UNIVERSITÉ DU QUÉBEC À MONTRÉAL

ESSAIS EN MACRO FINANCE

THÈSE

PRÉSENTÉE

COMME EXIGENCE PARTIELLE

DU DOCTORAT EN ÉCONOMIQUE

PAR

FONTON MOFFO AHMADOU MUSTAPHA

DECEMBRE 2023

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de cette thèse se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.04-2020). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

La réalisation de cette thèse a été possible grâce au concours de plusieurs personnes à qui je voudrais témoigner toute ma gratitude.

Je voudrais tout d'abord adresser toute ma reconnaissance à mon directeur de thèse, le professeur Dalibor Stevanovic, pour sa disponibilité, sa patience, son assistance et surtout son sens de l'écoute. Sans cet appui multiforme aussi bien moral qu'intellectuel et financier, je ne serai pas rendu à ce niveau. Il m'a également inculqué la rigueur qui est un préalable à tout succès. Je lui dit un grand merci pour tout.

Je désire aussi remercier les professeurs du département d'économie qui ont contribué de diverses manières à mon encadrement, par des cours magistraux, des conseils et des commentaires sur mes travaux lors des séminaires internes. Tout particulièrement, merci aux professeurs Nicholas Lawson et Alain Paquet pour les conseils. Merci également à Kevin Moran, professeur à l'Université Laval qui a accepté d'être examinateur externe de ma thèse.

Je voudrais exprimer ma profonde gratitude à Martine Boisselle et Karine Fréchette qui ont créé un environnement convivial au département et qui ne ménagent aucun effort pour trouver des solutions à nos problèmes. Je ne saurai oublier Lorraine Brisson qui bien que retraitée aura contribué à l'édification de cet environnement durant une bonne partie de mon parcours doctoral. Je voudrais également exprimer ma reconnaissance envers mes amis et collègues de doctorat qui m'ont apporté leur soutien moral et intellectuel tout au

long du programme. Un merci particulier à William Ewane et à Franck Atsiga Mballa pour leur écoute.

Enfin, je veux dire merci à mon épouse Rose et nos enfants, Mustapha Junior et Habiba Soraya, pour le soutien, et la chaleur indispensables ingrédients pour surmonter le stress au cours de ce parcours doctoral. Rose, c'est toi qui m'a encouragé à me lancer dans ce défi et tu as eu raison de me dire que je vais y arriver. Tu t'es occupée des enfants, tu as été compréhensive, patiente et tu m'as énormément soutenu durant les moments de stress ce qui m'a permis de maintenir le bon cap durant cette thèse. Rien de tout cela n'aurait été possible sans ton appui, Merci!

DEDICACE

A ma défunte grand-mère, MOUNCHILI AWA, qui est partie très tôt mais qui a continué à veiller sur nous et à ma mère, SONFACK MARIE THERESE, pour tous les sacrifices qu'elle a fait pour nous.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

RÉSUMÉ

Les techniques d'apprentissage automatique sont devenues au fil du temps de puissants outils de prévision, particulièrement utiles dans des domaines tels que la macroéconomie et la finance en raison de leur capacité à améliorer la qualité des prévisions relativement aux méthodes traditionnelles. L'importance des techniques d'apprentissage automatique dans ces domaines a été bien documentée dans des recherches antérieures (Goulet Coulombe et al. 2022, Fraisse and Laporte 2022).

Cette recherche doctorale comprend trois chapitres. Chaque chapitre examine dans quelle mesure les modèles d'apprentissage automatique combinés à une richesse d'information peuvent approfondir notre compréhension du rôle du secteur bancaire dans la stabilité macroéconomique et financière.

Le chapitre 1, « Le crédit bancaire et le cycle économique suivant la perspective des techniques d'apprentissage automatique », utilise une approche d'apprentissage automatique pour identifier les chocs d'offre de crédit bancaire dans un environnement riche en données. Dans un premier temps, nous réalisons un exercice de prévision hors échantillon afin de déterminer le modèle prévisionnel du ratio capital/actif des banquiers. Par la suite, nous construisons un choc de crédit global comme la moyenne pondérée des erreurs de prévision hors échantillon spécifiques à chaque banque. En introduisant ce choc dans un VAR, nous trouvons qu'une variation négative de notre métrique affecte inversement le volume et les prix du crédit, confirmant qu'elle identifie un choc d'offre de crédit bancaire. Ce choc impacte considérablement les variables macro-économiques clés, induisant

une baisse de la croissance du PIB, de l'inflation et une baisse prononcée du volume des crédits commerciaux et industriels.

Le chapitre 2, « Une approche d'apprentissage automatique dans les tests de stress des grandes banques américaines », évalue si les modèles d'apprentissage automatique améliorent l'analyse de risque dans les tests de stress bancaire relativement aux modèles linéaires. Les résultats mettent en évidence le double avantage des modèles d'apprentissage automatique. Indirectement, ils affinent la prévision des variables bancaires clés, notamment le revenu net avant provision (PPNR) et les charges nettes (NCO), par rapport aux modèles linéaires. Directement, les modèles d'apprentissage automatique fournissent une meilleure image des risques et de la vulnérabilité des banques en période de ralentissement économique relativement aux modèles linéaires.

Le chapitre 3, « Incertitude bancaire et cycle économique : cas de grandes banques américaines», présente une méthode pionnière d'évaluation de l'incertitude au niveau des banques. Cette métrique est dérivée d'erreurs de prévision du rendement sur actif (ROA) d'une banque, obtenues à partir d'un ensemble de modèles d'apprentissage automatique combinés à des ensembles de données bancaires et macro-économiques. L'incertitude bancaire est définie comme l'écart type de ces erreurs de prévision. En utilisant un VAR, il ressort qu'une hausse inattendue de l'incertitude bancaire entraîne un ralentissement économique et une détérioration des conditions de crédit, beaucoup plus importants que suite à une hausse similaire des indicateurs d'incertitude macro-économique et financière traditionnels.

Mots-clés : Technique d'apprentissage automatique; données massives; banque; stabilité financière; stabilité macro-économique.

INTRODUCTION

La Grande Récession de 2007-2009 a mis en exergue le rôle prépondérant des grandes banques dans l'amplification de la crise dans le secteur financier et sa propagation à toute l'économie réelle. Cette crise a renforcé la nécessité d'un plus grand contrôle du système bancaire et plus spécifiquement des grandes banques. En effet, au lendemain de la récession, plusieurs économies développées et émergentes ont rendu obligatoire et systématique des tests de stress bancaire [1]. Les tests de stress bancaire sont des exercices qui visent à vérifier si les banques sélectionnées sont assez capitalisées ou non pour faire face à des événements macroéconomiques défavorables tout en maintenant le financement de l'économie. Sur le plan international, la mise en place des accords de Bâle III a consacré l'introduction de plusieurs coussins en capital que les banques doivent détenir au délà du minimum de capital réglementaire requis. L'un de ces coussins a été fixé à 2.5% des actifs pondérés par les risques et est applicable à toutes les banques depuis 2019. Toutes ces mesures visent à mieux identifier les facteurs de risque macroéconomiques et financiers et les signes de vulnérabilité du système bancaire. Cela nécessite au préalable d'inclure globalement le secteur financier ou plus spécifiquement le secteur bancaire dans la construction des modèles macroéconomiques tel que préconisé par Quadrini (2011).

Cette thèse de doctorat qui regroupe trois chapitres, s'inscrit dans cette perspective et vise

[1] Depuis 2010, l'Autorité bancaire européenne réalise chaque année des tests de stress bancaire pour les États membres de l'Union européenne. La FED a rendu ces tests obligatoires aux États-Unis et a supervisé le processus. D'autres pays émergents, comme le Brésil, mènent également des tests de stress bancaire.

à apprécier dans quelle mesure, l'exploitation des techniques d'apprentissage automatique et la prise en compte de la richesse de l'information tant au niveau bancaire que macroéconomique peuvent permettre de mieux cerner les nuances de la relation entre le secteur bancaire et l'économie réelle dans le but de mieux comprendre le rôle du systeme bancaire dans la stabilité macroéconomique et financière.

Le premier chapitre est intitulé *Le crédit bancaire et le cycle économique suivant la perspective des techniques d'apprentissage automatique*. La crise des subprimes 2007-2009 a remis en selle la question du lien entre l'offre de crédit bancaire et l'économie dans sa globalité. En effet, le crédit bancaire aux USA a connu une contraction significative. Le taux de croissance trimestriel du volume du crédit bancaire est passé de 2% avant crise à -0.17 % durant la récession. Ce resserrement du crédit a constitué l'un des canaux de transmission de la crise financière à l'économie réelle mettant en lumière le rôle du crédit bancaire dans la régulation de l'équilibre macroéconomique. Cette question a aussi été analysée dans plusieurs études empiriques (p.ex., Bernanke et al. 1991, Bassett et al. 2014, Berrospide and Edge 2010 et Mésonnier and Stevanovic 2017). Cependant, toutes ces études afin d'identifier le choc d'offre de crédit utilisent soit uniquement les données agrégées, soit des modèles linéaires basés sur les données bancaires et agrégées limitées. Au cas où le véritable modèle serait non linéaire ou utiliserait plus d'information que fournie, cette approche prédominante dans la littérature questionne sur le risque d'omission d'information pertinente et de mauvaise spécification du modèle. Ce qui pourrait affecter l'identification de l'offre de crédit et affecter les analyses subséquentes. Pour réduire sensiblement ce risque, ce chapitre développe une approche originale d'identification du choc d'offre de crédit qui allie des informations sur le bilan des banques et des données macroéconomiques massives aux techniques d'apprentissage automatique. La méthodologie est structurée en deux parties. Dans une première partie, nous conduisons un exercice de prévision hors échantillon

utilisant différentes techniques d'apprentissage automatique et basé sur un panel des 100 plus grandes Américaines observées de 1986q3 à 2017q4. Ceci nous permet d'approximer le modèle prédictif du ratio du capital à l'actif du banquier comme celui présentant la plus faible valeur de l'erreur quadratique moyenne. Ensuite, nous définissons le choc de crédit agrégé comme la moyenne pondérée des erreurs de prévision spécifiques aux banques. Dans une deuxième partie, nous incluons le choc de crédit agrégé dans un VAR pour analyser ses effets sur le cycle économique.

Nous trouvons que suite à un choc de crédit agrégé négatif, le volume de crédit bancaire et l'écart de rendement entre les obligations BAA et les obligations de Trésor, matérialisant le prix du crédit, évoluent significativement dans des directions opposées. Ce résultat atteste que le choc de crédit agrégé identifie bien un choc d'offre de crédit bancaire. En outre, ce choc a des effets macroéconomiques importants. Il enclenche un ralentissement de la croissance du PIB, engendre l'inflation et conduit à une baisse substantielle du volume du crédit commercial et industriel. Cependant, le déclin du crédit commercial et industriel est plus prononcé et persistant que le ralentissement de la croissance du PIB.

Dans le deuxième chapitre intitulé *Une approche d'apprentissage automatique dans les tests de stress des grandes banques américaines*, nous évaluons si les techniques d'apprentissage automatique améliorent l'analyse des risques durant les tests de stress bancaire relativement aux modèles linéaires. Le choix des modèles linéaires comme référence se justifie par leur abondante utilisation dans les tests de stress bancaire (Voir par exemple, Grover and McCracken 2014, Guerrieri and Welch 2012, Hirtle et al. 2016, Liu et al. 2020). L'objectif de ce chapitre est d'apprécier si la flexibilité des techniques d'apprentissage automatique ainsi que leurs capacités à capturer différentes formes de non-linéarité et à s'accommoder de grandes bases de données, peuvent produire une meilleure caractérisation des risques et de la vulnérabilité des banques face aux chocs macroéconomiques et

financiers que les modèles linéaires. La construction des modèles de risque pour les tests de stress bancaire requiert au préalable de développer des prédictions conditionnelles de deux variables clés : le revenu net avant provision et les charges nettes [2] sous un scénario simulé de crise. Liu et al. (2020) soulignent que le développement de ces prévisions conditionnelles a pour fondation la construction des prévisions fiables sur la base des données observées des deux variables. Ainsi, nous structurons l'analyse en deux étapes. Dans une première étape, nous utilisons un panel de grandes banques dont l'actif total est supérieur en tout temps à 3 milliards de dollars, observées de 1986q3 à 2020q4 pour comparer les modèles basés sur les techniques d'apprentissage automatique aux modèles linéaires traditionnels (OLS, modèle auto-régressif, marche aléatoire). Le meilleur modèle est celui qui affiche la plus faible erreur quadratique moyenne hors échantillon.

Les résultats soulignent la supériorité du *Random Forest* et de *Adaptive Lasso* pour les prévisions de la charge nette et du revenu net avant provision respectivement. Ces résultats montrent que les techniques d'apprentissage automatique améliorent indirectement le test de stress bancaire en permettant une meilleure prévision de ces deux variables bancaires. Dans une deuxième étape, nous utilisons *Random Forest* et *Adaptive Lasso*, les deux meilleures techniques automatiques d'apprentissage pour simuler la charge nette et le revenu net avant provision sous des conditions macroéconomiques défavorables. Cet exercice permet de simuler sous les mêmes conditions le ratio de fonds propres de catégorie 1 (*T1CR*), l'un des principaux résultats du test de stress bancaire. *T1CR* se définit comme le total des fonds propres divisé par les actifs pondérés en fonction des risques. À l'instar de Covas et al. (2014), nous utilisons une approche en deux phases pour prédire ensuite

[2] Le revenu net avant provision est défini comme le ratio de la somme des revenus d'intérêt des revenus hors intérêt déduit de la somme des dépenses à l'actif trimestriel moyen. La charge nette se définit comme le ratio de la charge sur les prêts et locations déduite des remboursements sur ces prêts et locations à l'actif trimestriel moyen.

la densité de *T1CR*. Dans une première phase, nous utilisons la méthode de la régression Quantile Random Forest de Athey et al. (2019) pour estimer les quantiles conditionnels. Dans la deuxième phase, pour estimer la fonction de densité, nous estimons la distribution de student asymétrique de Azzalini and Capitanio (2003) qui permet d'approximer au mieux la fonction quantile estimée en première phase. Pour un souci de comparabilité, nous répliquons ce processus en utilisant le modèle linéaire à effets fixes pour simuler la charge nette et le revenu net avant provision et ensuite déduire la distribution de *T1CR*. Les résultats révèlent que la distribution de *T1CR* estimée par les techniques d,apprentissage automatique présente une asymétrie à gauche plus marquée pour les grandes banques à effet systémique que dans le cas linéaire. En simulant une détresse semblable à la Grande Récession, les modèles d'apprentissage automatique approximent mieux la densité de *T1CR* que le modèle linéaire. Ces résultats démontrent que les modèles d'apprentissage automatique permettent d'obtenir la meilleure image des vulnérabilités dans le système bancaire et du risque inhérent en période de ralentissement économique relativement aux modèles linéaires qui ont tendance à sous-estimer le risque systémique.

Dans le troisième chapitre intitulé *Incertitude bancaire et cycle économique aux USA*, nous introduisons une nouvelle approche de quantification de l'incertitude au niveau des banques basée sur les erreurs de prévision du rendement bancaire sur les actifs et implémentée par l'utilisation d'un ensemble de modèles d'apprentissage automatique combinés à des données bancaires granulaires et à un vaste ensemble de données macroéconomiques. En effet, le concept d'incertitude économique, défini par Watkins (1922) comme "l'incapacité des individus à prévoir les probabilités que des événements se produisent", est un sujet important dans la littérature économique depuis les travaux influents de Bloom (2009) . L'intérêt de ce sujet tire sa source de son lien avec le cycle économique, comme l'attestent la théorie et les recherches empiriques. Ces recherches empiriques mettent en lumière

systématiquement une forte augmentation de l'incertitude pendant les récessions. Ces observations sont valables aussi bien pour une incertitude mesurée par le biais des proxys (p.ex., Bloom 2009, Baker et al. 2016) que sur la base de modèles à l'instar de Jurado et al. (2015). Toutefois, l'abondante littérature sur l'incertitude se focalise majoritairement sur l'incertitude macroéconomique et occulte le rôle du secteur bancaire qui a été une plaque tournante de la Grande Récession. Ludvigson et al. (2021) montre que l'incertitude d'origine financière affecte l'activité économique contrairement à l'incertitude macroéconomique qui en est une résultante. Cependant, leur mesure d'incerti
tude est globale et non spécifique au secteur bancaire. Très peu d'études se penchent sur l'analyse de l'incertitude bancaire (par exemple, Soto 2021, Buch et al. 2015). Toutefois, ces études se limitent à l'analyse de l'impact de l'incertitude bancaire sur le secteur bancaire et n'étudient pas ses effets sur le cycle économique ou le secteur financier global.

Ce chapitre a pour but de combler ce déficit dans la littérature. L'analyse se décline en deux étapes. Premièrement, en utilisant un panel de grandes banques américaines observées de 1986q3 à 2020q4, nous construisons un ensemble de modèles d'apprentissage automatique dans l'optique de prévoir le rendement sur les actifs. Ensuite, nous construisons la moyenne des différentes prévisions afin d'obtenir une prévision moyenne. Nous utilisons cette prévision moyenne et les erreurs de prévision qui en découlent pour construire la mesure d'incertitude bancaire comme l'écart-type de ces erreurs de prévision. Cette approche présente plusieurs avantages. Elle capitalise sur divers modèles de données, atténue les risques de dépendance à un modèle, renforçant ainsi la robustesse des prévisions, et améliorant la précision des modèles individuels même dans le cas d'une instabilité des prévisions tel que démontré par Stock and Watson (2004). L'étude intègre des prédicteurs spécifiques aux banques et un vaste ensemble d'indicateurs macroéconomiques au rang de variables prédictives. Ainsi, cela purge le modèle des variations prévisibles, et garantit

que la mesure de l'incertitude incarne effectivement une véritable imprévisibilité.

Deuxièmement, nous intégrons la mesure de l'incertitude bancaire dans un VAR afin d'étudier son impact sur le secteur financier et le cycle économique. Les résultats montrent qu'une hausse de l'incertitude bancaire a d'importantes répercussions. Elle conduit à un important ralentissement économique et une détérioration des conditions de crédit. Nous trouvons aussi que les répercussions de notre mesure d'incertitude bancaire sur le cycle économique et le secteur financier sont nettement plus importantes que celles des autres mesures standard d'incertitude macroéconomique et de l'incertitude financière de Ludvigson et al. (2021). Ces résultats confirment la place centrale du secteur bancaire dans l'économie en prouvant qu'une incertitude bancaire peut générer une récession. En outre, ces résultats montrent également que l'incertitude financière qui est plus générale ne permet pas de cerner toutes les nuances du secteur bancaire, d'où l'intérêt d'une mesure d'incertitude plus spécifique au secteur bancaire.

CHAPTER I

ON BANK CREDIT AND BUSINESS CYCLE FROM A MACHINE LEARNING PERSPECTIVE

ABSTRACT

This paper proposes identifying the bank's credit supply shock using a machine learning approach in a data-rich setup. The analysis consists of two steps. First, we do a pseudo-out-of-sample forecasting exercise to approximate the bankers' predictive model of the capital-to-asset ratio. Then, we define the aggregate credit shock as the weighted average of bank-specific out-of-sample forecast errors. In the second step, we include the constructed shock in a VAR to assess its impact on the business cycle. We find that a one standard deviation negative aggregate credit shock moves credit volume and prices in the opposite direction, corroborating that it identifies a bank credit supply shock. Furthermore, this shock has significant macroeconomic effects. It triggers a slowdown in GDP growth, inflation, and a substantial commercial and industrial credit volume drop. However, the decline in commercial and industrial credit is more pronounced and persistent than the slowdown in GDP growth.

Keywords: Identification, machine learning, credit supply, shock , business cycle.
JEL classification: C53, C55, E32, E44, G01.

1.1 Introduction

The 2007-09 financial crisis underscored the intricate relationship between the real economy and the financial sector, which led to the most severe recession since 1929. Adrian and Shin (2010) highlighted the procyclicality of leverage as a key factor in accumulating risk. In economic upturns, firms amplify investments by increasing debt, elevating the leverage effect. Conversely, during downturns, the looming risk reduces this leverage. This peak exposure contributed to significant losses, especially for institutions holding devalued "toxic assets". The subsequent credit crunch intensified the crisis's impact on the real economy.

Following the crisis, Quadrini (2011) emphasized the criticality of embedding the financial sector within macroeconomic models to grasp business cycle nuances. In alignment with this, we introduce a distinctive microeconomic approach to isolate bank credit supply shock and gauge its macroeconomic implications. Initially, we use a pseudo-out-of-sample forecast exercise to mimic a banker's predictive model of the capital-to-asset ratio, evaluated by its out-of-sample mean square error (MSE)[1]. Our analysis unequivocally identifies Gradient Boosting—a nonlinear machine learning technique—as the superior forecasting model for approximating the capital-to-asset ratio. Given its predictive prowess, we construct the aggregate credit shock from the weighted average of bank-specific forecast errors related to this model. Our forecasting approach meticulously eradicates potential endogeneity from bank credit demand or inherent bank risk by integrating an extensive

[1]The capital-to-asset ratio is the ratio of equity capital to the bank's total assets. We approximate the banker's predictive model of the capital-to-asset ratio by the best forecast model of the capital-to-asset ratio. This model corresponds to the lowest out-of-sample mean square error (MSE) value. Let $y_{i,t}$, i = 1, ... n ; t=1...T be a variable of interest and $\hat{y}_{i,t}$, i = 1, ... n ; t=1...T its predicted values. MSE = $\frac{1}{n}\sum_{i,t}(y_{i,t} - \hat{y}_{i,t})^2$ where n is the test set sample size. The test set sample covers the period 2001q2-2017q4.

macroeconomic database with granular microeconomic insights. Thus, the conceived aggregate credit shock may epitomize an unanticipated variation in bank leverage, intrinsically tethered to credit supply. A positive fluctuation of this shock signifies a leverage surge, potentially prompting banks to augment their lending activities. In contrast, a negative manifestation implies a contractionary impetus, likely inducing a retrenchment in loan allocations.

To ascertain the macroeconomic repercussions of the aggregate credit shock, we embed it within a Vector Autoregression (VAR) model. This VAR encompasses two distinct blocks of variables: one portraying the real economy and the other representing the nuances of the financial and credit markets. Probing the impulse response functions, we discover that a negative variation of aggregate credit shock manifests tangible macroeconomic ramifications. Specifically, it instigates a peak decline in the annual real GDP growth rate of 1%, occurring predominantly two quarters after the shock, with its influence enduring for an entire fiscal year. Furthermore, the shock wields enduring effects on the growth trajectory of commercial and industrial credit volume. This volume experiences a maximum contraction of 3.2% annually, notably two quarters after the shock, with this negative trajectory persisting for roughly three years. In a reflexive policy response, the FED lowers its policy interest rate by a substantive 15 basis points within the same two-quarter horizon post-shock. Moreover, the yield spread between BAA[2] bonds and 10-year Treasury bonds witnesses an augmentation, peaking at 0.8% annually two quarters hence and sustaining elevated levels for over two years. Accordingly, this inverse dynamic between credit cost and volume outlines that the aggregate credit shock fluctuations translate into perturbations in credit supply, thus delineating it as a quintessential supply-side credit shock. We

[2] A rating classification by Moody's for medium-quality bonds embodying moderate risk.

implement various robustness checks to assess the strength of the results. These include substituting the FED's interest rate with a six-month Treasury bond yield, accommodating alternative lag structures of endogenous variables, and contemplating alternative proxies for economic activity. Across these varied scenarios, our core conclusions remain steadfast and compelling.

Despite our findings indicating the efficacy of our approach in identifying credit supply shocks, two pertinent questions arise. Firstly, what would be the outcome if we had constructed an aggregate capital ratio instead of bank-specific ratios, as Berrospide and Edge (2010) suggested? Addressing this, we replace the aggregate credit shock, our baseline measure by the aggregate capital ratio, within the VAR model. Surprisingly, a negative variation of this metric has no discernible effect on credit volume, reaffirming the importance of microdata in capturing credit supply shocks. Secondly, does the relationship between the capital-to-asset ratio and predictors, linear versus nonlinear, influence the determination of the aggregate credit supply shocks? To explore this, we emulate the Mésonnier and Stevanovic (2013) (MS) methodology, which employs a linear framework. When juxtaposed against our baseline, there are marked disparities in response functions. GDP and inflation show heightened responses in the linear (MS) model, whereas credit reaction is more robust in our baseline. This result underlines the criticality of recognizing the correct relational dynamics between variables. Misinterpretations could distort the perceived impact of the measure of credit supply shocks on pivotal macroeconomic and financial indices. Our agnostic approach appears favored in this context.

Our study offers two pivotal advances to existing literature. Firstly, we bridge microdata with a large macroeconomic database to derive the measure of bank credit supply shock, distinguishing our approach from prevalent trends. While most research either leans on aggregate data, like Berrospide and Edge (2010) and Bernanke et al. (1991) or marries

microdata with limited macro insights, as seen in works like Bassett et al. (2014) and Amiti et al. (2017), our method uniquely harnesses all havailable information. While Mésonnier and Stevanovic (2013) tread a similar path, their linear framework limits the depth of their analysis. Our second contribution lies in leveraging machine learning for more than mere forecasting, a trend spotted in studies such as Goulet Coulombe et al. (2022). We employ these techniques to estimate the optimal capital-to-asset ratio forecast model, enhancing the structural analysis of the credit supply shock's broader economic implications.

The rest of the paper is organized as follows. Section 1.2 explains the construction of our aggregate credit shock. Section 1.3 deals with the data. Section 1.4 presents aggregate credit shock and compares it to alternative measures. Section 1.5 presents the macroeconomic implication of the aggregate credit shock, and section 1.6 concludes.

1.2 Methodology

This section details our identification strategy and the benefit of the approach.

1.2.1 Identification strategy of bank's credit supply shock

Our identification strategy involves two steps. First, we estimate competing one-quarter ahead forecasting models of a bank's capital-to-asset ratio. Then, we compare models and derive the best forecasting model. Second, based on this model, we construct the shock.

1.2.1.1 Estimation of the forecasting models of capital-to-asset ratio

To construct a forecasting model for the capital-to-asset ratio[3], we employ both standard linear forecasting techniques (SL)[4] and machine learning techniques (ML)[5]. Let H represents the number of competing models. Formally, we present the forecasting model j (j=1...H) as follows:

$$Y_{i,t+1} = g^j(Z_{i,t}) + e_{i,t+1}, \tag{1.1}$$

where $Y_{i,t+1}$ is the value of the capital-to-asset ratio for bank i at quarter $t+1$ and $Z_{i,t}$ is the $K \times 1$ vector of predictors. This vector includes lagged dependant variables, bank balance sheet data, macroeconomic expectations, and the McCracken and Ng (2020) database of macroeconomic and financial series for ML models. For SL models, $Z_{i,t}$ is refined to a limited set of variables. Based on the data at quarter t, we postulate that it is sufficient for bankers to forecast the subsequent quarter's capital-to-asset ratio. This perspective is validated by existing literature such as Bassett et al. (2014) and Mésonnier and Stevanovic (2017). From our estimation, we determine:

$$\hat{Y}^j_{i,t+1} = \hat{g}^j(Z_{i,t}); \tag{1.2}$$

---

[3]We adopt the capital-to-asset ratio as our measure of capital due to its intrinsic link with leverage. The banker constrained by the new requirements of Basle III should predict the capital-to-asset ratio to ascertain its conformity with the regulation.

[4]For SL techniques, we use the random walk without drift, autoregressive model, and pooled OLS.

[5]Under ML techniques, we consider Lasso, Adaptive Lasso, Elastic net, Ridge, Gradient boosting, Random Forest, and Neural Network Regression (NN).

where $\hat{Y}^j_{i,t+1}$ symbolizes the forecast of bank i's capital-to-asset ratio for quarter $t+1$ using the forecast model j. The distinctive trait of ML techniques is their capacity to estimate $\hat{g}^j$, aiming to minimize the out-of-sample Mean Square Error (MSE) while simultaneously conducting regularization processes. Regularization's primary merits, steered by hyperparameter selection, are its ability to curtail data overfitting[6] and reduce model complexity. The latter often leads to forecasts that surpass those produced by traditional techniques due to practical bias-variance trade-offs. In the following subsection, we provide the method we chose for optimal hyperparameter choice.

1.2.1.2 Choice of hyperparameters

Hyperparameter selection is foundational to the efficacy of machine learning (ML) techniques. We pinpoint optimal hyperparameters using K-fold cross-validation and ensure compatibility with our panel data structure by cross-validating across both time series and cross-sectional dimensions. To elucidate, for a specified model and a given hyperparameter combination, we segment a subsample into k groups, each of roughly equal size. Adopting group j ($1 \leqslant j \leqslant k$) as our validation set, we train the model on the remaining $k-1$ groups. We then generate predictions for the capital-to-asset ratio within group j, allowing us to compute the associated mean square error (MSE^j):

$$MSE^j = \frac{1}{N} \sum_{i,t} (Y_{i,t+1} - \hat{Y}^j_{i,t+1})^2.$$

[6]Data overfitting emerges when the function $\hat{g}^j$ estimated by the forecasting technique mirrors the target variable almost flawlessly, erasing in-sample residual variability. This results in a deceptively low in-sample MSE that may hide subpar out-of-sample prediction performance.

Here, N denotes the number of banks present in the validation set, while i and t represent the cross-sectional and temporal dimensions, respectively. Reiterating this exercise for all k folds, we derive an aggregate forecasting metric for the stipulated hyperparameters:

$$MSE = \frac{1}{N} \sum_{1}^{k} MSE^{l}.$$

This procedure is replicated for every conceivable hyperparameter combination. The hyperparameters producing the minimal MSE are deemed optimal. We subsequently train the model on the entire subsample using these optimal hyperparameters. It is paramount to highlight that the estimation procedure for standard linear (SL) models differs markedly from their ML counterparts.

