DSSR

Discussion Paper No. 145

Generative AI: The Transformative Impact of ChatGPT on Systemic Financial Risk in Chinese Banks

Yikai Zhao, Runyu Dai, and Jun Nagayasu

March 4, 2025

Data Science and Service Research
Discussion Paper

Center for Data Science and Service Research
Graduate School of Economic and Management
Tohoku University
27-1 Kawauchi, Aobaku
Sendai 980-8576, JAPAN

Generative AI: The Transformative Impact of ChatGPT on Systemic Financial Risk in Chinese Banks

Yikai Zhao* Runyu Dai† Jun Nagayasu‡

March 4, 2025

Abstract

We investigate the impact of ChatGPT, a generative AI (GenAI) application, on the systemic financial risk of Chinese banks. Using a sample of 42 publicly traded banks and employing regression discontinuity and regression discontinuity difference-in-differences methodologies, we assess the immediate effects following the launch of ChatGPT on November 30, 2022. Our findings reveal an immediate and significant increase in systemic financial risk, measured by $\Delta CoVaR$. Robustness checks, including placebo tests, alternative risk measures, and varying sample windows, confirm the reliability of these results. Mechanism analysis highlights that transitional challenges during GenAI adoption exacerbate systemic vulnerabilities. Smaller banks, rural commercial banks, and banks with higher nonperforming loan ratios face heightened risks, while large state-owned banks remain relatively insulated. These findings underscore the double-edged nature of disruptive innovations such that GenAI integration poses short-term risks to financial stability even if GenAI has transformative potential.

Key Words: AI; ChatGPT; Systemic financial risk; Chinese Banks
JEL Classification: G21, G32, O33

*Tohoku University, Graduate School of Economics and Management, 27-1 Kawauchi, Aoba-Ku, Sendai-city, Miyagi, 980-8576, JAPAN; Email: zhao.yikai.e8@tohoku.ac.jp.

†Corresponding author. Tohoku University, Graduate School of Economics and Management, 27-1 Kawauchi, Aoba-Ku, Sendai-city, Miyagi, 980-8576, JAPAN; Email: dai.runyu.a3@tohoku.ac.jp or dai.runyu.r1@alumni.tohoku.ac.jp.
This research was supported by the Japan Society for the Promotion of Science (Grant-in-Aid for Research Activity Start-up 24K22603).

‡Tohoku University, Graduate School of Economics and Management, 27-1 Kawauchi, Aoba-Ku, Sendai-city, Miyagi, 980-8576, JAPAN; Tel: +81 22 795 6265; Fax: +81 22 795 6270; Email: jun.nagayasu.d8@tohoku.ac.jp.
The research was supported by the Japan Society for the Promotion of Science (Grant-in-Aid for Scientific Research (B) 24K00268).

1 Introduction

In the long journey of artificial intelligence (AI) development, the emergence of ChatGPT undoubtedly marks a significant milestone. As a breakthrough in natural language processing (NLP) technology, it heralds the era of large language models (LLMs). Training with large amounts of texts using machine learning techniques, such as ChatGPT, an LLM-based generative AI (GenAI), has reached unprecedented sophistication in allowing GenAI tools to understand and generate natural language. Unlike traditional AI, which is typically tailored to narrow and predefined tasks, GenAI exhibits a transformative ability to generalize across diverse applications, from creating dynamic customer interactions to making complex financial decisions (Kong et al., 2024). This generative capacity separates GenAI from the conventional approach, enabling flexible adaptation and autonomous innovation across industries (Liu et al., 2023).

Indeed, since late 2022, ChatGPT's public debut in November, GenAI has become a central topic for CEOs and senior leaders across industries, symbolizing a tipping point in the technology's trajectory. As highlighted in a Boston Consulting Group report (Rabener et al., 2024), GenAI may not improve the financial decisions of firms in the short run. This is because only about 2% of the organizations have a fully developed GenAI talent strategy, exposing a critical gap between rapid adoption of technology and the readiness to use it effectively while ensuring safety and security (Dubey et al., 2024; Rabener et al., 2024; Singh, 2024).

This fast-growing interest also raises concerns about GenAI, particularly its potential to increase financial instability under immature regulatory frameworks and unrefined applications. In trading, GenAI's rapid analytical and execution capabilities present significant challenges. According to the Financial Times, GenAI might enable advanced manipulation techniques or intensify crowded trades during normal market conditions, and thus, potentially amplify market volatility in times of stress, and such dynamics could pose serious threats to the overall stability of the financial system.[1] Beyond trading, the use of GenAI also carries notable operational risks. For example, the emergence of shadow AI, unauthorized or non-controlled applications, entails severe compliance issues (Bickford et al., 2023). Sensitive data leaks through prompt engineering, a task in which GenAI excels, are

[1]Online website, https://www.ft.com/content/d4d212a8-c63a-4b00-9f4c-e06ed59f9279. Accessed January 1, 2025.

becoming increasingly common. Moreover, GenAI could facilitate fraudulent activities, such as aiding criminals in impersonating customers, forging checks, or using voice imitation to execute social engineering attacks that trick employees into bypassing security protocols (Stokel-Walker & Van Noorden, 2023).

Although previous studies have explored the potential benefits and risks of GenAI (Bertomeu et al., 2023; Bloom et al., 2024; Wu et al., 2024), little is known about its immediate impact on systemic financial risk, particularly in the context of emerging markets where regulatory frameworks and technological infrastructure may still be maturing. We attempt to fill this gap by investigating how the introduction of ChatGPT, a prominent GenAI application, has influenced systemic financial risk within the Chinese banking sector. Analysis using regression discontinuity (RD) and regression discontinuity difference-in-differences (RD-DID) designs, and based on a sample of 42 publicly listed banks from October 19, 2022, to January 12, 2023, reveals that the introduction of ChatGPT led to a significant short-term increase in systemic financial risk within the Chinese banking sector, as measured by $\Delta CoVaR$. This finding remains robust through a series of checks, such as the donut hole test, the placebo test, alternative risk measures, and different time windows, providing strong evidence of the immediate market impact of GenAI innovations.[2]

Moreover, our mechanism analysis further reveals that the indirect impact stems from the transitional risks of adopting GenAI technologies. Banks with more prior experience in AI-related technologies exhibited significantly smaller responses to ChatGPT's launch than their counterparts with fewer AI capabilities. This trend can be attributed to banks that lack AI adoption and face greater

[2]Some may question that ChatGPT is not directly accessible in mainland China. However, we argue that using ChatGPT's launch as an exogenous shock is reasonable for three key reasons. First, the launch of ChatGPT represents not only a technological breakthrough but also a landmark event in the era of GenAI. Financial markets are highly sensitive; even if the technology itself is not directly accessible, market participants can price it or adjust their expectations about its potential impact (Duan et al., 2019; Murinde et al., 2022). This is particularly relevant because China, the second largest economy in the world, is deeply interconnected with global markets (Dong et al., 2016). Second, ChatGPT's release spurred domestic companies such as Baidu and Alibaba to quickly announce similar AI products, introducing competitive dynamics that could influence China's financial system through changes in market expectations (Wang & Liang, 2024). The resulting changes in the technological and business ecosystem have substantial implications (Qian, 2024). A case in point is China Banking Corporation has launched its GenAI solution, CHIB GPT to catch up with rapid AI adoption in the financial sector (Online website https://www.chinabank.ph/latest-news-chinabank-leverages-generative-ai-launches-chib-gpt-to-boost-employee-productivity. Accessed January 1, 2025.) Lastly, the timing of the ChatGPT launch was independently determined by OpenAI and is unrelated to China's financial markets. This independence aligns with the exogeneity assumption of our study, and rigorous methods are applied to control for other confounding factors, ensuring the robustness of our conclusions.

transitional pressures to improve their technological infrastructure in response to the GenAI stimulus. These efforts introduce additional systemic risks during transition, compounding financial system vulnerabilities. This finding on indirect impacts also contributes to the literature on how incumbents succeed in the face of disruptive technologies, such as Ansari et al. (2016) and Roy et al. (2018), underscoring the importance of preparedness and adaptability in mitigating transitional risks. It offers new insights into the dynamic interaction between technological disruption and financial stability.

Overall, this study makes several important contributions to the literature. First, it extends the research framework on the determinants of banks' systemic risk by incorporating external technological uncertainty and shocks into the analytical system. Thus, we address the limitations of existing studies that focus predominantly on macroeconomic policy shocks and economic uncertainty (Calmes & Théoret, 2014; Duan et al., 2022). Second, methodologically, this study innovatively employs a regression discontinuity design to effectively identify the causal effects of external technological shocks on systemic risk. At the empirical level, utilizing large-scale bank-level data, this causal inference study provides the first systematic quantification of both the magnitude and transmission mechanisms of the impact of external technological shocks on the systemic risk of Chinese banks. These findings offer novel empirical evidence to inform regulatory authorities in their efforts to design more effective risk prevention policies.

The rest of the paper is organized as follows. Section 2 surveys the literature on GenAIs influence on systemic risk and develops the core research hypothesis. Section 3 explains the research framework. Section 4 assesses ChatGPTs impact on bank systemic risk, while Section 5 examines the heterogeneities among bank types. Robustness checks are conducted in Section 6, and Section 7 investigates the underlying mechanism. Finally, Section 8 summarizes the findings.

2 Literature Review and Hypothesis Development

AI has become increasingly integrated into our societies and economies. Its ability to rapidly process vast amounts of information enhances efficiency and decision-making at both individual and organizational levels. AI also boosts labor productivity, creates business opportunities, and increases profits while helping mitigate political, security, and financial risks. Building on this trend, ChatGPT and other GenAI technologies are increasingly recognized as important drivers of social and economic development.

Despite GenAIs potential benefits, such technology may also trigger risks in financial system. Beckmann and Hark (2024) found that the U.S. banking sector experienced negative market reactions, as reflected in cumulative abnormal returns, following ChatGPTs launch in short-term. Empirical evidence indicates that disruptive technologies often have significant impacts on traditional industries, potentially leading to market turbulence and systemic risks (Arenas et al., 2023; Danneels, 2004). Similarly, the potential widespread adoption of GenAI tools in the financial system could also heighten operational risks, including market concentration (Aghion et al., 2018), increased interconnectedness among firms (Fernández, 2019), and risks tied to institutions deemed "too big to fail" (Leitner et al., 2024). Thus, the sudden and unforeseen emergence of ChatGPT at the end of 2022 has generated significant attention in financial markets. In the rest of this section, we focus on exploring the short-term impact of GenAI technologies on systemic risk based on existing research and gradually establish our hypotheses.

2.1 The formation of systmemic financial risks

The formation of systemic financial risks is an essential topic in financial economics, primarily stemming from the high interconnectedness of financial institutions and the propagation of external shocks within the system. Key drivers of systemic risk include asset price fluctuations, shifts in market sentiment, and liquidity contractions (Bai et al., 2025; Salisu et al., 2022). Allen and Gale (2000) summarized that the interbank credit market is a critical channel for risk transmission, where a liquidity crisis at a single institution can rapidly spread through asset correlations and financial networks, affecting the entire financial system. Adrian and Brunnermeier (2016) and Brunnermeier et al. (2020) further highlighted that external shocks can propagate quickly within the financial system through asset correlations, liquidity shortages, and information asymmetry, with the extent of their impact varying depending on economic cycles and the specific characteristics of banks. In this context, the emergence of GenAI can be seen as a typical external shock, potentially affecting the financial system through similar transmission mechanisms.

2.2 GenAI and information authenticity

Although GenAI can enhance the risk management capabilities of financial institutions and improve operational efficiency, its widespread adoption can also introduce new challenges to the financial industry, such as data privacy concerns and potential cybersecurity threats (Alawida et al., 2024; Gupta et al., 2023; Wach et al., 2023). For example, GenAI models could be maliciously exploited to compromise the information security of banking systems by embedding backdoors or injecting malicious code. Additionally, such technologies could be used to spread false or synthetic information, triggering market panic and disrupting investor behavior (Emett et al., 2024; Keeley, 2023). Given that defensive technologies and response mechanisms may still be insufficient in the short term following the introduction of ChatGPT, such attacks could negatively impact the stability of financial markets in the near future.

2.3 Regulation lag and risk management gaps

Regulatory gaps may exacerbate the vulnerabilities associated with the initial adoption of GenAI technologies. Zetzsche et al. (2017) underscored that delays in regulation can increase uncertainty within the financial system. In the case of ChatGPT and other GenAI technologies, the regulatory and supervisory framework remains underdeveloped, leaving banks without clear compliance guidelines, thereby amplifying systemic risks (Remolina, 2024). Furthermore, GenAI relies on vast datasets and complex machine learning models, increasing information asymmetry and system opacity, which could hinder effective risk management.

2.4 Workforce structure and financial institutions

The sudden emergence of GenAI may pose significant challenges to workforce structures, including those in the banking sector (Eisfeldt & Schubert, 2024). Unlike industrial robots that primarily replace routine tasks, AI often substitutes for tasks performed by high-skilled workers (Bloom et al., 2024). David (2024) noted that AI has the potential to empower lower-skilled workers, enabling them to perform tasks traditionally conducted by high-skilled experts. Eisfeldt et al. (2023) emphasized that when AI replaces core tasks rather than supplementary ones, shareholders typically benefit from increased firm value, while high-skilled workers, especially those in the banking industry, may face stagnating wages or even job displacement. Several factors may drive GenAI to bring about a similar trend. First, GenAI is expected to significantly enhance service efficiency, such as by automating credit assessments and risk analysis, thereby reducing reliance on human labor. Second, the high substitutability of administrative and clerical tasks makes banks more inclined to adopt GenAI to reduce labor costs. Third, the widespread application of machine learning and algorithmic tradings powered by GenAI may reduce the demand for analysts and traders, accelerating the automation of banking operations. Consequently, the benefits of adopting GenAI are likely to flow mainly to capital owners, while high-skilled employees could see limited wage growth or even job losses. This shift in income distribution favoring capital owners may exacerbate income inequality, weaken middle-class purchasing power, and lead to declining market demand. Tian and Nagayasu (2024) pointed out that such economic instability could further erode consumer confidence and worsen market expectations, thereby amplifying financial system vulnerabilities and increasing systemic risk.

