: **Examiner Access to Private Information and Their Influence on Future Risk-Taking and Investments**

This table analyzes how the part of examiner-specific effects that is more likely related to the access to private information is related with future bank risk-taking behavior and investment decisions. All specifications include controls for examination year, bank, and examiner fixed effects. We employ three measures of examiner-specific effects. The original measure is the difference between the actual rating and the rating predicted by a model using all available public information about the bank. Additionally, we introduce two decomposed measures: the positive component of the examiner-specific effect when the actual rating is harsher than the model-predicted rating (ESE_t^+), and the negative component when the actual rating is more lenient than the model-predicted rating (ESE_t^-). T-statistics are shown below the parameter estimates in parentheses and are calculated using robust standard errors clustered by bank. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Year $t+1$										
	Log(Assets)			Loan/Assets		Equity/Assets		Tier1 Ratio		NPL/Assets	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
ESE_t	-0.017***										
	(4.329)										
ESE_t^+		-0.056***		-0.013**		0.003***		0.005***		0.000	
		(6.640)		(2.061)		(5.230)		(3.759)		(0.020)	
ESE_t^-			0.012**		0.002		-0.000		0.002		0.001**
			(2.003)		(0.415)		(0.908)		(1.548)		(2.326)
Examination Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Bank FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Examiner FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	19,808	15,185	19,346	15,184	19,345	15,185	19,346	7,064	8,397	7,243	8,633
Adj. R-squared	0.974	0.976	0.977	0.745	0.732	0.781	0.777	0.771	0.766	0.714	0.694

Appendix: Description of Variables Used in the Analysis

Table A1: Variable Definition

The table provides variable definitions and specifies the source of data and the formula to calculate each variable using FR Y-9C mnemonics.

Variable	Definition	Source	Formula using FR Mnemonics
Composite rating	RFI/C(D) overall composite assessment	Federal Reserve	
Total Assets	Total assets (or natural log of total assets)	FR Y-9C, FR Y-9SP	BHCK2170, BHSP8519
Equity/Assets	Ratio of equity capital to total assets	FR Y-9C, FR Y-9SP	BHCK3210/BHCK2170, BHSP3210/BHSP8519
Securities/Assets	Ratio of total securities to total assets	FR Y-9C, FR Y-9SP	(BHCK1754 + BHCK1773)/BHCK2170, BHSP0390/BHSP8519
Loans/Assets	Ratio of total loans to total assets	FR Y-9C, FR Y-9SP	BHSP2122/BHCK2170, BHSP0390/BHSP8519
ROA	Ratio of net income to total assets (annual)	FR Y-9C	4×BHCK4340_qtr/BHCK2170, BHSP4340/BHSP8519
Trading Assets/Assets	Ratio of trading assets to total assets	FR Y-9C	BHCK3545/BHCK2170
Allowance for Loans/Assets	Ratio of allowance for loans to total assets	FR Y-9C	BHCK3123/BHCK2170
Leverage Ratio	Tier 1 leverage ratio	FR Y-9C	BHCK7204 (-2014) is replaced with BHCA7204 (2015-now)
Double Leverage Ratio	Ratio of lead bank's equity to parent's equity	FR Y-9C	BHCK3000/BHCK3210
Tier 1 Capital Ratio	Ratio of Tier 1 Capital to risk-weighted assets	FR Y-9C	BHCA8274/BHCAA223
Residential RE Loans	Ratio of residential RE loans to total loans	FR Y-9C	(BHDM5367 + BHDM5368 + BHDM1797)/BHCK2122
Commercial RE Loans	Ratio of commercial RE loans to total loans	FR Y-9C	(BHCKF158+BHCKF159 + BHDM1460 + BHCKF160+BHCKF161)/BHCK2122
C&I Loans	Ratio of C&I loans to total loans	FR Y-9C	(BHCK1763 + BHCK1764)/BHCK2122
Consumer Loans	Ratio of consumer loans to total loans	FR Y-9C	(BHCKB538+BHCKB539 + BHCKK137+BHCKK207)/BHCK2122
Depository Inst. Loans	Ratio of loans to foreign govts and official institutions to total loans	FR Y-9C	(BHCK2081)/BHCK2122
Agriculture Loans	Ratio of agricultural production loans to total loans	FR Y-9C	(BHCK1590)/BHCK2122
Foreign Inst. Loans	Ratio of loans to depository institutions to total loans	FR Y-9C	(BHCK1292+BHCK1296)/BHCK2122