1.2.1.3 SL models specificity

Specifically, the SL models we estimate are defined as:

$$Y_{i,t+h} = \alpha + \rho Y_{i,t} + \epsilon_{i,t+h} \,, \tag{1.3}$$

$$Y_{i,t+h} = Y_{i,t} + \epsilon_{i,t+h} \,, \tag{1.4}$$

$$Y_{i,t+h} = \alpha + \rho Y_{i,t} + Z'_{i,t}\beta + \epsilon_{i,t+h}, \tag{1.5}$$

where Equation (1.3) denotes the autoregressive model, Equation (1.4) represents the random walk without drift, and Equation (1.5) represents the pooled OLS. Unlike ML models, $Z_{i,t}$ the regressor vector comprises a limited set of variables, including microeconomic pre-

dictors, macroeconomic expectations, and a limited number of macroeconomic variables. Parameters α, ρ and β are estimated by OLS. We choose random walk without drift as the benchmark model to check if it is valuable to forecast the capital-to-asset ratio.

1.2.1.4 Pseudo out-of-sample forecasting exercise

We choose the expanding window estimation technique. Then, we split the data sample initially into two: the estimation and validation set, which extends from 1986q3 to 2000q1, and the test set, which covers 2000q2-2017q4. After choosing hyperparameters, we estimate the model in the first subsample and predict the capital-to-asset ratio one quarter ahead. We run the process iteratively by expanding the estimation and validation set by one quarter until we have predicted the dependent variable for the last quarter of the test set sample 2017q4. This methodology is analogously applied when constructing linear standard forecast models. After model development, we embark on a comparative analysis. We first evaluate their forecasting performance by computing their relative mean square error. Let us denote by MSE^j the out-of-sample mean square error of model j, MSE the out-of-sample mean square error of the benchmark model, and $RMSE^j$ the relative mean square error of model j, we can formulate :

$$RMSE^j = \frac{MSE^j}{MSE}.$$

The best forecasting model is the one with the lowest $RMSE$. To be confident that the superiority of the best model is not due to pure chance, we run the Diebold Mariano test (DM) between that model and the others individually.

1.2.1.5 Construction of the shock

Let us denote by $\hat{Y}_{i,t+1}^{opt}$ the prediction of the capital-to-asset ratio one quarter ahead relative to the best forecasting model. From (1.1) and (1.2), we first derive forecast error, our bank-specific shock, $\hat{e}_{i,t+1}$ for each quarter of the test set sample as:

$$\hat{e}_{i,t+1} \quad = Y_{i,t+1} - \hat{Y}_{i,t+1}^{opt}. \tag{1.6}$$

A positive value of $\hat{e}_{i,t+1}$ means that a bank's capital-to-asset ratio is higher than predicted during that quarter. A negative value reflects a bank capital-to-asset ratio lower than anticipated. Second, we determine a measure of credit supply shock by averaging $\hat{e}_{i,t}$ depending on the size of the bank[7]. Concretely, if we denote by e_t the aggregate credit shock at quarter t, we can write:

$$e_t = \sum_{i=1}^{N_t'} w_{i,t-1}\hat{e}_{i,t} \ , \tag{1.7}$$

where $w_{i,t-1}$ is the weight of bank i in $t-1$ in relation with the total assets of the $N_t' = min(N_{t-1}, N_t)$ banks present in the sample at $t-1$ and at t. The weighted average of bank-specific shock computed in (1.7) is robust to the idiosyncratic movement of some banks. It also integrates by construction the difference in the size of banks in our sample and their differential impact on credit supply shock. Since our measure of shock originates from the out-of-sample exercise, we compute it for the period 2000q2-2017q4, corresponding to the test set sample. By construction, aggregate credit shock is a shock on the bank's leverage. A positive value of e_t is a positive leverage shock that could bring banks to expand loans.

---

[7] Mesonnier and Stevanovic (2013) used this shock aggregation strategy.

A negative value, on the contrary, constrains the bank's capacity to invest and grant credit, resulting in credit restriction. There are many potential drivers of this shock. It could be driven, for example, by an unanticipated adverse event hitting the bank's solvency, like the assets price bubble collapse in the recent Great Recession. It could also come from surprise government actions affecting banks' leverage. For instance, The US rescue plan during the Great Recession resulted in the massive injection of capital into large systemic banks. As we control for extensive macroeconomic information and bank balance sheet characteristics and remove the anticipated component, it is reasonable to suggest that the measure is unrelated to credit demand and risk factors. Romer and Romer (2004) have used the same strategy to construct the monetary policy shock measure. They removed endogenous and anticipatory movements from the federal funds rate to derive a measure of monetary shock.

To enlighten the originality of our identification strategy, we need to clarify how this approach may improve the identification of bank supply credit shock.

1.2.2 How our methodological approach could improve the identification of bank credit supply shock?

We can state the true unknown forecast model of bank capital-to-asset ratio in the population as in Equation (1.8) :

$$Y_{i,t+1} = g^*() + \epsilon_{i,t+1} \, . \tag{1.8}$$

$$Y_{i,t+1} = g() + e_{i,t+1} \tag{1.9}$$

$$Y_{i,t+1} \quad = \quad \hat{g}() + \hat{e}_{i,t+1} \, . \tag{1.10}$$

Based on the theory, the literature, and intuition, the econometrician sets the specification in Equation (1.9) and estimates it in Equation (1.10). From Equations (1.8), (1.9), and (1.10), we can decompose the forecast error as follows:

$$Y_{i,t+h} - \hat{Y}_{i,t+h} \quad = \quad \underbrace{g^*() - g()}_{approximation\ error} \quad + \quad \underbrace{g() - \hat{g}()}_{estimation\ error} \quad + \epsilon_{t+h} \tag{1.11}$$

Our methodological approach by exploring various flexible forecasting techniques allows capturing the model $(g^*())$ whether linear or complex, reducing the approximation error in (1.8) and better approximating the banker systematic response $(g^*())$. Furthermore, considering all potential micro and macro predictors reduces the estimation error in (1.8). Therefore, our approach leads to better identification of the credit supply shock. Identifying the shock with micro and macro data is the core of two papers close to ours: Mésonnier and Stevanovic (2017) and Mésonnier and Stevanovic (2013). Mésonnier and Stevanovic (2017) take advantage of panels of large US BHC to construct a measure of bank credit supply as the residue of the regression with the capital-to-asset ratio as the dependent variable. They specify g as a linear function of limited micro and macro predictors, rendering their model too restrictive. If the suitable model is complex, their approach will not approximate properly $(g^*())$ and, ultimately, the credit measure. In addition, if the accurate model is not sparse, their method will entail estimation error and biased credit shock measure. Although built with a massive macroeconomic database, Mésonnier and Stevanovic's (2013) model is restrictive as it assumes a linear form for g(). They construct the shock similarly to our paper at the micro-level. However, by not exploring other functional forms, their shock could be affected by misspecification and, therefore, needing to

be correctly specified.

1.3 Data

This paper is based on an update and extension to the panel of the 100 largest American banks holding companies[8], initially constructed by Mésonnier and Stevanovic (2013). Our sample now spans from 1986q3 to 2017q4. As depicted in the top panel of Figure 1.1, there is a pronounced co-movement between the total assets of all banks and those of the banks in our sample. This synchronization can likely be attributed to the dominant size and influence of the banks we consider, which, on average, represent 65% of all banks' total assets, with surges reaching up to 70%. Intriguingly, there is a noticeable decline in the total assets and the proportionate weight of the selected banks in 2006. This anomaly can be traced back to a landmark regulatory shift in the first quarter of 2006. With the modification in the law concerning bank holding companies, all subsidiaries boasting assets above 1 billion USD transitioned to full-fledged bank holding companies. This regulatory move precipitated an average of 30% contractions in the impacted banks' cumulative assets during that year. Given our analytical approach, which combines machine learning techniques tailored for panel data with time series methods, we use two distinct datasets: bank-specific records and broader aggregate data.

[8]For detailed bank listings, refer to Appendix D. In the text, we will interchange the terms bank holding company (BHC) and bank. We focus on BHC instead of commercial banks since the leverage decision is taken at the headquarters level and affects all subsidiaries.

1.3.1 Bank specific information

We have extracted specific data from the Consolidated Financial Report of the Chicago FED (FRY-C9 form) to comprehensively analyze the microeconomic characteristics impacting a bank's capital-to-asset ratio. Building on the framework of Mésonnier and Stevanovic (2013), we have structured our dataset around several key variables that influence the capital-to-asset ratio:

One-quarter lagged value of capital-to-asset ratio ($capitalratio1$) gauges the inertia inherent in the capital-to-asset ratio.

Bank size ($size$) defined by the logarithm of the bank's total assets offers insights into the bank's prominence.

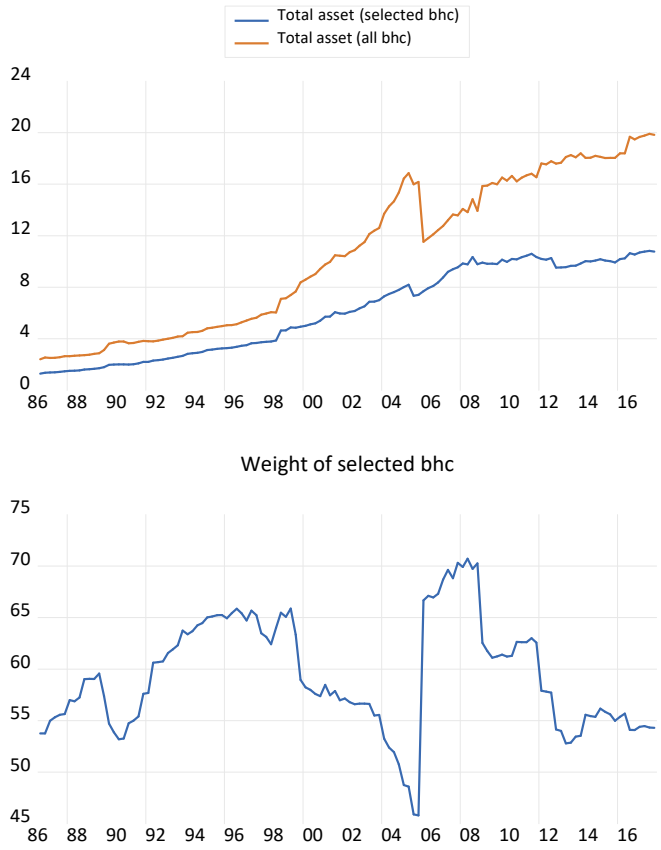
Return on Assets (ROA), calculated as the ratio of the bank's net income to its total assets, serves as a barometer for the bank's profitability and indicates the net profit per monetary unit of total assets.

Net Charge Off (NCO) represents the value derived from the charge-off on allowance for loans and lease losses minus recoveries, divided by total assets. It acts as a reflection of the banker's perception of risk tied to his commitments.

The Real Estate Loan Ratio ($REloan$) measures the proportion of mortgage loans to total assets, and the **Commercial and Industrial Credit Loan Ratio** ($CIloan$) calculates the ratio of commercial and industrial credit loan volume to total assets. Supplementing these variables, we have incorporated two dummy variables — $merge$ and $status$ — sourced from Mésonnier and Stevanovic's (2013) database and the FRED Chicago website. The $merge$ variable indicates whether a bank undergoes merging in a given quarter, while $status$ provides insights into any shifts in the bank's official status. Table 1.1 offers a detailed breakdown of the descriptive statistics tied to the dependent and the variables

mentioned above.[9] Notably, except for $size$, all variables display pronounced heterogeneity, evident from the ratio of their standard deviations relative to their means.

Figure 1.1 Evolution over time of total assets of banks



Note : Total assets is expressed in billions USD.

[9] Appendix C describes the variable computation and highlights the relationship between the capital-to-asset-ratio and other important accounting items.

Table 1.1 Summary statistics for some bank specific information

	mean	sd	p25	p50	p75
Capital ratio	8.54	5.87	6.63	7.77	9.15
size	17.18	1.43	16.08	16.92	17.97
NCO	1.17	1.53	0.28	0.67	1.45
CIloan	16.00	8.24	10.90	15.75	20.62
REloan	26.34	13.26	18.22	26.20	33.65
ROA	2.66	3.77	1.15	2.31	3.76
merge	0.03	0.16	0.00	0.00	0.00
status	0.12	0.32	0.00	0.00	0.00

1.3.2 Aggregate data

We consider all the macroeconomic and financial series of the McCracken and Ng (2020) database. We can classify these 250-time series into five categories, namely: (1) indicators of economic activity, (2) indicators of inflation, (3) interest rates, (4) indicators related to the credit and money market, and (5) other financial indicators and asset prices. We render all these series stationary before any operation according to the procedure described in the database. We furthermore include as potential predictors of *capital ratio* two variables reflecting macroeconomic expectation: (1) *expectation of the growth rate of the real economy* and (2) *expectation of change in monetary policy*. We measure the former by the average value of GDP growth rate expectations over one year. In contrast, we compute the latter as the average expected change over a one-year horizon of the interest rate on Treasury bonds with a three-month maturity. We derive the data on these two expectation variables from the statistical survey of forecasters conducted by the Federal Reserve of Philadelphia.

1.4 Measure of bank's aggregate credit shock

1.4.1 Performance of different capital-to-asset ratio forecasting models

Slightly more than a third of the models for predicting the bank's capital-to-asset ratio have predictive powers superior to the random walk (Table 1.2). The ML predictions' mean and median improve the random walk's predictive power by 3% and 6%, respectively. They are, in turn, dominated by Gradient Boosting. The latter improves the predictive capacity of the random walk by 12% and stands out as the best forecasting model. Analysis of the prediction performance according to the size of the bank[10] shows that the Gradient boosting remains the best model for forecasting the capital-to-asset ratio for the smallest and largest banks (group 1 and group 4). Regarding banks of intermediate size (group 2 and group 3), no model can better predict the bank's capital-to-asset ratio than the random walk and the Gradient Boosting. These two models display the same forecasting performance for these groups of banks. These facts can be observed in examining Figure 1.2. Indeed, Gradient Boosting ensures a suitable adjustment of the capital-to-asset ratio for four banks belonging to the four respective groups. Table 1.3 presents the results of the DM test between the Gradient Boosting and each of the competing forecast models of capital-to-asset ratio. The entries in this table confirm Gradient Boosting's forecast superiority relative to other models in forecasting capital-to-asset ratios.

---

[10] We divide the banks into four groups defined by the three quartiles of the average total assets of a bank on the test set. The quartiles are Q1=1.72e+07, Q2=4.93e+07, and Q3=1.95e+08; the unit is a thousand US dollars. The first group comprises banks whose average total assets on the test set are less than Q1. The second group includes banks whose average total assets are greater than or equal to Q1 and less than Q2; the third group includes banks whose average total assets are greater than or equal to Q2 and less than Q3. The last group consists of banks whose average total assets exceed Q3. $RMSE_i$, i=1...4 provides information on the mean square error of the different models for the banks of group i. Appendix F presents the list of banks by size.

Table 1.2 RMSE of different forecasting models

Model	RMSE	$RMSE_1$	$RMSE_2$	$RMSE_3$	$RMSE_4$
Random walk	1	1	1	0.33	0.26
Model of Mesonnier and Stevanovic(2013)	0.94	0.79	1.03	1.09	1.00
Pooled OLS	3.27	2.95	2.81	3.39	4.08
Lasso	1.00	0.84	1.03	1.12	1.11
Adaptive Lasso	0.94	0.82	**1.00**	1.09	1.00
Elastic net	1.09	0.84	1.13	1.18	1.28
Ridge	1.33	1.21	1.32	1.30	1.54
PCA	2.45	2.63	2.19	2.12	3.04
Auto regressive model	0.94	0.79	**1.00**	1.06	1.00
Random Forest	0.97	0.84	**1.00**	1.03	1.04
Gradient Boosting	**0.88**	**0.71**	**1.00**	**1.00**	**0.92**
NN1	1.51	1.32	1.58	1.58	1.81
NN2	1.58	1.50	1.74	1.21	1,69
NN3	1.27	1.03	1.42	1.42	1.50
NN4	1.30	1.05	1.55	1,42	1.54
Forecasts Average	0.97	0.81	1.03	1.06	1.04
Median forecasts	0.94	0.79	**1,00**	1.03	1.00

Note: The term in bold relates to the best model. The first column provides the MSE of the benchmark model i.e the random walk without drift normalized to 1. The other columns denotes the RMSE of models relative to the benchmark. NN1 is the neural network with one layers and 32 neurons. NN2 is the neural network with two layers containing respectively 32 and 16 layers. NN3 represents the neural network with three layers having respectively 32, 16 and 8 neurons. Last, NN4 denotes the neural network with four layers containing repectively 32, 16, 8 and 4 neurons. Average forecasts and Median forecasts denote respectively the average and median of ML forecasts.

Figure 1.2 Predicted and factual values of the capital-to-asset ratio for four banks belonging respectively to the four groups

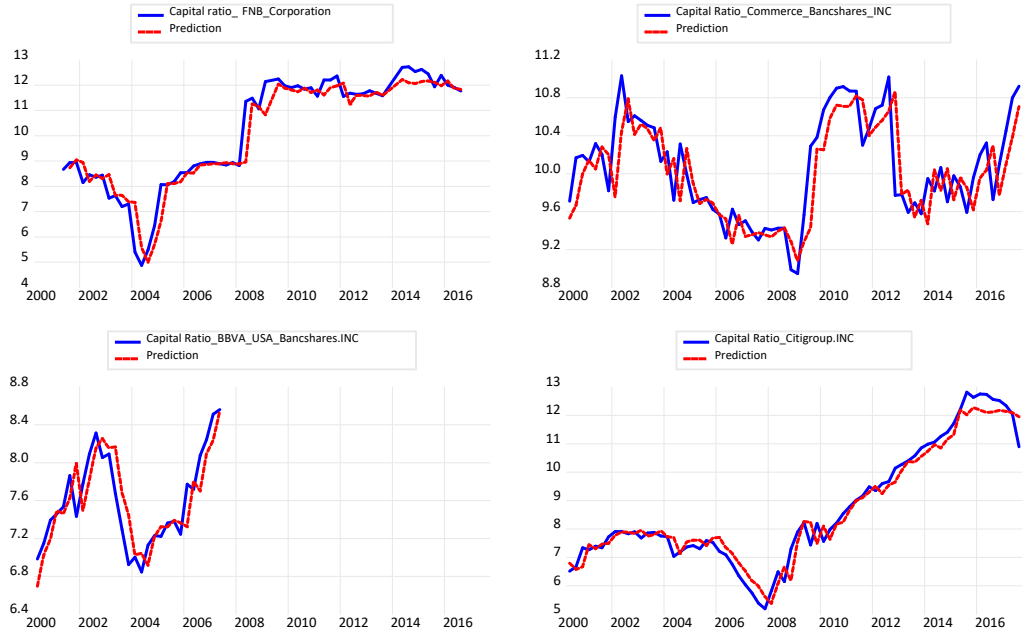


Table 1.3 Result of DM test

Model	Test statistics	p-value	significance
Random walk	-1.28	0.09	*
Model of Mésonnier and Stevanovic (2013)	-1.97	0.02	**
Pooled OLS	-18.93	0.00	***
Lasso	-3.90	0.00	***
Adaptive Lasso	-2.15	0.01	**
Elastic net	-6.97	0.00	***
Ridge	-11.56	0.00	***
PCA	-16.93	0.00	***
Autoregressive model	-1.83	0.03	**
Random Forest	-2.29	0.01	**
NN1	-15.47	0.00	***
NN2	-16.72	0.00	***
NN3	-10.57	0.00	***
NN4	-11.18	0.00	***
Forecasts average	-2.93	0.00	***
Forecasts median	-1.69	0.04	**

Note: NN1 is the neural network with one layers and 32 neurons. NN2 is the neural network with two layers containing respectively 32 and 16 layers. NN3 represents the neural network with three layers having respectively 32, 16 and 8 neurons. Last, NN4 denotes the neural network with four layers containing repectively 32, 16, 8 and 4 neurons. Average forecasts and Median forecasts denote respectively the average and median of ML forecasts. * significant to 10%; **significant to 5%; *** significant to 1%.

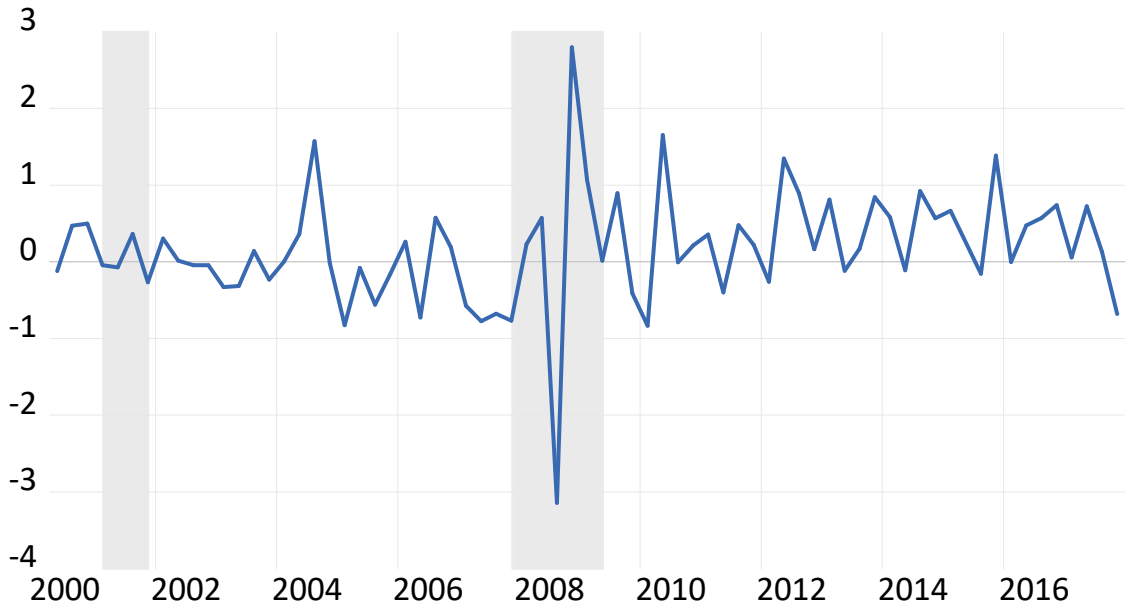
1.4.2 The cyclical behavior of the aggregate credit shock series

We derive the subsequent series of the aggregate credit shocks based on the best capital-to-asset ratio forecasting model, the Gradient Boosting. This measure has a nearly zero mean and a standard deviation of 0.22. Figure 1.3 plots the variation of the measure over time. The measure has approximately zero mean and a standard deviation of 0.72, which unveils some variability. The most pronounced adverse shocks materialize in 2006q2 and span from 2007q2 to 2007q4, immediately preceding the Great Recession. Another notable downturn occurred in 2008q3, during the recession itself,[11] followed by 2010q1. In contrast, significant positive shocks were observed in 2004q3, 2008q4, 2009q1 and 2009q2. The U.S. government's bailout initiative during the Great Recession, which prioritized the

[11] Recession periods are demarcated in gray on the Figure.

recapitalization of systemic banks, may elucidate the pronounced positive shifts in the aggregate credit shock in late 2008 and early 2009. The aggregate shock displays a negative skewness of -0.17 and a positive and significant kurtosis of 9.43, which suggests more extreme negative values than positive ones. These substantial adverse shocks, especially those manifesting just before or during the onset of the Great Recession, underscore the instrumental role of the bank credit channel in transmitting the repercussions of the subprime financial crisis to the broader economy.

Figure 1.3 Fluctuation of the aggregate credit shock



1.4.3 Aggregate credit shock and other measures of bank credit supply

We critically assess four prominent credit supply indicators as described in extant literature. Our foremost indicator is the **risk premium on corporate bonds** (*prime*), as elaborated by Gilchrist and Zakrajšek (2012). This metric gauges the appeal of corporate bonds to investors.[12] A surge in the value of *prime* implies worsening credit conditions, while a diminishing value indicates a relaxation of financial constraints.

The second indicator, the **Net Percentage Tightening** (*Npt*), represents the net percentage of banks tightening credit conditions. The FED quarterly survey, which involves a representative sample of large banks, inquires about changes in their lending standards. Specifically, banks report any tightening of credit conditions from the preceding quarter. This metric calculates the difference between the percentages of banks that tightened lending conditions and those that eased them. An increase in the measure could imply a decline in credit supply, while a decrease would mean an increase.

The **aggregate Capital Ratio** is defined as the ratio of aggregate capital to aggregate assets, as characterized by Berrospide and Edge (2010).

Finally, the **Mésonnier and Stevanovic (2013) shock** (*MS shock*) is an adaptation to forecasting of their methodological approach[13]. Their original work constructs a shock inferentially by estimating a panel regression model. The model's dependent variable is the bank's capital-to-assets ratio, and the explanatory variables combine microeconomic data and extensive macroeconomic variables. They aggregate the residuals from this model to

---

[12] The risk premium for a bond is derived from linear regression, with the yield spread of a corporate bond as the dependent variable and the bond's attributes and the associated firm's characteristics as explanatory variables. The overarching aggregate risk premium is ascertained by computing the quarterly arithmetic mean of the risk premiums across diverse corporate bonds.

[13] We provide details on the Mésonnier and Stevanovic (2013) methodology in appendix B.

create the aggregate shock.

By their constructions, the aggregate capital ratio and *MS shock* are shocks to the leverage effect of banks. A positive change of one standard deviation of one of these shocks should give banks more leeway and encourage them to increase the volume of loans. A negative variation should have the opposite effect.

Figure 1.4 draws the aggregate credit shock alongside the above indicators. It shows that the lagged values of *capitalratio* and the contemporary *risk premium* values are negatively correlated[14]. This same observation emerges from the graphical comparison between the aggregate capital ratio and *Npt*. Figure 1.5 illustrates a positive correlation between aggregate credit shock and MS shock and between aggregate credit shock and capital ratio. Statistics confirm this graphical trend by placing the correlation between aggregate credit shock and *MS shock* at 0.77 and then between aggregate credit shock and the aggregate capital ratio at 0.24.

1.5 Macroeconomic implications

1.5.1 Estimation and identification of the VAR model

We estimate a Vector Autoregression (VAR), which includes seven endogenous variables[15]. The real GDP growth rate, denoted as *GDP*, is derived from the log difference in Real GDP across successive quarters. The inflation rate, termed *Inflation*, is computed from

[14]The aggregate credit shock is contemporanously correlated to all other measures of credit supply. However, the correlation is strong with MS Schock (0.72) and weak for Npt and aggregate capital ratio.

[15]For a more in-depth exposition of the VAR modeling, recursive identification method, descriptive statistics, and visual depictions of the variables discussed, refer to Appendix E. The only shortcoming of this approach is that VAR is linear, whereas the shock is built upon nonlinear models.

Figure 1.4 Aggregate credit shock, Risk premium and Npt

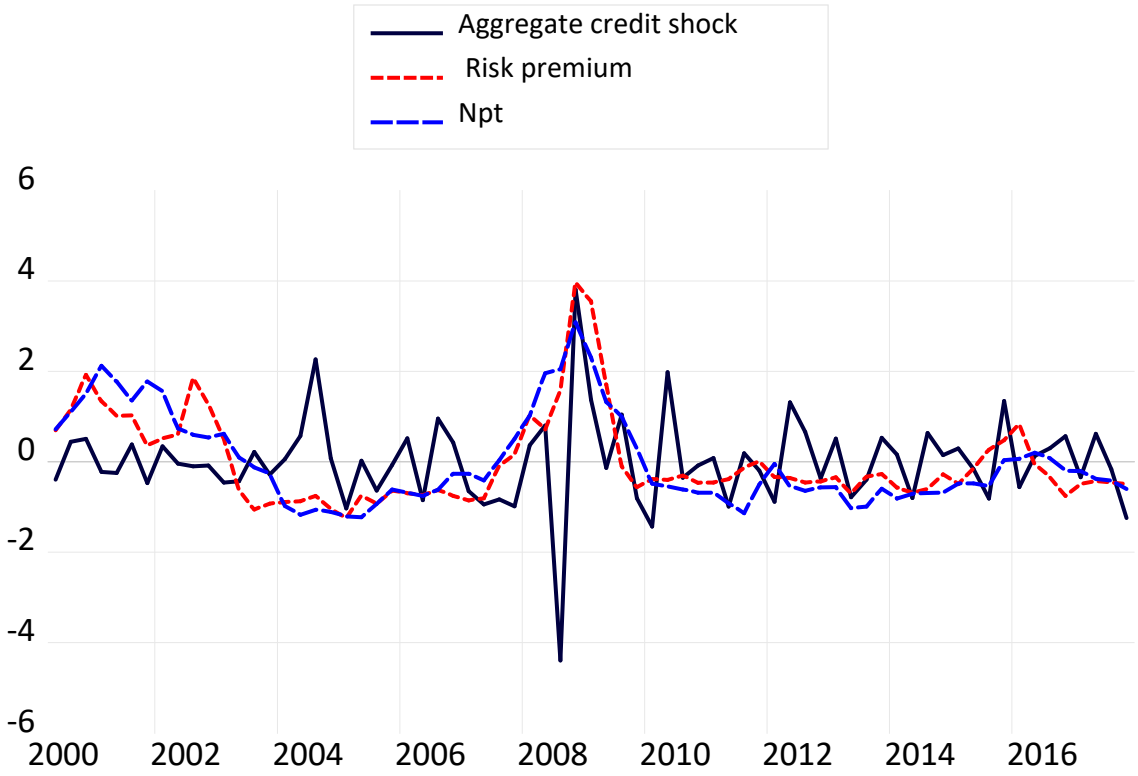
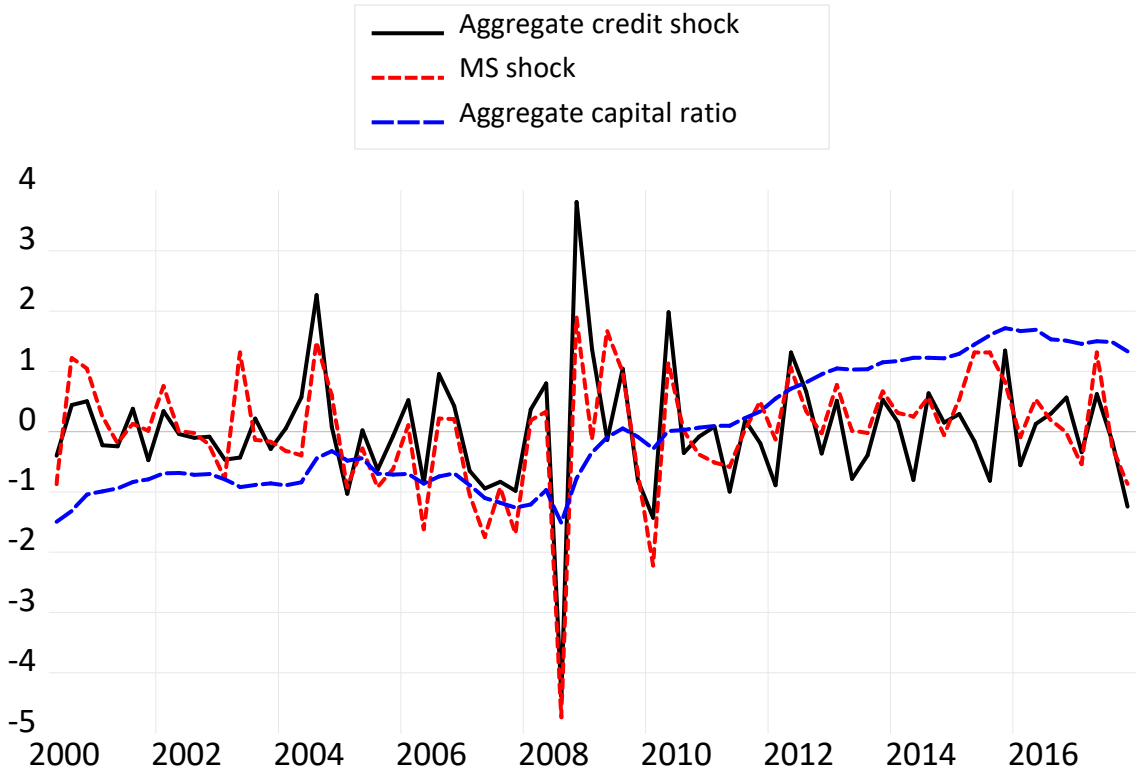


Figure 1.5 Aggregate credit shock, MS shock and aggregate capital ratio



the log difference in the Consumer Price Index (CPI) between two consecutive quarters. The FED policy interest rate, represented as *TFED*, captures the FED's monetary stance. The growth rate of credit volume, denoted as *Volume of Credit*, is obtained from the log difference in the volume of commercial and industrial loans extended by banks over adjacent quarters. The aggregate credit shock, labeled as *Shock*, captures unforeseen economic disturbances as previously constructed. The corporate bond risk premium, referenced by Gilchrist and Zakrajšek (2012) and termed *Risk Premium*, signifies the additional return investors demand for holding corporate bonds relative to risk-free assets. Lastly, the *yield gap* measures the differential between returns on BAA corporate bonds and 10-year Treasury bonds and indicates credit market conditions and investor sentiment.