Furthermore, the adoption of GenAI forces financial institutions to undergo significant structural reforms, potentially exacerbating systemic risks. Deploying GenAI often requires substantial investments in technological infrastructure and workforce reskilling (Babina et al., 2024). During this transition phase, banks with varying levels of technological readiness face differing degrees of operational and financial pressures. Institutions lacking sufficient AI capabilities may experience increased liquidity constraints and operational disruptions as they attempt to adapt to these advancements (Vives, 2019). Such asymmetries in preparedness could worsen systemic financial risks, particularly during periods of market stress.

Therefore, we conjecture that even though GenAI has the potential to improve efficiency and productivity, its sudden adoption also introduces risks that could destabilize China's financial systems if not managed carefully, especially in the early stage. We thus propose the following research hypothesis.

Hypothesis: *The launch of ChatGPT, as a representative breakthrough in GenAI, will significantly increase the systemic risk of Chinese banks in the short term.*

3 Research Design

This section introduces the proxy for financial systemic risk and a statistical method to evaluate the short-term impact of ChatGPT. We describe the data used herein and present the preliminary evidence of a significant change (i.e., discontinuity) in the data around the cut-off point determined by the timing of ChatGPT introduction.

3.1 Financial systemic risk

While researchers have advocated several pragmatic definitions and measures of financial systemic risk, we primarily follow one of the most recognized measures, $\Delta CoVaR$, developed by Adrian and Brunnermeier (2016). $\Delta CoVaR$ quantifies the change in the value at risk (VaR) of the entire financial system when a particular institution is in distress compared with when it is in its normal state. It relies primarily on the conditional VaR ($CoVaR$), which refers to the VaR of the entire financial system conditional on a specific institution in a particular state (q). To illustrate, we start by defining the VaR of bank i at time t, along with the $CoVaR$ of the financial system (sys) conditional on bank i:

$$\mathbf{P}\left(X_t^i \leqslant VaR_{q,t}^i\right) = q, \quad \mathbf{P}\left(R_t^s \leqslant CoVaR_{q,t}^{sys|X^i} | X_t^i = VaR_{q,t}^i\right) = q, \tag{1}$$

where R_t^s and X_t^i represent the return of the system and bank i, respectively. In simple terms, $CoVaR$ reflects the tail-dependency between the entire financial system and a specific bank. Considering the asymmetric effect on the returns' volatility shift, we assume that $\left(R_t^s, X_t^i\right)$ follow a bivariate dynamic conditional correlation–generalized autoregressive conditional heteroskedasticity (DCC–GARCH) process. Appendix A.1 provides the formal definitions of the DCC–GARCH model. The univariate volatility for the system and bank i are denoted by $\sigma^2_{s,t}$ and $\sigma^2_{i,t}$, respectively. The dynamic correlation between the system and bank i is denoted by $\rho_{si,t}$. We use $q = 5\%$ to denote a distressed state and $q = 50\%$ to denote a normal state. Then, $\Delta CoVaR^{\mathrm{sys}|X^i}_{q,t}$ measures the change in the risk of the financial system due to the distress of bank i at time t, as given by the formula

$$\Delta CoVaR^{sys|X^i}_{q=5\%,t} = CoVaR^{sys|X^i}_{q=5\%,t} - CoVaR^{sys|X^i}_{q=50\%,t}. \quad (2)$$

Here, the system's VaR conditional on bank i ($CoVaR^{\mathrm{sys}|X^i}_{q,t}$) is measured by

$$CoVaR^{\mathrm{sys}|X^i}_{q,t} = \rho^{si}_t \frac{\sigma^s_t}{\sigma_{i,t}} VaR^i_{q,t} + \sigma^s_t \left(1 - \rho^{si^2}_t\right)^{1/2} G^{-1}(q), \quad (3)$$

where the parameters, $\rho^{si}_t, \sigma_{s,t}$ and $\sigma_{i,t}$, from DCC-GARCH model, and $G(\cdot)$ is the conditional distribution of the system's return. See more detailed derivations of (3) in Adrian and Brunnermeier (2016).

It is worth noting that in Section 6, we perform a robustness check of our study using alternative empirical methods to estimate *CoVaR*, specifically Quantile-based *CoVaR* (*Q-CoVaR*) and Copula-based *CoVaR* (*C-CoVaR*). Furthermore, we include another widely used risk measure, Marginal Expected Shortfall (MES) (Acharya et al., 2017).[3]

3.2 Data and methodology

Our sample comprises 42 banks listed on the Shanghai Stock Exchange and the Shenzhen Stock Exchange. These banks constitute the CSI Bank Index (code: 399986). The complete alignment between our sample banks and the index components ensures comprehensive coverage of China's listed banking sector. To examine the impact of ChatGPT's launch (November 30, 2022) on systemic risk, we use $\Delta CoVaR$ and the sample spans from October 19, 2022, to January 12, 2023, that is, a $\pm$30-trading-day window around the event date. To construct the daily $\Delta CoVaR$ for each bank,

[3] We do not consider SRISK (Brownlees & Engle, 2017), a systemic risk measure that quantifies the expected capital shortfall of an institution, conditional on prolonged and severe market declines. This exclusion is due to two key reasons. First, our study relies on daily data for a relatively short period, whereas SRISK requires quarterly corporate financial data. Second, the sample period lacks significant market downturns, making SRISK less suitable to capture systemic capital shortfalls in this context.

we employ a rolling window approach with $\pm$3-month around the target date and using return data from three months before and after each target date.[4] The daily return data are calculated using the closing price growth rate and comes from the CSMAR database. We specifically focus on a narrow time frame ($\pm$30-trading-day) to capture the immediate market response to ChatGPT's introduction while minimizing confounding effects from other events. This short-run analysis is particularly crucial for two reasons: First, it enables timely risk assessment and potential intervention; second, it provides cleaner identification compared to medium-run windows, where multiple factors could influence systemic risk measures.

Our benchmark analysis employs the RD model as it effectively capitalizes on the exogenous shock of ChatGPT's launch to precisely assess its immediate impact on the systemic financial risk of Chinese banks. By focusing on the discontinuity around the event, the model effectively isolates the effects of ChatGPT, ensuring robust causal inference. Specifically, banks' systemic risk before launch serves as the control group, while the risk after launch is used to create the treated group. Assuming other conditions remain constant, a sudden and significant change in systemic risk after the launch of ChatGPT, with continuity around other time points, suggests that this change is attributable to the sudden emergence of GenAI.

The RD model provides a more precise measure of local average treatment effects and can effectively minimize the influence of extraneous factors on the results. As noted in Hausman and Rapson (2018) and Meng and Yu (2023), when using time as a cut-off point, we must consider that temporal variations in treatment effects may lead to biased RD estimates if not properly accounted for.[5] Therefore, we employ a two-step regression approach following Greenstone et al. (2022) and Meng and Yu (2023). First, we residualize the dependent variable on a series of control variables and fixed effects, and then use the residual of financial systemic risk as the dependent variable and regress it on the RD indicator. In detail, the first step is

[4] For instance, the $\Delta CoVaR$ calculation for October 19, 2022, incorporates return data from July 19, 2022, to January 19, 2023.

[5] Hausman and Rapson (2018) referred to the RD design with time as the running variable as RDiT (RD in time). As documented in their review, RDiT has been widely used in empirical research. For example, Anderson (2014) applied RDiT to estimate the effects of transit strikes using the days from the onset of the strike as the running variable. Similarly, Bento et al. (2014) used RDiT to evaluate the impact of the Clean Air Vehicle Sticker policy on travel time with date of policy as the running variable and suggested that RD with time could serve as an alternative research design when the construction of appropriate control groups proves challenging.

$$\Delta CoVaR_{i,t} = \beta_0 + \beta_1 ConVars_{i,t} + \rho_i + \omega_d + \theta_m + \gamma_{im} + \varepsilon_{i,t}, \tag{4}$$

where i indexes banks, t denotes trading dates, and m represents months. $ConVars_{i,t}$ is a vector of daily control variables that captures broader financial conditions and market environment: the three-month Treasury yield as a proxy for market spread, the difference between the one-year AAA corporate bond rate and China's one-year Treasury yield to capture market credit risk, the term spread (10-year Chinese Treasury yield minus three-month Chinese Treasury yield) to control for economic cycles, the percentage change in bank market value, and a dummy variable to capture the impacts from the relaxation of COVID-19 epidemic on December 7, 2022. These variables are summarized in Table 1. This table also reports p values of the Levin–Lin–Chu (LLC) panel unit root test with a time trend, suggesting that the data are stationary. We further add fixed effects. ρ_i captures bank fixed effects, θ_m controls for month fixed effects, ω_d is for weekday fixed effects, and γ_{im} represents bank-by-month fixed effects to account for any time-varying bank characteristics such as financial conditions and unobserved heterogeneity in banks' risk exposure over time. These fixed effects help us control for bank-specific seasonality patterns in risk measures, varying regulatory requirements across banks and periods, potential differences in banks' business cycle sensitivity, and changing financial market conditions that might heterogeneously affect banks.

After controlling for the various factors in Equation (4), we perform the RD analysis on the residual systemic risk measure. The second stage regression can be expressed as

$$\begin{aligned} \Delta CoVaR_{i,t} - \Delta\widehat{CoVaR}_{i,t} = \alpha_0 &+ \alpha_1 D\left(t \geq ChatGPT_{i,t}\right) + \alpha_2 f\left(t - ChatGPT_{i,t}\right) \\ &+ \alpha_3 D\left(t \geq ChatGPT_{i,t}\right) \times f\left(t - ChatGPT_{i,t}\right) + \eta_{i,t}. \end{aligned} \tag{5}$$

Here, $\Delta CoVaR_{i,t} - \Delta\widehat{CoVaR}_{i,t}$ represents the residual systemic risk after removing the effects of control variables and the fixed effects from Equation (4). $D\left(t \geq ChatGPT_{i,t}\right)$ is a dummy variable that equals 0 before ChatGPT's launch, and 1 after it. The coefficient α_1 is of interest as it indicates whether an immediate discontinuity exists due to the launch of ChatGPT. The term $t - ChatGPT_{i,t}$ represents the distance in time from the event date, and $f\left(t - ChatGPT_{i,t}\right)$ is a polynomial function that controls differences before and after the launch date.

[Table 1 about here.]

3.3 Statistics and RD plots

Table 2 presents the summary statistics for the key variables. The sample consists of 2,562 bank-day observations and is strongly balanced (42 banks $\times$ 61 days $= 2,562$). Of particular interest are our risk measures. The $\Delta CoVaR$ has a mean of 0.013, with a substantial variation measured by the standard deviation of 0.006. Its distribution exhibits positive skewness and excess kurtosis, showing a heavy right-tail characteristic indicating more frequent extreme positive values than that expected in a normal distribution. After controlling for various factors, the residual $\Delta CoVaR$ is centered at zero by construction and shows even more pronounced heavy-tailed characteristics, with a high kurtosis around 5 and positive skewness around 0.4. This leptokurtic distribution suggests that extreme systemic risk events still exist even after removing the influence of market conditions and bank characteristics, with the residuals showing a strong tendency for positive outliers.

[Table 2 about here.]

Before we employ the nonparametric estimation for the RD model with the residual of $\Delta CoVaR$ as the dependent variable, we first determine any discontinuity in residual $\Delta CoVaR$ over time. Figure 1 presents the RD plots with different polynomial functions, where Figure 1a uses a first-order polynomial, and Figure 1b uses a second-order polynomial. In both subplots, we observe a significant increase in the residual $\Delta CoVaR$ value after the launch of ChatGPT in the short run. This suggests that the introduction of ChatGPT has an immediate and notable impact on the systemic risk captured by $\Delta CoVaR$. To ensure the validity of the RD design, we also verify that covariates exhibit no significant jumps at the cut-off point. This verification ensures that the observed discontinuity in residual $\Delta CoVaR$ is not driven by systematic differences in covariates around the cut-off, which supports causal inference. The RD plots of other covariates are shown in Figure A.1, and we find that the discontinuities at the cut-off are not statistically significant at the 5% level.

[Figure 1 about here.]

4 Benchmark Results

We now perform global polynomial RD regressions, which measure the overall trend of the dependent variable across the entire data range by fitting high-order polynomials to capture nonlinear features at the cut-off point. Unlike local polynomial RD, global polynomial regression uses observations from the entire sample period rather than just around the discontinuity in order to estimate the treatment effect. Specifically, we set the bandwidth to 30 days on each side, including all observations within this window. Table 3 reports the estimation results for different polynomial orders and kernel functions, with residual $\Delta CoVaR$ as a dependent variable. The z-statistics and bias-corrected p values show that the effects of RD are statistically significant at the 5% level. Considering column 1, the RD estimate of 0.0025 indicates that, following ChatGPT's launch, market systemic risk increased significantly by around 0.4 standard deviation of $\Delta CoVaR$ (0.0025/0.006 $\approx$ 0.42). This is a noteworthy phenomenon, as it reflects a substantial increase in interconnectedness and potential contagion risk among banks after ChatGPT's launch. It also highlights the market volatility triggered by ChatGPTs introduction and the profound impact of technological innovation on the financial system.