Loans HHI	Herfindahl-Hirschman Index based on loan shares	FR Y-9C	$(BHCKB538+4BHCKB539/BHCK2122)^2$ + $((BHDM5367 + BHDM5368 + BHDM1797)/BHCK2122)^2$ + $((BHCKF158\_BHCKF159 + BHDM1460 + BHCKF160\_BHCKF161)/BHCK2122)^2$ + $((BHCK1763 + BHCK1764)/BHCK2122)^2$
Assets Herfindahl-Hirschman Index	Herfindahl-Hirschman Index based on asset shares for credit card loans, residential real estate loans, commercial real estate loans, commercial and industrial loans, investment securities, and trading assets	FR Y-9C	$(BHCKB538\_4BHCKB539/BHCK2170)^2$ + $((BHDM5367 + BHDM5368 + BHDM1797)/BHCK2170)^2$ + $((BHCKF158\_BHCKF159 + BHDM1460 + BHCKF160\_BHCKF161)/BHCK2170)^2$ + $((BHCK1763 + BHCK1764)/BHCK2170)^2$ + $((BHCK1754 + BHCK1773)/BHCK2170)^2$ + $(BHCK3545/BHCK2170)^2$
Risk-Weighted Assets/Assets	Ratio of risk-weighted assets to total assets	FR Y-9C	BHCAA223/BHCK2170
NPL/Assets	Ratio of nonperforming loans to total assets	FR Y-9C	BHCKA223/BHCK2170. Note: BHCKA223 (-2014) is replaced with BHCAA223 (2015-now).
NPL/Loans	Ratio of nonperforming loans to loans	FR Y-9C	(BHCK1407 + BHCK1403 + BHCK3506 + BHCK3507)/BHCK2122. Note: BHCK5525 (1990-2017) is replaced with BHCK1407 (-1990 & 2018-now), and BHCK5526 (1990-2017) is replaced with BHCK1403 (-1990 & 2018-now)
Loan Loss Reserve/Loans	Ratio of loan loss reserves to total loans	FR Y-9C	BHCK3123/BHCK2122
Loan Loss Reserve/NPL	Ratio of loan loss reserves to nonperforming loans	FR Y-9C	BHCK3123/(BHCK1407 + BHCK1403 + BHCK3506 + BHCK3507). Note: BHCK5525 (1990-2017) is replaced with BHCK1407 (-1990 & 2018-now), and BHCK5526 (1990-2017) is replaced with BHCK1403 (-1990 & 2018-now)
NPL-Residential Real Estate	Ratio of nonperforming residential real estate loans to total residential real estate loans. Exclude BHCs where the category is ¡1% of total loans.	FR Y-9C	(BHCK5399 + BHCKC237 + BHCKC239 + BHCK5400 + BHCKC229 + BHCKC230)/(BHDM5367 + BHDM5368 + BHDM1797)
NPL-Commercial Real Estate	Ratio of nonperforming commercial real estate loans to total commercial real estate loans. Exclude BHCs where the category is ¡1% of total loans.	FR Y-9C	(BHCKF174+BHCKF175 + BHCKF176+BHCKF177 + BHCK3500 + BHCK3501 + BHCKF180 + BHCKF181 + BHCKF182 + BHCKF183)/(BHCKF158 + BHCKF159 + BHDM1460 + BHCKF160 + BHCKF161)

NPL-Commercial and Industrial	Ratio of nonperforming commercial and industrial loans to total commercial and industrial loans. Exclude BHCs where the category is ¡1% of total loans.	FR Y-9C	(BHCK1607 + BHCK1608)/(BHCK1763 + BHCK1764)
NPL-Consumer	Ratio of nonperforming consumer loans to total consumer loans. Exclude BHCs where the category is ¡1% of total loans.	FR Y-9C	(BHCKB576 + BHCKK214 + BHCKK217 + BHCKB577 + BHCKK215 + BHCKK218)/(BHCKB538 + BHCKB539 + BHCKK137 + BHCKK207). Note: BHCK5384 (-2000) is replaced with BHCKB576 (2001-now), and BHCK5385 (-2000) is replaced with BHCKB577 (2001-now)
Monthly Stock Return (1 year) (%)	Average monthly returns over 12 months	CRSP	
Stock Return Volatility	Std Dev of monthly returns over 12 months	CRSP	
Average Stock Turnover	Average turnover (volume/shares outstanding) over 12 months	CRSP	
Market-to-Book Ratio	Ratio of market capital to book equity	CRSP, FR Y-9C	(PRC×SHROUT)/BHCK3210
Monthly Stock Return (1 quarter) (%)	Average monthly returns over 3 months	CRSP	