We choose one as the optimal lag in estimating the VAR using the BIC criteria. Consistent with the approaches of Berrospide and Edge (2010), Bassett et al. (2014), and Mésonnier and Stevanovic (2017), we integrate a set of financial variables into the VAR model, supplementing those related to the economic cycle. At the macroeconomic level, we identify the bank credit supply shock as a variation equivalent to one standard deviation of *Shock* orthogonal to the other variables within the VAR framework. We employ the conventional recursive method based on the order, as mentioned above, of endogenous variables for this identification. As part of our approach, we postulate that *GDP* and *Inflation* remain unaffected by the immediate impact of *Shock*. This assumption mirrors the method Berrospide and Edge (2010) used, which, when extracting a credit supply shock—assumed to reflect a change in the aggregate capital ratio to banks' total aggregate assets—also enforced a restriction barring macroeconomic variables from responding to the concurrent innovation. This assumption holds water as the modification of investment and production plans transpires over an extended period, making instantaneous adjustments within a single quarter implausible. Additionally, our restrictions rule out the possibility of *TFED* reacting

instantaneously to *Shock*. This constraint is anchored in the notion that the FED's monetary policy adheres to a Taylor reaction function[16], with response exclusively dedicated to shifts in inflation and the output gap. Furthermore, even if central banks occasionally engage in financial stability matters, monetary authorities' delayed awareness of banking market trends does not invalidate our restriction. *Yield gap* and *Risk premium* undergo immediate shifts due to oscillations in other market variables within the VAR. The agile nature of financial markets in reacting to macroeconomic shifts and the frequent availability of economic data (often daily) accounts for the swift adjustments in these two financial metrics. Furthermore, we evaluate uncertainty around the point estimate of the response function of endogenous variables to shock on aggregate credit by constructing a confident interval based on the bootstrap method.

1.5.2 Effects of a negative variation of one standard deviation of the aggregate credit shock

Figure 1.6 below illustrates the macroeconomic variables' responses to an unexpected negative variation of one standard deviation to *Shock*. This fluctuation immediately triggers a decrease of 0.64 points in *Shock*, which then increases gradually to return to its long-term level after four quarters. *GDP* declines significantly to a maximum of 0.25% quarterly (i.e., 1% on an annual basis) two quarters after variation of *Shock*. It remains at a level below its initial level for a year. Gilchrist and Zakrajšek (2012) also find a rapid response of a slightly higher amplitude of *GDP* (0.5% quarterly) to a financial shock to the bond

[16] *A Taylor reaction function*, introduced by economist John Taylor in 1993, delineates a monetary policy rule correlating the central bank's benchmark interest rate with inflation rates and the output gap: $i_t = \pi_t + \rho + \theta_\pi(\pi_t - \hat{\pi}_t) + \theta_y(y_t - \hat{y}_t)$. Here, i_t represents the key interest rate at time t, π_t stands for the inflation rate, $\hat{\pi}_t$ is the central bank's targeted inflation rate, ρ denotes the long-term real interest rate, and y_t and $\hat{y}_t$ signify the actual and potential GDP at time t, respectively.

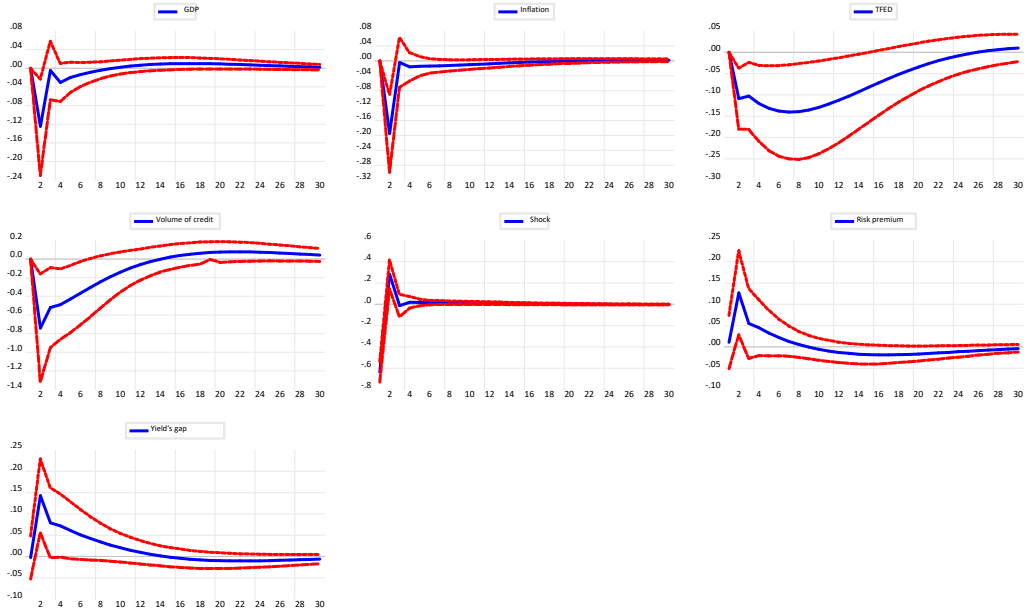
risk premium identified as a credit shock.

Volume of Credit shrinks one quarter later and reaches its lowest value in the second quarter, i.e., a quarterly decrease of 0.8 %. Unlike *GDP*, the shock has lasting effects on *Volume of Credit*, which remains below its long-term level for about 12 quarters. Mésonnier and Stevanovic (2017) and Berrospide and Edge (2010) find similar results regarding the magnitude and the persistence of a financial shock on the growth rate of the volume of credit.

Inflation response is negative and short because it only spreads over four quarters. In the second quarter, *Inflation* reaches its minimum value following the fluctuation in *Shock*, i.e., a maximum quarterly decline of 0.32%. An easing of monetary policy results in a ten basis point reduction in *TFED* for two quarters.

Risk premium and *Yield gap* significantly increase after *Shock* fluctuation and register their most significant increases two quarters after (1% maximum). Both indicators remain above their respective long-term values for more than two years. The dichotomous responses of credit quantity (*Volume of Credit*) and credit pricing metrics *(Risk premium* and *Yield gap*) in the wake of *Shock* perturbation suggest its characterization as a credit supply shock.

Figure 1.6 Response functions of macroeconomic variables to a negative variation of one standard deviation of *Shock*



Note: Confidence intervals are constructed for a threshold of 90%.

1.5.3 Transmission mechanisms of the aggregate credit shock

In the prior segment, we discerned the pronounced implications of *Shock* variation across all VAR components. Now, the focal inquiry shifts to uncover the transmission conduit of *Shock* to *GDP*.[17] To unravel this mechanism, we estimate the reduced equation specific to *GDP*. The outcomes of this estimation find a place in Table 1.4. A salient observation from Table 1.4 reveals that the antecedent value of *Shock* remains inconsequential

[17] *GDP* is a predominant variable in the economic cycle. We can, therefore, restrict the question of the transmission of *Shock* to this variable.

in shaping *GDP*, which emphasizes the non-existence of a direct transmission pathway of *Shock* to *GDP*. Further dissection indicates that *Volume of credit* adjustment mechanism is not the medium of *Shock* transmission, given its negligible influence on *GDP*. The spotlight thus shifts to *Risk premium*, emerging as the solitary variable that bears significant weight in explicating *GDP* variations. Consequently, it stands out as the exclusive channel that ferries the implications of *Shock* to *GDP*. Drilling deeper into this finding, a negative perturbation in *Shock* incites a reflexive action from banks, manifested as a curtailment in their engagement with corporate bonds. Given the magnitude and stature of these financial institutions, such a move invariably reverberates through the financial market's credit supply. As the credit reservoir recedes, it ushers in a spike in the risk premium. This surge, in turn, casts shadows on investment and consumption appetites, culminating in a downtrend in *GDP*.

Table 1.4 Estimate of the reduced *GDP* equation

	(1) GDP_t
GDP_t-1	0.0142 (0.1272)
Inflation_t-1	-0.0372 (0.1655)
TFED_t-1	0.0225 (0.0723)
Volume of credit_t-1	0.0109 (0.0144)
Shock_t-1	0.1852 (0.1744)
Risk premium_t-1	-0.5819* (0.3104)
Yield's gap_t-1	0.1545 (0.3331)
_cons	0.0193 (1.0317)
N	70
R^2	0.364
F	3.7270

Standard errors in parentheses
$^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$

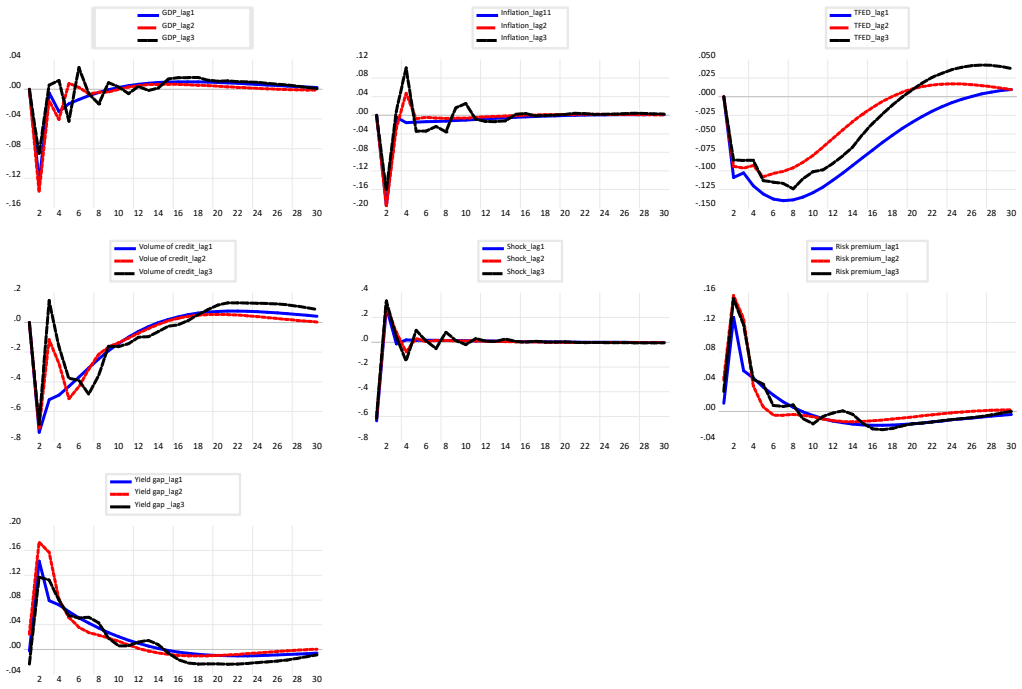
1.5.4 Robustness analysis

We evaluate the robustness of the VAR analysis according to three criteria: (1) the successive use of two lags and three lags in the estimation of the VAR, (2) the use of the growth rate of the industrial production index as an alternative measure of economic activity replacing *GDP* , (3) the replacement of *TFED* with the interest rate on 6-month Treasury bonds of maturity (*TB6M*).

1.5.4.1 Estimation of the VAR model with additional lags of endogenous variables

Our analysis involves a sequential estimation of the VAR by integrating 2 and 3 lags of the endogenous variables. As evidenced in Figure 1.7, the response functions arising from these two specifications are congruent in sign and pattern, with those emanating from the baseline model anchored on a single lag. Such consistency reinforces the resilience and robustness of our VAR estimation against alterations in the lag selection for endogenous variables.

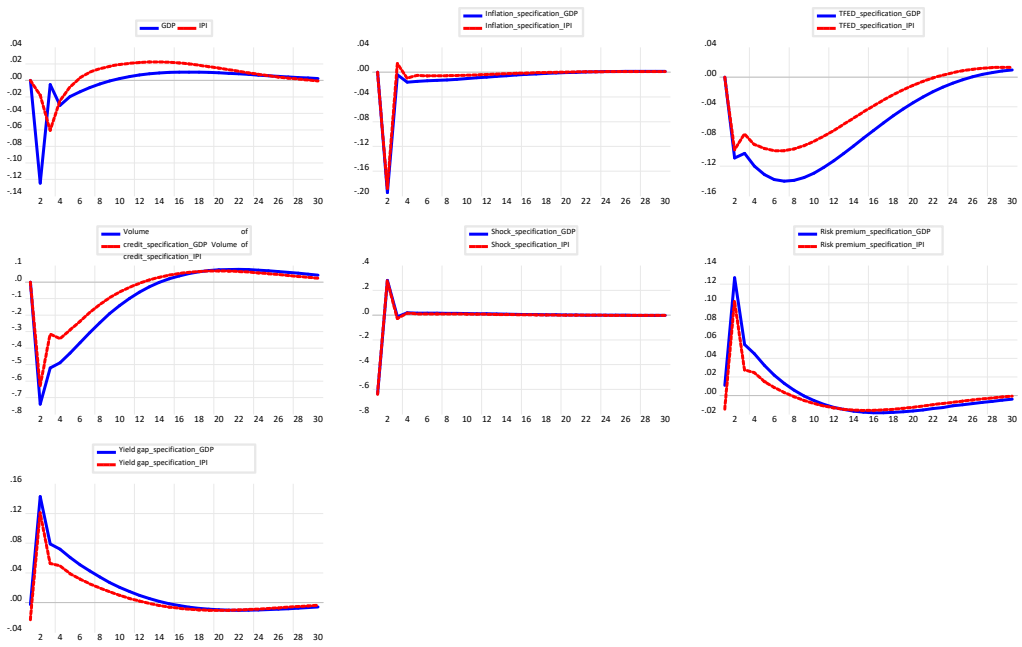
Figure 1.7 Response functions of macroeconomic variables to a negative variation of one standard deviation of *Shock* for lags 1, 2 and 3



1.5.4.2 Estimation of the VAR model with an alternative specification of the variable providing information on the state of the economic cycle

In our extended analysis, we opt for a substitution in our measure of economic activity. Specifically, we transition from *GDP* to another benchmark – the growth rate of the industrial production index, denoted as *IPI*. The ensuing response functions, illustrated in Figure 1.8, indicate that this modification does not perturb the established patterns. Thus, irrespective of whether we employ *GDP* or *IPI* as our economic activity gauge, the observed dynamics remain consistent, further attesting to the robustness of our findings.

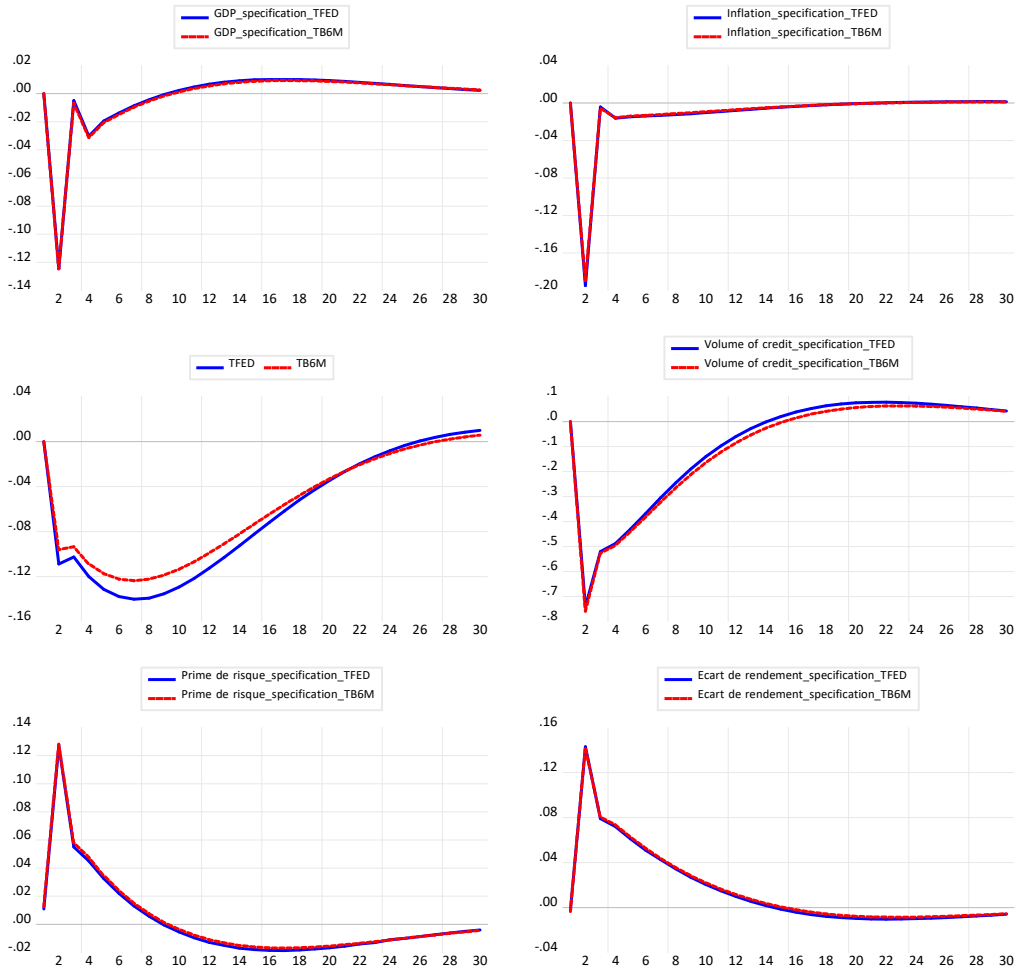
Figure 1.8 Response functions of macroeconomic variables to a negative variation of one standard deviation of the aggregate shock according to two measures of economic activity



1.5.4.3 Estimation of the VAR model with TB6M as the interest rate reflecting the stance of monetary policy

The *TB6M*, akin to *TFED*, serves as an indicator of monetary policy direction. However, it is noteworthy that during the Great Depression, the FED made a pivotal move by reducing the benchmark interest rate, *TFED*, to a zero threshold (as detailed in Appendix E.2) and maintained it at this nadir for an extended duration. Given this historical context, evaluating whether such a monumental event skews the response functions is imperative. To achieve this, we substitute *TFED* with *TB6M* in our model. The insights from Figure 1.9 indicate that substituting these monetary indicators does not introduce any distortions in the response functions. This finding underscores that the system's dynamics remain consistent irrespective of whether *TFED* or *TB6M* is used as a monetary policy measure.

Figure 1.9 Response functions of macroeconomic variables to a negative variation of one standard deviation of *Shock* according to *TB6M* and *TFED*



We have established the robustness of our VAR estimation to the three criteria quoted above. However, one question remains: Does the identification method used to extract credit supply shock matter? This question is the object of the following subsection.

1.5.5 Does the identification method matter in extracting aggregate credit shock ?

Reassessing the nature and influence of credit supply shocks, we juxtapose our baseline model with two alternative indicators. The first alternative stems from Mésonnier and Stevanovic's (2013) model, termed *shockMS*. At the same time, the second is anchored in Berrospide and Edge (2010) approach, rooted in the aggregate capital ratio. Upon examination, Figure 1.10 underscores that all models exhibit responses in the same direction, but their magnitudes and nuances are discernibly different. Specifically, *GDP* and *Inflation* sensitivities appear more pronounced under the *shockMS* model compared to the baseline. In contrast, the *Volume of credit* showcases heightened reactivity within the baseline when juxtaposed with *shockMS*. This empirical observation finds further reinforcement in Table 1.5. The table demonstrates that *shockMS* accounts for 13% of *GDP* fluctuations across horizons 2, 4, and 8, while our baseline model's *Shock* elucidates only about 5% on average. However, this narrative reverses when delving into the *Volume of credit*. A plausible explanation for this disparity lies in the inherent methodological distinctions between the two models. While both models harness similar datasets, they interpret them through distinct lenses. The *shockMS* model operates on an implicit linearity assumption between a bank's capital-to-asset ratio and its predictors. In contrast, our analysis, harnessing the capabilities of Gradient Boosting, posits a more intricate nonlinear relationship. Such a distinction may be pivotal in understanding the varied findings, suggesting that the intrinsic nonlinearity of certain relationships plays a crucial role in shaping outcomes.

The response of key macroeconomic indicators to variations in the aggregate capital ratio is notably subdued. As depicted in Figure 1.11, movements in the aggregate capital ratio seem to be inconsequential to *GDP*, *Inflation*, and *Volume of credit*. This observation is further substantiated by Table 1.5, which indicates that the aggregate capital

ratio imparts only a peripheral influence across varying horizons on the oscillations of macroeconomic markers tethered to real economic activity. Interestingly, the apparent detachment of the aggregate capital ratio from the *Volume of credit* underscores its inability to capture the nuances of bank credit supply shocks—such finding ushers in two pivotal takeaways. Firstly, the indispensability of microeconomic data becomes evident. Without delving into granular, micro-level datasets, capturing the subtleties and intricacies of bank supply credit shocks becomes a formidable challenge. Secondly, it outlines the primacy of the methodological framework. In essence, when engaging in the structural analysis of macroeconomic implications of banking shocks, an agnostic stance, complemented by suitable predictors and functional framework, proves paramount.

Figure 1.10 Response functions of macroeconomic variables to a negative variation of one standard deviation of the aggregate credit shock (baseline model) and shockMS

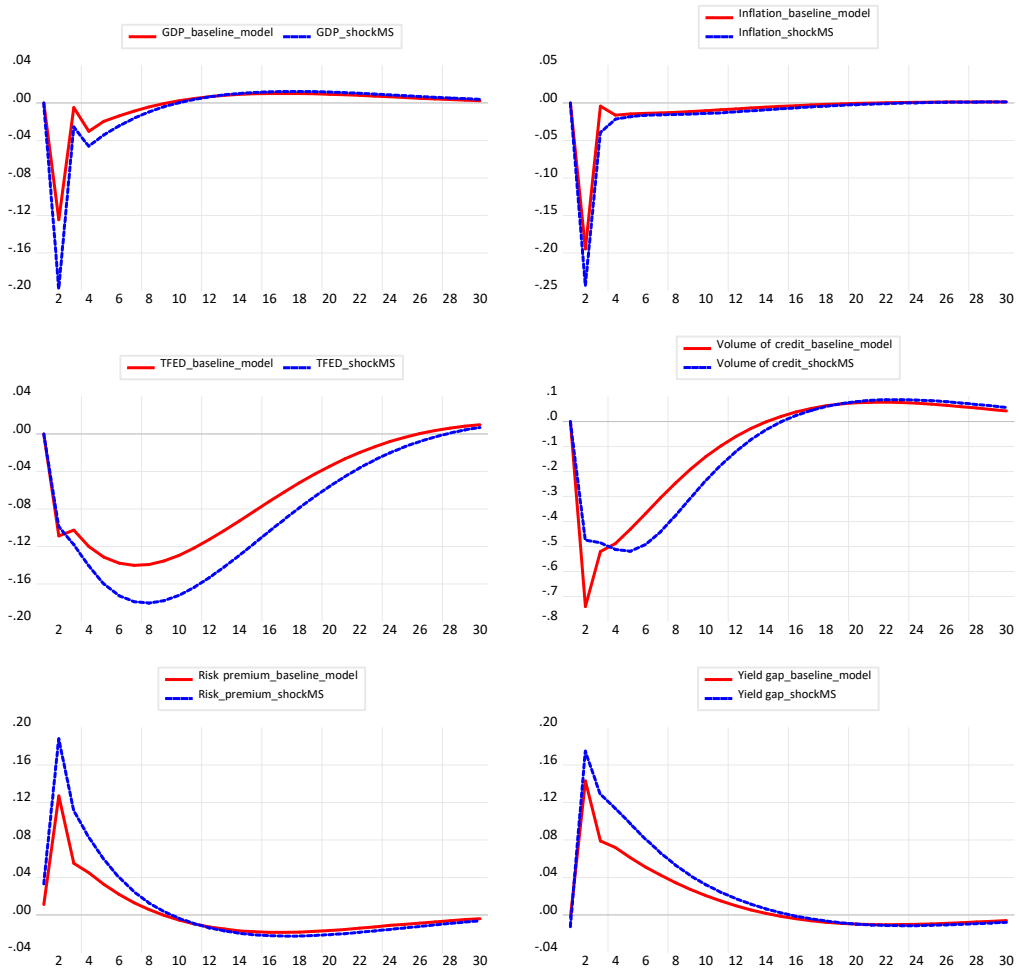
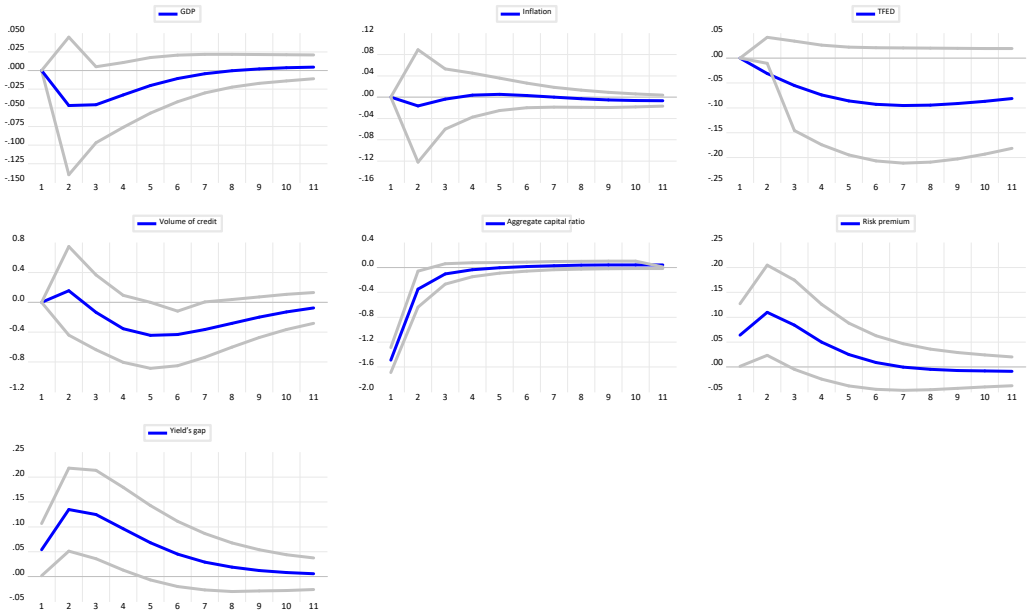


Table 1.5 Decomposition of the variance according to *Shock*, *shockMS* and aggregate capital ratio (in %)

	horizon	baseline model	ShockMS model	aggregate capital ratio model
GDP	h=2	5.4	13.3	0.8
	h=4	5.1	13.0	1.7
	h=8	5.0	13.0	1.7
Inflation	h=2	12.5	18.3	0.1
	h=4	11.8	18.1	0.1
	h=8	11.7	17.9	0.1
TFED	h=2	4.3	3.6	0.3
	h=4	5.5	6.6	1.5
	h=8	6.7	10.4	2.8
Volume of credit	h=2	1.6	1.2	0.2
	h=4	5.4	3.9	0.9
	h=8	6.1	6.7	3.3
Risk premium	h=2	5.7	12.7	6.1
	h=4	4.9	13.2	6.4
	h=8	4.5	13.0	5.8
Yield gap	h=2	8.4	13.2	8.3
	h=4	7.3	13.8	10.4
	h=8	6.8	14.3	9.2

Figure 1.11 Response functions of macroeconomic variables to a negative variation of the aggregate capital ratio



1.6 Conclusion

We use an original approach to better define the shock to the bank credit supply identified as a shock to the leverage effect. This approach allows us to use the financial data related to bank balance sheets and all the macroeconomic series to build an aggregate credit shock based on the best forecasting model of the bank capital-to-asset ratio, denoted *Shock*. By including *Shock* in a VAR of variables describing the real economy and those related to the financial market, we find that a negative variation of one standard deviation of *Shock* significantly affects the business cycle. However, the impact on industrial and commercial credit growth rates is more pronounced and persistent than the GDP growth. The VAR model that we consider is robust to alternative lags in its estimation, to an alternative measure of economic activity replacing the GDP growth rate, and to replacing the FED's key interest rate with the interest rate on 6-month Treasury bonds. We identify only one transmission channel for this shock: the risk premium. A negative variation in *Shock* by affecting the leverage of large banks diverts them from corporate bonds. The resulting decline in the supply of credit combined with an increase in the risk premium reduces investment prospects and affects GDP.