[Table 3 about here.]

However, global polynomial regression may be sensitive to the data distribution far from the cut-off point. To further validate this finding and improve the robustness of the analysis, we employ the local polynomial regression method, which focuses on data near the cut-off point to more accurately capture the short-term effects of ChatGPT's launch on systemic financial risk. Table 4 presents the results of these local polynomial RD regressions. We use the bandwidth selection method of Calonico et al. (2014), employing both the mean square error optimal bandwidth selector (MSE) and the coverage error rate (CER). Besides, Gelman and Imbens (2019) highlights that high-order polynomials should be avoided in the RD setting due to their instability. Following Fu and Gu (2017), we adopt a low-order polynomial, specifically a linear local and quadratic specification. The results for higher-order polynomials are reported in Table A.1.

As shown in Table 4, generally speaking, the local polynomial RD effect is close to the global RD effect in Table 3. The RD estimates are slightly larger with quadratic polynomial regressions in Table 4a and smaller with linear regressions in Table 4b. All estimates are significant at the 5% level, except in columns 4 and 5, where they are significant at the 10% level. We observe no notable change in significance before and after bias adjustment, indicating robustness. The optimal bandwidth varies with different settings; however, the choice of bandwidth selection method does not heavily affect estimated RD effects, as evidenced by the similarity between the results in columns 1 and 4, 2 and 5, as well as columns 3 and 6. We also use three kernel functions to construct the local polynomial estimators, namely, triangular, epanechnikov, and uniform, and the RD effects are close to each other. Thus, the RD results are relatively insensitive to bandwidth selection and exhibit strong consistency and robustness across different kernel functions. That is, local polynomial models effectively capture the distributional characteristics near the cut-off point, further supporting the conclusion that the launch of ChatGPT had a significant causal impact on systemic financial risk.

[Table 4 about here.]

Using time as the running variable may pose challenges. For example, while the dependent variable in our analysis is the residual $\Delta CoVaR$, which accounts for and removes seasonal variations, unobserved factors tied to time that could influence systemic risk, such as year-end effects. To address these concerns, following Fu and Gu (2017), Xue et al. (2023), and Persson and Rossin-Slater (2024), we implement the RD-DID design, comparing changes over the same calendar period across different years, thus controlling for potential time-related biases and ensuring a more robust causal inference. We center the analysis around November 30, 2021, using the 30 trading days before and after this date as the control group ($treat_{i,y=2021} = 0$), and the 30 trading days before and after November 30, 2022, as the treatment group ($treat_{i,y=2022} = 1$).[6] The event date is the month and day of ChatGPTs launch. We denote the dummy variable $D(t \geq c)$ based on c, where c represents November 30. Referring to Persson and Rossin-Slater (2024), the following model is constructed:

[6] The time range for the year prior to the event spans from October 19, 2021, to January 12, 2022. For simplicity, we uniformly represent it as $y = 2021$.

$$\Delta CoVaR_{ity} - \Delta \widehat{CoVaR}_{ity} = \beta_0 + \alpha_0 treat_{iy} + \alpha_1 D(t \geq c) + \alpha_2 treat_{iy} \times D(t \geq c) + \alpha_3 D(t \geq c) \times f(t-c) + \alpha_4 f(t-c) + \varepsilon_{ity}, \tag{6}$$

where i denotes banks, t represents dates, and y refers to a year (2021 or 2022). We calculate the financial systemic risk of banks in 2021 using the same methodology as that elaborated in Section 3. Residual $\Delta CoVaR$ is then derived from Equation (5). α_2 is the coefficient of interest, capturing the effect of ChatGPTs launch on systemic risk. $f(t-c)$ is a flexible function of the running variable, the day centered around November 30.

[Table 5 about here.]

Table 5 documents the results of the RD-DID method. The comparison focuses on differences in systemic risk before and after the event within 30 trading days in 2022, relative to the corresponding period in 2021. This table shows that the magnitude and significance of the coefficient $treat_{iy} \times D(t \geq c)$ remain consistent across different polynomial orders. In all regressions, the coefficient is significantly positive at the 1% level, indicating that the launch of ChatGPT has a robust and statistically significant positive effect on systemic financial risk. Thus, the intervention introduced by ChatGPTs deployment increased the vulnerability of the financial system, captured by the residual $\Delta CoVaR$. The consistency of the results across polynomial orders further supports the stability of this finding and reduces concerns about model specification biases.[7]

5 Heterogeneous Results

Thus far, we have presented evidence of an increase in systemic risk immediately following the introduction of ChatGPT in the banking sector. This section replicates the analysis while accounting for heterogeneity among banks. Generally, as summarized below, we observe a similar effect of

[7] To further verify whether there is a discontinuity at the November 30, 2021 cut-off, we plot the RD effect and corrected p values for 2021 in the Figure A.2. The p values exceed 0.05, confirming no significant effect of RD at the cut-off point of 2021. Parallel trends and dynamic policy effect graphs are included in the Figure A.3 for robustness checks.

ChatGPT across different groups of banks, although some interesting but often insignificant discrepancies exist in their responses.

5.1 Bank types

Different types of banks, such as state-owned, city commercial, and rural commercial banks, may exhibit varying levels of systemic risk and responses to external shocks, such as the launch of ChatGPT. These differences stem from operational focus, customer base, market exposure, and regulatory environment. For example, state-owned banks typically have more stable funding structures and stronger government backing, whereas rural commercial banks may be more vulnerable to market fluctuations and external disruptions.[8]

Table 6a presents the results of the heterogeneity analysis by bank types. We follow the approach of Cleary (1999) and Lu et al. (2019) to compute the empirical p values for the differences between the groups. Specifically, we use a bootstrap resampling method with 100 iterations to generate a series of empirical samples, which are then grouped to obtain the empirical distribution of the differences in regression coefficients between groups. The results are presented in the final row of the table. State-owned banks in column 1 have fewer RD effects, with rural commercial banks having the largest impact. The difference in the coefficients of effect of RD between columns 1 and 2 is -0.0008, with the empirical p value (in parentheses) being 0.360. Thus, while the systemic risk of city commercial banks is higher for GenAI, the result is not statistically significant. Comparing columns 1 and 3, the difference in the coefficients of the RD effect between state-owned banks and rural commercial banks is -0.0019, with a corresponding empirical $p$value of 0.030. This result is consistent with the expectation that state-owned banks benefit from more stable funding structures and stronger government backing, which buffers them from short-term shocks. In contrast, rural commercial banks exhibit significantly larger RD effects because they serve localized markets with limited diversification and, as a result, are particularly exposed to tail risks during periods of

[8] The information about banks in this section, including bank types, state-owned shareholding ratio, nonperforming loan ratio, and bank assets, all come from the CSMAR database. We further divide the sample into two groups based on the median level of state ownership share: low and high. We obtain a similar result (Table A.4) that banks with lower levels of government control are more exposed to systemic risk, with more significant discrepancies between the two groups.

technological transition.

5.2 Nonperforming loan ratios

Considering the potential influence of credit risk on systemic vulnerability, we further divide the sample into two groups based on the median ratio of nonperforming loans (NPL) to total loans: the low NPL group and the high NPL group. This classification allows us to examine whether banks with higher credit risk are more susceptible to systemic shocks. The findings reported in Table 6b reveal that the RD coefficients for the high NPL group (column 1) are higher than those for the low NPL group (column 2). Although the difference in the RD coefficients between the high NPL group (column 1) and low NPL group (column 2) is statistically insignificant based on the empirical p value, the observed pattern remains consistent with expectations. The higher RD coefficient for the high NPL group suggests that banks with higher NPL ratios experienced an increase in systemic risk following the launch of ChatGPT. This finding implies that preexisting credit risk may exacerbate the impact of disruptive technological shocks.

5.3 Bank size

To explore how bank size influences the impact of ChatGPT's initiation on systemic financial risk, we categorize the sample into two groups based on total assets: large banks and small banks. Bank size can play a dual role in shaping the responses to systemic risk. On the one hand, larger banks may face greater challenges in adopting new technologies because of their complexity and rigid organizational structure. However, they typically have more resources, such as skilled talent and advanced infrastructure, to support this transition. On the other hand, smaller banks may be more flexible and adaptive during transformation but often lack the resources and expertise necessary to fully capitalize on technological advancements. Table 6c presents the results of this analysis. The difference between the two groups is reported in the last row, along with the empirical p value. The findings show that the RD coefficients of large banks (column 2) are smaller than those of small banks (column 1). The difference was 0.0010 with p value of less than 0.1. That is, while large banks benefit from their resource advantages, small banks face heightened systemic risk, possibly because of their limited capacity to manage technological transformation challenges.

[Table 6 about here.]

6 Robustness Checks

Given the findings in the previous section that immediate responses to ChatGPT are often statistically homogeneous among banks, we perform several robustness checks from different perspectives. These robustness checks cover the validity of our a priori assumptions in the RD estimation and the unobservable definition of systemic financial risk.

6.1 Donut hole test

The first robustness check is the donut hole test, which considers the potential biases introduced by observations around the cutoff point. These biases could arise from data irregularities such as noise, measurement errors, or the strategic behavior of market participants. Although we assume that the launch of ChatGPT is an exogenous shock, some investors may have anticipated its launch and reacted accordingly. This raises the potential for manipulation near the cutoff value. We adopt the donut hole test, as in Barreca et al. (2011) to address this issue.

Figure 2 plots the results of the robustness check using the donut hole test for different trading-day exclusions around the cut-off. Specifically, we first determine the optimal bandwidth using the MSE criterion and perform the analysis sequentially, excluding observations within 1%, 5%, 10%, 15%, and 25% of the trading days around the cutoff using the triangular kernel. The leftmost vertical line in the figure represents the baseline regression (column 1 in Table 4a), and the subsequent points from left to right show the point estimates and 95% confidence intervals after excluding observations within the varying percentages of trading days. In Figure 2, the estimated RD effect remains positive and the confidence intervals across all specifications consistently exclude zero. The estimates are comparable to those from the baseline regression without any exclusion. These results suggest that the estimated RD effects are robust when observations close to the cut-off value are excluded. The positive jump in systemic risk resulting from ChatGPT's launch is not driven by potential manipulation or anomalies immediately around the cutoff, further reinforcing the validity of our findings.

[Figure 2 about here.]

6.2 Placebo test

To further validate the robustness of our results, we perform an additional analysis using false policy discontinuities. By introducing these artificial cut-offs, we examine whether the observed RD effects are specific to the actual cutoff value. Specifically, we assume that ChatGPT was launched three, five, and seven trading days after the actual launch date, and three, five, and seven trading days before the actual launch date.

Figure 3 reports the results of the placebo tests, assuming that the policy intervention occurred on different dates. The leftmost point represents the result of the baseline regression, followed by the estimates and corresponding 95% confidence intervals for fake policy cut-offs occurring seven trading days before (-7), five days before (-5), three days before (-3), three days after ($+3$), five days after ($+5$), and seven days after ($+7$) the actual launch date. Except for the baseline regression result, the confidence intervals for all other placebo estimates include zero, indicating that the observed effects are not statistically significant at the 5% level when these false cut-offs are applied. This supports our initial findings, confirming that the identified significant effects are specific to the actual intervention date, corresponding to the launch of ChatGPT.

[Figure 3 about here.]

6.3 Bandwidth sensitivity

We conduct a sensitivity test for the bandwidth selection, considering that the choice of bandwidth may affect the robustness of the RD estimates. The bandwidth determines the range of observations on either side of the cutoff, directly influencing the estimated effect and significance level. If the bandwidth is chosen incorrectly, the results may depend heavily on a specific subset of data, thereby undermining the credibility of the analysis. A sensitivity analysis of the bandwidth selection ensures the robustness of our findings.

Specifically, we scale the MSE-optimal bandwidth by factors of 0.5, 0.75, 1.25, 1.5, and 1.75 to assess the sensitivity of the RD estimates to variations in bandwidth selection. The results obtained using these scaled bandwidths and the optimal bandwidth are plotted sequentially from left to right in Figure 4. Each point represents the RD estimate accompanied by a corresponding 95% confidence interval. RD point estimates remain consistent with the baseline regression for all bandwidth specifications, indicating that the observed treatment effect is not sensitive to changes in the sample inclusion range. Importantly, the 95% confidence intervals for all estimates consistently excluded zero, thus reinforcing the statistical significance of the results. These findings demonstrate that the magnitude and significance of the RD estimates remain stable even when the bandwidth is either substantially narrowed or widened.

[Figure 4 about here.]

6.4 Alternative measures

To further ensure the robustness of our findings, we employ three alternative measures of financial systemic risk: *Q-CoVaR*, *C-CoVaR*, and MES following Adrian and Brunnermeier (2016), Mensi et al. (2017), Acharya et al. (2017), Brownlees and Engle (2017), and Cincinelli et al. (2022). The *Q-CoVaR* focuses on how the risk profile of the entire financial system changes when a specific institution moves from its normal state to a stressed state by leveraging quantile regression. The *C-CoVaR* enhances systemic risk measurements by incorporating copula functions to model the dependency structure between individual institutions and financial systems. Unlike traditional methods that assume linear relationships, the copula approach captures complex nonlinear dependencies and joint tail behaviors. The MES measures an individual institutions expected losses when the financial system is in distress. By focusing on conditional losses during systemic crises, the MES provides insight into the extent to which each institution contributes to systemic vulnerability under extreme market conditions.