Online Appendix A: Robustness to Alternative Estimation Methods

Our baseline methodology employs a two-step procedure: first estimating a benchmark model of supervisory ratings using observable financial information and prior supervisory ratings, then analyzing deviations between assigned and predicted ratings as examiner-specific effects (ESE). While this approach transparently separates the estimation of observable-information benchmarks from the analysis of supervisory information production, a potential concern is that results could be artifacts of the two-step procedure. We verify robustness to alternative specifications that estimate rating equations and test information access effects in a single step.

OA.1 One-Step Specifications

As described in Section 5, our one-step robustness approach directly includes information access variables in the rating equation (equation 10). For each hypothesis, we implement four parallel specifications:

1. **One-step rating equation:** Estimate equation (10) from Section 5 directly, including information access variables (distance, off-site exam, public status) alongside controls and fixed effects.

2. **Absolute deviation test:** Extract residuals from the one-step rating equation and regress their absolute value on information access proxies:

   $$|\hat{\varepsilon}_{i,t}| = \gamma_1 \text{Info Access}_{i,t} + \gamma_2 \text{Controls}_{i,t} + \delta_i + \mu_t + \eta_{i,t}$$

   This tests whether rating deviations vary with information access without constructing ESE as an intermediate variable.

3. **Variance test:** Regress squared residuals on information access proxies:

   $$\hat{\varepsilon}^2_{i,t} = \lambda_1 \text{Info Access}_{i,t} + \lambda_2 \text{Controls}_{i,t} + \delta_i + \mu_t + \nu_{i,t}$$

This tests whether conditional variance differs by information access.

4. **Two-step ESE:** Our baseline approach from Tables 6–8 in the main text, reproduced for comparison.

OA.2 Results

Table OA1 Panel A tests whether rating deviations vary with examiner-bank distance and examination modality. Specification choices follow Table 6 in the main text, using year fixed effects without bank fixed effects because distance is largely time-invariant within banks. Results show strong consistency across methods. Distance Above Median coefficients range from -0.013^{**} to -0.022^{***}, with the one-step absolute deviation test (-0.015^{***}) and two-step ESE approach (-0.022^{***}) differing by 47%. Offsite Exam coefficients range from -0.024^{***} to -0.026^{**}, within 8% of each other. All specifications indicate that off-site examinations are associated with 10–12% smaller rating deviations relative to the sample mean.

Table OA1 Panel B examines public trading status. One-step and two-step approaches yield nearly identical results: -0.032^{***} (absolute deviation), -0.032^{***} (variance), and -0.034^{*} (two-step ESE). The maximum difference across methods is 6%. These results confirm that public banks—where market-based information is more readily available—exhibit 14–15% smaller examiner-specific deviations.

Table OA1 Panel C tests interaction effects between bank opacity and examination modality. All nine coefficients show consistency: Public main effect -0.063^{***} (one-step) versus -0.068^{***} (two-step), within 8%; Offsite main effect -0.041^{***} (one-step) versus -0.043^{***} (two-step), within 5%; Public×Offsite interaction $+0.049^{***}$ (one-step) versus $+0.054^{***}$ (two-step), within 10%. The positive interaction confirms that off-site examinations compress information differences more strongly for opaque (private) banks than transparent (public) banks. For private banks, off-site exams reduce rating deviations by 4.1–4.3 percentage points. For public banks, this effect is attenuated to near zero ($-0.041 + 0.049 \approx 0$), consistent with supervision adding value primarily when private information is most valuable.

Across all specifications, one-step and two-step approaches yield quantitatively similar estimates. The coherent pattern of stronger information effects for opaque banks and on-site examinations supports our interpretation that examiner-specific effects reflect supervisory information production.