To analyze to which extent the methodological approach can affect the identification of shock and response functions, we consider two alternative methods for identifying credit supply shock: the approach of Mésonnier and Stevanovic (2013) and the approach of Berrospide and Edge (2010). On this last point, the impact of the ratio of aggregate capital to aggregate assets fluctuation over the economic cycle is insignificant. This result reminds us of the limits of identifying such a shock based solely on aggregate data. In addition, using microeconomic and big macroeconomic data while constraining the relationship between the capital-to-asset ratio and its predictors to be linear does not allow for

the recovery of its functions of responses in the case of a more complex relationship. The perfect illustration is the difference observed between the response function of the baseline model and that of the specification of Mésonnier and Stevanovic (2013). It is, therefore, essential to use all available information before imposing a way to combine them to identify a credit supply shock and correctly identify its macroeconomic effects.

CHAPTER II

A MACHINE LEARNING APPROACH IN STRESS TESTING US BANK HOLDING COMPANIES

ABSTRACT

This paper assesses the utility of machine learning (ML) techniques combined with comprehensive macroeconomic and microeconomic datasets in enhancing risk analysis during stress tests. The analysis unfolds in two stages. I initially benchmark ML's efficacy in forecasting two pivotal banking variables, net charge-off (NCO) and pre-provision net revenue (PPNR), against traditional linear models. Results underscore the superiority of Random Forest and Adaptive Lasso models in this context. Subsequently, I use these models to project PPNR and NCO for selected bank holding companies under adverse stress scenarios. This exercise feeds into the Tier 1 common equity capital (T1CR) densities simulation. T1CR is the equity capital ratio corrected by some regulatory adjustments to risk-weighted assets. Crucially, findings reveal a pronounced left skew in the T1CR distribution for globally systemically important banks vis-à-vis linear models. By mirroring distress akin to the Great Recession, ML models elucidate intricate macro-financial linkages and fortify risk assessment in downturns.

Keywords: Machine learning, Big data, Forecasting, Scenarios, Stress-test.
JEL classification: C53, C55, E44, G17, G18, G32.

2.1 Introduction

In the aftermath of the 2007-2009 financial crisis, the Federal Reserve (FED) introduced an enhanced supervision framework for large banks and complex financial institutions, which became a pivotal element of the Dodd-Frank Wall Street Reform and Consumer Protection Act.[1] As part of this framework, the FED conducts an annual stress test, which evaluates how distressed macroeconomic and financial conditions may impact the regulatory capital of bank holding companies (BHCs). This assessment aims to determine whether BHCs, with total consolidated assets of at least 10 billion USD, have enough capital to absorb losses during challenging situations while still providing loans to households and businesses.

This paper focuses on the top-down approach to stress tests, which comprises two stages. In the first stage, the FED constructs forecasting models for pre-provision net revenue (PPNR)[2] and net charge-offs (NCO)[3], two crucial variables that influence the dynamics of a bank's equity capital. In the second stage, these models are used to simulate PPNR and NCO under stressful economic scenarios over a given stress test horizon. The ultimate goal of this exercise is to compute the Tier 1 common equity capital (T1CR) for each selected BHC during the same period. T1CR, the ratio of common equity Tier 1 capital

---

[1] The Dodd-Frank Act, a critical regulatory measure for the American financial system since the mid-1930s, was implemented in July 2010 by President Obama to address the crisis and mitigate future upheavals.

[2] PPNR is the sum of a bank's interest and non-interest income, less interest and non-interest expenses scaled by the bank's average quarterly assets.

[3] NCO represents banks' charges for loan and lease losses, adjusted for recoveries and scaled by the bank's average quarterly assets.

(CET1)[4] to risk-weighted assets is the final output of the stress test. To assess the capital positions of the selected BHCs, the FED compares T1CR to a threshold defined by the Basel III committee throughout the stress test horizon.

I aim to go beyond the traditional point forecasting of T1CR. Specifically, I seek to investigate whether integrating machine learning techniques (ML) with comprehensive macroeconomic and microeconomic data sets can improve risk analysis in stress tests by offering a more accurate estimation of the T1CR distribution compared to standard linear models. This investigation is of significant policy interest as it examines potential avenues for improving the models used in stress tests, ultimately contributing to more robust risk assessment methodologies. Building robust models that accurately predict PPNR and NCO under hypothetical macroeconomic scenarios is essential to capture the risk in stress tests. The accurate prediction of these variables is crucial for constructing a reliable estimation of the T1CR distribution, allowing risk analysis. Liu et al. (2020) underscore the importance of the initial step of this process, which involves creating reliable forecasting models for PPNR and NCO using observed macroeconomic and financial variables. These models are the foundation for subsequent analysis and decision-making within the stress test framework. In light of this, I structure the analysis into two main steps.

In the first step, I investigate to what extent ML and big data can improve the accuracy of PPNR and NCO forecasts compared to standard linear models such as the fixed effect linear model, autoregressive model, random walk model, and pooled OLS model. Previous research, including the work of Goulet Coulombe et al. (2022), has emphasized the ability of ML to capture nonlinearities, making it a valuable tool in macroeconomic forecasting.

---

[4] CET1 is the equity capital corrected by regulatory adjustments. It is the core capital and the highest quality capital immediately available to absorb unexpected shocks.

In the same vein, exploring whether ML techniques can outperform traditional forecasting methods when applied to bank-level variables is essential. In addition, selecting the proper set of predictors enables a more accurate assessment of the financial health of banks and the risks they might encounter.

For this investigation, I use a panel of 100 large American BHCs observed from the third quarter of 1986 to the fourth quarter of 2019. I construct various out-of-sample forecasting models, encompassing both linear and ML approaches. The set of predictors aligns with those used by the FED in their stress tests. Given its standard usage by the FED in projecting PPNR and NCO components, I choose the fixed effect linear model as the benchmark for forecasting. Following this, I compute each model's relative out-of-sample mean square error (RMSE), a measure obtained by dividing a particular model's out-of-sample mean square error by that of the fixed effect linear model. The comparison of RMSE values allows the identification of the model offering the most accurate forecasts, indicated by the lowest RMSE. This comparative analysis underscores the added value of machine learning methodologies within this context. In addition, I challenge the significance of expanding the macroeconomic database and integrating BHC balance sheet characteristics beyond the variables used in the FED stress scenarios when forecasting PPNR and NCO. To this end, I sequentially construct ML forecasting models for PPNR and NCO, gradually expanding the set of predictors to include a larger macroeconomic database and subsequently combining this extensive macrodata with a wide range of microdata. Using the RMSE criteria, I compare the performance of ML models that incorporate this expanded macroeconomic database against models that strictly rely on the predictors defined in the FED stress scenarios. Moreover, I juxtapose ML models that use extensive macroeconomic and microeconomic data against models that rely solely on enlarged macroeconomic information. This multi-faceted comparison provides valuable insights into whether expanding the

pool of predictors results in a noticeable relative improvement in the forecast accuracy of ML models compared to the top-performing linear model.

In the second step, I use the best-performing ML forecast models for NCO and PPNR to generate conditional predictions under severely adverse scenarios defined by the FED. From these predictions, I compute point forecasts of T1CR and simulate T1CR distribution for each selected BHC throughout the stress test horizon. Therefore, using ML modeling, I can effectively characterize the banking sector's vulnerability and assess capital adequacy risks. For comparison purposes, I replicate the same analysis using a fixed effect linear model as the forecasting model for NCO and PPNR. In addition, by assessing the density forecast accuracy of both the ML model and the fixed effect linear model during the Great Recession and comparing the distribution of T1CR for globally systematically important banks, I may gain a deeper understanding of their respective abilities to capture and predict risks in a challenging economic environment. This approach provides valuable insights into the potential impact of adverse scenarios on the financial stability of individual BHCs and the banking system as a whole.

The main contribution of this paper is to showcase the value of integrating ML modeling into the risk analysis of stress tests, both indirectly and directly. Indirectly, the paper explores various models and demonstrates that ML models improve the forecast accuracy of crucial banking variables. It also sheds light on the importance of expanding macroeconomic information in improving NCO and PPNR forecast accuracy over some time horizons. This improvement in accuracy can have significant implications for assessing and managing risk in the banking sector. Directly, the paper highlights that ML modeling also leads to an improved density forecast of T1CR under stressful conditions compared to linear models. Many prior studies do not consider ML modeling in predicting NCO and PPNR (see, for example, Liu et al. 2020, Guerrieri and Welch 2012, Hirtle et al. 2016).

Indeed, while some studies have applied machine learning techniques to forecast PPNR and NCO, such as the works by Barth et al. (2019) and Kapinos and Mitnik (2016), they do not consider the broader implications for risk analysis and financial stability.

The rest of the paper is organized as follows. Section 2.2 provides a brief overview of the FED stress test framework. Section 2.3 presents the methodology. Section 2.4 contains the description of the data. Section 2.5 delivers results. Section 2.6 concludes.

2.2 An overview of the FED stress test framework

Before the financial crisis, stress tests were typically discretionary and limited to specific institutions.[5] In the wake of the Great Recession, monetary authorities in developed and emerging countries established rules that made stress tests mandatory for large banks and complex financial institutions.[6] The primary goal of these stress tests is to mitigate the impact of potential economic downturns on the financial system and the real economy. Since 2011, the FED in the USA has been conducting annual stress tests on large banks and financial institutions, following a provision of the Dodd-Frank Wall Street Reform and Consumer Protection Act. The main objective of this exercise is to ensure that these institutions maintain sufficient capital to withstand severely adverse economic conditions without disrupting the economy's financing. The FED uses two complementary tools for its stress tests: the Dodd-Frank Act Stress Test (DFAST) and the Comprehensive Capital Analysis and Review (CCAR). These tools assess the capital position of BHCs with assets

[5] See Lederman (1990) and Hirtle and Lehnert (2015), for examples, of the sectorial stress test in the USA before the Great Recession.

[6] Since 2010, the European Bank Authority has conducted yearly stress tests for European Union member states. The FED made these tests mandatory in the USA and oversaw the process. Other emerging countries, such as Brazil, also conduct regular stress tests.

exceeding ten billion USD. They project critical revenue and loss variables under three scenarios: severely adverse, adverse, and baseline. The severely adverse and adverse scenarios represent hypothetical distressed macroeconomic and financial conditions designed to measure the solidity and resilience of large BHCs. On the other hand, the baseline scenarios rely on unconditional forecasts of scenario variables over a specific time horizon. DFAST and CCAR are closely aligned, employing identical scenario variables, models, and methodologies to produce comparable outcomes.

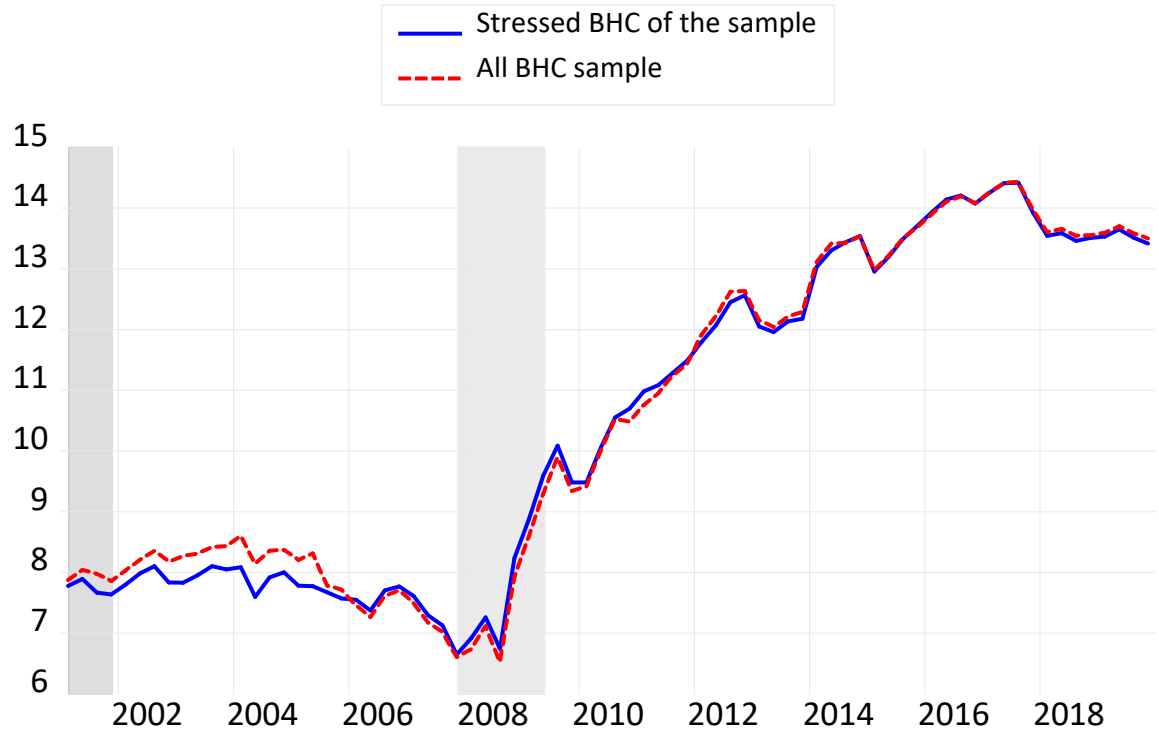
However, the computation of regulatory capital differs between the two approaches. DFAST uses planned capital distributions, such as dividend payments and common stock repurchases, as defined by the FED, while CCAR uses the bank's planned capital distribution. To pass a stress test, a participating bank should have its projected tier 1 common equity capital ratio (T1CR)[7] remaining greater than or equal to 4.5% throughout the nine-quarter forecast horizon. Figure 2.1 illustrates the evolution of the aggregate T1CR over time, defined as the aggregate common equity tier 1 capital to the aggregate risk-weighted asset. The graph indicates that during the Great Recession, the aggregate T1CR declined, reaching a minimum level close to the regulatory requirement of 4.5%. However, there was a reversal in the trend after the turmoil, with the T1CR steeply increasing and showing moderate stability in 2018. This observation suggests a continuous improvement in banks' capital position since the crisis, with the 4.5% T1CR constraint remaining far from binding. Apart from the T1CR, Basel III sets the minimum total regulatory capital ratio[8]

[7]T1CR is the ratio of common equity tier 1 capital to risk-weighted assets. Mapping book equity capital into common equity tier 1 requires subtracting regulatory adjustments, which include goodwill and other intangibles, deferred tax assets, cash flow from hedge reserves, and gains on sales related to securitization transactions.

[8]The total regulatory capital ratio is the sum of core equity tier 1 and tier 2 capital to the risk-weighted asset. Tier 2 capital is the second layer of capital, comprising revaluation reserves, hybrid instru-

at 8%.

Figure 2.1 Aggregate T1CR



Note: The hatched bands represent recessions

---

ments, and subordinated term debt.

2.3 Econometric approach

This section describes how I build the forecast models of two banking measures, i.e., PPNR and the NCO, and use them in the stress test.

2.3.1 Forecasting methodology

2.3.1.1 Machine learning techniques

I apply the machine learning techniques to the following models:

$$Y_{i,t+h} = \hat{g}(Z_{i,t}) + e_{i,t+h} , \tag{2.1}$$

where h represents the forecast horizon ranging from one to nine quarters, in line with the range used by the FED in stress test exercises on BHCs. $Y_{i,t+h}$ denotes the value of either PPNR or NCO of BHC i at quarter $t + h$, $e_{i,t+h}$ represents the forecast error and $Z_{i,t}$ is a $K \times 1$ regressor vector. The h-step ahead forecast of variable $Y_{i,t+h}$ for BHC i given the information available at t is given by $\hat{Y}_{i,t+h}$=$\hat{g}(Z_{i,t})$. I opt for a direct forecast approach for h=2 to 9, as it has been shown in various studies that it is more robust to misspecification compared to iterated forecasts (Chevillon and Hendry 2005,Marcellino et al. 2006, Schorfheide 2005). Given a lack of confidence in a particular model, I explore several models to find the most suitable. I estimate $\hat{g}$ using three specifications combining parametric and nonparametric machine learning techniques. The parametric techniques I consider include "Lasso" regression, "Adaptive Lasso" regression, "Ridge" regression, "Elastic Net" regression, Principal Component Analysis (PCA). For nonparametric techniques, I include Gradient Boosting, Random Forest and Neural Network. Appendix A

provides comprehensive details on these parametric and nonparametric machine learning techniques.

In the first specification, $Z_{i,t}$ includes only lagged dependent variables, some bank-specific predictors, and the macroeconomic series used by the FED in their stress tests. The second specification expands $Z_{i,t}$ to include lagged dependent variables, microeconomic predictors from the first specification, and all macroeconomic and financial series from the FRED database. The last specification incorporates additional BHC balance sheet characteristics and the predictors from the previous specification. The primary objective of all machine learning techniques is to estimate $\hat{g}$, which minimizes the out-of-sample Mean Squared Error (MSE) while simultaneously applying regularization procedures. Regularization serves a dual purpose. Firstly, it prevents the overfitting of the data, where the estimated function $\hat{g}$ fits the variable of interest too closely, leading to poor out-of-sample prediction performance. Secondly, it reduces the model complexity and often results in better forecast performance than traditional techniques through bias-variance trade-offs. The choice of appropriate hyperparameters is crucial in regularization procedures. These hyperparameters include penalization coefficients in "Lasso," "Elastic Net," and "Ridge" regressions, as well as the number of trees and variables selected randomly at each step for Random Forest.

I determine the optimal hyperparameters through K-fold cross-validation. Since the sample is a panel dataset, I perform cross-validation across time series and cross-sectional dimensions. Specifically, I divide the subsample into k groups of approximately equal size for a given model and a combination of hyperparameters. Group j where ($1 \leqslant j \leqslant k$) is treated as the validation set, and I estimate the model on the remaining $k-1$ groups. Then, I predict the variable of interest PPNR or NCO for group j. The Mean Squared

Error (MSE^j) is computed as:

$$MSE^j = \frac{1}{N}\sum_{i,t}(Y_{i,t+1} - \hat{Y}^j_{i,t+1})^2,$$

where N represents the number of banks in the validation set, i denotes the cross-sectional dimension, and t represents the time dimension. This process is repeated k times, treating each of the remaining $k-1$ groups as a validation sample. Finally, I compute the average forecasting performance for a given set of hyperparameters as follows:

$$MSE = \frac{1}{N}\sum_{1}^{k} MSE^l,$$

where l ranges from 1 to k. I repeat this process for all potential combinations of hyperparameters and select the ones that yield the lowest MSE, thus deriving the estimation model.

2.3.1.2 Estimation of standard linear forecasting models

In addition to machine learning models, I also consider several standard linear forecast models: the autoregressive model, the random walk model without drift, pooled OLS model, and the fixed effect linear model. They can be defined as follows:

$$Y_{i,t+h} = \alpha + \rho Y_{i,t} + \epsilon_{i,t+h}\,, \tag{2.2}$$

$$Y_{i,t+h} = Y_{i,t} + \epsilon_{i,t+h}\,, \tag{2.3}$$

$$Y_{i,t+h} = \alpha + \rho Y_{i,t} + Z'_{i,t}\beta + \epsilon_{i,t+h}, \tag{2.4}$$

$$Y_{i,t+h} = \alpha_i + \rho Y_{i,t} + Z'_{i,t}\beta + \epsilon_{i,t+h}. \tag{2.5}$$

In these equations, $Z_{i,t}$ represents the regressor vector of the first specification. Parameters α, ρ, and β are estimated using OLS, while α_i represents the bank fixed effect. Equation (2.5) is estimated using the Least Squares Dummy Variable (LSDV) estimator. It is worth noting that the estimation of equation (2.5) can be biased in the presence of lagged dependent variables, as discussed in Galvao Jr (2011). However, this bias diminishes in a long panel dataset like ours. Although I define the fixed effect linear model as the benchmark, the other linear models also serve as natural benchmarks. These linear models are widely used in the literature on bank stress tests, as highlighted in studies such as Covas et al. (2014), Hirtle et al. (2016), and Liu et al. (2020). By comparing the machine learning models to these linear models, I can assess the performance and contribution of ML models in the context of bank stress tests.

2.3.1.3 Pseudo out-of-sample forecasting exercise

I use the expanding window estimation technique for the forecasting exercise. Initially, I divide the data sample into two parts, i.e., the estimation and validation sample, which spans from 1986q3 to 2000q1, and the test set, which covers the period from 2000q2 to 2019q4. After selecting the hyperparameters, I estimate the model using the entire first subsample for each machine learning (ML) method concerning all three specifications. The goal is to forecast the dependent variable $Y_{i,t+h}$, h quarters ahead. I iteratively run the process by expanding the estimation and validation sample by one quarter until I have predicted the dependent variable for the final quarter of the test set, 2019q4. I follow a similar procedure to build linear standard forecast models (SL) but estimate them based only on the first specification. Once I construct various forecast models, I need to evaluate

their forecast accuracy.

Let M denotes the number of competing forecast models, h represents the forecast horizon (h=1...9), and j indicates the specification ranging from 1 to 3. To evaluate the forecast model K (K=1...M) at the forecast horizon h within specification j, I calculate its relative out-of-sample mean square error ($RMSE_j^K(h)$), which is defined as :

$$RMSE_j^K(h) = \frac{\sum_{i=1}^{N}\sum_{t=T_{min}}^{T_{max}}(Y_{i,t+h} - \hat{Y}_{i,t+h}^K(j))^2}{\sum_{i=1}^{N}\sum_{t=T_{min}}^{T_{max}}(Y_{i,t+h} - \hat{Y}_{i,t+h}^B)^2} .$$

Here, N, T_{min}, and T_{max} represent the number of banks, the start date, and the end date of the test set sample, respectively. $Y_{i,t+h}$ denotes the value of the variable of interest for bank i at quarter $t+h$ while $\hat{Y}_{i,t+h}^K(j)$ represent predicted value of $Y_{i,t+h}$ h quarters ahead relative to model K and given information at quarter t. $\hat{Y}_{i,t+h}^B$ represents the predicted value of $Y_{i,t+h}$ h quarters ahead relative to the benchmark model and given information at quarter t. The numerator represents the out-of-sample mean square error $(MSE_j^K(h))$ for model K at horizon h within specification j. At the same time, the denominator refers to the same quantity as the numerator but relative to the benchmark model.

I identify the best model within each specification and for a given forecast horizon by following a two-step process.[9] First, I determine the best ML and SL model[10]. Second, I compare the best SL model with the best ML model. The model with the lowest RMSE is the superior model. I use the Diebold-Mariano test, adapted for panel data, to rank them

[9]The best model given the forecast horizon has the lowest RMSE. Here, I proceed in two steps to gauge the eventual superiority of ML models over SL models.

[10]An ML model K is the best ML forecast model for specification j relative to a forecast horizon h if for any ML model L estimated in the specification and different from M $RMSE_j^K(h) \leqslant RMSE_j^L(h)$. Similarly, SL model N is the best SL forecasting model if for any other SL model P different from N, $RMSE_j^N(h) \leqslant RMSE_j^P(h)$.

formally. Diebold and Mariano (2002) provide the test (the Diebold Mariano test) adapted for panel data, allowing the models to be ranked. In my case, I use this test to compare the forecast errors of the best SL and ML models.[11] Besides determining the best forecast model, this comparison exercise helps assess the relevance of ML techniques in forecasting banking variables by comparing the best ML and SL models under specification 1. Indeed, given that both models use the same limited information set, any difference in forecasting performance can be attributed to the forecasting technique itself. Last, I compare ML models across different specifications to assess the contribution of expanding macro and microdata. Specifically, I compare out-of-sample RMSE of ML models across specifications.

2.3.2 Stress test methodology

This section describes the methodological approach of the stress test. First, it outlines how I map the predictions of PPNR and NCO under severely adverse scenarios to the prediction of T1CR. Then, I present a formal framework of the stress test exercise.

2.3.2.1 Mapping PPNR and NCO predictions into the T1CR

The stress test is a conditional forecast exercise. Therefore, I use the best PPNR and NCO forecasting models to predict the two banking variables under the severely adverse scenarios defined by the FED. PPNR and provision on loan losses affect the regulatory capital through their direct action on book equity capital. Since, practically, banks adjust their provisions to respond to the variation in charge-off on loan and lease losses, I assume,

[11] Detailed information about the Diebold-Mariano test can be found in Appendix F.

as Covas et al. (2014), that provision is equal to net charge-off. Moreover, I keep Covas et al. (2014) assumption on the evolution of the book equity capital amount that can be expressed as follow:

$$K_{i,t} = K_{i,t-1} + (1-\tau) \times [PPNR_{i,t} \times \overline{Asset_i} - NCO_{i,t} \times \overline{Asset_i}] - \overline{Equity\ payout_i}\ . \tag{2.6}$$

$K_{i,t}$ is the dollar amount of book equity capital of BHC i at quarter t. τ is the marginal tax rate set at 21%, its value in 2018q4, the quarter before the stress test horizon 2019q1-2020q1. $PPNR_{i,t}$ and $NCO_{i,t}$ are, respectively, the PPNR and NCO of BHC i at quarter t. The principle behind the stress test is to ensure that large BHCs maintain their capacity to intermediate even when faced with adverse shocks. To translate this principle, I follow the approach of Covas et al. (2014) and assume that the total assets of BHC i at quarter t denoted as $(Asset_{i,t})$, remain constant throughout the stress horizon and equal to its value in 2018q4 $(\overline{Asset_i})$. I also assume, as Covas et al. (2014), that the aggregate equity payout of BHC i, which includes dividends paid on common and preferred stocks and the re-purchase of Treasury shares, remains constant. I set its value to its corresponding value in 2018q4. By assuming a constant equity payout, I consider the costliness for a bank to raise new equity during a severe recession. The prediction of the common equity tier 1 capital, which is the main focus of the stress test, is determined by the aforementioned predicted book equity capital, the predicted regulatory adjustments denoted as $RADJUST_{i,t}$, and the predicted risk-weighted assets denoted as $RWA_{i,t}$. For the same reasons mentioned earlier, I assume that $RADJUST_{i,t}$ and $RWA_{i,t}$ are constants throughout the stress test horizon and are respectively equal to their values in 2018q4 denoted by $\overline{RADJUST_i}$ and

$\overline{RWA_i}$. Formally, the common equity tier 1 capital (T1CR) can be calculated as follows :

$$T1CR_{i,t} = \frac{K_{i,t} - \overline{RADJUST_i}}{\overline{RWA_i}}$$

.

2.3.2.2 Density forecast of T1CR

I study the entire distribution of T1CR instead of focusing only on its mean as usual in the literature. Predicting the distribution of T1CR helps to compute the probability of violating the minimum requirements, one central indicator in stress tests.[12] Therefore, this approach provides a more detailed picture of the banks' vulnerability and risk in an unfavorable macroeconomic environment than the one relying solely on point estimates. Like Covas et al. (2014), I use a two-step approach to forecast the density of T1CR under distressed economic conditions.

In the first step, I use Athey et al. (2019) Quantile Random Forest method to infer conditional quantiles. I adopt a recursive window-expanding estimation technique commonly used in forecasting exercises to perform this estimation. The Quantile Random Forest approach is a nonparametric method that mitigates the issues of misspecification bias. It leverages the splitting tree rules of Random Forest and adapts them to capture quantile heterogeneity effectively. The conditional quantile is chosen to minimize the weighted

[12] See, for example, Covas et al. (2014).

absolute values of errors as stated below:

$$\hat{q}_\tau(y_{it+1}/x_{i,t}) = argmin_{\beta(x_{i,t})}\Bigg\{\sum_{i=1}^{N}\sum_{t=1}^{T}\Big[(\tau(y_{i,t+1}-\beta(x_{i,t}))1(y_{i,t+1}\geqslant\beta(x_{i,t})) + (1-\tau)(\beta(x_{i,t})-y_{i,t+1})1(y_{i,t+1}<\beta(x_{i,t}))\Big]\Bigg\}.$$

N and T denote the data's cross-sectional and time series dimensions. $y_{i,t}$ is the variable of interest, and $x_{i,t}$ is the vector of the forcing variables, the same that the FED uses in its severely adverse scenario. 1(.) denotes an indicator function, τ is the percentile, and $\hat{q}_\tau(y_{it+1}/x_{i,t})$ represents the τ percent quantile. I forecast the 1, 5, 50, and 95 percent conditional quantiles through the stress horizon.

In the second step, to recover the density function, I fit a skewed t-distribution of Azzalini and Capitanio (2003) with the constrain of matching it to the estimated quantile function:

$$f(y;\mu;\sigma;\alpha;\nu) \quad = \frac{2}{\sigma}t\Big(\frac{y-\mu}{\sigma};\sigma\Big)T\left(\alpha\frac{y-\mu}{\sigma}\sqrt{\frac{\nu+1}{\nu+(\frac{y-\mu}{\sigma})^2}};\nu+1\right). \tag{2.7}$$

t(.) denotes the PDF of the Student-t distribution, and T(.) is its corresponding CDF. The four parameters μ, σ, α, and ν pin down the distribution's location, scale, shape, and fatness, respectively. The intuition is to perturb a standard symmetric t-distribution PDF by its CDF, adding an asymmetry through the slant parameter α. For example, if $\alpha = 0$ and $\nu < \infty$, f reduces to standard symmetric t-distribution. The case in which $\alpha = 0$ and $\nu = \infty$ results in a standard normal distribution, while the case in which $\alpha \neq 0$ and $\nu = \infty$ leads to a skewed normal distribution. $\alpha \neq 0$ and $\nu < \infty$ corresponds to a skewed t distribution. This class of distributions offers greater flexibility compared to the standard t-distribution, which has only two parameters. It allows for different shapes and can

accommodate both symmetric and asymmetric environments. Due to this flexibility, this family of density functions has been extensively used in the growth and risk literature to examine the asymmetric effects of financial conditions on growth rates. Notable works such as Adrian et al. (2019) and Plagborg-Møller et al. (2020) have considered this class of functions. In the banking stress test, I use this class of functions to account for potential asymmetry in the distribution of T1CR under disastrous macroeconomic conditions. Covas et al. (2014) previously highlighted the heavy left tail of the density forecast for aggregate T1CR at the end of the stress period, indicating evidence against the symmetry of the T1CR distribution during turmoil.