Table 7 presents the results obtained using these alternative measures, where the dependent variable is the residual systemic risk values derived from Equations (4) and (5). The RD estimation adopts a linear specification with a triangular kernel and the bandwidth selected based on the MSE criterion. The higher-order estimates and results obtained using alternative kernels are presented in Table A.2. The robust p values for the RD effect are consistently below 10%: The identified causal relationship between the intervention (e.g., ChatGPT's launch) and systemic financial risk is statistically significant and robust to the choice of systemic risk measurement methods. These findings underscore the reliability of our results and highlight the consistent impact of the intervention on the various methodological frameworks.

[Table 7 about here.]

6.5 Different sample ranges

The baseline regression uses a sample range of 60 trading days before and after the cutoff. To test the robustness of our results, we change the sample range. Specifically, we use 20, 40, 50, and 60 trading days before and after the cutoff as alternative sample ranges. The motivation for this approach is to examine whether the RD estimates are sensitive to including data further away from the cutoff, as data further from the threshold may introduce noise or bias in the estimates due to changing market conditions and extra shocks.

Table 8 presents the results using different sample ranges, with the dependent variable being the residual $\Delta CoVaR$. The estimated RD coefficients remain consistent with the baseline regression and are robust across all sample ranges. Moreover, the robust p values for all the estimates are below 0.05. Thus, the choice of sample range does not drive the RD effect and is robust to variations in the observation window length. This finding further supports the validity of the causal relationship between interventions and systemic financial risk.

[Table 8 about here.]

7 Mechanism

One potential mechanism underlies the challenges that banks face during their transitions, particularly because they are encouraged to adopt more AI-related technologies. Specifically, if ChatGPT accelerates banks adoption of AI-driven tools, it may increase systemic financial risk, especially during the transition period (Arenas et al., 2023; Palmié et al., 2020). This is because the integration of AI solutions often requires substantial adjustments to business operations, which can exacerbate short-term vulnerabilities. If this hypothesis holds, then institutions with limited prior experience in AI adoption are likely to encounter higher risks following the launch of ChatGPT.

Although existing research suggests that adopting AI technologies in the financial sector can mitigate risks through improved credit assessment and operational efficiency, the initial phase of disruptive innovation, particularly during the immature stages of adoption, may lead to increased tail risks. Advanced AI systems, such as those inspired by ChatGPT, may introduce unanticipated challenges, including technical failures, data privacy concerns, and operational complexities. Moreover, institutions with a stronger focus on AI adoption are better positioned to adapt to the evolving market landscape and to deploy such tools more effectively. By contrast, banks that have paid less attention to AI-related tools face greater pressure to catch up, potentially leading to a more turbulent transition process (Vives, 2019). This divergence in adaptation speed could amplify systemic risks by creating an uneven playing field within the financial system.

To investigate whether the business transformation associated with banks adoption of AI technologies serves as a channel through which ChatGPT increases financial systemic risk in the short term, we focus on the potential risks inherent in this transformation process. The integration of AI technologies often requires significant organizational adjustments, which may exacerbate short-term vulnerabilities and operational risks. To test this mechanism, we conduct a textual analysis of banks annual reports. We calculate the average frequency of AI-related keywords over the three years prior to the event from 2019 to 2021 for each bank to ensure consistency and reduce the impact of annual fluctuations.

Based on these values, we classify banks into high- and low AI-related focus groups using the median, tertiles, and quartiles as thresholds. We then perform RD estimations separately for each group. If the effect of RD is significantly greater for banks in the high AI-related group, indicating that institutions less prepared for AI adoption face greater risks during the transition period, this would validate our hypothesis that the transformation risks associated with AI adoption are a critical mechanism through which ChatGPT increases systemic risk in the short term.

[Table 9 about here.]

Table 9 presents the results of the mechanism analysis. Columns 1 and 2 divide the sample into the top 50% (high AI-related) and bottom 50% (low AI-related) based on the median frequency of AI-related keywords. Columns 3 and 4 categorize banks into high, medium, and low AI-related focus groups, with the top third (above the 67th percentile) representing high AI-related banks and the bottom third (below the 33rd percentile) representing low AI-related banks. Columns 5 and 6 divide banks into quartiles, with the top 25% (above the 75th percentile) classified as high AI-related and the bottom 25% (below the 25th percentile) classified as low AI-related. A comparison of the two columns within each sub-column reveals that the RD effect is consistently larger for the low AI-related group and statistically significant at the 5% level, whereas the high AI-related group shows a significant but smaller value, equal to approximately half of the low AI-related group. Comparing columns 1, 3, and 5, we observe that the RD effect increases as AI-related focus decreases.

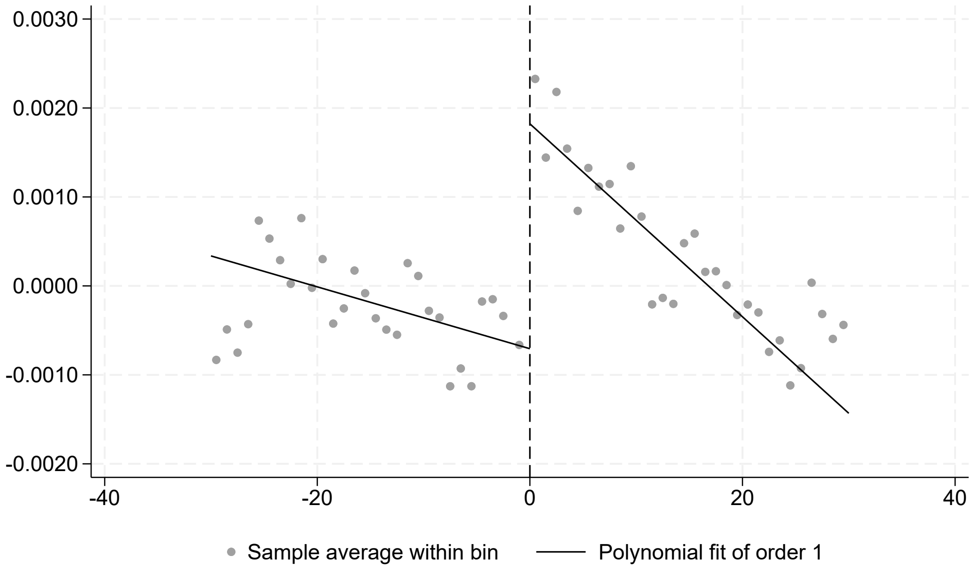
8 Conclusion

This study investigates the implications of the systemic financial risk of ChatGPT's launch in the Chinese banking sector, and offers significant insights into how disruptive technologies influence financial stability. Using RD and RD-DID methodologies, we find that the introduction of ChatGPT leads to a marked short-term increase in systemic risk, as evidenced by a significant increase in $\Delta CoVaR$ values. This underscores the potential of GenAI technologies to increase interconnectivity and risk spillovers among financial institutions during their initial deployment phases. The main conclusions were robust across multiple dimensions. Through robustness checks, including placebo tests, donut hole tests, alternative systemic risk measures (e.g., Copula-based *CoVaR*, MES), and varying sample windows, our results consistently reveal that ChatGPT's launch has a causal and significant impact on systemic financial risk.

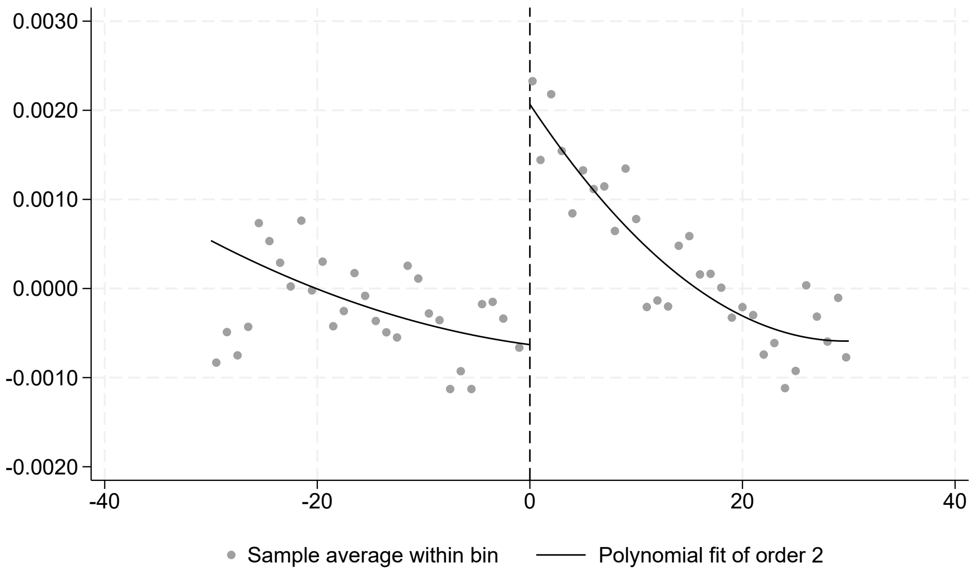
A heterogeneity analysis further uncovers disparities in how different types of bank experience increase systemic risk. Although such discrepancies are often statistically insignificant, smaller banks, rural commercial banks, and banks with higher nonperforming loan ratios tend to exhibit greater susceptibility, likely due to limited resources and higher baseline vulnerabilities. By contrast, large state-owned banks with stronger financial buffers and government backing are shielded from these risks. These findings emphasize the importance of tailoring risk-mitigation strategies to specific bank characteristics, ensuring that weaker institutions receive targeted support.

Finally, our mechanism analysis reveals that the observed increase in systemic risk is closely related to the transitional challenges banks face when adopting advanced AI technologies. This indirect effect is evidenced by the fact that banks with lower readiness to adopt AI or weaker technological capacity are disproportionately affected, facing higher operational and adjustment costs. This highlights the economic importance of reducing technological disparities between institutions, as smoother transitions improve the stability of individual banks and mitigate the risk of contagion within the financial system. These findings have significant political implications for future studies. Although GenAI has immense potential for improving operational efficiency, its deployment must be accompanied by safeguards to mitigate transitional vulnerabilities. Regulatory authorities should encourage financial institutions to strengthen their technological infrastructure and risk-management capabilities to better accommodate such innovations. Furthermore, policies promoting the equitable adoption of AI in banks of different sizes and ownership structures could reduce systemic disparities and improve overall financial resilience.

Figures



(a) Linear polynomial



(b) Quadratic polynomial

Note: The optimal number of bins was chosen based on a data-driven process using the integrated mean-squared error-optimal evenly spaced method with spacing estimators. The global polynomial order for Figure 1a is 1, and for Figure 1b is 2. Standard errors are clustered at the bank level using a triangle kernel function.

Figure 1. RD Plots of Residual $\Delta CoVaR$

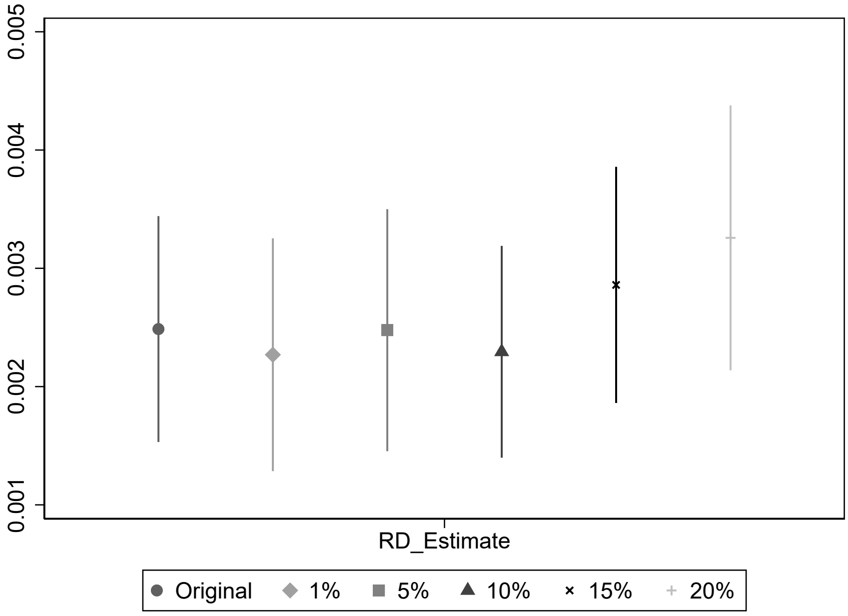


Figure 2. Results of Donut Hole test

Note: This figure plots the results of the Donut Hole test, showing RD local non-parametric estimates and their corresponding 95% confidence intervals after excluding portions of the sample around the cut-off. From left to right, the results represent the baseline estimates after excluding 1%, 5%, 10%, 15%, and 25% of the sample. A linear specification with a triangular kernel is used for the RD estimation. The bandwidths are selected based on the MSE criterion.