Table OAA1: **Robustness: Information Access Effects on Rating Deviations**

This table presents alternative specifications testing whether rating deviations vary with examiner-bank distance and examination modality. Results are provided in three panels based on the measure of information access used: Panel A examines distance and off-site effects; Panel B examines public-trading status; Panel C examines interaction effects. Within each panel, columns test different specifications. Column 1 estimates the rating equation directly (equation 10 from Section 5), including information access variables. Column 2 tests whether absolute residuals from Column 1 vary with information access (one-step dispersion test). Column 3 uses squared residuals (variance test). Column 4 reproduces the two-step ESE approach from Tables 6–8. All specifications include year fixed effects without bank fixed effects, as distance and public status are time-invariant within banks. Standard errors are clustered by bank. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Distance and Off-Site Effects

	Rating	|Residual|	Residual^2	|ESE| (2-Step)
	(1)	(2)	(3)	(4)
Distance Above Median	-0.020***	-0.015***	-0.013**	-0.022***
	(0.008)	(0.006)	(0.007)	(0.007)
Offsite Exam	-0.068***	-0.024***	-0.040***	-0.026**
	(0.012)	(0.008)	(0.011)	(0.010)
Examination Year FE	YES	YES	YES	YES
Controls	YES	YES	YES	YES
Observations	22,754	22,754	22,754	22,754
Adj. R-squared	0.741	0.082	0.055	0.040

Panel B: Public Trading Status

	Rating	|Residual|	Residual^2	|ESE| (2-Step)
	(1)	(2)	(3)	(4)
Publicly Traded	-0.018	-0.032***	-0.032***	-0.034***
	(0.011)	(0.009)	(0.010)	(0.011)
Examination Year FE	YES	YES	YES	YES
Controls	YES	YES	YES	YES
Observations	22,754	22,754	22,754	22,754
Adj. R-squared	0.740	0.083	0.054	0.039

Panel C: Interaction Effects

	Rating	|Residual|	Residual^2	|ESE| (2-Step)
	(1)	(2)	(3)	(4)
Publicly Traded	-0.052***	-0.063***	-0.067***	-0.068***
	(0.018)	(0.012)	(0.014)	(0.015)
Offsite Exam	-0.083***	-0.041***	-0.057***	-0.043***
	(0.013)	(0.009)	(0.014)	(0.012)
Publicly Traded × Offsite Exam	0.051**	0.049***	0.057***	0.054***
	(0.020)	(0.014)	(0.017)	(0.017)
Examination Year FE	YES	YES	YES	YES
Controls	YES	YES	YES	YES
Observations	22,754	22,754	22,754	22,754
Adj. R-squared	0.741	0.083	0.055	0.040

Online Appendix B: Robustness to Control Set Specification

Table OAB1: **Bank Opacity and Rating Deviations: Core Specifications**

This table presents robustness tests examining the relationship between the absolute Examiner-Specific Effect on Rating and whether the bank is publicly traded. The specifications use baseline control variables only, providing a streamlined test of the main hypothesis. The dependent variable is the absolute difference between the model-predicted and actual rating, where larger values indicate greater divergence from predictions based on publicly available information. Column (1) includes no controls, Column (2) adds examination year fixed effects, and Column (3) incorporates baseline bank-level controls. T-statistics are shown below the parameter estimates in parentheses and are calculated using robust standard errors clustered by bank. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Absolute Examiner-Specific Deviation		
	(1)	(2)	(3)
Publicly Traded	-0.077^{***}	-0.063^{***}	-0.021^{**}
	(9.265)	(7.689)	(2.193)
Examination Year FE		YES	YES
Controls			YES
Observations	22,754	22,754	22,754
Adj. R-squared	0.004	0.039	0.053

Table OAB2: **Information Access and Examination Modality: Parsimonious Specifications**

This table presents core specifications examining how examiner access to private information affects deviations between actual supervisory ratings and model-predicted ratings. The dependent variable is the absolute examiner-specific effect, defined as the absolute difference between the assigned rating and the rounded model-predicted rating. The specifications test interactions between bank opacity (publicly traded status) and examination modality (off-site indicator) using baseline controls. Columns (1)–(2) include examination year fixed effects, while Columns (3)–(4) add baseline bank-level controls and distance-bin indicators, respectively. All specifications use standard errors clustered at the bank level. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels.

	Absolute Examiner-Specific Deviation			
	(1)	(2)	(3)	(4)
Publicly Traded	-0.125^{***}	-0.104^{***}	-0.047^{***}	-0.043^{***}
	(9.228)	(7.735)	(3.333)	(3.111)
Offsite Exam	-0.048^{***}	-0.029^{***}	-0.050^{***}	-0.047^{***}
	(5.394)	(3.234)	(5.148)	(4.882)
Publicly Traded × Offsite Exam	0.063^{***}	0.061^{***}	0.039^{**}	0.036^{**}
	(3.534)	(3.472)	(2.225)	(2.042)
Examination Year FE		YES	YES	YES
Controls			YES	YES
Distance Bin FE				YES
Observations	22,754	22,754	22,754	22,754
Adj. R-squared	0.005	0.040	0.054	0.056

Online Appendix C: Parameter Stability Analysis

A central methodological concern in our analysis is whether time-series variation in average examiner-specific effects (as shown in Figure 2) reflects genuine shifts in supervisory stringency or merely parameter instability in our forecasting model. Because the benchmark model used to generate ESE is estimated over the full 2005–2023 sample period, it imposes a constant relationship between observable bank characteristics and supervisory ratings across multiple distinct macroeconomic and regulatory regimes.