To fit the t-skewed distribution, I adopt the framework developed by Adrian et al. (2019). For each quarter of the stress horizon and for each stressed bank, I estimate the four parameters $\mu_{i,t+1}$, $\sigma_{i,t+1}$, $\alpha_{i,t+1}$, and $\nu_{i,t+1}$. The estimation is performed by minimizing the squared distance between the estimated conditional quantiles and the quantiles of the t-distribution. Formally, I can state the following optimization problem :

$$\begin{aligned}\{\mu_{i,t+1}, \sigma_{i,t+1}, \alpha_{i,t+1}, \nu_{i,t+1}\} = argmin_{\mu,\sigma,\alpha,\nu} \sum_{\tau} \Big[\hat{q}_{\tau}(y_{it+1}/x_{i,t}) \\ - F^{-1}(\tau; \mu, \sigma, \alpha, \nu) \Big]^2 ;\end{aligned} \tag{2.8}$$

where $\tau = 1, 5, 50, 95$ are the percentiles, and F is the CDF of the t-skewed distribution defined in (2.7).

2.4 Data

I use a panel of 100 BHCs, each with total assets consistently exceeding three billion USD, to implement the methodology. The panel data spans from 1986q3 to 2019q4, covering

significant events such as the implementation of Basel I, II, and III regulations, the 1990 recession, and the recent subprime crisis (see Appendix D for BHCs list). As a result of mergers, acquisitions, and bank failures, the panel is unbalanced—only 20% of BHCs operate throughout the period. In line with the Federal Reserve's stress-testing practices, I focus on BHCs instead of commercial banks. Unlike the FED, which only considers banks with total assets exceeding 10 billion USD, I include a larger number of banks in my sample. Extending sample aims to enhance forecast accuracy by constructing a more extensive dataset with broader temporal and cross-sectional dimensions.

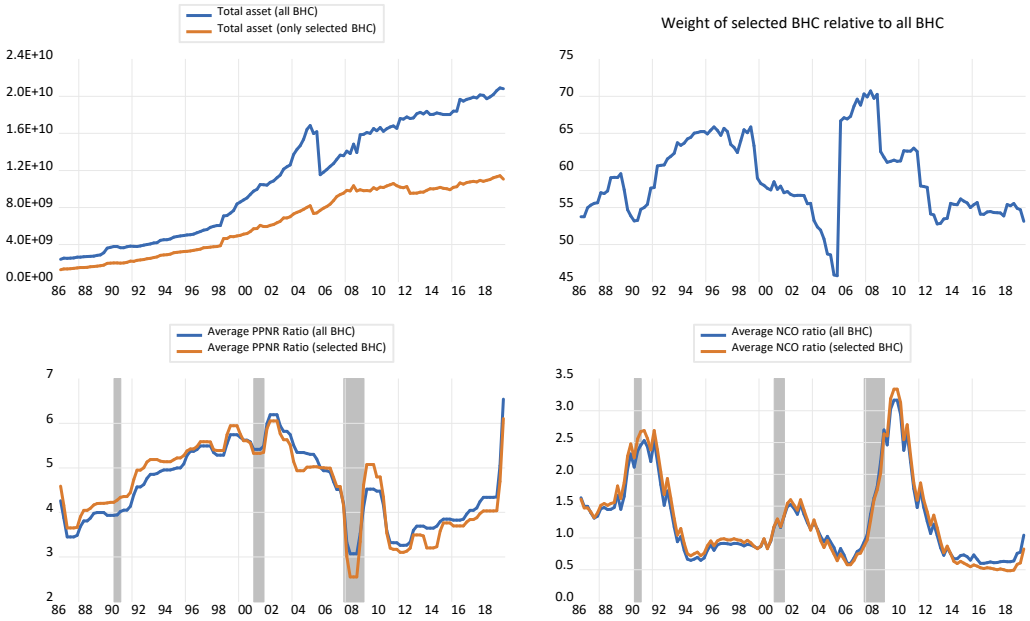
Figure 2.2 illustrates key data insights. The top left panel shows strong co-movement between the total assets of all BHCs and the sampled BHCs, indicating an accurate representation of the broader BHC population in the sample. Moreover, the top right panel reveals that the sampled BHCs' total assets constitute over 55% of all BHCs' total assets, peaking at 70%. These figures underscore the representativeness of the sample. However, it is important to note a contraction in the total assets of the sampled BHCs relative to all BHCs starting in 2006. This reduction can be ascribed to a regulatory change requiring BHC subsidiaries with total assets over 1 billion USD to become autonomous. This change led to the disappearance of several BHCs from our sample, thereby reducing the total assets of the sampled BHCs.

In the analysis, I focus on two key variables: (1) Net Charge-Off (NCO), defined as the annualized charge-off for loan and lease losses, minus recoveries for these losses, scaled by a quarterly average of total assets; and (2) Pre-Provision Net Revenue (PPNR), defined as the annualized sum of net interest and non-interest income divided by the quarterly av-

erage of total assets. These computations draw upon data from the FR Y-9C form of the Consolidated Financial Statements for BHCs, sourced from the Chicago Federal Reserve. Cyclical trends for average PPNR and NCO are evident in the bottom panel of Figure 2.2. Average PPNR decreases during recessions while NCO increases, and vice versa, during recoveries. These inverse trends, significant within the context of stress test literature, suggest that stressful macroeconomic conditions can erode a bank's capital position, potentially impeding the effective functioning of the BHC. The goal of stress testing is to assess the impact of such weakened capital positions.

I use two types of predictors—bank-specific and macroeconomic—across three different specifications to forecast PPNR and NCO accurately. The bank-specific predictors are balance sheet variables drawn from the FR Y-9C form. The macroeconomic predictors, consisting of macroeconomic and financial variables, are sourced from the FRED QD database. These data are transformed into stationary series following the procedures specified in the FRED database. I use the entire series from the FRED QD database as macroeconomic predictors in the second and third specifications, aiming to capture a broad spectrum of potential influences on the variables of interest. However, for the first specification, I intentionally limit the macroeconomic predictors to those defined as FED stress scenario variables (as outlined in Table 2.1) to precisely assess the effect of these stress indicators and ML modeling in improving PPNR and NCO forecast accuracy. Additionally, to ensure that the sample closely mirrors the banks likely to be of interest in real-world stress tests, I refine it to include only the 13 BHCs that were part of the 2017, 2018, and 2019 stress tests (as listed in Table 2.2). These BHCs represent 92% of the total assets of all BHCs in the sample, emphasizing large banks' relevance in stress testing scenarios.

Figure 2.2 A brief overview of some data



Note: (1) The hatched bands represent recessions.
(2) The average PPNR and NCO are computed as the aggregate PPNR and NCO ratio to total quarterly average assets.

Table 2.1 List of FED stress test scenarios macroeconomic variables

FRED mnemonic	Variable name	Scale	Variable definition
GPC1	GDP	percentage change	US real GDP growth
UNRATE	UNRATE	level	US unemployment rate
CPIAUCSL	Inflation rate	percentage change	U.S CPI inflation
TB3MS	TB3MS	level	U.S. 3-month Treasury rate
GS5	GS5	level	U.S. 5-year Treasury yield
GS10	GS10	level	U.S. 10-year Treasury yield
MORTGAGE30US	Mortgagerate	level	U.S. mortgage rate
SP 500	SP500	percentage change	U.S. stock market index growth
USSTHPI	USSTHPI	percentage change	U.S. House Price Index growth(all transactions)
VXOCLSx	VXOCLSx	level	VIX implied volatility index

Table 2.2 Subsample of BHC included in 2017, 2018 and 2019 stress test

Bank holding companes	Total Asset in 2018q4[a]	Weight in 2018q4[b]
Bank of America Corporation	2.36e+09	21.40
Bancwest corporation	20695678	0.19
BBVA USA Bancshares Inc	90947174	0.83
Citigroup INC	1.92e+09	17.41
Comerica Inc	70906003	0.64
Fifth Third Bancorp	1.46e+08	1.33
JP Morgan chase and company	2.62e+09	23.82
Keycorp	1.40e+08	1.27
Huttington Bancshares Inc	1.09e+08	0.99
Northern trust corporation	1.32e+08	1.20
PNC Financial Services INC	3.82e+08	3.47
State street corporation	2.45e+08	2.22
Wells Fargo et Company	1.89e+09	17.22
All	1.03e+10	91.99

Note: Total consolidated assets (in thousand dollars).
Weight represents the ratio of bank's total asset to all sampled BHC total asset in 2018q4.

2.5 Results

This section discusses the results of the empirical analysis. I start with some forecasting exercises based on observed data, then check the results' robustness and determine the main drivers of PPNR and NCO. The last subsection is devoted to the empirical stress test.

2.5.1 Forecasting Exercise

2.5.1.1 Out-of-sample performance of NCO's predictions models

Tables 2.3, 2.4, and 2.5 present the out-of-sample relative mean square error (RMSE) of NCO forecasting models in specifications (1), (2), and (3). As Table 2.3 suggests, most machine learning (ML) models outperform the benchmark model in NCO forecasting, as indicated by RMSE values lower than one across various horizons. Specifically, the Diebold-Mariano test (DM) highlights that the Random Forest model significantly sur-

passes the pooled OLS - the most effective standard linear (SL) model - across the nine forecast horizons in the specification (1) at a 1% significance level. These findings indicate that the Random Forest model offers superior forecasting across all horizons in the specification (1). This superior performance suggests the existence of nonlinearity in the bank's loss model, which linear models may fail to capture adequately. The implication is that ML models like Random Forest, capable of handling complex, nonlinear relationships, may yield more accurate and robust forecasting in economic scenarios like stress testing.

Tables 2.4 and 2.6 illustrate how incorporating a broader macroeconomic database instead of limited macro data can impact model rankings and predictive accuracy. As Table 2.4 indicates, Random Forest outperforms most SL models across most horizons, except for horizons 3 and 8. Despite Gradient Boosting emerging as the best model in Horizon 4, the Diebold-Mariano (DM) test does not establish its superiority over Random Forest. Table 6 reveals that excluding horizons 3 and 8, Random Forest – when paired with a comprehensive macroeconomic database – improves the RMSE of Pooled OLS by an average of 49%. This finding contrasts with an average improvement of 39% when Random Forest uses limited macroeconomic data. These results emphasize the advantages of extending macroeconomic data beyond the Federal Reserve's stress test scenario: it bolsters the forecasting performance of leading ML models in predicting NCO by approximately 10% across seven of the nine forecast horizons. Interestingly, Table 2.5 and Table 2.6 suggest that the inclusion of more bank-specific data does not enhance the predictive accuracy of Random Forest or other ML models. On the contrary, it tends to deteriorate the performance of top ML models relative to the most efficient SL model. This result implies that

additional bank-specific data may not offer predictive value beyond what the macroeconomic data and initial bank characteristics already provide.

Table 2.3 Out-of-sample RMSE of NCO's forecast models in specification (1) (all banks)

Model	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	h=9
Benchmark model	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*
Pooled OLS	0.62	0.68	0.64	0.71	0.59	0.66	0.66	0.71	0.60
Autoregressive model	2.05	1.93	2.08	3.20	2.05	1.04	1.17	3.08	1.27
Random walk	2.39	1.64	1.61	3.19	1.61	1.03	1.03	1.53	1.21
Lasso	1.51	0.69	0.64	0.68	0.57	0.64	0.65	0.69	0.60
Adaptive Lasso	0.61	0.60	0.65	0.68	0.58	0.66	0.67	0.68	0.62
Ridge	0.59	0.62	0.53	0.97	0.65	0.59	0.53	0.70	0.68
Elastic net	0.64	0.61	0.63	0.68	0.57	0.67	0.65	0.69	0.60
Principal component	2.50	1.71	1.63	3.23	0.58	1.17	1.21	0.62	0.58
NN1	3.39	1.09	0.72	1.58	0.72	0.72	0.72	1.26	0.78
NN2	6.45	1.11	0.67	1.16	0.64	0.70	0.64	0.93	0.76
NN3	4.14	1.17	0.69	1.29	0.65	0.80	0.61	0.91	0.81
Gradient boosting	0.97	0.93	0.71	0.52	0.55	0.43	0.38	0.67	0.45
Random Forest	**0.39** ***	**0.45****	**0.51****	**0.45****	**0.35****	**0.31****	**0.36****	**0.36****	**0.39****

Note: Machine learning and standard linear models include the same variables FED uses in the bank's stress test. (2) The first line represents the benchmark model's mean square error (MSE) normalized to 1, while the other lines present the relative MSE of different forecast models. (3) The figures in bold relate to the best forecast model. No star: best linear models outperform the best machine learning model or have the same performance as it for the DM test, *: best machine learning model outperforms best standard linear model for the DM test at 10% level, **: 5% level; ***: 1% level. NN1 designates the neural network with one hidden layer that has 32 neurons. NN2 is the neural network with two hidden layers containing 32 and 16 neurons, respectively. NN3 is a neural network with three hidden layers having 32, 16, and 8 neurons, respectively.

Table 2.4 Out-of-sample RMSE of NCO's forecast models in specification (2) (all banks)

Model	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	h=9
Benchmark model	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*
Pooled OLS	0.62	0.68	**0.64**	0.71	0.59	0.66	0.66	0.71	0.60
Autoregressive model	2.05	1.93	2.08	3.20	2.05	1.04	1.17	3.08	1.27
Random walk	2.39	1.64	1.61	3.19	1.61	1.03	1.03	1.53	1.21
Lasso	1.88	3.60	0.97	0.58	1.12	1.59	0.76	0.49	0.67
Adaptive Lasso	2.17	1.28	0.83	0.52	1.17	4.73	0.63	0.54	0.60
Ridge	1.23	2.08	0.80	1.23	1.35	1.50	0.56	0.55	0.66
Elastic net	1.73	3.57	0.78	0.58	1.13	1.69	0.71	0.48	0.66
Principal component	1.13	1.02	0.79	1.39	0.89	0.69	0.67	0.85	0.82
NN1	3.48	1.71	1.74	7.16	1.64	0.87	0.94	1.85	0.98
NN2	3.92	1.96	2.05	8.68	1.91	1.02	1.21	2.38	1.20
NN3	5.08	2.54	2.57	10.55	2.43	1.29	1.38	2.70	1.43
Gradient boosting	0.42	1.05	0.90	**0.39*****	0.28	0.21	0.28	**0.39*****	0.40
Random Forest	**0.28*****	**0.39*****	**0.64**	0.45	**0.22*****	**0.17*****	**0.27*****	0.40	**0.28*****

Note: (1) Machine learning models are estimated with an extensive macroeconomic database.(2) The first line represents the benchmark model's mean square error (MSE) normalized to 1, while the other lines present the relative MSE of different forecast models. (3) The figures in bold relate to the best forecast model. (4) No star: Best linear models outperform the best machine learning model or have the same performance as it for the DM test *: Best machine learning model outperforms best standard linear model for the DM test at 10% level; **: 5% level; ***: 1% level. NN1 designates the neural network with one hidden layer that has 32 neurons. NN2 is the neural network with two hidden layers containing 32 and 16 neurons, respectively. NN3 is a neural network with three hidden layers having 32, 16, and 8 neurons, respectively.

Table 2.5 Out-of-sample RMSE of NCO's forecast models in specification (3) (all banks)

Model	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	h=9
Benchmark model	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*
Pooled OLS	0.62	0.68	**0.64**	**0.71**	**0.59**	0.66	0.66	0.71	**0.60**
Autoregressive model	2.05	1.93	2.08	3.20	2.05	1.04	1.17	3.08	1.27
Random walk	2.39	1.64	1.61	3.19	1.61	1.03	1.03	1.53	1.21
Lasso	1.69	2.83	0.86	0.84	1.16	1.17	0.80	**0.67**	0.80
Adaptive Lasso	2.09	7.99	0.84	**0.71**	1.20	0.94	0.66	**0.67**	0.69
Ridge	1.42	1.05	0.76	1.61	1.38	1.51	0.57	0.76	0.66
Elastic net	1.55	2.79	0.84	0.90	1.17	1.25	0.74	0.70	0.76
Principal component	1.16	1.10	0.98	2.06	1.07	0.79	0.83	1.04	0.87
NN1	5.20	2.57	2.63	10.84	2.50	1.33	1.42	2.78	1.47
NN2	4.05	1.99	2.04	8.45	1.94	1.04	1.12	2.14	1.17
NN3	5.20	2.59	2.63	10.84	2.50	1.33	2.78	0.91	1.49
Gradient boosting	**0.45****	0.61	0.75	1.13	0.68	0.48	0.55	0.84	0.81
Random Forest	0.47	**0.47****	0.69	1.26	0.65	**0.41*****	**0.54*****	0.80	0.62

Note: (1) Machine learning models include an extensive macroeconomic database and enriched BHC characteristics. (2) The first line represents the benchmark model's mean square error (MSE) normalized to 1, while the other lines present the relative MSE of different forecast models.(3) The figures in bold relate to the best forecast model. (4) no star: Best linear models outperform the best machine learning model or have the same performance as it for the DM test *: Best machine learning model outperforms best standard linear model for the DM test at 10% level; **: significance at 5%; ***: significance at 1%. NN1 designates the neural network with one hidden layer that has 32 neurons. NN2 is the neural network with two hidden layers containing 32 and 16 neurons, respectively. NN3 is a neural network with three hidden layers having 32, 16, and 8 neurons, respectively.

Table 2.6 Percentage variation of best ML relative to best SL model forecast accuracy of NCO under the three specifications (all banks)

Measure	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	h=9
$PC_1(h)$	-37.10	-33.82	-20.31	-36.62	-40.68	-53.03	-45.45	-49.30	-35.00
$PC_2(h)$	-54.84	-42.65	0.00	-45.07	-62.71	-74.24	-59.09	-45.07	-53.33
$PC_3(h)$	-27.42	-30.88	7.81	0.00	10.17	-37.88	-18.18	-5.63	0.00

Note: (1) $PC_j(h)$, j=1,2,3 denotes the percent variation of best ML relative to best SL forecast accuracy for different forecast horizons h and for specification j. A negative value means that using the best ML reduces the RMSE of the best SL of PC_j, improving its forecast accuracy. A positive value on contrary denotes a deterioration of forecast accuracy in case of using best ML model insted of best SL model.

2.5.1.2 Out-of-sample performance of PPNR's predictions models

Table 2.7, 2.8, 2.9, and 2.10 display the out-of-sample RMSE of models across specifications (1), (2), and (3), as well as the percentage variation of the best ML RMSE compared to the best SL RMSE for each specification. Table 2.7 indicates that the Adaptive Lasso, a linear ML model, demonstrates superior forecasting performance for PPNR across horizons 1 to 8, exhibiting the lowest RMSE. Its forecasting performance surpasses the leading SL models (Pooled OLS or Autoregressive) at a 1% significance level. However, in horizon 9, although the Adaptive Lasso's RMSE remains lower than that of the top-performing SL model, the performance difference is not statistically significant. Other ML models, excluding Principal Component, show potential for superior performance, with RMSE values under one across most forecast horizons. Nevertheless, these models still lag behind the best SL model across all horizons.

Table 2.8 underscores that by broadening the macroeconomic database beyond the Federal Reserve scenario, three machine learning models stand out across all nine forecast horizons: Adaptive Lasso, Random Forest, and Gradient Boosting. Consistently, Table 2.10 indicates that including additional macroeconomic information only boosts the accuracy of the top-performing ML model for the more distant horizons (5, 6, and 7). In these instances, the improvement in RMSE of the best ML model over the top SL models is more pronounced than in the first specification. Contrastingly, Table 2.9 and Table 2.10 reveal that integrating more bank balance sheet data can negatively impact forecast accuracy, leading to an increase in the Adaptive Lasso's RMSE along with other ML models, in comparison to specification (2). Moreover, under specification (3), the top SL model surpasses the best ML model in specific forecast horizons. These results underline the im-

portance of judicious variables and information source selection for each forecast horizon. While adding macroeconomic data can improve the performance of leading ML models in specific horizons, incorporating excessive bank-specific data can prove counterproductive.

Table 2.7 Out-of-sample RMSE of PPNR's forecast models under specification (1) (all banks)

Model	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	h=9
Benchmark model	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	1
Pooled OLS	0.53	0.50	0.51	0.53	0.51	0.53	0.55	0.63	0.60
Autoregressive model	4.24	1.61	1.31	1.58	0.72	0.53	0.53	0.63	0.40
Random walk	4.53	2.79	2.63	3.28	2.22	1.92	2.00	2.24	1.79
Lasso	0.47	0.45	0.47	0.47	0.47	0.46	0.47	0.54	0.50
Adaptive Lasso	**0.35*****	**0.34*****	**0.39*****	**0.39*****	**0.40*****	**0.40*****	**0.43*****	**0.40*****	**0.38**
Ridge	1.00	0.70	0.67	0.78	0.65	0.61	0.61	0.71	0.66
Elastic net	0.53	0.48	0.51	0.47	0.53	0.52	0.54	0.62	0.60
Principal component	4.47	2.75	2.47	2.67	1.99	1.69	1.76	1.83	1.60
NN1	1.12	0.75	0.71	0.72	0.71	0.73	0.74	0.79	0.87
NN2	1.24	0.80	0.78	0.75	0.72	0.67	0.54	0.90	0.73
NN3	1.47	0.77	0.80	0.78	0.74	0.62	0.71	0.92	0.56
Gradient boosting	0.82	0.41	**0.39**	0.58	0.53	0.80	0.51	0.63	0.49
Random Forest	0.53	0.45	0.47	0.58	0.46	0.43	0.49	0.54	0.44

Note: (1) Machine learning and standard linear models include the same variables FED uses in the bank's stress test. (2) The first line represents the benchmark model's mean square error (MSE) normalized to 1, while the other lines present the relative MSE of different forecast models. (3) The figures in bold relate to the best forecast model. No star: best linear models outperform the best machine learning model or have the same performance as it for the DM test, *: best machine learning model outperforms best standard linear model for the DM test at 10% level, **: 5% level; ***: 1% level. NN1 designates the neural network with one hidden layer that has 32 neurons. NN2 is the neural network with two hidden layers containing 32 and 16 neurons, respectively. NN3 is a neural network with three hidden layers having 32, 16, and 8 neurons, respectively.

Table 2.8 Out-of-sample RMSE of PPNR's forecast models under specification (2) (all banks)

Model	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	h=9
Benchmark model	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	1
Pooled OLS	0.53	0.50	0.51	0.53	0.51	0.53	0.55	0.63	0.60
Autoregressive model	4.24	1.61	1.31	1.58	0.72	0.53	0.53	0.63	0.40
Random walk	4.53	2.79	2.63	3.28	2.22	1.92	2.00	2.24	1.79
Lasso	0.53	0.45	0.76	0.50	0.40	0.41	0.44	0.45	0.41
Adaptive Lasso	**0.41*****	0.39	1.84	**0.39*****	**0.33*****	**0.34*****	0.68	**0.42*****	0.44
Ridge	1.02	0.70	1.00	0.78	0.70	0.63	0.62	0.70	0.65
Elastic net	0.59	0.66	1.41	0.50	0.43	0.52	0.67	0.60	0.42
Principal component	1.24	0.77	0.71	0.78	0.61	0.54	0.59	**0.42**	0.53
NN1	1.53	0.91	0.76	0.83	0.68	0.56	0.58	0.69	0.58
NN2	1.41	0.86	0.83	0.78	0.69	0.54	0.56	0.64	0.56
NN3	1.18	0.95	0.82	0.89	0.71	0.52	0.57	0.71	0.55
Gradient boosting	0.65	**0.34*****	**0.41*****	0.53	0.40	0.48	0.51	0.53	0.44
Random Forest	0.59	0.39	0.43	0.64	0.42	**0.34**	**0.39*****	0.46	**0.35*****

Note: (1) Machine learning models are estimated with an extensive macroeconomic database. (2) The first line represents the benchmark model's mean square error (MSE) normalized to 1, while the other lines present the relative MSE of different forecast models. (3) The figures in bold relate to the best forecast model. (4) No star: Best linear models outperform the best machine learning model or have the same performance as it for the DM test *: Best machine learning model outperforms best standard linear model for the DM test at 10% level; **: 5% level; ***: 1% level. NN1 designates the neural network with one hidden layer that has 32 neurons. NN2 is the neural network with two hidden layers containing 32 and 16 neurons, respectively. NN3 is a neural network with three hidden layers having 32, 16, and 8 neurons, respectively.

Table 2.9 Out-of-sample RMSE of PPNR's forecast models under specification (3) (all banks)

Model	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	h=9
Benchmark model	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	1
Pooled OLS	0.53	0.50	**0.51**	0.53	0.51	0.53	0.55	0.63	0.60
Autoregressive model	4.24	1.61	1.31	1.58	0.72	0.53	**0.53**	0.63	**0.40**
Random walk	4.53	2.79	2.63	3.28	2.22	1.92	2.00	2.24	1.79
Lasso	0.53	0.55	0.82	0.50	0.43	0.45	0.66	**0.51*****	0.45
Adaptive Lasso	**0.47*****	**0.45*****	1.88	**0.39*****	**0.39*****	**0.43*****	0.98	0.53	0.53
Ridge	1.03	0.80	1.10	0.78	0.74	0.70	0.70	0.80	0.80
Elastic net	0.59	0.75	1.45	0.50	0.44	0.64	0.78	0.69	0.46
Principal component	1.24	0.89	0.80	0.92	0.67	0.60	0.59	0.63	0.55
NN1	1.53	0.90	0.78	0.83	0.70	0.60	0.71	0.85	0.60
NN2	1.41	0.88	0.80	0.78	0.72	0.58	0.72	0.90	0.58
NN3	1.18	0.97	0.84	0.89	0.74	0.55	0.75	0.95	0.57
Gradient boosting	0.82	0.52	**0.51**	0.67	0.57	0.55	0.62	0.72	0.58
Random Forest	0.65	0.55	0.61	0.92	0.61	0.61	0.64	0.77	0.61

Note : (1) Machine learning models include an extensive macroeconomic database and enriched BHC characteristics. (2) The first line represents the benchmark model's mean square error (MSE) normalized to 1, while the other lines present the relative MSE of different forecast models. (3) The figures in bold relate to the best forecast model. (4) No star: Best linear models outperform the best machine learning model or have the same performance as it for the DM test *: Best machine learning model outperforms best standard linear model for the DM test at 10% level; **: significance at 5%; ***: significance at 1%. NN1 designates the neural network with one hidden layer that has 32 neurons. NN2 is the neural network with two hidden layers containing 32 and 16 neurons, respectively. NN3 is a neural network with three hidden layers having 32, 16, and 8 neurons, respectively.

Table 2.10 Percentage variation of the best ML relative to the best SL model forecast accuracy of PPNR under the three specifications (all banks)

Measure	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	h=9
$PC_1(h)$	-33.96	-32.00	-23.53	-26.41	-21.57	-24.53	-18.87	-36.51	-5.00
$PC_2(h)$	-22.64	-32.00	-19.61	-26.41	-35.29	-35.85	-26.41	-33.33	-12.50
$PC_3(h)$	-11.32	-10.00	0.00	-26.41	-23.53	-18.87	11.32	-20.63	12.50

Note : $PC_j(h)$, j=1,2,3 denotes the percent variation of best ML relative to best SL forecast accuracy for different forecast horizons h and for specification j. A negative value means that using the best ML reduces the RMSE of the best SL of PC_j, improving its forecast accuracy. A positive value on contrary denotes a deterioration of forecast accuracy in case of using best ML model.

2.5.1.3 Robustness analysis

This subsection evaluates the robustness of the top-performing model selection by focusing solely on the sub-sample of large BHCs selected for the empirical stress test.

Forecasting NCO on the sub-sample of large BHC selected for the empirical stress test

Tables 2.11, 2.12, and 2.13 delineate the out-of-sample forecasting performance of various models across the three specifications. On the other hand, Table 2.14 illustrates the percentage variation in RMSE between the top ML model and the most efficient SL model across these specifications. The Random Forest consistently emerges as the best forecasting model across all three specifications, significantly outperforming the Pooled OLS - the leading SL model - in nearly all forecast horizons. Interestingly, Table 2.14 reveals a more pronounced forecasting performance of the Random Forest relative to the best SL model when forecasting net charge-offs (NCO) for larger banks. For instance, in specification (1), when applied solely to the largest banks, Random Forest improves the RMSE of top SL model by an average of 48.74% , compared to a 39% improvement when applied to the entire sample. Similar patterns are observable in other specifications. One potential explanation could be the relatively poorer performance of the best SL model in forecasting NCO for larger BHCs. Furthermore, the average RMSE of the Random Forest model in specification (2) trails behind the best SL model's RMSE by 57.84% across some horizons. These findings imply that expanding macroeconomic information can enhance the forecasting accuracy of the top ML model for specific forecast horizons. As mentioned

previously, when considering the whole sample, broadening the BHC characteristics does not necessarily improve the performance of the best ML model, which underscores the need for a more targeted selection of predictors.

Table 2.11 Out-of-sample relative mean square error of NCO's forecast models under specification (1) (only banks selected for the empirical stress test)

Model	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	h=9
Benchmark model	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*
Pooled OLS	0.69	0.80	0.75	0.77	0.75	0.79	0.78	0.82	0.72
Autoregressive model	2.02	3.21	3.76	3.40	4.63	1.88	2.41	4.19	3.08
Random walk	2.38	1.81	1.97	1.91	1.68	1.04	0.99	1.58	2.58
Lasso	0.72	0.80	0.73	0.77	0.67	0.75	0.77	0.78	0.72
Adaptive Lasso	0.69	0.79	0.76	0.73	0.67	0.79	0.79	0.77	0.76
Ridge	0.62	0.71	0.55	0.91	0.73	0.66	0.57	0.74	0.82
Elastic net	0.72	0.71	0.72	0.73	0.67	0.76	0.75	0.78	0.72
Principal component	2.70	1.87	1.61	1.64	0.67	1.10	1.07	0.67	0.67
NN1	3.39	1.60	0.86	1.55	0.84	0.94	0.84	1.31	0.93
NN2	6.45	1.54	0.76	1.05	0.72	0.87	0.72	0.89	0.94
NN3	4.14	1.73	0.80	1.18	0.75	1.01	0.70	0.95	1.04
Gradient boosting	1.26	1.54	0.93	0.55	0.84	0.59	0.45	0.94	0.72
Random Forest	**0.25** ***	**0.49****	**0.59****	**0.30****	**0.31****	**0.25****	**0.30****	**0.49****	**0.54****

Note : (1) Machine learning models and standard linear models include the same variables that FED uses in the bank's stress test. (2) The first line represents the mean square error (MSE) of the benchmark model normalized to 1, while the other lines present the relative MSE of different forecast models. (3) The figures in bold relate to the best forecast model. No star : best linear models outperform the best machine learning model or have the same performance as it for DM test, *: best machine learning model outperforms best standard linear model for the DM test at 10% level , **: 5% level ; ***: 1% level. NN1 designates the neural network with one hidden layer that has 32 neurons. NN2 is the neural network with two hidden layers containing 32 and 16 neurons, respectively. NN3 is a neural network with three hidden layers having 32, 16, and 8 neurons, respectively.