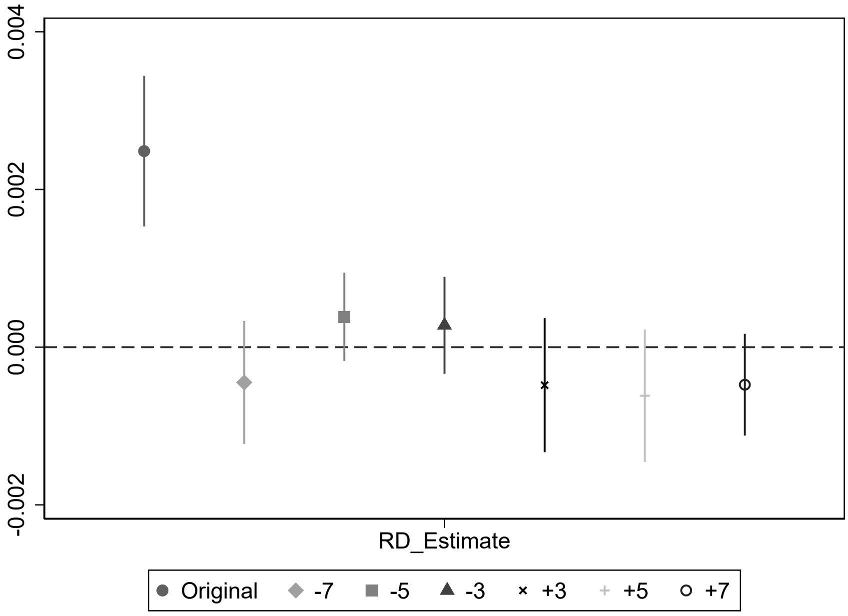


Figure 3. Results of Placebo Test

Note: This figure presents the results of the placebo tests, where we introduce artificial cut-offs to examine whether the observed RD effects are specific to the actual intervention date (the launch of ChatGPT). The leftmost point represents the baseline regression result using the actual cut-off, while the subsequent points correspond to placebo tests assuming the policy intervention occurred seven trading days before (-7), five trading days before (-5), three trading days before (-3), three trading days after (+3), five trading days after (+5), and seven trading days after (+7). Each point shows the RD estimate with the corresponding 95% confidence interval. A linear specification with a triangular kernel is used for the RD estimation. The bandwidths are selected based on the MSE criterion.

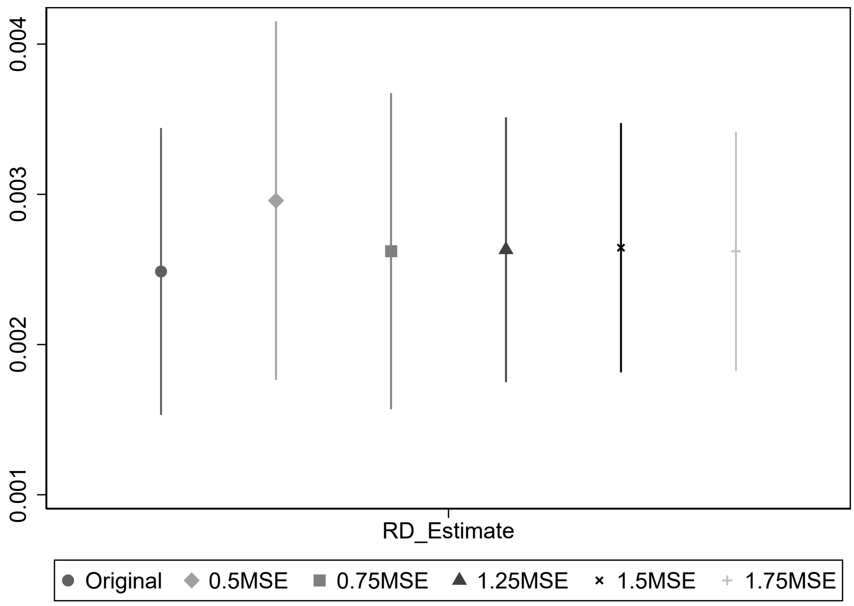


Figure 4. Results of Bandwidth Sensitivity

Note: This figure illustrates the sensitivity of the RD estimates to bandwidth selection. The horizontal axis represents the scaling factor applied to the MSE-optimal bandwidth, ranging from 0.5 to 1.75. Specifically, "0.5MSE" indicates that the bandwidth used is 50% of the MSE-optimal bandwidth, while "0.75MSE," "1.25MSE," "1.5MSE," and "1.75MSE" indicate 75%, 125%, 150%, and 175% of the MSE-optimal bandwidth, respectively. The leftmost point represents the baseline regression result using the MSE-optimal bandwidth. Each point shows the RD estimate with the corresponding 95% confidence interval. A linear specification is used for the RD estimation with a triangular kernel. The bandwidths are selected based on the MSE criterion.

Tables

Table 1. Control Variables Used in Equation (4)

Variables	Definition	Frequency	Data source	Unit-root test (p value)
Y3MChina	three-month Treasury yield, indicating market spread	Daily	CEIC	0.0000
Credit	one-year AAA corporate bond rate minus China's one-year Treasury yield, indicating credit risk	Daily	CEIC	0.0000
CycChina	10-year Chinese Treasury yield minus three-month Chinese Treasury yield, indicating economic cycle	Daily	CEIC	0.0000
MV	Bank market value growth rate	Daily	CSMAR	0.0000
COVID19	Set to 1 if the date is December 7, 2022, or later; otherwise 0	Daily	-	-

Note: The unit-root test is conducted using LLC, incorporating both a time trend and the removal of individual fixed effects since our data is panel data. We report the p value in the of the LLC test.

Table 2. Descriptive statistics

	Obs.	Mean	SD	Min	p50	Max	Kurtosis	Skewness
Y3MChina	2,562	1.938	0.191	1.617	2.022	2.220	1.591	-0.369
Credit	2,562	0.510	0.138	0.274	0.522	0.796	2.114	-0.089
CycChina	2,562	0.868	0.135	0.633	0.840	1.091	1.668	0.035
MV	2,562	0.000	0.014	-0.097	0.000	0.095	8.802	0.739
COVID19	2,562	0.426	0.495	0.000	0.000	1.000	1.089	0.298
$\Delta CoVaR$	2,562	0.013	0.006	0.003	0.012	0.032	4.672	1.248
$Residual\Delta CoVaR$	2,562	0.000	0.002	-0.005	0.000	0.006	5.341	0.433

Note: This figure presents the descriptive statistics of the variables. All continuous variables are winsorized at the 1% and 99% levels to mitigate the influence of extreme values and confirmed to be free of unit roots. $\Delta CoVaR$ is the dependent variable in Equation (4), while $Residual\Delta CoVaR$ is the dependent variable in Equation (5).

Table 3. Global Polynomial RD Estimation

	(1)	(2)	(3)	(4)	(5)	(6)
	ResΔ*CoVaR*	*Res*Δ*CoVaR*	*Res*Δ*CoVaR*	*Res*Δ*CoVaR*	*Res*Δ*CoVaR*	*Res*Δ*CoVaR*
RD effect	0.0025	0.0025	0.0027	0.0027	0.0025	0.0025
	(8.089)	(8.552)	(5.540)	(5.537)	(4.489)	(4.482)
Robust p-value	0.000	0.000	0.000	0.000	0.000	0.000
Order	1	1	2	2	3	3
Kernel	Triangular	Epanechnikov	Triangular	Epanechnikov	Triangular	Epanechnikov
Obs.	2,562	2,562	2,562	2,562	2,562	2,562

Note: This table reports the results of global polynomial RD effects. The values in parentheses are unadjusted z-statistics. "Robust p values" are calculated using bias-corrected standard errors following Calonico et al. (2017). *order* denotes the order of the global polynomial, and *Kernel* represents the kernel function. Standard errors are clustered at the bank level.

Table 4. Local Polynomial RD Estimation

(a) Linear Polynomial

	(1)	(2)	(3)	(4)	(5)	(6)
	ResΔCoVaR	*ResΔCoVaR*	*ResΔCoVaR*	*ResΔCoVaR*	*ResΔCoVaR*	*ResΔCoVaR*
RD effect	0.0025	0.0024	0.0025	0.0026	0.0026	0.0030
	(5.102)	(4.911)	(4.701)	(4.883)	(4.847)	(5.343)
Robust p-value	0.000	0.000	0.000	0.000	0.000	0.000
Order	1	1	1	1	1	1
Kernel	Triangular	Epanechnikov	Uniform	Triangular	Epanechnikov	Uniform
Bandwidth selector	MSE	MSE	MSE	CER	CER	CER
Obs.	2,562	2,562	2,562	2,562	2,562	2,562
Bandwidth	10.79	9.81	6.62	8.65	7.86	5.31
Eff # of left	420	378	252	336	294	210
Eff # of right	462	420	294	378	336	252

(b) Quadratic Polynomial

	(1)	(2)	(3)	(4)	(5)	(6)
	ResΔCoVaR	*ResΔCoVaR*	*ResΔCoVaR*	*ResΔCoVaR*	*ResΔCoVaR*	*ResΔCoVaR*
RD effect	0.0032	0.0029	0.0021	0.0027	0.0027	0.0035
	(4.497)	(4.336)	(3.553)	(3.635)	(3.581)	(4.860)
Robust p-value	0.001	0.000	0.000	0.067	0.067	0.050
Order	2	2	2	2	2	2
Kernel	Triangular	Epanechnikov	Uniform	Triangular	Epanechnikov	Uniform
Bandwidth selector	MSE	MSE	MSE	CER	CER	CER
Obs.	2,562	2,562	2,562	2,562	2,562	2,562
Bandwidth	7.58	10.16	12.35	5.88	5.47	6.53
Eff # of left	294	420	504	210	210	252
Eff # of right	336	462	546	252	252	294

Note: p denotes the order of the local polynomial, and h represents the bandwidth. The values in parentheses are unadjusted z-statistics. "Robust p values" are calculated using bias-corrected standard errors following Calonico et al. (2017). "Bandwidth" refers to the bandwidth used for the estimation of the regression function around the cut-off on each side. The last two rows indicate the effective observations to the left and right of the cutoff, respectively. Standard errors are clustered at the bank level using the uniform kernel function.

Table 5. Results of RD-DID

	(1) $Res\Delta CoVaR$	(2) $Res\Delta CoVaR$	(3) $Res\Delta CoVaR$	(4) $Res\Delta CoVaR$
$D(t \geq c)$	0.0008***	0.0009***	0.0005**	0.0010***
	(9.827)	(4.471)	(2.689)	(5.846)
$treat$	-0.0001*	-0.0001*	-0.0001*	-0.0001*
	(-1.772)	(-1.772)	(-1.771)	(-1.771)
$treat \times D(t \geq c)$	0.0003***	0.0003***	0.0003***	0.0003***
	(5.380)	(5.379)	(5.378)	(5.378)
Order	1	2	3	4
F value	32.456	55.275	47.282	56.844
Adj R-squared	0.065	0.065	0.070	0.078
Obs.	5,002	5,002	5,002	5,002

Note: This table presents the RD-DID estimation results of ChatGPT's impact on financial systemic risk. Bank 001227 (Lanzhou Bank) was excluded due to missing data for 2021, resulting in a total of 5,002 observations ($41 \times 61 \times 2$). "Order" indicates the degree of $f(t-c)$. Values in parentheses are t-statistics. Significance levels are reported as follows: $*p < 0.10$, $**p < 0.05$, $***p < 0.001$. Standard errors are clustered at the bank level.

Table 6. Heterogeneous Analysis

(a) Bank Types

	State-owned	City commercial	Rural commercial
	(1)	(2)	(3)
	$Res\Delta CoVaR$	$Res\Delta CoVaR$	$Res\Delta CoVaR$
RD effect	0.0016	0.0024	0.0036
	(2.215)	(4.080)	(3.482)
Robust p-value	0.037	0.001	0.005
Order	1	1	1
Kernel	Triangular	Triangular	Triangular
Bandwidth selector	MSE	MSE	MSE
Obs.	366	1,586	610
Diff value [Empirical p-value]		-0.0008[0.360]	-0.0019[0.030]

(b) NPL ratio

	Low NPL ratio	High NPL ratio
	(1)	(2)
	$Res\Delta CoVaR$	$Res\Delta CoVaR$
RD effect	0.0022	0.0028
	(3.047)	(4.623)
Robust p-value	0.019	0.000
Order	1	1
Kernel	Triangular	Triangular
Bandwidth selector	MSE	MSE
Obs.	1,281	1,281
Diff value [Empirical p-value]		-0.0006[0.160]

(c) Bank size

	Small size	Big size
	(1)	(2)
	$Res\Delta CoVaR$	$Res\Delta CoVaR$
RD effect	0.0031	0.0020
	(4.145)	(3.964)
Robust p-value	0.001	0.001
Order	1	1
Kernel	Triangular	Triangular
Bandwidth selector	MSE	MSE
Obs.	1,281	1,281
Diff value [Empirical p-value]		0.0010[0.070]

Note: This table presents the results of the heterogeneity analysis. The values in parentheses are unadjusted z-statistics. "Robust p values" are calculated using bias-corrected standard errors following Calonico et al. (2017). A linear specification with a triangular kernel is used for the RD estimation. The bandwidths are selected based on the MSE criterion. Standard errors are clustered at the bank level. "Diff value" is the difference in RD effect between columns 1 and 2. The corresponding empirical p value is shown in brackets based on bootstrap resampling (100 iterations).