If the informational content of financial ratios changes systematically during crises—or if supervisors fundamentally reweight certain metrics in response to macroeconomic conditions—these structural breaks could manifest as time-varying deviations from model predictions even if examiner behavior remains constant. This appendix directly addresses this concern by testing whether the core relationships documented in Section 6.3 (information access and rating deviations) remain stable across time periods.

We divide our sample into four distinct periods that capture major shifts in the supervisory and macroeconomic environment:

- **Period 1 (2006–2008):** Pre-Financial Crisis
- **Period 2 (2009–2015):** Crisis and Post-Crisis Recovery
- **Period 3 (2016–2019):** Post-Dodd-Frank Regulation (Pre-COVID)
- **Period 4 (2020–2023):** COVID-19 Pandemic Era

To test parameter stability, we estimate an augmented version of our baseline information access model (Table 6) that includes interaction terms between period indicators and our two key information access proxies: examiner-bank distance above median and offsite examination status. The specification is:

$$
\begin{aligned}
|\text{ESE}_{i,t}| =& \beta_1 \text{Distance Above Median}_{i,t} + \sum_{p=2}^{4} \beta_{1p} \text{Distance}_{i,t} \times \text{Period}_p \\
&+ \beta_2 \text{Offsite}_{i,t} + \sum_{p=2}^{4} \beta_{2p} \text{Offsite}_{i,t} \times \text{Period}_p \\
&+ \text{Controls}_{i,t} + \mu_t + \eta_{i,t}
\end{aligned}
$$

where Period_p are dummy variables for periods 2–4, with Period 1 (Pre-Crisis) serving as the baseline. We include year fixed effects and baseline controls (log assets, equity/assets, securities/assets, loans/assets, ROA), and cluster standard errors by bank.

This specification allows us to compute period-specific effects through linear combinations and test whether coefficients differ significantly across time periods using joint F-tests.

Table OAC1 presents the interaction model estimates. Consistent with our baseline findings, the main effects of both distance and offsite examinations are negative in the Pre-Crisis period, indicating that greater information access is associated with larger deviations from observable-information benchmarks.

The interaction terms reveal differential patterns across the two information access measures:

- **Distance effects** show no significant time variation. All interaction coefficients are small in magnitude and statistically insignificant, and the joint test fails to reject coefficient equality across periods ($F = 0.49$, $p = 0.692$).
- **Offsite examination effects** exhibit significant time variation, driven primarily by crisis periods. The positive and significant coefficient on the Period 2 interaction (Crisis/Recovery) indicates that the negative effect of offsite examinations attenuates substantially during and after the Global Financial Crisis. Similarly, the COVID-era coefficient is negative, reinforcing the baseline offsite effect during this disruption period.

Table OAC2 presents period-specific coefficients computed as linear combinations from the interaction model. Panel A shows that the distance effect remains remarkably stable across all four periods, with coefficients ranging from -0.012 to -0.028 and no economically meaningful pattern across time. This stability provides strong evidence that our core finding—that geographic proximity enables greater acquisition of soft information—is not an artifact of parameter instability.

Panel B reveals a more nuanced pattern for offsite examinations. The effect is strongest during the Pre-Crisis period (-0.052, p = 0.004) and the COVID era (-0.073, p = 0.001), when offsite examinations compressed rating deviations by 5–7 percentage points. In contrast, the effect is substantially weaker and statistically insignificant during the Crisis/Recovery period (-0.011, p = 0.371) and the Post-Regulation period (-0.022, p = 0.227).

Joint tests confirm significant time variation in offsite effects (F = 3.02, p = 0.029), with pairwise comparisons showing that the COVID period differs significantly from the Pre-COVID Post-Regulation period (F = 3.88, p = 0.049).