Table 2.12 Out-of-sample relative mean square error of NCO's forecast models under specification (2) (only banks selected for the empirical stress test)

Model	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	h=9
Benchmark model	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*
Pooled OLS	0.69	0.80	**0.75**	0.77	0.75	0.79	0.78	0.82	0.72
Autoregressive model	2.02	3.21	3.76	3.40	4.63	1.88	2.41	4.19	3.08
Random walk	2.38	1.81	1.97	1.91	1.68	1.04	0.99	1.58	2.58
Lasso	2.77	6.36	1.52	0.59	1.81	2.45	1.02	0.54	1.00
Adaptive Lasso	3.33	2.16	1.25	0.50	1.92	7.70	0.83	0.63	0.88
Ridge	1.00	1.50	0.90	0.70	1.20	2.00	0.70	0.62	0.90
Elastic net	2.60	6.34	1.18	0.59	1.81	2.62	0.92	0.54	0.37
Principal component	1.31	1.33	1.06	1.64	1.04	0.78	0.75	0.74	0.94
NN1	3.50	2.00	1.10	1.00	1.30	1.10	0.70	1.00	0.97
NN2	7.00	1.70	1.00	0.90	1.20	1.20	0.74	0.75	0.99
NN3	5.14	1.90	1.30	0.95	1.00	1.40	0.77	0.77	1.10
Gradient boosting	0.54	1.97	1.62	0.44	0.31	0.25	0.31	0.46	0.53
Random Forest	**0.20** ***	**0.54****	1.13	**0.32****	**0.17****	**0.12****	**0.30****	**0.39****	**0.27****

Note : (1) Machine learning models are estimated with extensive macroeconomic database.(2) The first line represents the mean square error (MSE) of the benchmark model normalized to 1, while the other lines present the relative MSE of different forecast models. (3) The figures in bold relate to the best forecast model. (4) No star : Best linear models outperform the best machine learning model or have the same performance as it for the DM test *: Best machine learning model outperforms best standard linear model for the DM test test at 10% level ; **: 5% level ; ***: 1% level.NN1 designates the neural network with one hidden layer that has 32 neurons. NN2 is the neural network with two hidden layers containing 32 and 16 neurons, respectively. NN3 is a neural network with three hidden layers having 32, 16, and 8 neurons, respectively.

Table 2.13 Out-of-sample relative mean square error of NCO's forecast models under specification (3) (only banks selected for the empirical stress test)

Model	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	h=9
Benchmark model	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*
Pooled OLS	0.69	0.80	0.75	0.77	0.75	0.79	0.78	0.82	0.72
Autoregressive model	2.02	3.21	3.76	3.40	4.63	1.88	2.41	4.19	3.08
Random walk	2.38	1.81	1.97	1.91	1.68	1.04	0.99	1.58	2.58
Lasso	2.51	4.80	1.21	0.82	1.65	1.65	0.92	0.76	1.01
Adaptive Lasso	3.21	3.00	1.24	0.59	1.71	6.70	0.74	0.75	0.63
Ridge	0.98	1.70	0.89	0.69	1.10	1.50	0.60	0.70	0.70
Elastic net	2.36	7.31	1.18	0.86	1.67	1.78	0.79	0.80	0.94
Principal component	1.46	1.23	1.14	2.18	1.11	0.78	0.84	0.95	0.92
NN1	3.00	2.10	1.12	1.30	1.30	1.20	0.60	1.40	0.97
NN2	6.00	1.50	1.02	1.10	1.20	1.10	0.70	0.90	0.99
NN3	4.50	1.75	1.40	1.20	1.10	1.40	0.75	0.80	1.10
Gradient boosting	0.44	0.77	0.94	0.73	0.63	0.41	**0.46**	0.56	0.89
Random Forest	**0.24** ***	**0.49****	**0.70**	**0.44****	**0.53****	**0.32****	0.50	**0.52****	**0.56****

Note : (1) Machine learning models include extensive macroeconomic database and enriched BHC characteristics .(2) The first line represents the mean square error (MSE) of the benchmark model normalized to 1, while the other lines present the relative MSE of different forecast models.(3) The figures in bold relate to the best forecast model. (4) no star : Best linear models outperform the best machine learning model or have the same performance as it for the DM test *: Best machine learning model outperforms best standard linear model for the DM test test at 10% level; **: significance at 5% ; ***: significance at 1%.NN1 designates the neural network with one hidden layer that has 32 neurons. NN2 is the neural network with two hidden layers containing 32 and 16 neurons, respectively. NN3 is a neural network with three hidden layers having 32, 16, and 8 neurons, respectively.

Table 2.14 Percentage variation of best ML relative to best SL model forecast accuracy of NCO under the three specifications (only banks selected for the empirical stress test)

Measure	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	h=9
$PC_1(h)$	-63.77	-38.75	-21.33	-61.04	-58.67	-68.35	-61.54	-40.24	-25.00
$PC_2(h)$	-71.01	-32.50	20.00	-58.44	-77.33	-84.81	-61.54	-52.44	-62.50
$PC_3(h)$	-65.22	-38.75	-6.67	-42.86	-29.33	-59.49	-41.02	-36.58	-22.22

Note : $PC_j(h)$, j=1,2,3 denotes the percent variation of best ML relative to best SL forecast accuracy for different forecast horizons h and for specification j. A negative value means that using the best ML reduces the RMSE of the best SL of PC_j, improving its forecast accuracy. A positive value on contrary denotes a deterioration of forecast accuracy in case of using best ML model insted of best SL model.

Forecasting PPNR on the sub-sample of large BHC selected for the empirical stress test

As shown by the results from Table 2.15, the best forecast model of PPNR for the large banks throughout most of the horizons in the specification (1) is the Adaptive Lasso. The usefulness of additional macroeconomic information is limited to very few forecast horizons, in which relative improvement of the ML best model compared to the SL best model is higher than in the first specification (see Tables 2.16 and 2.18). Lastly, as revealed in Tables 2.17 and 2.18, the supplementary bank-specific data do not lead to forecast amelioration of PPNR relative to the previous specifications.

Table 2.15 Out-of-sample relative mean square error of PPNR's forecast models under specification (1) (only banks selected for the empirical stress test)

Model	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	h=9
Benchmark model	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*
Pooled OLS	0.37	0.39	1.02	0.40	0.37	0.38	0.39	0.47	0.45
Autoregressive model	3.79	2.61	2.44	3.20	1.95	1.59	0.61	1.94	1.55
Random walk	4.05	2.98	2.79	4.00	2.47	2.10	2.21	2.80	2.16
Lasso	0.32	0.34	0.33	0.37	0.33	0.34	0.35	0.40	0.39
Adaptive Lasso	**0.26*****	**0.26*****	**0.29*****	**0.31*****	0.29	0.31	**0.31****	**0.32*****	**0.31*****
Ridge	0.74	0.55	0.50	0.60	0.48	0.45	0.45	0.53	0.51
Elastic net	0.37	0.36	0.37	0.37	0.39	0.38	0.40	0.47	0.46
Principal component	4.16	2.98	2.75	3.46	2.21	1.87	1.97	2.35	1.88
NN1	0.89	0.61	0.60	0.51	0.47	0.73	0.63	0.61	0.70
NN2	1.05	0.41	0.65	0.51	0.50	0.45	0.50	0.70	0.63
NN3	1.74	1.70	0.70	0.70	0.55	0.70	0.71	0.80	0.50
Gradient boosting	0.53	0.27	0.27	0.43	0.35	0.54	0.36	0.46	0.35
Random Forest	0.31	0.29	0.31	0.40	**0.19*****	**0.29*****	0.35	0.40	0.34

Note : (1) Machine learning models and standard linear models include the same variables that FED uses in the bank's stress test. (2) The first line represents the mean square error (MSE) of the benchmark model normalized to 1, while the other lines present the relative MSE of different forecast models. (3) The figures in bold relate to the best forecast model. No star : best linear models outperform the best machine learning model or have the same performance as it for DM test, *: best machine learning model outperforms best standard linear model for the DM test at 10% level , **: 5% level ; ***: 1% level. NN1 designates the neural network with one hidden layer that has 32 neurons. NN2 is the neural network with two hidden layers containing 32 and 16 neurons, respectively. NN3 is a neural network with three hidden layers having 32, 16, and 8 neurons, respectively.

Table 2.16 Out-of-sample relative mean square error of PPNR's forecast models under specification (2) (only banks selected for the empirical stress test)

Model	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	h=9
Benchmark model	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*
Pooled OLS	0.37	0.39	1.02	0.40	0.37	0.38	0.39	0.47	0.45
Autoregressive model	3.79	2.61	2.44	3.20	1.95	1.59	0.61	1.94	1.55
Random walk	4.05	2.98	2.79	4.00	2.47	2.10	2.21	2.80	2.16
Lasso	0.39	0.30	0.63	0.37	0.29	0.30	0.32	0.36	0.34
Adaptive Lasso	**0.31****	**0.26*****	1.77	**0.28*****	**0.24*****	0.27	0.57	**0.34*****	0.44
Ridge	0.77	0.55	0.50	0.58	0.45	0.45	0.47	0.50	0.65
Elastic net	0.41	0.43	1.23	0.38	0.42	0.43	0.48	0.51	0.51
Principal component	1.16	0.50	0.81	1.00	0.95	0.61	1.97	0.39	0.60
NN1	0.92	0.61	0.80	0.47	0.40	0.63	0.65	0.57	0.55
NN2	1.20	0.41	0.75	0.45	0.43	0.35	0.55	0.70	0.60
NN3	1.85	1.70	0.90	0.60	0.50	0.60	0.72	0.74	0.53
Gradient boosting	0.46	0.33	**0.29*****	0.36	0.42	0.33	0.35	0.36	0.32
Random Forest	0.36	**0.26*****	0.31	0.43	0.36	**0.23*****	**0.28*****	0.35	**0.28*****

*Note : (1) Machine learning models are estimated with extensive macroeconomic database.(2) The first line represents the mean square error (MSE) of the benchmark model normalized to 1, while the other lines present the relative MSE of different forecast models. (3) The figures in bold relate to the best forecast model. (4) No star : Best linear models outperform the best machine learning model or have the same performance as it for the DM test *: Best machine learning model outperforms best standard linear model for the DM test test at 10% level ; **: 5% level ; ***: 1% level. NN1 designates the neural network with one hidden layer that has 32 neurons. NN2 is the neural network with two hidden layers containing 32 and 16 neurons, respectively. NN3 is a neural network with three hidden layers having 32, 16, and 8 neurons, respectively.*

Table 2.17 Out-of-sample relative mean square error of PPNR's forecast models under specification (3) (only banks selected for the empirical stress test)

Model	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	h=9
Benchmark model	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*
Pooled OLS	0.37	0.39	1.02	0.40	0.37	0.38	0.39	0.47	0.45
Autoregressive model	3.79	2.61	2.44	3.20	1.95	1.59	0.61	1.94	1.55
Random walk	4.05	2.98	2.79	4.00	2.47	2.10	2.21	2.80	2.16
Lasso	0.35	0.35	0.69	0.37	0.36	**0.30*****	0.46	**0.39*****	**0.36*****
Adaptive Lasso	**0.30*****	**0.30*****	1.75	**0.29*****	**0.25*****	**0.30*****	0.77	0.41	0.50
Ridge	0.75	0.59	0.50	0.55	0.45	0.48	0.57	0.55	0.70
Elastic net	0.39	0.43	1.19	0.37	0.29	0.39	0.48	0.50	**0.36*****
Principal component	1.16	0.91	0.87	1.09	0.72	0.70	0.64	0.76	0.60
NN1	0.90	0.70	0.81	0.50	0.50	0.65	0.70	0.70	0.55
NN2	1.10	0.55	0.75	0.60	0.40	0.37	0.59	0.75	0.60
NN3	1.80	1.60	0.90	0.65	0.60	0.62	0.72	0.79	0.53
Gradient boosting	0.84	0.37	**0.31*****	0.43	0.32	0.32	**0.36**	0.43	0.37
Random Forest	0.47	0.39	0.49	0.77	0.47	0.43	0.52	0.70	0.55

Note : (1) Machine learning models include extensive macroeconomic database and enriched BHC characteristics .(2) The first line represents the mean square error (MSE) of the benchmark model normalized to 1, while the other lines present the relative MSE of different forecast models.(3) The figures in bold relate to the best forecast model. (4) no star : Best linear models outperform the best machine learning model or have the same performance as it for the DM test *: Best machine learning model outperforms best standard linear model for the DM test test at 10% level; **: significance at 5% ; ***: significance at 1%. NN1 designates the neural network with one hidden layer that has 32 neurons. NN2 is the neural network with two hidden layers containing 32 and 16 neurons, respectively. NN3 is a neural network with three hidden layers having 32, 16, and 8 neurons, respectively.

Table 2.18 Percentage variation of best ML relative to best SL model forecast accuracy of PPNR under the three specifications (only banks selected for the empirical stress test)

Measure	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	h=9
$PC_1(h)$	-29.73	-33.33	-71.57	-22.50	-48.65	-23.68	-20.51	-31.91	-31.11
$PC_2(h)$	-16.21	-33.33	-71.57	-30.00	-35.14	-39.47	-28.20	-27.66	-37.78
$PC_3(h)$	-18.92	-23.08	-69.61	-27.50	-32.43	-21.05	-7.69	-17.02	-20.00

Note : $PC_j(h)$, j=1,2,3 denotes the percent variation of best ML relative to best SL forecast accuracy for different forecast horizons h and for specification j. A negative value means that using the best ML reduces the RMSE of the best SL of PC_j, improving its forecast accuracy. A positive value on contrary denotes a deterioration of forecast accuracy in case of using best ML model insted of best SL model.

2.5.2 Empirical stress test

In this subsection, I use results about forecasting to undertake a pseudo-stress test under severely adverse scenarios. The jump-off date for the conditional forecast is 2018q4, and the stress test period evaluation spans from 2019q1 to 2021q1. The bank's specific forcing variables are assumed to hold constant values equal to their pre-stress levels (2018q4) over the stress horizon. In turn, macroeconomic forcing variables follow the trajectories defined by the FED in its severely adverse scenarios (Table 2.19).

Table 2.19 Severely adverse scenarios in 2019q1-2021q1

Date	GDP	UNRATE	Inflation rate	TB3MS	GS5	GS10	Mortgagerate	SP500	USSTHPI	VXOCLS
2018q4	1.31	3.83	1.29	2.32	2.88	3.03	4.78	-22.95	-0.18	43.1
2019q1	-5.0	4.7	1.2	0.3	0.3	0.8	3.9	-36.6	-3.0	44.4
2019q2	-9.4	6.3	1.6	0.2	0.5	0.9	4.2	-19.4	-3.1	43.1
2019q3	-7.2	7.5	1.7	0.1	0.6	1.0	4.4	-9.8	-3.7	39.2
2019q4	-5.0	8.4	1.8	0.1	0.6	1.1	4.5	-3.5	-4.4	34.9
2020q1	-3.8	9.2	1.9	0.1	0.7	1.2	4.3	4.6	-4.6	30.5
2020q2	-1.5	9.7	1.8	0.1	0.7	1.2	4.2	6.9	-4.2	27.3
2020q3	-0.3	10.0	2.0	0.1	0.7	1.2	4.1	7.1	-4.4	25.3
2020q4	2.9	9.9	2.0	0.1	0.7	1.2	3.9	8.5	-2.6	23.5
2021q1	3.6	9.7	2.1	0.1	0.9	1.5	3.9	5.4	-0.7	22.5

2.5.2.1 Predicting capital adequacy measure under hypothetical severely adverse scenarios

In this section, I run the stress test using two distinct models. The first model (Best ML), which acts as the baseline, employs the Adaptive Lasso and Random Forest - the top ML forecasting models - to construct conditional predictions of PPNR and NCO. Subsequently, these predictions are used to derive T1CR. The second model (FE-LM) relies on a linear framework - specifically, a fixed-effect linear model - to forecast both banking variables (PPNR and NCO), and, subsequently, T1CR. Table 2.20 presents the minimum predicted T1CR values throughout the stress test horizon for both models and, for comparison, includes the minimum historical T1CR values during the Great Recession (G.R). The " All " row reflects aggregate T1CR, calculated as the ratio of aggregate common equity capital to aggregate risk-weighted assets.

All the banks in the two frameworks meet the minimum T1CR threshold of 4.5%, thus passing the stress test. However, during the last global financial crisis, two banks - Bancwest Corporation and PNC Financial Services INC - would have fallen below this threshold, indicating a potential failure in the stress test. While the mandatory rule for minimum T1CR requirement was not in effect during the crisis, these results highlight the relatively low capitalization of banks prior to the Great Recession. For example, the average T1CR before the crisis was 7.13, nearly half its average value just one quarter before the stress test period. Post-crisis, Basel III reforms have required banks to improve their capital positions.[13] One such reform, the capital conservation buffer, has been enforceable since 2019

[13]These reforms raised the minimum requirement for different regulatory capital and introduced other capital buffers, improving the capital position of banks.

and aims to add an extra layer of resilience to the banking system in the face of unexpected and significant losses. The minimum capital buffer was set at 2.5% of total risk assets on top of the common equity tier 1 capital (CET1), raising the minimum T1CR for large BHCs subject to a stress test to 7%. Since BHCs subject to stress tests are large banks, many also meet the 7% minimum T1CR requirement for both methods throughout the stress test horizon. The aggregate T1CR reaches a minimum of 7.70% for the Best ML and 8.22% for the FE-LM model, comfortably above the 7% treshold. However, the impact of severely adverse conditions is not uniform across all banks for both models. For instance, Northern Trust Corporation is less affected by the simulated crisis scenarios than Bank of America Corporation for the Best ML model. Conversely, JP Morgan Chase and Company experiences a more pronounced impact from the simulated adverse economic conditions than Well Fargo in the FE-LM model. Business model differences could explain the crisis's disparate effects on these banks.

Table 2.20 Predicted T1CR and historical values

Bank	min T1CR on the stress horizon		$T1CR_{2018q4}$	min T1CR for G.R	$T1CR_{2007q3}$
	Best ML	FE-LM			
Bank of America Corporation	7.44	8.28	13.66	6.17	6.63
Bancwest Corporation	6.53	4.50	11.01	3.95	4.62
BBVA USA Bancshares INC	5.90	7.83	12.05	7.81	8.56
Citigroup INC	7.36	10.44	14.83	6.10	7.13
Comerica INC	7.85	6.61	10.25	6.70	6.96
Fifth Third Bancorp	7.77	8.97	11.25	5.82	6.57
JP Morgan Chase and Company	7.61	6.01	13.67	7.25	7.42
Keycorp	6.05	5.05	10.56	5.94	6.43
Huntington Bancshares INC	8.70	4.55	10.63	6.22	7.24
Northern Trust Corporation	13.48	10.44	14.50	8.57	9.46
PNC Financial Services INC	7.61	8.37	12.01	4.47	6.28
State Street Corporation	17.30	12.39	17.55	9.06	11.61
Well Fargo	7.62	9.01	13.48	6.29	7.68
All	7.70	8.22	13.51	6.64	7.13

2.5.2.2 The Analysis of banks' vulnerabilities : Best ML versus FE-LM modeling

In the following analysis, I will delve deeper into the T1CR distribution by also focusing on the range of possibilities that it represents for each bank. Table 2.21 details the probability of each bank violating specific T1CR thresholds in 2021q1, according to both models. The "All" row represents the probability of the aggregate T1CR falling below-given thresholds for both models, effectively capturing the average risk of violation for all banks. As the results indicate, there is, on average, no risk of the banks not meeting the 4.5% minimum T1CR threshold according to both models. However, this average needs to capture the nuances across different banks and models. For example, the Best ML model indicates a 9% chance of Bank of America violating the 4.5% minimum requirement, while it shows a 0% chance of such a violation for State Street Corporation. Similarly, Citigroup INC has a 7% risk of not meeting the rule according to the Best ML model, but this risk drops to zero with the FE-LM model. When considering the 7% minimum requirement, the picture changes somewhat. Only a few banks are predicted to have no chance of violating

this threshold. Nevertheless, the overall probability of violating the 7% requirement is estimated at 2% for both models. This result aligns with the findings of Covas et al. (2014), who estimated a 2.5% chance of breaking an 8% T1CR minimum requirement for both the fixed effect quantile and fixed effect linear models. This analysis suggests that at the 4.5% threshold, most banks have a negligible risk of failing to meet the minimum T1CR rule according to both models. However, at 7%, there is a generally higher risk of banks not meeting the requirement.

Table 2.21 Probability of violating different tresholds in 2021q1

Bank	minimum T1CR requirement: 4.5%		minimum T1CR requirement: 7%	
	Best ML	FE-LM	Best ML	FE-LM
Bank of America Corporation	0.09	0.00	0.45	0.01
Bancwest Corporation	0.00	0.51	0.23	0.93
BBVA USA Bancshares INC	0.01	0.07	0.46	0.39
Citigroup INC	0.07	0.00	0.44	0.00
Comerica INC	0.06	0.00	0.39	0.38
Fifth Third Bancorp	0.07	0.00	0.40	0.00
JP Morgan Chase and Company	0.04	0.00	0.32	0.38
Keycorp	0.00	0.25	0.44	0.92
Huntington Bancshares INC	0.00	0.57	0.01	0.97
Northern Trust Corporation	0.00	0.00	0.00	0.00
PNC Financial Services INC	0.08	0.00	0.47	0.09
State Street Corporation	0.00	0.00	0.00	0.00
Well Fargo	0.08	0.00	0.39	0.00
All	0.00	0.00	0.02	0.02

Note : The probability to violate the given treshold is the probability of having T1CR strictly below that value.

Table 2.22 provides a more in-depth look at the two models' conditional distribution of T1CR at the end of the stress test horizon. The probability density function (PDF) of T1CR is left-skewed, indicating an asymmetric distribution. This pattern is consistent with some studies (such as Covas et al. 2014; Hirtle et al. 2016) that find a left-skewed T1CR PDF under crisis scenarios. A closer examination reveals that for almost all global systemically important banks (G-SIBs)[14], the left tail of the T1CR distribution is more substantial for the Best ML model than the FE-LM model. This fact is visually evident in Figure 2.3, which plots the probability densities of T1CR for the four G-SIBs. Covas et al. (2014) found a similar phenomenon, where the left tail of the T1CR at the end of their

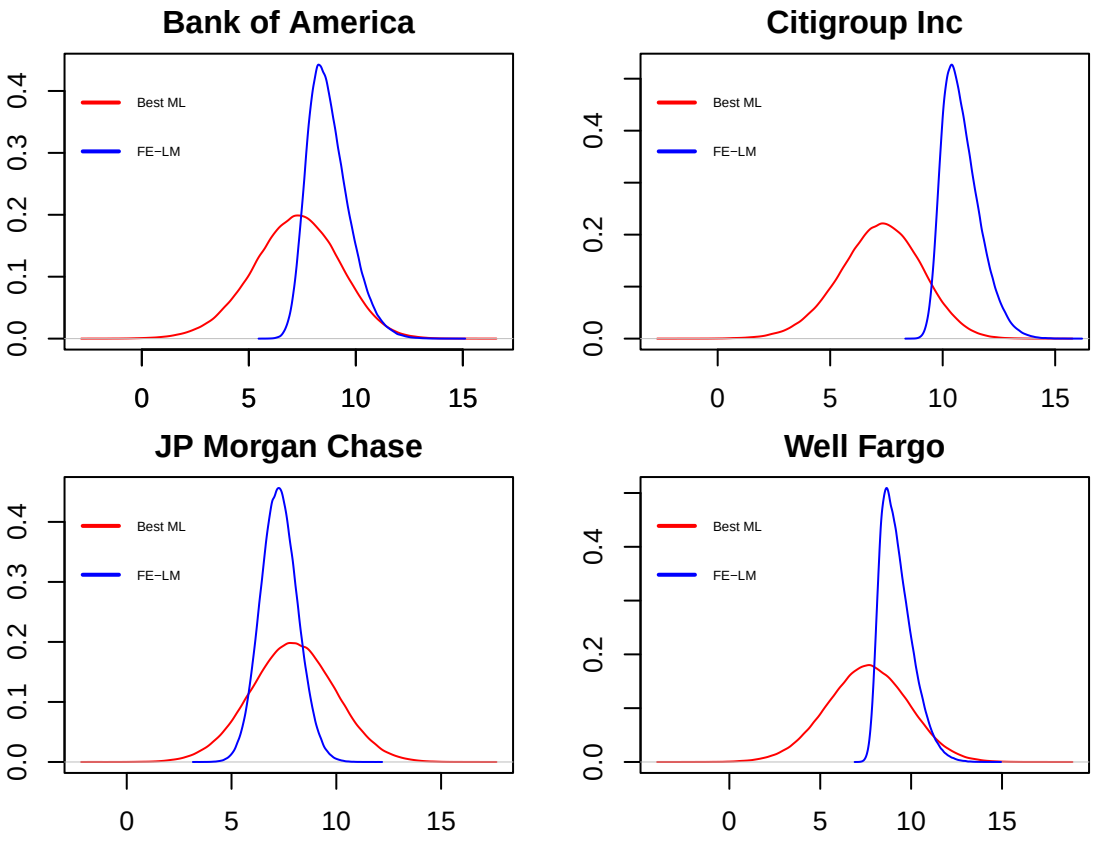
[14] A global systemically important bank is a financial institution whose failure increases systemic risk and may trigger a financial crisis. The following banks in our sample figure out in the list for the year 2019 of the globally systemically banks established annually by the Financial Stability Board in coordination with the Basel committee: Bank of America Corporation, Citigroup INC, JP Morgan Chase and Company, State Street Corporation, and Well Fargo. Bank of America Corporation, Citigroup INC, JP Morgan Chase and Company, and Well Fargo are the larger banks in our sample, weighing 80% of the sampled bank's total assets.

stress-test period (2013q4) was denser for their nonlinear baseline model than for the fixed effect linear model for most banks. One plausible explanation for this could be a heavier right tail in the conditional distribution of NCO under the Best ML model compared to the FE-LM model, as shown in Figure 2.4. When looking at PPNR, another component of T1CR, Figure 2.5, does not depict a distinct difference between the left tail distributions for the two models. These findings suggest that the linear framework may underestimate the risk of significant losses for G-SIBs during economic downturns compared to a framework combining nonlinear models for loss predictions and linear models for revenue predictions. Consequently, linear models might paint a more optimistic picture of banking stability and minimize the perceived systemic risk during economic downturns.

Table 2.22 Predicted conditional percentiles of T1CR in the stress test horizon in 2021q1

Bank	1^{st} percentile		5^{th}percentile		Median		95^{th} percentile	
	Best ML	FE-LM	Best ML	FE-LM	Best ML	FE-LM	Best ML	FE-LM
Bank of America Corporation	-0.34	6.73	7.40	7.63	8.55	8.95	10.02	10.36
Bancwest Corporation	6.24	1.05	6.72	2.87	7.43	4.58	8.47	7.31
BBVA USA Bancshares INC	4.43	-0.34	6.09	7.05	6.79	8.38	9.16	9.99
Citigroup INC	-0.40	9.30	7.40	9.79	8.47	10.98	9.89	12.27
Comerica INC	-0.70	4.95	7.35	6.09	8.32	7.04	9.80	8.23
Fifth Third Bancorp	-0.34	7.05	7.38	7.85	8.39	9.04	9.99	10.36
JP Morgan Chase and Company	-0.50	4.95	7.46	6.16	8.58	7.17	10.36	9.06
Keycorp	4.47	2.18	6.03	3.80	6.79	6.10	9.39	7.61
Huntington Bancshares INC	6.75	1.49	7.63	3.02	8.79	4.52	10.33	6.84
Northern Trust Corporation	11.10	9.30	12.03	9.79	13.60	10.98	15.34	12.26
PNC Financial Services INC	-0.40	4.41	7.39	7.42	8.39	8.68	9.96	10.33
State Street Corporation	12.43	9.06	14.02	11.59	17.44	12.96	20.02	15.53
Well Fargo	-0.60	7.54	7.46	8.08	8.56	9.29	10.30	10.64
All	6.74	6.74	7.29	7.40	8.87	9.42	12.25	12.40

Figure 2.3 Probability Densities of T1CR for the largest G-SIBs bank in 2021q1



Note : These figures show the estimated conditional skewed t PDF at the end of the stress test period for the four largest banks of the sample.