Table 7. Alternative Measures for Financial Systemic Risk

	(1) *ResQ*-Δ*CoVaR*	(2) *ResC*-Δ*CoVaR*	(3) *Res*MES
RD effect	0.0019	0.0020	0.0004
	(2.710)	(9.680)	(3.186)
Robust p-value	0.031	0.000	0.004
Order	1	1	1
Kernel	Triangular	Triangular	Triangular
Bandwidth selector	MSE	MSE	MSE
Obs.	2,562	2,562	2,562

Note: This table documents results using different systemic risk measurements. Column (1)-(3) stand for *Q-CoVaR*, *C-CoVaR* and MES, respectively. For the *C-CoVaR*, the optimal copula was determined according to Akaike information criterion values of several copula candidates, with the Student's t copula selected as the best fit. The MES was calculated using a DCC–GARCH model, capturing dynamic conditional correlations among institutions and the market. The values in parentheses are unadjusted z-statistics. "Robust p values" are calculated using bias-corrected standard errors following Calonico et al. (2017). A linear specification with a triangular kernel is used for the RD estimation. The bandwidths are selected based on the MSE criterion. Standard errors are clustered at the bank level.

Table 8. Different Sample Ranges

	(1) *Res*Δ*CoVaR*	(2) *Res*Δ*CoVaR*	(3) *Res*Δ*CoVaR*	(4) *Res*Δ*CoVaR*
RD effect	0.0030	0.0025	0.0025	0.0026
	(4.641)	(5.075)	(5.353)	(6.464)
Robust p-value	0.000	0.000	0.000	0.000
Order	1	1	1	1
Kernel	Triangular	Triangular	Triangular	Triangular
Bandwidth selector	MSE	MSE	MSE	MSE
Sample range	[-20, +20]	[-40, +40]	[-50, +50]	[-60, +60]
Obs.	1,722	3,402	4,242	5,082

Note: This table documents the results of the RD estimation using different sample ranges around the cut-off. The dependent variable is the residual $\Delta CoVaR$. The sample ranges refer to the number of trading days before and after the cut-off included in the analysis. For example, [-20, +20] indicates that the sample consists of 20 trading days before and 20 after the cut-off. The values in parentheses are unadjusted z-statistics. "Robust p values" are calculated using bias-corrected standard errors following Calonico et al. (2017). A linear specification with a triangular kernel is used for the RD estimation. The bandwidths are selected based on the MSE criterion. Standard errors are clustered at the bank level.

Table 9. Mechanism Analysis

	50th vs 50th		33th vs 67th		25th vs 75th	
	(1)	(2)	(3)	(4)	(5)	(6)
	ResΔCoVaR	*ResΔCoVaR*	*ResΔCoVaR*	*ResΔCoVaR*	*ResΔCoVaR*	*ResΔCoVaR*
RD effect	0.0031	0.0019	0.0037	0.0020	0.0043	0.0023
	(3.860)	(3.092)	(3.838)	(2.517)	(4.481)	(2.350)
Robust p-value	0.003	0.005	0.002	0.020	0.000	0.036
Order	1	1	1	1	1	1
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular
Bandwidth selector	MSE	MSE	MSE	MSE	MSE	MSE
Obs.	1,342	1,220	854	854	671	610

Note: This table documents the results of the mechanism analysis, examining the impact of ChatGPT on financial systemic risk through banks' transformation risks due to the adoption of AI technologies. The values in parentheses are unadjusted z-statistics. "Robust p values" are calculated using bias-corrected standard errors following Calonico et al. (2017). A linear specification with a triangular kernel is used for the RD estimation. The bandwidths are selected based on the MSE criterion. Standard errors are clustered at the bank level.

References

Acharya, V. V., Pedersen, L. H., Philippon, T., & Richardson, M. (2017). Measuring systemic risk. *The review of financial studies*, *30*(1), 2–47.

Adrian, T., & Brunnermeier, M. K. (2016). Covar. *The American Economic Review*, *106*(7), 1705.

Aghion, P., Jones, B. F., & Jones, C. I. (2018, January). *Artificial intelligence and economic growth*. University of Chicago Press. http://www.nber.org/chapters/c14015

Alawida, M., Abu Shawar, B., Abiodun, O. I., Mehmood, A., Omolara, A. E., & Al Hwaitat, A. K. (2024). Unveiling the dark side of chatgpt: Exploring cyberattacks and enhancing user awareness. *Information*, *15*(1), 27.

Allen, F., & Gale, D. (2000). Financial contagion. *Journal of Political Economy*, *108*(1), 1–33.

Anderson, M. L. (2014). Subways, strikes, and slowdowns: The impacts of public transit on traffic congestion. *American Economic Review*, *104*(9), 2763–2796.

Ansari, S., Garud, R., & Kumaraswamy, A. (2016). The disruptor's dilemma: Tivo and the us television ecosystem. *Strategic Management Journal*, *37*(9), 1829–1853.

Arenas, L., Gil-Lafuente, A. M., & Reverter, J. B. (2023). The impact of disruptive technology on banking under switching volatility regimes. *Technological and Economic Development of Economy*, *29*(4), 1264–1290.

Babina, T., Fedyk, A., He, A., & Hodson, J. (2024). Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics*, *151*, 103745.

Bai, C., Duan, Y., & Goodell, J. W. (2025). Multi-media sentiment to systemic risk: Evidence from covid-19. *International Review of Economics & Finance*, *97*, 103745.

Barreca, A. I., Guldi, M., Lindo, J. M., & Waddell, G. R. (2011). Saving babies? revisiting the effect of very low birth weight classification. *The Quarterly Journal of Economics*, *126*(4), 2117–2123.

Beckmann, L., & Hark, P. F. (2024). Chatgpt and the banking business: Insights from the us stock market on potential implications for banks. *Finance Research Letters*, *63*, 105237.

Bento, A., Kaffine, D., Roth, K., & Zaragoza-Watkins, M. (2014). The effects of regulation in the presence of multiple unpriced externalities: Evidence from the transportation sector. *American Economic Journal: Economic Policy*, *6*(3), 1–29.

Bertomeu, J., Lin, Y., Liu, Y., & Ni, Z. (2023). Capital market consequences of generative ai: Early evidence from the ban of chatgpt in italy. *Available at SSRN 4452670*.

Bickford, J., Cegiela, R., King, J., Lucas, K., Pardasani, N., Rabener, E., Rehberg, B., Riemer, S., Strauss, M., Sugihara, J., & Widowitz, M. (2023, September). Banking on generative ai: Maximizing the financial services opportunity [White Paper]. https://www.bcg.com

Bloom, D. E., Prettner, K., Saadaoui, J., & Veruete, M. (2024). *Artificial intelligence and the skill premium* (tech. rep.). National Bureau of Economic Research.

Brownlees, C., & Engle, R. F. (2017). Srisk: A conditional capital shortfall measure of systemic risk. *The Review of Financial Studies*, *30*(1), 48–79.

Brunnermeier, M., Rother, S., & Schnabel, I. (2020). Asset price bubbles and systemic risk. *The Review of Financial Studies*, *33*(9), 4272–4317.

Calmes, C., & Théoret, R. (2014). Bank systemic risk and macroeconomic shocks: Canadian and us evidence. *Journal of Banking & Finance*, *40*, 388–402.

Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2017). Rdrobust: Software for regression-discontinuity designs. *The Stata Journal*, *17*(2), 372–404.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, *82*(6), 2295–2326.

Cincinelli, P., Pellini, E., & Urga, G. (2022). Systemic risk in the chinese financial system: A panel granger causality analysis. *International Review of Financial Analysis*, *82*, 102179.

Cleary, S. (1999). The relationship between firm investment and financial status. *The Journal of Finance*, *54*(2), 673–692.

Danneels, E. (2004). Disruptive technology reconsidered: A critique and research agenda. *Journal of Product Innovation Management*, *21*(4), 246–258.

David, H. (2024). Applying ai to rebuild middle class jobs. *NBER Working Paper*, (w32140).

Dong, Y., Firth, M., Hou, W., & Yang, W. (2016). Evaluating the performance of chinese commercial banks: A comparative analysis of different types of banks. *European Journal of Operational Research*, *252*(1), 280–295.

Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of big data–evolution, challenges and research agenda. *International Journal of Information Management*, *48*, 63–71.

Duan, Y., Fan, X., & Wang, Y. (2022). Economic policy uncertainty and bank systemic risk: A cross-country analysis. *Pacific-Basin Finance Journal*, *75*, 101828.

Dubey, S. S., Astvansh, V., & Kopalle, P. K. (2024). Express: Generative ai solutions to empower financial firms. *Journal of Public Policy & Marketing*, 07439156241311300.

Eisfeldt, A. L., & Schubert, G. (2024). *Ai and finance* (tech. rep.). National Bureau of Economic Research.

Eisfeldt, A. L., Schubert, G., Zhang, M. B., & Taska, B. (2023). The labor impact of generative ai on firm values. *Available at SSRN 4436627*.

Emett, S. A., Eulerich, M., Pickerd, J. S., & Wood, D. A. (2024). Short and synthetically distort: Investor reactions to deepfake financial news. *Available at SSRN 4869830*.

Fernández, A. (2019). Artificial intelligence in financial services. *Economic Bulletin*, (JUN). https://ideas.repec.org/a/bde/journl/y2019i6daan7.html

Fu, S., & Gu, Y. (2017). Highway toll and air pollution: Evidence from chinese cities. *Journal of Environmental Economics and Management*, *83*, 32–49.

Gelman, A., & Imbens, G. (2019). Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics*, *37*(3), 447–456.

Greenstone, M., He, G., Jia, R., & Liu, T. (2022). Can technology solve the principal-agent problem? evidence from chinas war on air pollution. *American Economic Review: Insights*, *4*(1), 54–70.

Gupta, M., Akiri, C., Aryal, K., Parker, E., & Praharaj, L. (2023). From chatgpt to threatgpt: Impact of generative ai in cybersecurity and privacy. *IEEE Access*.

Hausman, C., & Rapson, D. S. (2018). Regression discontinuity in time: Considerations for empirical applications. *Annual Review of Resource Economics*, *10*, 533–552.

Keeley, N. G. (2023). Cash, credibility, and conversion: The influence of synthetic media on investment behavior. *arXiv preprint arXiv:2306.05033*.

Kong, Y., Nie, Y., Dong, X., Mulvey, J. M., Poor, H. V., Wen, Q., & Zohren, S. (2024). Large language models for financial and investment management: Applications and benchmarks. *Journal of Portfolio Management*, *51*(2).

Leitner, G., Singh, J., van der Kraaij, A., & Zsámboki, B. (2024). The rise of artificial intelligence: benefits and risks for financial stability. *Financial Stability Review*, *1*. https://ideas.repec.org/a/ecb/fsrart/202400012.html

Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., He, H., Li, A., He, M., Liu, Z., et al. (2023). Summary of chatgpt-related research and perspective towards the future of large language models. *Meta-Radiology*, 100017.

Lu, Y., Wang, J., & Zhu, L. (2019). Place-based policies, creation, and agglomeration economies: Evidence from chinas economic zone program. *American Economic Journal: Economic Policy*, *11*(3), 325–360.

Meng, X., & Yu, Y. (2023). Does the russia-ukraine conflict affect gasoline prices? *Energy Economics*, *128*, 107113.

Mensi, W., Hammoudeh, S., Shahzad, S. J. H., & Shahbaz, M. (2017). Modeling systemic risk and dependence structure between oil and stock markets using a variational mode decomposition-based copula method. *Journal of Banking & Finance*, *75*, 258–279.

Murinde, V., Rizopoulos, E., & Zachariadis, M. (2022). The impact of the fintech revolution on the future of banking: Opportunities and risks. *International Review of Financial Analysis*, *81*, 102103.

Palmié, M., Wincent, J., Parida, V., & Caglar, U. (2020). The evolution of the financial technology ecosystem: An introduction and agenda for future research on disruptive innovations in ecosystems. *Technological Forecasting and Social Change*, *151*, 119779.

Persson, P., & Rossin-Slater, M. (2024). When dad can stay home: Fathers' workplace flexibility and maternal health. *American Economic Journal: Applied Economics*, *16*(4), 186–219.

Qian, J. (2024). Unleashing generative ai: Funding implications and insights from china. *Journal of Asian Public Policy*, 1–24.

Rabener, E., Riemer, S., Widowitz, M., Strauss, M., Tan, C., Rehberg, B., Kleppe, A., Bickford, J., Hilbers, P., Ogata, K., Sugihara, J., Lucas, K., Leoni, M., & Kataeva, N. (2024, April). Transformations edge: The state of genai in global financial institutions [A BCG benchmarking survey]. https://media-publications.bcg.com/Transformations-Edge-The-State-of-GenAI-in-Global-Financial-Institutions.pdf

Remolina, N. (2024). Mapping generative ai regulation in finance and bridging regulatory gaps. *Journal of Financial Transformation, Forthcoming*.

Roy, R., Lampert, C. M., & Stoyneva, I. (2018). When dinosaurs fly: The role of firm capabilities in the avianizationof incumbents during disruptive technological change. *Strategic Entrepreneurship Journal*, *12*(2), 261–284.

Salisu, A. A., Demirer, R., & Gupta, R. (2022). Financial turbulence, systemic risk and the predictability of stock market volatility. *Global Finance Journal*, *52*, 100699.

Singh, B. (2024). Generative artificial intelligence: Prospects for banking industry. *International Journal of Research in Engineering, Science and Management*, *7*(3), 83–86.

Stokel-Walker, C., & Van Noorden, R. (2023). What chatgpt and generative ai mean for science. *Nature*, *614*(7947), 214–216.

Tian, J., & Nagayasu, J. (2024, October). *AI and Financial Systemic Risk in the Global Market* (TUPD Discussion Papers No. 55). Graduate School of Economics and Management, Tohoku University. https://ideas.repec.org/p/toh/tupdaa/55.html

Vives, X. (2019). Digital disruption in banking. *Annual Review of Financial Economics*, *11*(1), 243–272.