These results address the parameter stability concern in three ways:

First, the stability of distance effects across all periods demonstrates that our core information-production mechanism is not driven by time-varying relationships between bank fundamentals and ratings. If model misspecification were driving our results, we would expect systematic variation in coefficients across macroeconomic regimes. Instead, the relationship between geographic proximity and supervisory information production remains constant.

Second, the time-varying pattern in offsite examination effects is economically interpretable and consistent with shifts in supervisory practice rather than measurement error. Offsite examinations showed reduced stringency most strongly during crisis periods (GFC and COVID), precisely when supervisors faced operational constraints and when regulatory tolerance for forbearance may have been highest. The attenuation during normal times (Periods 2 and 3) suggests that as supervisory operations normalized, the information disadvantage of offsite work became less pronounced—potentially due to improved technology or examiner adaptation.

Third, the crisis-period amplification of offsite effects provides quasi-experimental validation of our information-production interpretation. During COVID-19, when on-site examinations became physically impossible for extended periods, the sharp increase in the offsite effect (-0.073, the largest across all periods) reflects a genuine shift in how examiners accessed information rather than spurious correlation. This natural experiment strongly supports our conclusion that examination modality matters through its effect on information acquisition.

Importantly, this time-series pattern also reconciles with Figure 2. The temporal variation in average examiner-specific effects reflects both genuine shifts in supervisory stringency over the business cycle and systematic changes in how information access constraints bind. Crisis periods exhibit both more lenient average ratings (Figure 2) and stronger offsite effects (Table OAC2), consistent with supervisors relaxing standards when direct information acquisition is most difficult.

Overall, the parameter stability analysis confirms that our main findings are robust to time-varying relationships between bank fundamentals and ratings. While offsite examination effects vary across periods in an economically interpretable pattern, distance effects remain stable, validating our core mechanism. The time-series variation in supervisory assessments documented in Section 6.5 reflects genuine shifts in examiner stringency rather than model misspecification.

Table OAC1: **Time-Varying Information Access Effects: Interaction Model**

This table tests whether the relationship between information access measures and examiner-specific rating deviations varies across time periods. The dependent variable is the absolute examiner-specific effect. We interact two information access proxies—examiner-bank distance above median and offsite examination status—with dummy variables for three post-baseline periods: Crisis/Recovery (2009–2015), Post-Regulation (2016–2019), and COVID Era (2020–2023). The baseline period is Pre-Crisis (2006–2008). The specification includes year fixed effects and baseline controls (log assets, equity/assets, securities/assets, loans/assets, ROA). Standard errors are clustered by bank. T-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Absolute Examiner-Specific Deviation
	(1)
Distance Above Median	
Distance Above Median	−0.014
	(0.992)
Distance × Crisis/Recovery (2009–2015)	−0.013
	(0.784)
Distance × Post-Regulation (2016–2019)	−0.014
	(0.755)
Distance × COVID Era (2020–2023)	0.002
	(0.097)
Offsite Examination	
Offsite Exam	−0.052***
	(2.882)
Offsite × Crisis/Recovery (2009–2015)	0.041**
	(2.095)
Offsite × Post-Regulation (2016–2019)	0.031
	(1.349)
Offsite × COVID Era (2020–2023)	−0.020
	(0.739)
Examination Year FE	YES
Controls	YES
Observations	22,754
Adj. R-squared	0.040

Table OAC2: **Period-Specific Information Access Effects**

This table reports period-specific effects of information access measures on absolute examiner-specific deviations, computed as linear combinations from the interaction model in Table OAC1. Panel A shows the effect of examiner-bank distance above median; Panel B shows the effect of offsite (vs. onsite) examinations. Coefficients represent the estimated effect within each period. Standard errors and test statistics are computed using the delta method. The joint tests examine whether all period interactions jointly equal zero, indicating parameter stability. T-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Period-Specific Effects			
	Pre-Crisis (2006-2008)	Crisis/Recovery (2009-2015)	Post-Reg (2016-2019)	COVID Era (2020-2023)
Panel A: Distance Above Median				
Coefficient	-0.014	-0.027^{***}	-0.028^{**}	-0.012
	(0.992)	(2.633)	(2.119)	(0.855)
Panel B: Offsite Examination				
Coefficient	-0.052^{***}	-0.011	-0.022	-0.073^{***}
	(2.882)	(0.895)	(1.209)	(3.315)
Parameter Stability Tests				
Distance: Joint test (all periods)	F = 0.49, p = 0.692			
Offsite: Joint test (all periods)	F = 3.02, p = 0.029			
Examination Year FE	YES			
Controls	YES			
Observations	22,754			