2.5.2.3 Predicting capital shortfalls under distressed macroeconomic and financial conditions

To ascertain the capital gap for a given threshold, I quantify the discrepancy between the average predicted capital and the indispensable minimum capital required to prevent the breach of the specified criteria. I establish the aggregate capital shortfall by compiling

Figure 2.4 Probability Densities of NCO for the largest G-SIBs banks in 2021q11

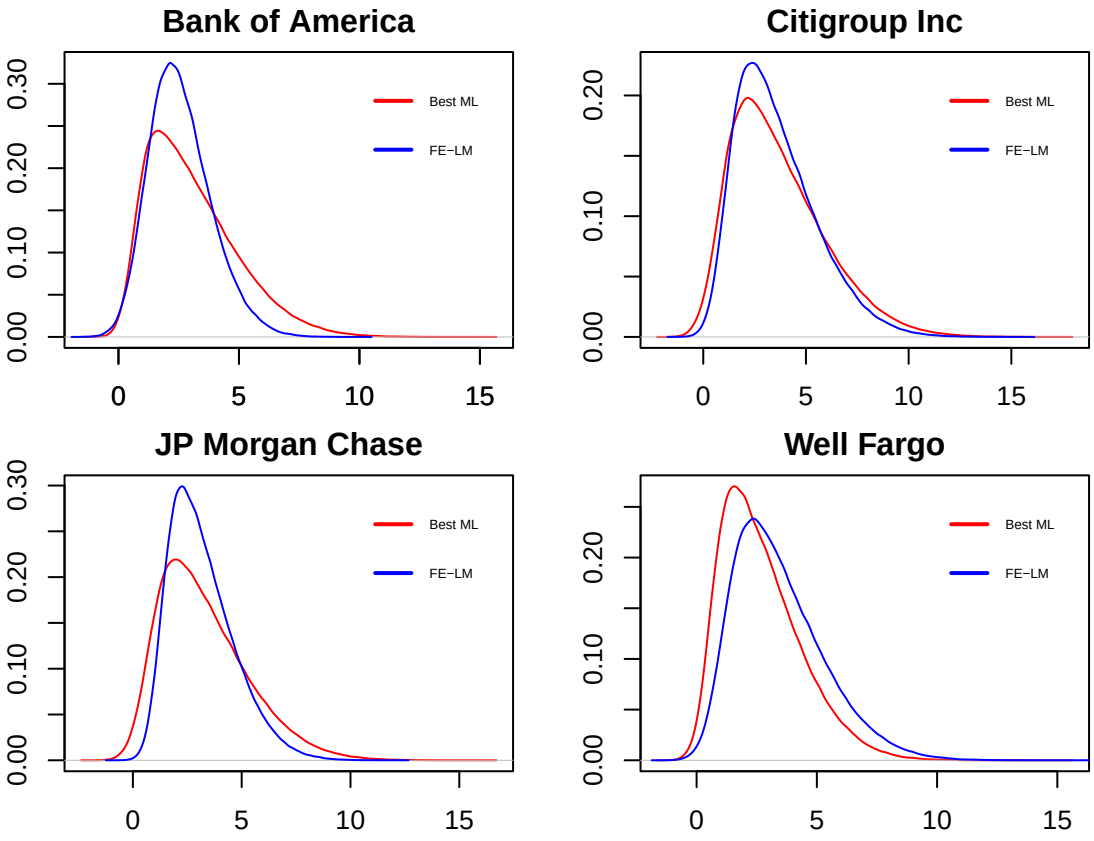
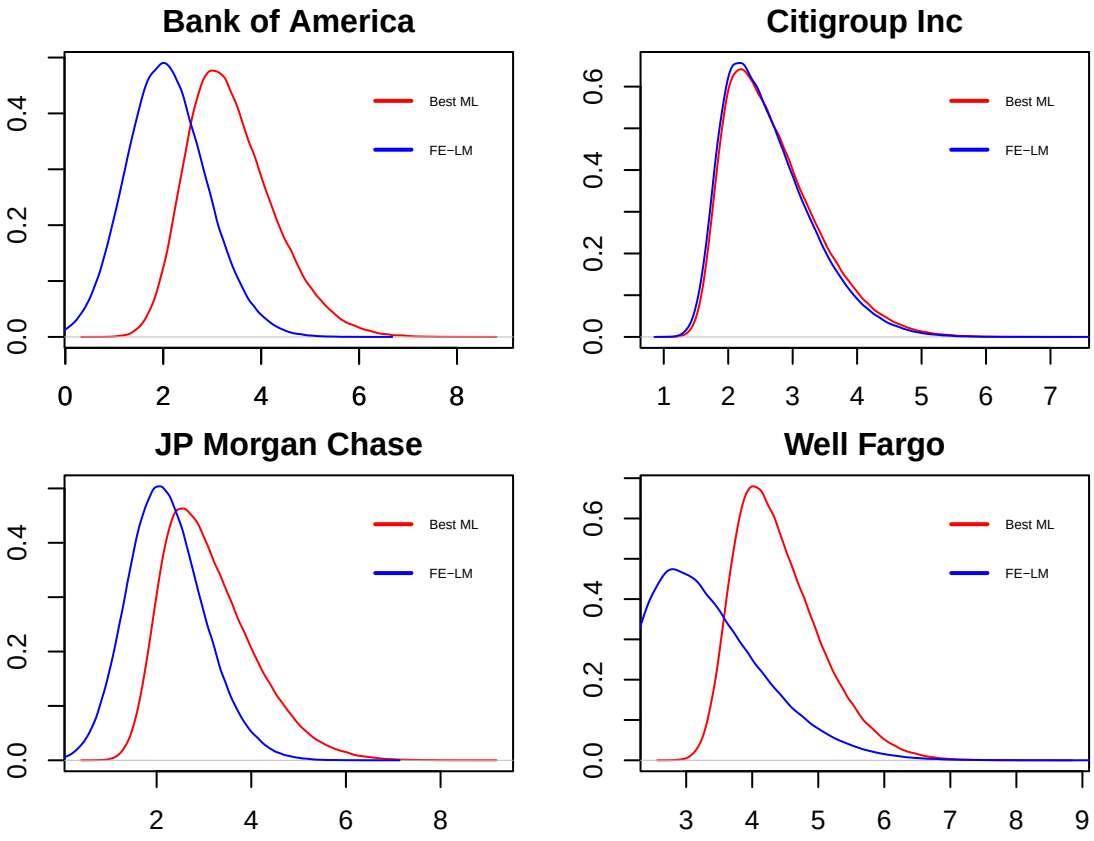


Figure 2.5 Probability Densities of PPNR for the largest G-SIBs banks in 2021q1



the capital gaps across all banks. This calculation is illustrated in Table 2.23. At the end of the stress assessment period, with a minimum capital requirement set at 4.5%, there is no aggregate accumulated capital gap. However, the aggregate capital shortfall reached 1 billion USD during the Great Recession. Raising the threshold to 7% significantly inflates the aggregate cumulative capital shortfall to 2 billion USD under the Best ML model and a staggering 28 billion USD under the FE-LM model. In stark contrast, the capital gap surged to 85 billion USD under the same 7% threshold during the Great Recession. These results underscore the enhanced stability of the financial system post-Great Recession. This improvement is attributable, in part, to more rigorous regulations and increased frequency of financial oversight exercises. Thus, this analysis offers a robust argument for the efficacy of regulatory reform in mitigating financial system vulnerabilities.

Table 2.23 The cumulated capital gap at the end of the Great Recession[c](2009q2) and the stress test period (2021q1) (billion of USD dollars)

Bank	Stress test horizon				Great Recession	
	4.5% T1CR requirement		7% T1CR requirement		4.5% T1CR requirement	7% T1CR requirement
	Best ML	FE-LM	Best ML	FE-LM		
Bank of America Corporation	0.00	0.00	0.00	0.00	0.00	22.30
Bancwest Corporation	0.00	0.00	0.06	0.90	1.02	11.50
BBVA USA Bancshares INC	0.00	0.00	0.95	0.00	0.00	0.00
Citigroup INC	0.00	0.00	0.00	0.00	0.00	17.90
Comerica INC	0.00	0.00	0.00	0.26	0.00	0.65
Fifth Third Bancorp	0.00	0.00	0.00	0.00	0.00	4.46
JP Morgan Chase and Company	0.00	0.00	0.00	15.20	0.00	0.00
Keycorp	0.00	0.00	1.52	4.66	0.00	3.29
Huntington Bancshares INC	0.00	0.00	0.00	7.40	0.00	0.66
Northern Trust Corporation	0.00	0.00	0.00	0.00	0.00	0.00
PNC Financial Services INC	0.00	0.00	0.00	0.00	0.03	9.63
State Street Corporation	0.00	0.00	0.00	0.00	0.00	0.00
Well Fargo	0.00	0.00	0.00	0.00	0.00	15.30
All	0.00	0.00	2.53	28.42	1.05	85.69

Note : During the Great Recession, T1CR requirements were not set. We use these thresholds to evaluate the capital gap prevailing at that period if the conditions were applicable.

2.5.2.4 Comparing predicted T1CR PDF accuracy between Best ML and FE-LM models during the Great Recession

I use two evaluation methods to compare the two models: interval forecast evaluation and direct density forecast evaluation. In particular, I concentrate on one quarter-ahead forecast.

Interval forecast evaluation

Evaluating interval forecasts to rank predicted densities is a common approach in the literature (see Gneiting and Raftery 2007, Goulet Coulombe et al. 2022). I use three criteria: empirical coverage rate, average interval length, and average interval score.

Empirical Coverage Rate: This measures how often the actual value of T1CR falls within

the predicted interval, calculated as $(1-\alpha)100\%$, where α signifies the significance level. The predictive quantiles at level $\frac{\alpha}{2}$ and $1-\frac{\alpha}{2}$ denote the lower and upper limits of the interval, labeled as l and u, respectively. I set α at 0.1 for my evaluation. A higher (or lower) coverage rate than 90% implies that the predicted interval is overly wide (or narrow). The most accurate model maintains an empirical coverage rate closest to 90%.

Average Interval Length: This is the difference between the upper and lower limits of the interval. I ascertain the average length across all forecasted intervals.

Average Interval Score: Introduced by Gneiting and Raftery (2007), this criterion gauges the quality of interval forecasts, tracing its roots back to **?**. The interval score, denoted as $S_{\alpha}^{int}(l,u,x)$, where x is a realized value of T1CR, is calculated as follows:

$$S_{\alpha}^{int}(l,u,x) \quad = (u-l) + \tfrac{2}{\alpha}(l-x)1x < l + \tfrac{2}{\alpha}(x-u)1x > u. \tag{2.9}$$

Here, 1(.) represents the indicator function. According to this metric, the superior model yields the smallest score value. Intuitively, the interval score encourages a narrower interval forecast, rewarding precision and accuracy. This intuition can be traced to Engle (1982), who advocated for dynamic interval forecasts surrounding point predictions. He argued that these intervals should contract during stable periods and expand during turbulent times due to the inherent stability leading to more accurate forecasts. The score also penalizes forecasters for missing the interval, with the severity of the penalty varying with α.

When comparing models, the Best ML approach more accurately approximates the T1CR density during the 2007-2009 financial crisis in terms of empirical coverage rate and aver-

age interval score, as illustrated in Table 2.24. However, it is worth noting that the FE-LM model exhibits a shorter average interval length than the Best ML model. This observation implies that, on average, the FE-LM model predicts a more narrowly distributed left tail of the T1CR density during the Great Recession compared to the Best ML model. Thus, while both models have their strengths, the Best ML model's superior performance in both empirical coverage rate and average interval score suggests it is generally more effective in modeling and predicting T1CR density during financial crises. However, the thinner left tail produced by the FE-LM model can be advantageous under certain circumstances and should be noticed.

Table 2.24 Comparison of one step ahead 90% interval forecasts between ML and linear modeling)

Model	Empirical coverage rate(%)	Average interval lenght	Average interval score
Best ML modeling	80.95	3.34	7.59
FE-LM modeling	50.00	2.20	9.37

Direct density forecast evaluation

Guided by Goulet Coulombe et al. (2022) methodology, I employ a scoring method to gauge the accuracy of density forecasts. The score is given as:

$$S_{i,t+1} \quad = -log(\hat{f}(y_{i,t+1}/X_{i,t})). \tag{2.10}$$

In this equation, $\hat{f}(y_{t+1})$ represents the predicted conditional density, evaluated at the realized value $y_{i,t+1}$. The indices i and t are the cross-sectional and time dimensions of the panel, respectively, while $X_{i,t}$ is the set of information available at quarter t. The model with the lowest average score is considered the most accurate, as a lower score signals

a higher mass closer to the realized value. Table 2.24 reveals the average score for each selected bank during the 2007-2009 financial crisis, showing the Best ML modeling with a lower average score than the FE-LM modeling. To assess the comparative ranks of the two models, I employ the Diebold-Mariano test (DM test), modifying the definition of the loss differential as follows:

$$d_{i,t} \quad = S_{i,t}^{ML} - S_{i,t}^{LIN}; \tag{2.11}$$

where $S_{i,t}^{ML}$ and $S_{i,t}^{LIN}$ denote the scores for the Best ML and FE-LM modeling for bank i at quarter t, respectively. Following Iacopini et al. (2023) interpretation, $d_{i,t}$ serves the same function as in the original DM test. I test the null hypothesis of equal forecast accuracy $H0 : E(di, t)$=0 against the alternative hypothesis $H_1 : E(d_{i,t})$<0, favoring the Best ML model. I perform this test by regressing $d_{i,t}$ on a constant, where a zero constant corresponds to H_0, and a negative constant signifies the superiority of the Best ML model over FE-LM modeling for predicting T1CR density during the Great Recession. Table 2.25 displays a significantly negative constant at the 1% significance level, suggesting that the Best ML model is superior to the FE-LM model in approximating T1CR density during the 2007-2009 financial crisis.

Table 2.25 Score for selected stressed BHC throughout the 2007-2009 financial crisis

Bank holding companes	Score for ML modeling (S_{ML})	Score for linear modeling (S_L)	Difference (S_{ML}-S_L)
Bank of America Corporation	2.17	77.84	-75.67
Bancwest corporation	1.47	8.21	-6.74
BBVA USA Bancshares Inc	1.18	39.94	-38.76
Citigroup INC	3.40	52.15	-48.75
Comerica Inc	3.37	127.73	-124.36
Fifth Third Bancorp	1.96	38.00	-36.04
JP Morgan chase and company	1.58	29.99	-28.41
Keycorp	2.20	8.23	-6.03
Huttington Bancshares Inc	1.99	7.19	-5.20
Northern trust corporation	1.86	2.19	-0.33
PNC Financial Services INC	1.50	27.25	-25.76
State street corporation	2.47	2.65	-0.18
Wells Fargo et Company	1.23	26.18	-24.95
All	2.03	37.04	**-35.01*****

Note : *** indicates that the ML modeling is better than the Linear modeling in predicting T1CR densities at 1% significant level.

2.6 Conclusion

This paper thoroughly explores the benefits of integrating machine learning (ML) methodologies with comprehensive macroeconomic and bank-specific data, intending to optimize risk analysis during stress tests. The research unfolds in two sequential stages, yielding substantive findings contributing to our risk analysis understanding.

In the first stage, I establish the superior efficacy of Random Forest, a nonparametric and nonlinear ML technique, over Pooled OLS and other ML models in predicting Net Charge-Offs (NCO) across nine forecast horizons. This finding resonates with Covas et al. (2014)'s emphasis on the nonlinear relationship between bank losses and predictors and further elaborates on the utility of Random Forest in enhancing the Relative Mean Squared Error (RMSE) of linear models in NCO forecasting for stress tests. Nonetheless, the predictive power of NCO forecasting was further improved, by approximately 17%, across six of nine forecast horizons when I expanded the macroeconomic database. Conversely, sup-

plementing bank balance sheet information has no beneficial impact and, in some cases, impedes the predictive capabilities of Random Forest. Regarding Pre-Provision Net Revenue (PPNR), my findings lean towards Adaptive Lasso, a linear ML model that outperforms standard linear models across all forecast horizons. The linear relationship between PPNR and predictors persists, yet the regularization imparted by Adaptive Lasso triggers a notable improvement over linear models. While using a larger macroeconomic database enhances the best ML models' performance on horizons 5, 6, and 7, additional bank balance sheet information had a detrimental effect on the performance of ML models.

A critical revelation of the paper lies in analyzing the Tier 1 Capital Ratio (T1CR) distribution in the second stage. The left tail of the T1CR distribution is considerably heavier when ML models are employed compared to linear models, especially for globally systemically important banks. Moreover, ML models offer a superior approximation of T1CR density, particularly under strained macro-financial conditions. These findings underline the risk of over-optimism and systemic risk underestimation when exclusively relying on linear models. Thus, ML models, with their ability to incorporate nonlinearity and complexity, provide a more holistic perspective on the interconnectedness of the banking sector and potential systemic risks.

This paper underpins the benefits of ML models in enhancing stress test risk analysis and asserts the significance of an extensive macroeconomic database for NCO and PPNR forecasts. Nonetheless, the study does not delve into the direct influence of additional predictors on the accuracy of T1CR density forecast, paving the way for future research to pinpoint additional predictors for NCO and PPNR models and assess their implications for T1CR forecasts. This avenue of investigation can deepen our understanding of the direct effects of these predictors on the efficacy of stress test methodologies.

CHAPTER III

BANK-LEVEL UNCERTAINTY AND THE BUSINESS CYCLE: EVIDENCE FROM LARGE US BANK HOLDING COMPANIES

ABSTRACT

This paper introduces a novel method to quantify bank-level uncertainty from forecast errors of a bank's return on asset (ROA) derived from an ensemble of machine learning models combined with granular bank data and an extensive macroeconomic dataset. From various ROA forecasts, I compute the optimal forecast errors from the average prediction across models. Then, I define the bank-level uncertainty measure as the standard deviation of these forecast errors. Using Vector Autoregression (VAR) analysis, the paper shows that this measure significantly influences both business cycles and credit markets. Unanticipated spikes in bank-level uncertainty lead to notable economic downturns and worsened credit conditions, exceeding the predictive power of conventional macroeconomic and financial uncertainty metrics. These findings advocate for targeted regulatory action in the banking sector to enhance financial stability.

Keywords: Big data, Bank, Uncertainty, Fluctuation, Business cycle.
JEL classification: D81, C55, E32, E44, G14.

3.1 Introduction

The concept of *economic uncertainty* , defined by Watkins (1922) as "people's inability to forecast the likelihood of events happening," has been a prominent topic in economic literature since the influential work of Bloom (2009). The interest in this subject stems from its connection to the business cycle, as supported by both theory and empirical evidence. In a general equilibrium setting, an increase in uncertainty may depress investment, consumption, and GDP through the real option effect[1], the precautionary saving effect, or an increase in financial frictions[2]. In addition, empirical research consistently demonstrates a sharp increase in uncertainty during recessions. This evidence holds whether uncertainty is measured using free model measures (e.g., Bloom 2009, Baker et al. 2016) or model-based measures as in Jurado et al. (2015) (referred to as JLN).

While there is a wealth of literature focusing on macroeconomic uncertainty originating in the real sector of the economy, overshadowing the potential role of the financial industry as a driver, Ludvigson et al. (2021) differentiate between financial uncertainty and macroeconomic uncertainty, showing that the former acts as an impulse for economic activity. Nevertheless, measuring uncertainty within the financial industry, particularly at the bank level, and examining its impact on the business cycle remain relatively new and understudied research areas. Few studies focus on bank-level uncertainty (see, for example, Soto 2021 and Buch et al. 2015). However, these studies limit their scope to the banking sector and do not investigate the business cycle effect of bank-level uncertainty.

---

[1] Bloom (2009) simulates the impact of uncertainty on investment and employment under the hypothesis of fixed cost. He shows that the rise in uncertainty increases the real option value of inaction, leading most production units to temporarily freeze hiring, resulting in a drop in employment and investment.

[2] Valencia (2017) presents a model in which commercial banks insure themselves in the face of increasing uncertainty by contracting loan supply.

This paper pioneers a multi-faceted approach to quantify bank-level uncertainty, capitalizing on granular data from large banking institutions and a comprehensive macroeconomic dataset. The central research question probes the endogeneity of bank-level uncertainty, investigating whether it operates solely as an internal financial sector dynamic or wields broader economic ramifications. The analysis is particularly pertinent to regulatory policy considerations, elucidating a potential alternative conduit via which systemic instability can reverberate through the real economy. Since financial crises often manifest within the banking sector before cascading into broader markets, the bank-level uncertainty metric might serve as an anticipatory indicator of economic instability, outperforming existing generic measures.

I operationalize the construction of the bank-level uncertainty using a panel dataset encompassing 104 large Bank Holding Companies (BHCs) with total assets exceeding 3 billion USD from 1986q3 to 2020q4. I compute the metric as the cross-sectional standard deviation of the unpredictable component of return on assets (ROA).[3] In developing a predictive framework for ROA, the paper leverages an ensemble of both nonlinear and linear machine learning models, taking the average of their one-quarter-ahead forecasts. This ensemble approach has several merits. It capitalizes on diverse data patterns, mitigates the risks of model dependency, thereby fortifying forecast robustness, and improves individual model accuracy even in the face of forecast instability (Fraisse and Laporte 2022). The study incorporates bank-specific covariates and an expansive set of macroeconomic indicators for the predictor variables. This comprehensive data inclusion serves dual purposes. First, it purges the model of predictable variations, and second, it ensures that the uncertainty measure effectively embodies genuine unpredictability. Given that this metric relies on

[3]This choice of the profit variable aligns with Jurado et al. (2015), who construct firm-level uncertainty based on the conditional volatility of unforecastable profit components.

out-of-sample forecasting, its calculation begins from 2000q2, which marks the initiation of the test set and extends through 2020q4.

This paper advances the literature in two key ways. First, I demonstrate that bank-level uncertainty is a byproduct of economic downturns and a significant driver of business and financial cycles, extending its influence beyond the scope examined in existing studies like Ludvigson et al. (2021). Second, I innovate methodologically by employing machine learning techniques alongside granular bank and macroeconomic data to construct a more robust measure of bank-level uncertainty. This approach addresses limitations in traditional linear frameworks and leverages the predictive power of machine learning, as validated by studies like Goulet Coulombe et al. (2022) and Stock and Watson (2004).

The rest of the paper is organized as follows. Section 3.2 presents the econometric framework under which I construct the measure. Section 3.3 contains the description of the data. Section 3.4 provides an estimation of the bank-level uncertainty. Section 3.5 discusses the impact of the uncertainty measure on the business cycle and financial sector, and section 3.6 concludes.

3.2 Econometric Framework

I now turn to describe the construction of the uncertainty measure. The econometric framework involves two steps. The first step consists of constructing various machine learning forecasting models, whereas in the second step, I present the derivation of the bank-level uncertainty.

3.2.1 Forecasting exercise

3.2.1.1 Estimating machine learning models

Let us write the model as follows:

$$Y_{i,t+1} = \hat{g}(Z_{i,t}) + e_{i,t+1}\ , \tag{3.1}$$

where $Y_{i,t+1}$ denotes the value of the return to the asset (ROA) of BHC i at quarter $t+1$, $e_{i,t+1}$ represents the forecast error and $Z_{i,t}$ is a $K \times 1$ regressor vector. $Z_{i,t}$ comprises lagged dependent variable, microeconomic predictors, and all macroeconomic and financial series of the FRED database. The one step ahead forecast of variable $Y_{i,t+1}$ for BHC i given the information available at t is given by $\hat{Y}_{i,t+1}$=$\hat{g}(Z_{i,t})$. I estimate $\hat{g}$ considering different parametric and nonparametric machine learning techniques.[4] The specificity of all of the machine learning techniques is to estimate $\hat{g}$, which minimizes the out-of-sample mean square error (MSE)[5] while simultaneously carrying out regularization procedures.[6] The

---

[4]As parametric techniques, I consider "Lasso" regression, "Adaptive Lasso" regression, "Ridge" regression, "Elastic Net" regression, and Principal component analysis. I include Gradient boosting, Random Forest, and Neural Networks for nonparametric techniques. I provide complete details on parametric and nonparametric machine learning techniques in this paper in appendix A.

[5]The MSE of a model is defined by:

$$MSE = \frac{1}{N}\sum_{i,t}(Y_{i,t+1} - \hat{Y}_{i,t+1})^2;$$

where N denotes the number of banks present in the test set, i the cross-sectional dimension and t the time dimension.

[6]The right choice of hyperparameters is at the core of regularization procedures. These include, for example, the penalization coefficients in the "Lasso," Elastic Net," "Ridge" regressions, and the number of trees and the number of variables selected randomly at each step for Random Forest.

regularization has a dual objective. First, it can avoid over-training the data.[7] Second, it reduces the complexity of the model and often ensures a forecast performance higher than traditional forecast techniques by the channel of bias-variance arbitrage. The choice of the correct hyperparameters is the core of the ML techniques. I determine the optimal hyperparameters by the K-fold cross-validation. Since the sample is a panel, I cross-validate along time series and cross-section dimensions. Specifically, I divide a subsample of estimation and validation samples into k groups of approximately equal size for a given model and a combination of hyperparameters. I consider group j ($1 \leqslant j \leqslant k$) as the validation set, and I estimate the model on the remaining k-1 groups. Then, I predict the return to the asset in group j and compute its mean square error (MSE^j). I redo the exact exercise k times by treating each of the remaining k-1 groups as a validation sample. In the end, I compute an average forecasting performance for the given hyperparameters as

$$MSE = \frac{1}{N} \sum_{1}^{k} MSE^l$$

I repeat the same exercise for all potential combinations of hyperparameters. I choose the hyperparameters yielding the lowest value of MSE. Thus, I estimate the model with the optimal hyperparameters in the entire subsample.

3.2.1.2 Constructing individual and average forecasts

In the first step, I build individual forecasts for each technique by applying the expanding window technique. Therefore, I split the data sample initially into two. The first subsample

[7] Over-training of the data leads the function $\hat{g}$ estimated by the forecasting technique to approach the variable of interest almost perfectly and eliminate the residual variability in-sample. This phenomenon results in a generally low mean square error in-sample, which masks poor prediction performance out of the sample.

comprises the estimation and validation sample and extends from 1986q3 to 2000q1, and the second subsample, the test set, covers 2000q2-2020q4. I use the model estimated in the first subsample to forecast $Y_{i,t+1}$, one quarter ahead. I run the process iteratively by expanding the first subsample by one quarter until I have predicted ROA for the last quarter of the test set sample 2020q4. I build the average forecast as the mean of individual forecasts. Concretely, let us denote by $\hat{Y}^{j}_{i,t+1}$, the forecast of ROA relative to model j, B the number of ML models. The average forecast of ROA, $\hat{Y}^{avg}_{i,t+1}$, is defined as

$$\hat{Y}^{avg}_{i,t+h} = \frac{\sum_{j=1}^{B} \hat{Y}^{j}_{i,t+1}}{B} . \tag{3.2}$$

I employ this methodology to capture the predictive power of multiple models, thereby improving the robustness and reliability of the forecasts.

3.2.2 Building a time varying bank-level uncertainty

3.2.2.1 Derivation of the measure

Building on Equation (3.2), I calculate a measure of uncertainty, denoted as U_t, for each quarter t in the test set. This measure is formulated as follows:

$$U_t = \sqrt{\frac{1}{N_{t+1}} \sum_{i=1}^{N_{t+1}} \left[Y_{i,t+1} - \hat{Y}^{avg}_{i,t+1} \right]^2} , \tag{3.3}$$

where i refers to individual banks and N_{t+1} represents the total number of banks in operation during quarter t+1. To express the previous Equation in terms of forecast errors, let $\hat{\epsilon}_{i,t+1}$ denotes the forecast error for bank i in quarter t+1 relative to the average forecast.

Using this notation, Equation (3.3) can be rewritten as :

$$U_t = \sqrt{\frac{1}{N_{t+1}} \sum_{i=1}^{N_{t+1}} \hat{\epsilon}^2_{i,t+1}} \quad . \tag{3.4}$$

It is evident from this formulation that U_t is the cross-sectional conditional standard deviation of the forecast errors relative to the average forecast. As in JLN, this measure gauges the volatility of the unpredictable Return on Assets (ROA) components. The estimate of the bank-level uncertainty is calculated using out-of-sample forecasts, specifically for the period 2001q2-2020q4. A higher value of U_t is expected to correspond to the period of higher bank-level uncertainty as it translates to more bankers' difficulty in predicting ROA. On the contrary, the measure's lower value should reflect a calm period. Characterizing the volatility through the dispersion of firm profit is at the cornerstone of a large body of literature focusing on micro-based uncertainty (See, for example, Bloom et al. 2018, JLN, Buch et al. 2015). These papers show that these measures capture uncertainty and are perfectly countercyclical. Accordingly, according to this literature, I choose ROA as a variable whose dispersion of unpredictable components could reflect the fluctuation of uncertainty.

3.2.2.2 Decomposition of the time varying bank-level uncertainty

Let us consider the true but unknown forecast model of ROA for the population as represented by Equation (3.5):

$$Y_{i,t+1} = g^*(Z_{i,t}) + \epsilon_{i,t+1} \, . \tag{3.5}$$

Here, g^* is the true forecasting model of ROA for one quarter ahead based on the available information at time t. $Z_{i,t}$ represents the vector of predictors, while $\epsilon_{i,t+1}$ denotes the true, unknown forecast error. Informed by theory, prior literature, and intuitive reasoning, I specify the forecasting model g as shown in Equation (3.6), and estimate it through $\hat{g}$ in Equation (3.7) :

$$Y_{i,t+1} = g(Z_{i,t}) + e_{i,t+1} \tag{3.6}$$

$$Y_{i,t+1} = \hat{g}(Z_{i,t}) + \hat{\epsilon}_{i,t+1} \, , \tag{3.7}$$

Let $\hat{Y}_{i,t+1}$ represents the estimated forecast model of ROA, which implies $\hat{Y}_{i,t+1}$= $\hat{g}(Z_{i,t})$. Building upon Equations (3.5), (3.6), and (3.7), the estimated forecast error $\hat{\epsilon}_{i,t}$ can be decomposed as follows:

$$\hat{\epsilon}_{i,t} = Y_{i,t+h} - \hat{Y}_{i,t+h} = \underbrace{g^*() - g()}_{approximation\ error} + \underbrace{g() - \hat{g}()}_{estimation\ error} + \epsilon_{i,t+1}$$

This decomposition reveals two types of errors: (1) the approximation error and (2) the estimation error. The approximation error measures the divergence between the true specification $(g^*())$ and the econometrician-defined specification $g()$. The estimation error arises from omitted variable bias. Averaging various machine learning (ML) forecasts mitigates the approximation error by capturing different data features, such as linearity and nonlinearity, as demonstrated by Stock and Watson (2004). As a consequence, this technique contributes to reducing the approximation error. Furthermore, using a comprehensive dataset that includes macroeconomic and microeconomic predictors minimizes the estimation error, contributing to a more accurate measure of uncertainty based on ROA's

unforecastable component. This approach fills a gap in the existing literature. For instance, JLN control for estimation error but ignores nonlinearities. Similarly, the work of Buch et al. (2015) is potentially subject to approximation and estimation errors due to its limited predictor set and linear model.