Wach, K., Duong, C. D., Ejdys, J., Kazlauskait, R., Korzynski, P., Mazurek, G., Paliszkiewicz, J., & Ziemba, E. (2023). The dark side of generative artificial intelligence: A critical analysis of controversies and risks of chatgpt. *Entrepreneurial Business and Economics Review*, *11*(2), 7–30.

Wang, S., & Liang, Z. (2024). What does the public think about artificial intelligence? an investigation of technological frames in different technological context. *Government Information Quarterly*, *41*(2), 101939.

Wu, Q., Zhuang, Q., Liu, Y., & Han, L. (2024). Technology shock of chatgpt, social attention and firm value: Evidence from china. *Technology in Society*, *79*, 102756.

Xue, J., Wang, L., Zheng, J., Li, Y., & Tan, Y. (2023). Can chatgpt kill user-generated q&a platforms? *Available at SSRN 4448938*.

Zetzsche, D. A., Buckley, R. P., Barberis, J. N., & Arner, D. W. (2017). Regulating a revolution: From regulatory sandboxes to smart regulation. *Fordham J. Corp. & Fin. L.*, *23*, 31.

A Appendix For Online Only

A.1 DCC-GARCH model

This section presents the detailed descriptions of the DCC-GARCH model we mentioned in Section 3.1. The log return of the daily closing prices of the CSI Bank Index (code: 399986) is used as the system return of the financial market. We assume that the returns ($\mathbf{R}_t^i$) of the financial system and bank i follow a bivariate GARCH process with the DCC-GARCH specification (Engle, 2002). Define a bivariate return $\mathbf{R}_t^i = (R_t^s, X_t^i)'$. We have

$$\mathbf{R}_t^i = \varepsilon_t^i,\ \varepsilon_t^i = \mathbf{\Sigma}_{i,t}^{1/2}\mathbf{z}_t, \tag{7}$$

where $\mathbf{z}_t$ follows the bivariate Student's t distribution and $\mathbf{\Sigma}_{i,t}$ is the conditional covariance (co-volatility) matrix of the error term ε_t with the form:

$$\mathbf{\Sigma}_{i,t} = \begin{pmatrix} \sigma_t^{s^2} & \sigma_t^{si} \\ \sigma_t^{is} & \sigma_{i,t}^2 \end{pmatrix} = \mathbf{D}_{i,t}\mathbf{C}_{i,t}\mathbf{D}_{i,t}. \tag{8}$$

Note that in (8), we assume the volatility of the system's return, $\sigma_t^{s^2}$, and the volatility of bank i's return, $\sigma_{i,t}^2$, both follow the standard GARCH(1,1) process. That is, for demeaned return X_t^i,

$$X_t^i = \varepsilon_{i,t},\ \ \varepsilon_{i,t} = z_{i,t}\sigma_{i,t}, \tag{9}$$

where $z_{i,t}$ follows $i.i.d.$ Student's t distribution and $\sigma_{i,t}$ satisfies:

$$\sigma_{i,t}^2 = \omega_i + \alpha_i\varepsilon_{i,t-1}^2 + \beta_i\sigma_{i,t-1}^2. \tag{10}$$

The GARCH(1,1) model of the system's return R_t^s is defined in the same way. Returning to (8), $\mathbf{D}_{i,t}$ is a diagonal matrix with diagonal elements $(\sigma_t^s, \sigma_{i,t})$. $\mathbf{C}_{i,t}$ is the dynamic conditional correlation matrix in the form of

$$\mathbf{C}_{i,t} = \begin{pmatrix} \rho_t^{si} & 1 \\ 1 & \rho_t^{is} \end{pmatrix} = \mathrm{diag}(\mathbf{Q}_{i,t})^{-1/2}\mathbf{Q}_{i,t}\mathrm{diag}(\mathbf{Q}_{i,t})^{-1/2},$$

where $\rho_t^{si} = \rho_t^{is}$ denotes the correlation coefficients between the system and bank i at time t. Furthermore, $\mathbf{Q}_{i,t}$ has the following dynamic structure:

$$\mathbf{Q}_{i,t} = (1-a-b)\mathbf{S}_i + a\eta_{i,t-1}\eta'_{i,t-1} + b\mathbf{Q}_{i,t-1}, \tag{11}$$

where $\eta_{i,t} = (R_t^s/\sigma_t^s,\ X_t^i/\sigma_{i,t})'$, $\mathbf{S}_i$ is the unconditional covariance matrix of $\eta_{i,t}$, and the DCC coefficients satisfy $a, b \geqslant 0$, $a+b<1$. The model is then estimated by standard QMLE procedures.

A.2 Quantile-based CoVaR Method

We employ the standard quantile regression method to estimate *CoVaR* (*Q-CoVaR*), following the methodology described in Adrian and Brunnermeier (2016). The quantile measure of $CoVaR_{q,t}^{sys|X^i}$

is given by

$$CoVaR_{q,t}^{sys|X^i} = VaR_{q,t}^{sys|X^i=VaR_q^i} = \hat{\alpha}_q^i + \hat{\beta}_q^i VaR_q^i \tag{12}$$

since the definition of value at risk implies that

$$CoVaR_{q,t}^{sys|X^i} = \hat{R}_q^{s|X^i} = \hat{\alpha}_q^i + \hat{\beta}_q^i X^i, \tag{13}$$

where the second equality means the system loss is predicted by the return of a particular bank i. Note that in (12), VaR_q^i is derived from the qth quantile of the losses of bank i. Subsequently, the systemic risk attributable to the distress of bank i can be expressed as:

$$\Delta CoVaR_{q,t}^{sys|X^i} = CoVaR_{q,t}^{sys|X^i} - CoVaR_{50\%,t}^{sys|X^i} = \hat{\beta}_q^i (VaR_q^i - VaR_{50\%}^i). \tag{14}$$

A.3 Copula-based CoVaR Method

We follow a typical Copula-based *CoVaR* (*C-CoVaR*) approach applied in Reboredo and Ugolini (2015) and Mensi et al. (2017). First, let us recall the definition of a bivariate copula. The joint distribution $F_{XY}(x,y)$ of two continuous random variables X and Y can be represented using a copula function $C(u,v)$:

$$F_{XY}(x,y) = C(u,v), \tag{15}$$

where $u = F_X(x)$ and $v = F_Y(y)$ are the marginal distribution functions of X and Y, respectively. A copula is, therefore, a multivariate function with uniform marginals that describes the dependence structure between the two random variables. Next, with the definition of generalized *CoVaR* in Girardi and Ergün (2013) that $\mathbf{P}\left(R_t^s \leqslant CoVaR_{q,t}^{sys|X^i} | X_t^i \leqslant VaR_{\alpha,t}^i\right) = q$, we have

$$\mathbf{P}\left(R_t^s \leqslant CoVaR_{q,t}^{sys|X^i},\ X_t^i \leqslant VaR_{\alpha,t}^i\right) = q\alpha. \tag{16}$$

Then *CoVaR* can be written in terms of copulas as:

$$\begin{aligned} C\left(F_{R_t^s}\left(CoVaR_{q,t}^{sys|X^i}\right), F_{X_t^i}\left(VaR_{\alpha,t}^i\right)\right) &= q\alpha, \\ 1 - F_{R_t^s}\left(CoVaR_{q,t}^{sys|X^i}\right) - F_{X_t^i}\left(VaR_{1-\alpha,t}^i\right) + C\left(F_{R_t^s}\left(CoVaR_{q,t}^{sys|X^i}\right), F_{X_t^i}\left(VaR_{1-\alpha,t}^i\right)\right) &= q\alpha, \end{aligned} \tag{17}$$

where $F_{R_t^s}$ and $F_{X_t^i}$ denote the marginal distribution functions of R_t^s and X_t^i, respectively. Given a specific copula representation, the *CoVaR* then can be computed in a standard two-step procedure. In the first step, for given q and α, we solve the above equations (17) to obtain the value of $F_{R_t^s}\left(CoVaR_{q,t}^{sys|X^i}\right)$. In the second step, with the value of $F_{R_t^s}\left(CoVaR_{q,t}^{sys|X^i}\right)$, we can obtain the *CoVaR* value as the quantile of the distribution of R_t^s by $F_{R_t^s}^{-1}\left(F_{R_t^s}\left(CoVaR_{q,t}^{sys|X^i}\right)\right)$. The generalized $\Delta CoVaR$ (Girardi & Ergün, 2013) is then calculated as

$$\Delta CoVaR_{q,t}^{sys|X^i} = \frac{CoVaR_{q,t}^{sys|X^i} - CoVaR_{q=50\%,t}^{sys|X^i}}{CoVaR_{q=50\%,t}^{sys|X^i}}. \tag{18}$$

Note that in this study, we evaluate seven widely used copula specificationsGaussian, Student's t, BB7, Clayton, Gumbel, Survival Clayton, and Survival Gumbelto find the best-fitting copula functions based on the Akaike Information Criterion (AIC). The Student's t copula is selected as the optimal copula specification.

A.4 Marginal Expected Shortfall

We introduce an additional measure of systemic risk, Marginal Expected Shortfall (MES), as proposed by Acharya et al. (2017) and extended by Brownlees and Engle (2017). In this study, MES represents the expected return of bank i, conditional on the system experiencing a loss exceeding its VaR at the q level. First, let us define Expected Shortfall (ES), which represents the expected loss conditional on exceeding the VaR, as follows:

$$ES_t^{sys} = E_{t-1}(X_t^i | R_t^s < VaR_q^{sys}) = \sum_{i=1}^{N} w_t^i E_{t-1}(X_t^i | R_t^s < VaR_q^{sys}), \tag{19}$$

where w_t^i is the weight of bank i in the system. The MES of bank i, which captures the systemic risk contribution of bank i, is then given by:

$$MES_t^i = \frac{\partial ES_t^{sys}}{\partial w_t^i} = E_{t-1}(X_t^i | R_t^s < VaR_q^{sys}). \tag{20}$$

Following Brownlees and Engle (2011) and Brownlees and Engle (2017), we estimate MES using a DCC-GARCH model, treating the returns of bank i and the system as a bivariate system. This approach has been widely adopted in the literature, as evidenced by Idier et al. (2014), Benoit et al. (2017) and Cincinelli et al. (2022). Building on the notations introduced in Section A.1, we model the demeaned return processes for bank i and the system as follows:

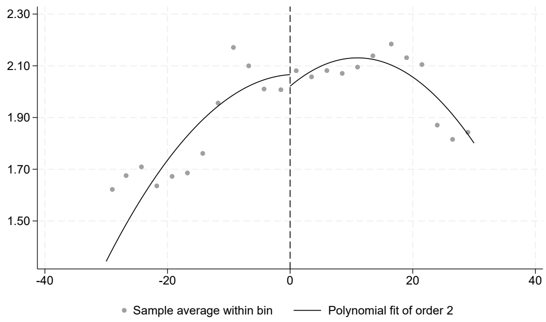
$$R_t^s = \sigma_t^s \eta_t^s, \;\; X_t^i \;\; = \sigma_{i,t}\eta_{i,t} = \sigma_{i,t}\rho_t^{is}\eta_t^s + \sigma_{i,t}(1-\rho_t^{is^2})^{1/2}\xi_{i,t}, \;\; (\eta_t^s, \xi_{i,t}) \sim F, \tag{21}$$

where the disturbance shocks $(\eta_t^s, \xi_{i,t})$ are independent and identically distributed over time, with zero mean, unit variance, and no covariance. Notably, $\xi_{i,t}$ is derived from a simple one-factor CAPM model, allowing for a time-varying *beta*. This model is defined as: $X_{i,t} = \beta_{i,t}R_t^s + v_{i,t}$, where $\beta_{i,t} = \sigma_t^{is}/\sigma_t^{s^2}$ and $v_{i,t} = \sigma_{i,t}^v \xi_{i,t}$. MES can then be explicitly expressed as follows:

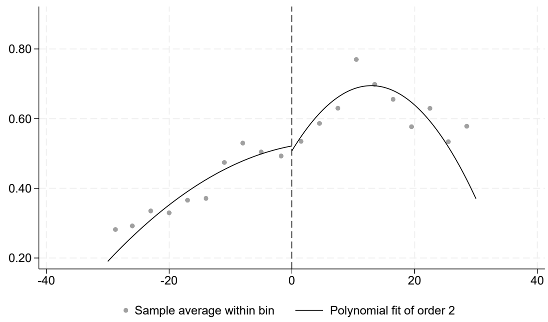
$$\begin{aligned} MES_{i,t-1} &= E_{t-1}(X_t^i | R_t^s < VaR_q^{sys}) \\ &= \sigma_{i,t} E_{t-1}(\eta_{i,t} | \eta_t^s < VaR_q^{sys}/\sigma_t^s) \\ &= \sigma_{i,t} E_{t-1}(\rho_t^{is}\eta_t^s + (1-\rho_t^{is^2})^{1/2}\xi_{i,t} | \eta_t^s < VaR_q^{sys}/\sigma_t^s) \\ &= \sigma_{i,t}\rho_t^{is} E_{t-1}(\eta_{i,t} | \eta_t^s < VaR_q^{sys}/\sigma_t^s) + \sigma_{i,t}(1-\rho_t^{is^2})^{1/2} E_{t-1}(\xi_{i,t} | \eta_t^s < VaR_q^{sys}/\sigma_t^s). \end{aligned} \tag{22}$$

Using the variance and correlation parameters estimated from the DCC-GARCH model, MES can be easily calculated by averaging the two disturbance shocks across all instances that satisfy the condition $\eta_t^s < VaR_q^{sys}/\sigma_t^s$ as noted in Brownlees and Engle (2011).

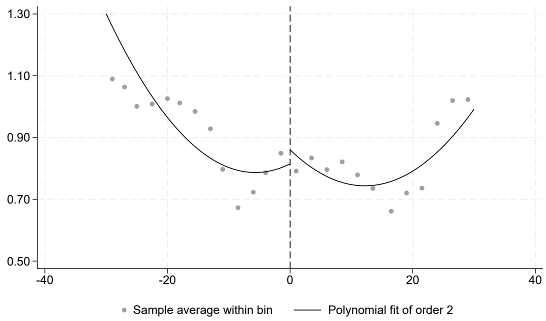
A.5 Tables and Figures



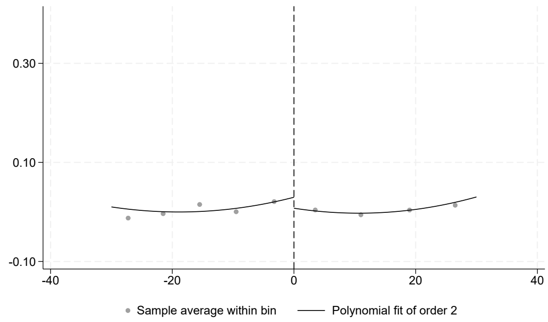
(a) Y3MChina, *p*-value=0.998



(b) Credit, *p*-value= 0.171

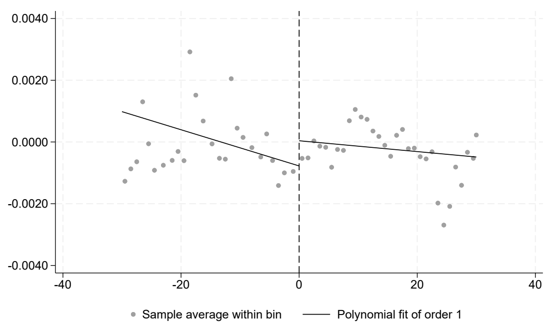


(c) CycChina, *p*-value= 0.257

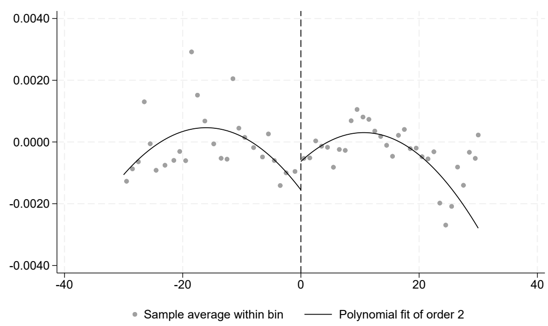


(d) MV, *p*-value=0.380

Figure A.1. RD Plots of Covariates



(a) With $Order = 1$, *p*-value=0.998



(b) With $Order = 2$, *p*-value=0.754

Figure A.2. RD Plots of Residual $\Delta CoVaR$ with Cutoff in November 30, 2021

Table A.1. Local Polynomial RD Estimation

(a) Quadratic Polynomial

	(1)	(2)	(3)	(4)	(5)	(6)
	ResΔ*CoVaR*	*Res*Δ*CoVaR*	*Res*Δ*CoVaR*	*Res*Δ*CoVaR*	*Res*Δ*CoVaR*	*Res*Δ*CoVaR*
RD effect	0.0035	0.0036	0.0035	0.0031	0.0029	0.0038
	(4.528)	(4.623)	(4.496)	(4.000)	(3.738)	(4.881)
Robust p-value	0.000	0.001	0.000	0.187	0.378	0.011
Order	3	3	3	3	3	3
Kernel	Triangular	Epanechnikov	Uniform	Triangular	Epanechnikov	Uniform
Bandwidth selector	MSE	MSE	MSE	CER	CER	CER
Obs.	2,562	2,562	2,562	2,562	2,562	2,562
Bandwidth	11.53	10.39	11.51	9.02	8.13	9.00
Eff # of left	462	420	462	378	336	378
Eff # of right	504	462	504	420	378	420

(b) Quartic Polynomial

	(1)	(2)	(3)	(4)	(5)	(6)
	ResΔ*CoVaR*	*Res*Δ*CoVaR*	*Res*Δ*CoVaR*	*Res*Δ*CoVaR*	*Res*Δ*CoVaR*	*Res*Δ*CoVaR*
RD effect	0.0038	0.0036	0.0026	0.0028	0.0025	0.0040
	(4.803)	(4.652)	(2.954)	(3.701)	(3.280)	(4.686)
Robust p-value	0.019	0.074	0.000	0.784	0.774	0.000
Order	4	4	4	4	4	4
Kernel	Triangular	Epanechnikov	Uniform	Triangular	Epanechnikov	Uniform
Bandwidth selector	MSE	MSE	MSE	CER	CER	CER
Obs.	2,562	2,562	2,562	2,562	2,562	2,562
Bandwidth	13.56	12.83	17.92	11.18	10.50	14.24
Eff # of left	546	504	714	462	420	588
Eff # of right	588	546	756	504	462	630

Table A.2. Alternative Measures for Financial Systemic Risk with Quadratic Polynomial

	(1)	(2)	(3)	(4)	(5)	(6)
	ResQ-ΔCoVaR	*ResQ-ΔCoVaR*	*ResC-ΔCoVaR*	*ResC-ΔCoVaR*	*ResMES*	*ResMES*
RD effect	0.0014	0.0014	0.0032	0.0032	0.0003	0.0004
	(1.755)	(1.628)	(13.781)	(13.668)	(2.702)	(2.690)
Robust p-value	0.118	0.192	0.000	0.008	0.008	0.008
Order	2	2	2	2	2	2
Kernel	Triangular	Epanechnikov	Triangular	Epanechnikov	Triangular	Epanechnikov
Bandwidth selector	MSE	MSE	MSE	MSE	MSE	MSE
Obs.	2,562	2,562	2,562	2,562	2,562	2,562

Table A.3. Different Sample Ranges with Quadratic Polynomial

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	ResΔCoVaR	*ResΔCoVaR*	*ResΔCoVaR*	*ResΔCoVaR*	*ResΔCoVaR*	*ResΔCoVaR*	*ResΔCoVaR*	*ResΔCoVaR*
RD effect	0.0032	0.0032	0.0028	0.0027	0.0026	0.0025	0.0026	0.0026
	(4.352)	(4.317)	(4.253)	(4.177)	(4.734)	(4.525)	(4.990)	(4.967)
Robust p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Order	2	2	2	2	2	2	2	2
Kernel	Triangular	Epanechnikov	Triangular	Epanechnikov	Triangular	Epanechnikov	Triangular	Epanechnikov
Bandwidth selector	MSE	MSE	MSE	MSE	MSE	MSE	MSE	MSE
Sample range	[-20, +20]	[-20, +20]	[-40, +40]	[-40, +40]	[-50, +50]	[-50, +50]	[-60, +60]	[-60, +60]
Obs.	1,722	1,722	3,402	3,402	4,242	4,242	5,082	5,082

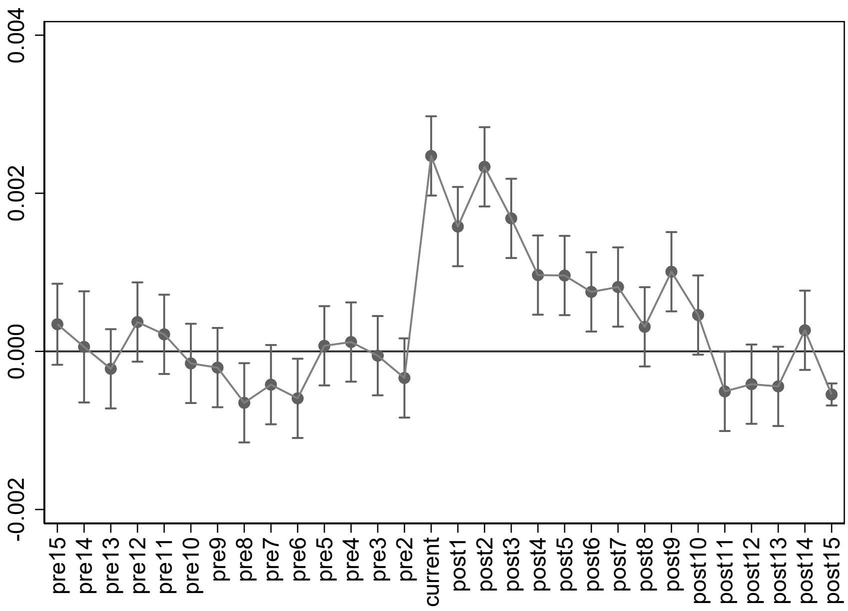


Figure A.3. Parallel Trends Assumption Test

Table A.4. Heterogeneous Analysis: State Ownership Share

	Low state ownership	High state ownership
	(1)	(2)
	ResΔ*CoVaR*	*Res*Δ*CoVaR*
RD effect	0.0034	0.0016
	(4.355)	(3.296)
Robust p-value	0.001	0.003
Order	1	1
Kernel	Triangular	Triangular
Bandwidth selector	MSE	MSE
Obs.	1,281	1,281
Diff value [Empirical p-value]		0.0018[0.030]

Note: This table presents the results of the heterogeneity analysis by state ownership. The values in parentheses are unadjusted z-statistics. "Robust p-values" are calculated using bias-corrected standard errors following Calonico et al. (2017). Banks are categorized into two groups: low state ownership in Column 1 and high state ownership in Column 2, based on the median level of state ownership share. A linear specification with a triangular kernel is used for the RD estimation. The bandwidths are selected based on the MSE criterion. Standard errors are clustered at the bank level. "Diff value" is the difference in RD effect between Columns 1 and 2. The corresponding empirical p-value is shown in brackets based on bootstrap resampling (100 iterations).

A.6 Banks in our sample

Agricultural Bank of China (601288.SH), Bank of Communications (601328.SH), Industrial and Commercial Bank of China (601398.SH), Postal Savings Bank of China (601658.SH), XD Construction Bank (601939.SH), Bank of China (601988.SH), Ping An Bank (000001.SZ), Lanzhou Bank (001227.SZ), Ningbo Bank (002142.SZ), Jiangyin Bank (002807.SZ), Zhangjiagang Bank (002839.SZ), Zhengzhou Bank (002936.SZ), Qingdao Bank (002948.SZ), Qingnong Commercial Bank (002958.SZ), Suzhou Bank (002966.SZ), Shanghai Pudong Development Bank (600000.SH), Huaxia Bank (600015.SH), China Minsheng Bank (600016.SH), China Merchants Bank (600036.SH), Wuxi Bank (600908.SH), Jiangsu Bank (600919.SH), Hangzhou Bank (600926.SH), Xi'an Bank (600928.SH), Nanjing Bank (601009.SH), Chongqing Rural Commercial Bank (601077.SH), Changshu Bank (601128.SH), Industrial Bank (601166.SH), Bank of Beijing (601169.SH), Xiamen Bank (601187.SH), Shanghai Bank (601229.SH), Rui Feng Bank (601528.SH), Changsha Bank (601577.SH), Qilu Bank (601665.SH), China Everbright Bank (601818.SH), Shanghai Rural Commercial Bank (601825.SH), Chengdu Bank (601838.SH), Zijin Bank (601860.SH), Zhejiang Commercial Bank (601916.SH), Chongqing Bank (601963.SH), Guiyang Bank (601997.SH), CITIC Bank (601998.SH), Suzhou Rural Bank (603323.SH).

References

Acharya, V. V., Pedersen, L. H., Philippon, T., & Richardson, M. (2017). Measuring systemic risk. *The review of financial studies*, *30*(1), 2–47.

Adrian, T., & Brunnermeier, M. K. (2016). Covar. *The American Economic Review*, *106*(7), 1705.

Benoit, S., Colliard, J.-E., Hurlin, C., & Pérignon, C. (2017). Where the risks lie: A survey on systemic risk. *Review of Finance*, *21*(1), 109–152.

Brownlees, C., & Engle, R. F. (2011). Volatility, correlation and tails for systemic risk measurement. *SSRN Electronic Journal*.

Brownlees, C., & Engle, R. F. (2017). Srisk: A conditional capital shortfall measure of systemic risk. *The Review of Financial Studies*, *30*(1), 48–79.

Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2017). Rdrobust: Software for regression-discontinuity designs. *The Stata Journal*, *17*(2), 372–404.

Cincinelli, P., Pellini, E., & Urga, G. (2022). Systemic risk in the chinese financial system: A panel granger causality analysis. *International Review of Financial Analysis*, *82*, 102179.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, *20*(3), 339–350.

Girardi, G., & Ergün, A. T. (2013). Systemic risk measurement: Multivariate garch estimation of covar. *Journal of Banking & Finance*, *37*(8), 3169–3180.

Idier, J., Lamé, G., & Mésonnier, J.-S. (2014). How useful is the marginal expected shortfall for the measurement of systemic exposure? a practical assessment. *Journal of Banking & Finance*, *47*, 134–146.

Mensi, W., Hammoudeh, S., Shahzad, S. J. H., & Shahbaz, M. (2017). Modeling systemic risk and dependence structure between oil and stock markets using a variational mode decomposition-based copula method. *Journal of Banking & Finance*, *75*, 258–279.

Reboredo, J. C., & Ugolini, A. (2015). Systemic risk in european sovereign debt markets: A covar-copula approach. *Journal of International Money and Finance*, *51*, 214–244.