3.3 Data

I use a panel dataset of 100 Bank Holding Companies (BHCs)[8] observed quarterly, each with total assets exceeding three billion USD, to operationalize the proposed methodology. Mésonnier and Stevanovic (2017) initially curated this panel dataset, which spans from the third quarter of 1986 to the fourth quarter of 2011. I further extend this coverage up to the fourth quarter of 2020 to incorporate more recent periods of economic uncertainty. The extended sample period encapsulates significant macroeconomic events characterized by heightened uncertainty. These events include but are not limited to the recession of 1990, the subprime mortgage crisis, and the economic downturn triggered by the COVID-19 pandemic. It is essential to mention that the panel is unbalanced due to the inherent dynamics of the banking industry—such as mergers, acquisitions, and bank failures. Specifically, only 20% of the BHCs in the sample are operational throughout the entire period under study.

I model ROA, which is defined as the annualized ratio of net bank income to total assets. It is one of the bank's widely used indicators of profitability. It provides information on the bank gains or losses a unity asset generates. I compute the dependent variable from the information enclosed in the FR Y-9C form of the Consolidated Financial Statements for BHC delivered quarterly by the Chicago FED. Table 3.1 presents the summary statistics for

[8]The complete list of BHCs is available in Appendix D

the bank's variables and shows that ROA is quite volatile as its mean accounts for 70.64% of its standard deviation. As bank-specific predictors, I consider several variables. *The lagged value of ROA* is used to capture dynamic effects. *Size* is represented by the natural logarithm of the bank's total assets and serves as a proxy for its overall size. *Capital Ratio*, computed as the bank's equity capital relative to its total assets, indicates the bank's capital adequacy. *NCO*, the ratio of net charge-offs on loans and leases to total assets, quantifies loan risk. *Cost* gauged through the ratio of total non-interest expenses to total assets captures the bank's efficiency. The model also includes *CILoan*, defined as the ratio of such loans to total assets, which measures a bank's exposure to consumer and investment loans. *RELoan* quantifies a bank's real estate loan exposure and is similarly defined as the ratio of real estate loans to total assets. Lastly, *Diversification*, defined as the ratio of total non-interest income to total income, serves as a measure to capture the extent of a bank's diversified activities.

Further examining Table 3.1, I find salient insights into large banking institutions' risk exposure and business models. Notably, banks exhibit a higher average exposure to real estate loans, constituting 26.71% of total assets, relative to consumer and investment loans, which account for 15.83%. In addition to asset allocation, the data underscore significant heterogeneity in the banks' operating models. Some institutions focus predominantly on interest-generating activities, while others diversify their income streams. Nonetheless, traditional lending activities remain the dominant source of income for most banks, contributing an average of 74.58% to total earnings. Turning to the size of the banks, the relatively low standard deviation compared to the mean in the size variable signifies uniformity among the large banks in our sample. This observation is noteworthy, as the study focuses on large banking institutions, and such uniformity corroborates the analytical focus on this specific segment. Finally, it is essential to note that the predictor variables

employed in this study are sourced from the Consolidated Financial Statements for Bank Holding Companies released by the Federal Reserve Bank of Chicago.

Table 3.1 Summary Statistics for Bank variables

	mean	sd	min	p50	max
ROA	2.55	3.61	-26.11	2.28	77.47
Size	17.19	1.44	14.93	16.93	21.94
Capital Ratio	8.62	5.70	-0.35	7.85	80.92
NCO	1.12	1.49	-1.23	0.64	18.50
Cost	9.30	9.01	0.32	8.08	175.35
CILoan	15.83	8.21	0.00	15.58	51.54
RELoan	26.71	13.50	0.00	26.34	83.96
Diversification	25.42	14.98	0.00	21.72	99.44

3.4 Measures of Uncertainty

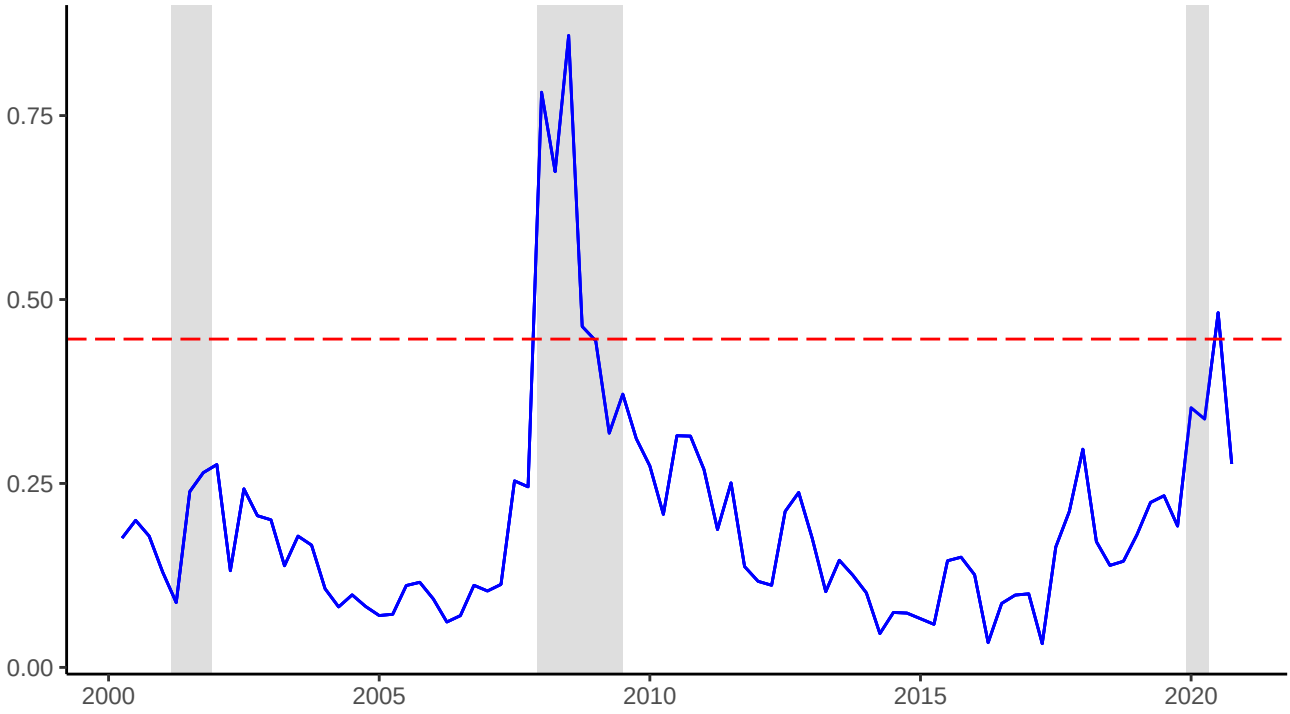
3.4.1 Estimates of bank-level uncertainty

Figure 3.1 illustrates the series of bank-level uncertainty (U_t) and notably underscores its countercyclical nature. A significant negative correlation is apparent between U_t and GDP growth ($\Delta lnGDP_t$),[9] as evidenced by an estimated correlation coefficient of -0.57. Uncertainty markedly increases during recessions and slowly dissipates as the turbulence wanes. This empirical observation aligns with findings from prior research on uncertainty, such as those by JLN, Buch et al. (2015), and Bloom (2009). I adopt the criteria Bloom (2009) established to delineate periods of heightened uncertainty. According to this criteria, periods of heightened uncertainty are those where the measure of uncertainty exceeds the mean of the time series plus 1.65 times its standard deviation. By this measure, two periods of

[9] I compute GDP growth rate as a four-quarter moving average of the quarterly growth rate, following JLN's approach.

intensified uncertainty emerge, corresponding chronologically to the Great Recession and the COVID-19 Recession. Remarkably, the uncertainty measure's pinnacle coincides with the Great Recession.

Figure 3.1 The bank-level uncertainty (U_t)



Notes: The vertical grey lines delimit the US NBER recession periods. The horizontal red line represents the sample mean of the measure of uncertainty plus 1.65 times its standard deviation.

Figure 3.2 delves into the relationship between bank-level uncertainty and shifts in financial and credit markets. It juxtaposes bank-level uncertainty with three pivotal indicators of financial market well-being: the National Financial Condition Index ($NFCI$), the spread of Moody's seasoned BAA bond yield over the 10-year Treasury constant maturity yield ($BAA10YM$), and the excess bond premium (EBP). The $NFCI$, a product of the Chicago FED, provides a weekly snapshot of financial conditions spanning money mar-

kets, debt, equity markets, and the banking sector. Elevated $NFCI$ values signify worsening financial health, whereas diminished values indicate amelioration. From a quarterly perspective, I average the weekly $NFCI$ values over three months. As for $BAA10YM$, an increased spread suggests strained credit market conditions, whereas a narrow spread signals a more buoyant credit environment. Furthermore, Gilchrist and Zakrajšek (2012) introduced the excess bond premium (EBP) to quantify the risk premium on corporate bonds. An upsurge in EBP values signifies a heightened risk premium, marking a dip in bond allure and a tightening of credit conditions. Conversely, diminishing EBP values indicate a relaxation in credit market conditions. Figure 3.2 underscores a concerted movement across the four metrics. As uncertainty escalates at the bank level, all three indicators mirror a concurrent financial and credit market health decline.

Figure 3.2 Bank-level uncertainty (U_t), $NFCI$, $BAA10YM$ and EBP

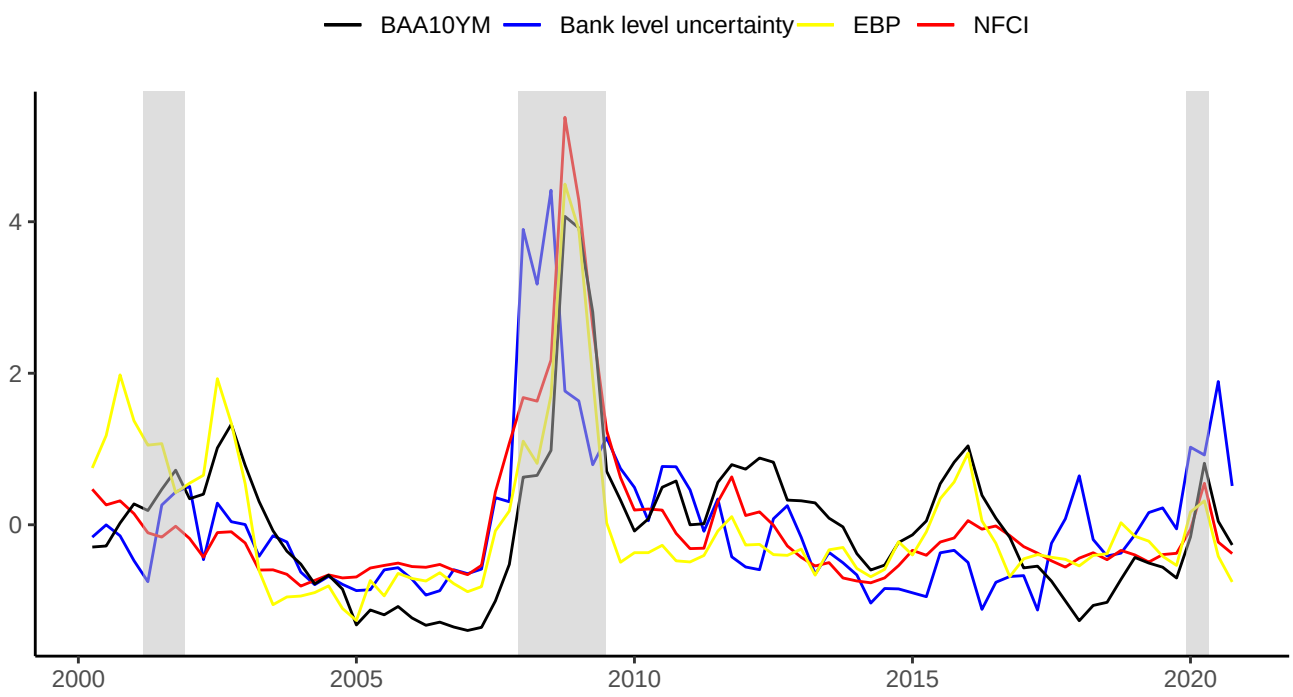


Table 3.2 delves deeper and showcases the outcomes of a two-sided Granger causality test between bank-level uncertainty and each financial condition metric. Notably, the table establishes that bank-level uncertainty (U_t) granger-causes $BAA10YM$, EBP, and $NFCI$ at a 1% significance level, but not vice versa. These results shed light on the potential interplay between bank-level uncertainty and broader financial nuances, echoing the findings of Gilchrist et al. (2014) regarding idiosyncratic uncertainty. Specifically, Gilchrist et al. (2014) posit that idiosyncratic uncertainty shocks instigate financial friction, sway credit supply, and engender countercyclical credit spreads. These findings intimate that fluctuations in bank-level uncertainty might be instrumental in decoding the fluid dynamics of financial markets.

Table 3.2 Granger causality test between the bank level uncertainty and the measures of financial condition

Granger test hypothesis H_0	Statistics	p value
U_t does not granger cause $BAA10YM$	9.40	0.00***
$BAA10YM$ does not granger cause U_t	0.68	0.69
U_t does not granger cause EBP	5.91	0.00***
EBP does not granger cause U_t	0.66	0.71
U_t does not granger cause $NFCI$	9.01	0.00***
$NFCI$ does not granger cause U_t	1.56	0.16

Note : * : p value of the granger test<0.1; ** : p values of the test <0.05; ***: pvalue of the test<0.01.

3.4.2 Descriptive Analysis of the bank-level uncertainty and the alternative measures

In assessing the nuances of bank-level uncertainty, I contrast it with three alternative dispersion-based measures in Table 3.3. These measures, for simplicity, are denoted as $U_t^{(1)}$, $U_t^{(2)}$ and $U_t^{(3)}$. Each of these measures offers a unique perspective on the multifaceted nature of uncertainty within the banking sector, grounded in their distinct computational methodologies:

$U_t^{(1)}$: This measure is computed as the quarterly cross-sectional standard deviation of the return on assets (ROA). Its conceptual foundation aligns with the firm-level uncertainty explored by Bloom (2009). Notably, it does not distinguish between the forecastable and non-forecastable components of a bank's profit, offering a broad-brush perspective on uncertainty.

$U_t^{(2)}$: Drawing inspiration from Buch et al. (2015), this measure is derived as the cross-sectional standard deviation of the forecast errors of ROA. The forecasting model that underpins this measure is minimalist and encompasses only bank and time-fixed effects as predictors. This measure offers a constrained view of uncertainty, focusing solely on

deviations from expected ROA based on these limited parameters.

$U_t^{(3)}$: This measure mirrors $U_t^{(2)}$ in its foundational definition. However, it embraces a broader analytical scope by integrating additional predictors.[10]

By juxtaposing U_t with these alternative measures, I aim to unravel the relative significance of various predictors in constructing bank-level uncertainty. This comparative analysis illuminates how each measure encapsulates distinct dimensions of uncertainty inherent in the banking sector.

A primary observation from Table 3.3 is the pronounced skewness and kurtosis of U_t compared to its counterparts. These statistics indicate that U_t is characterized by extreme values, potentially implying that this measure is adept at capturing episodes of heightened bank-level uncertainty. Such episodes are crucial for understanding systemic risks in the banking sector. The first-order autocorrelation coefficient of U_t stands at 0.77, corresponding to the slope of its AR1 model. Notably, this coefficient significantly surpasses those of the alternative proxies. As a result, the estimated half-life[11] of innovations to U_t spans roughly three-quarters, a duration that considerably outlasts the corresponding half-lives of alternative measures. This longer persistence indicates that U_t provides a more stable and enduring reflection of bank-level uncertainty, a vital feature for long-term risk assessment. Moreover, the correlation between bank-level uncertainty and economic activity is quite robust. For instance, the contemporaneous correlation with GDP growth, denoted by (corr(u_t, $\Delta lnGDPt$)) is -0.55. In contrast, the correlations for $U_t^{(1)}$, $U_t^{(2)}$, and $U_t^{(3)}$ are -0.16, -0.28, and -0.10. These findings underscore U_t's enhanced sensitivity to macroe-

[10]The considered macroeconomic variables include the GDP growth rate, CPI inflation, and the Federal funds rate, collectively characterizing the business cycle.

[11]Let us consider a stationary AR1 times series defined by $y_t = ry_{t-1} + \epsilon_t$. The half-life of the series is approximated by $\frac{-Ln(2)}{Ln(r)}$.

conomic dynamics, likely attributed to its integration of a comprehensive macroeconomic dataset and distinction of non-forecastable components.

Furthermore, Table 3.3 delineates cross-correlations between U_t and GDP growth. A peak cross-correlation of -0.49 with GDP growth emerges at a lead of t+1, while a lag of t-1 yields a maximum cross-correlation of -0.51. This pattern intimates a potential interaction between bank-level uncertainty and the business cycle; U_t could presage economic activity by a quarter and respond to business cycle shifts in the succeeding quarter. The table's concluding rows present the Granger causality test's null hypothesis and statistics, evaluating the causal interplay between uncertainty measures and GDP growth. The results indicate that U_t, $U_t^{(1)}$ and $U_t^{(2)}$ have predictive power for future GDP growth, significant at the 5% and 10% levels, respectively. These empirical observations bolster the hypothesis that bank-level uncertainty influences economic activity. However, a more profound, structurally-oriented inquiry is warranted to establish this causal relationship. Such an exploration could render compelling evidence to substantiate or refute this preliminary insight.

Table 3.3 Summary statistics of uncertainty's measures in quarterly frequency

Statistics	Uncertainty measured using :			
	$U_t^{(1)}$	$U_t^{(2)}$	$U_t^{(3)}$	The bank level uncertainty (U_t)
Mean, sd	1.18, 0.34	0.87, 0.41	2.19, 2.07	0.20, 0.15
Skewness, Kurtosis	0.40, 0.03	0.81, 0.13	2.35, 5.75	2.22, 6.17
AR1, Half Life	0.35, 0.67	0.19, 0.42	0.37, 0.69	0.77, 3.00
corr($U_t, \Delta lnGDP_t$)	-0.16	-0.28	-0.10	-0.57
corr($U_t, \Delta lnGDP_{t+4}$), corr($U_t, \Delta lnGDP_{t-4}$)	0.00, -0.17	-0.03, -0.19	0.09, -0.40	0.02, -0.27
$max_{k>0} corr(U_t, \Delta lnGDP_{t+k})$	-0.13	-0.22	0.09	-0.49
At lag k=	1	1	4	1
$max_{k<0} corr(U_t, \Delta lnGDP_{t+k})$	-0.20	-0.31	-0.40	-0.51
At lag k=	-1	-1	-4	-1
G.T : Uncertainty measure does not granger cause $\Delta lnGDP_t$	1.86*	2.02*	1.06	2.44**
G.T : $\Delta lnGDP_t$ does not granger cause the uncertainty measure	0.65	0.22	4.09***	0.78

Note : G.T designates the granger causality test. * : p value of the test<0.1; ** : p values of the test <0.05; ***: pvalue of the test<0.01 .

3.4.3 Bank-level uncertainty, financial uncertainty and macroeconomic uncertainty

I juxtapose the bank-level uncertainty measure with Ludvigson et al. (2021)'s financial uncertainty measure and three established macroeconomic uncertainty metrics: the JLN uncertainty, the VIX volatility index, and Baker et al. (2016)'s economic policy uncertainty.

JLN uncertainty : This influential measure of macroeconomic uncertainty, introduced by JLN, derives from the average volatilities of non-forecastable components across various macroeconomic and financial series. Using a data-rich approach, JLN construct predictive models for each series, extracting their
non-forecastable facets.

Ludvigson et al. (2021) financial uncertainty: Adopting a methodology akin to JLN, Ludvigson et al. (2021) hone in exclusively on financial series to construct their measure of financial uncertainty.

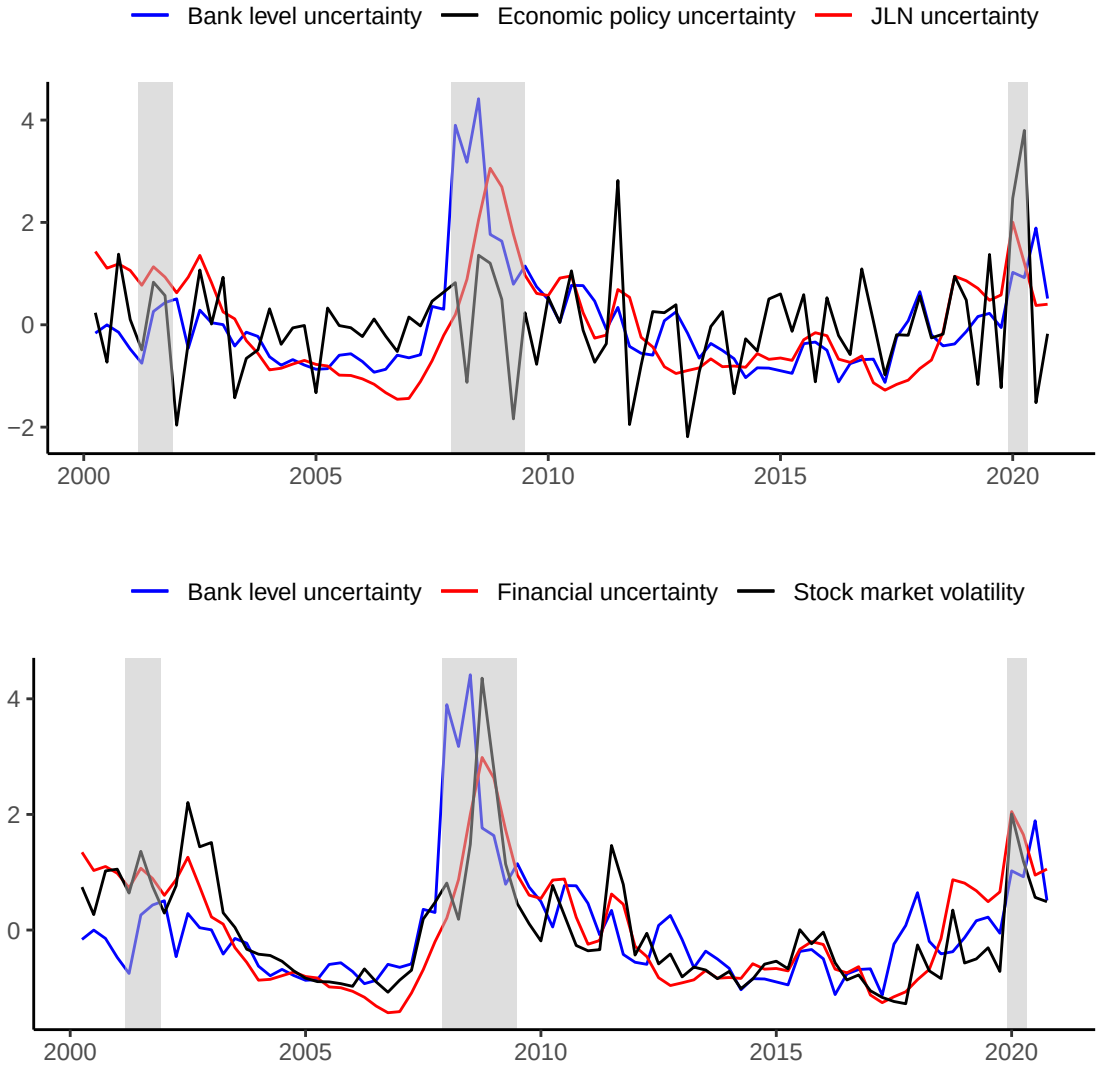
VIX Volatility Index: Serving as a barometer for US stock market volatility, the VIX index was championed by Bloom (2009) as a surrogate for macroeconomic uncertainty in his seminal work on the nexus between uncertainty and the business cycle.

Baker et al. (2016) Economic Policy Uncertainty: This measure quantifies policy uncertainty through a frequency analysis of specific keywords such as "economic" or "economy" paired with "uncertain" or "uncertainty" within leading US newspapers.

Empirical analysis reveals a robust and positive association between bank-level uncertainty and multiple financial and macroeconomic uncertainty indices. However, a notable exception is observed in the case of economic policy uncertainty. Specifically, correlation coefficients predominantly range between 0.56 and 0.62, emphasizing the strength of these relationships. In contrast, the correlation coefficient between bank-level and economic

policy uncertainty is more tempered, registering an estimated 0.22. Figure 3.3 visually represents these measures and highlights their pronounced countercyclical behavior. During contractionary phases of the economy, all five indices manifest an upward trajectory indicative of escalating uncertainty. Conversely, during stages of economic resurgence, these indices show a discernible decline, signaling mitigation in uncertainty levels. Such empirical patterns bolster the prevailing academic hypothesis: economic downturns invariably correlate with amplified uncertainty across diverse metrics.

Figure 3.3 Bank-level uncertainty , JLN uncertainty, VIX Stock market volatility, Economic policy uncertainty of Baker et al. (2016) and Financial uncertainty of Ludvigson et al. (2021)



3.5 VAR analysis

This section comprehensively analyzes the interaction between bank-level uncertainty and the business cycle. Additionally, it compares the macro-financial effects of the bank-level uncertainty shock with those of other uncertainty measures.

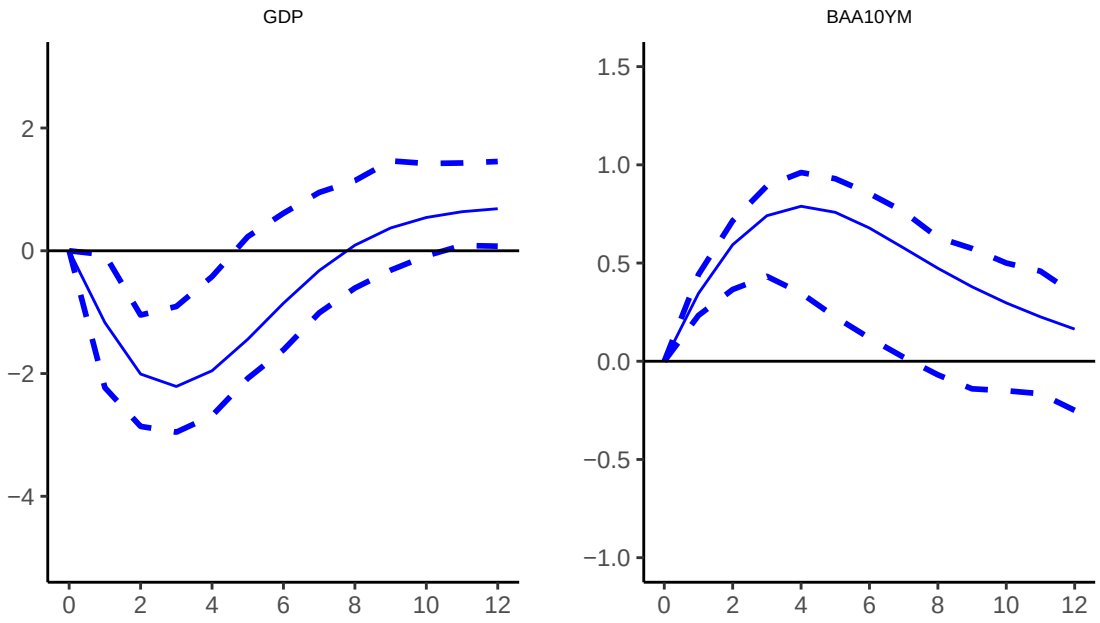
3.5.1 Bank-level uncertainty and macroeconomic and financial dynamic

I use a Structural Vector Autoregression (SVAR) framework to investigate the interrelation between bank-level uncertainty and macroeconomic and financial indicators fluctuations. The foundational structure of the baseline VAR influenced by JLN incorporates a set of 12 variables. Including a comprehensive variable set aims to counteract potential omitted variable bias—a concern raised by Carriero et al. (2018). The variables included in the baseline VAR model are as follows: (1) GDP growth rate (*GDP*), (2) unemployment rate (*UNRATE*), (3) real consumption growth rate (*CONSUMPTION*), (4) CPI inflation (*INFLATION*), (5) growth rate of orders (*ORDERS*), (6) real wage growth rate (*WAGE*), (7) hours of work (*HOURS*), (8) Federal funds rate (*FEDFUNDS*), (9) spread between BAA bond's yield and the 10-year maturity treasury bond yield (*BAA10YM*), (10) growth rate of M2 money stock (*M2REAL*), (11) bank-level uncertainty (U_t), and (12) financial uncertainty of Ludvigson et al. (2021) (U_t^F). I estimate the VAR using one lag based on information criteria. The data used in the paper spans from 2000q2 to 2020q4.

I use the Cholesky decomposition to derive the impulse response functions of the above variables to an innovation in bank-level uncertainty, subsequently described as a "shock." The decomposition respects the previously detailed variable ordering. Notably, by placing Ut^F as the concluding variable in this sequence, I postulate that any events occurring

within the banking sector during a quarter directly influence the broader financial industry, subsequently affecting the financial uncertainty metric. The incorporation of the financial uncertainty metric is deliberate. It illuminates the intricate dynamics between this and bank-level uncertainty, shedding light on the ripple effects of banking-sector frictions on the comprehensive financial system and the macroeconomy. Adhering to the analytical techniques of Bloom (2009) and JLN, the shock's intensity, on bank-level uncertainty, is calibrated at a value fourfold its standard deviation, allowing for streamlined comparative analysis. For a more focused and lucid depiction, Figure 3.4 singularly portrays the reactions of *GDP* and *BAA10YM* . Figure 3.4 elucidates the substantial macroeconomic and financial repercussions from a positive shock to bank-level uncertainty. In the aftermath of the shock, *BAA10YM* - a barometer of credit market conditions—registers a pronounced surge, sustaining this heightened level for ten quarters. Concurrently, *GDP* growth undergoes a contraction, shrinking by 2% on a quarterly basis. This downturn manifests two-quarters post-shock and persists beneath its long-term trajectory for roughly eight quarters. Such findings underscore that a surge in bank-level uncertainty precipitates an escalation in the external financial premium. This surge is evident from the amplified *BAA10YM* value, which indicates a steeper financing cost in capital markets and exerts a drag on GDP growth.

Figure 3.4 IRFs of *GDP* and *BAA10YM* to a positive shock on the bank level uncertainty



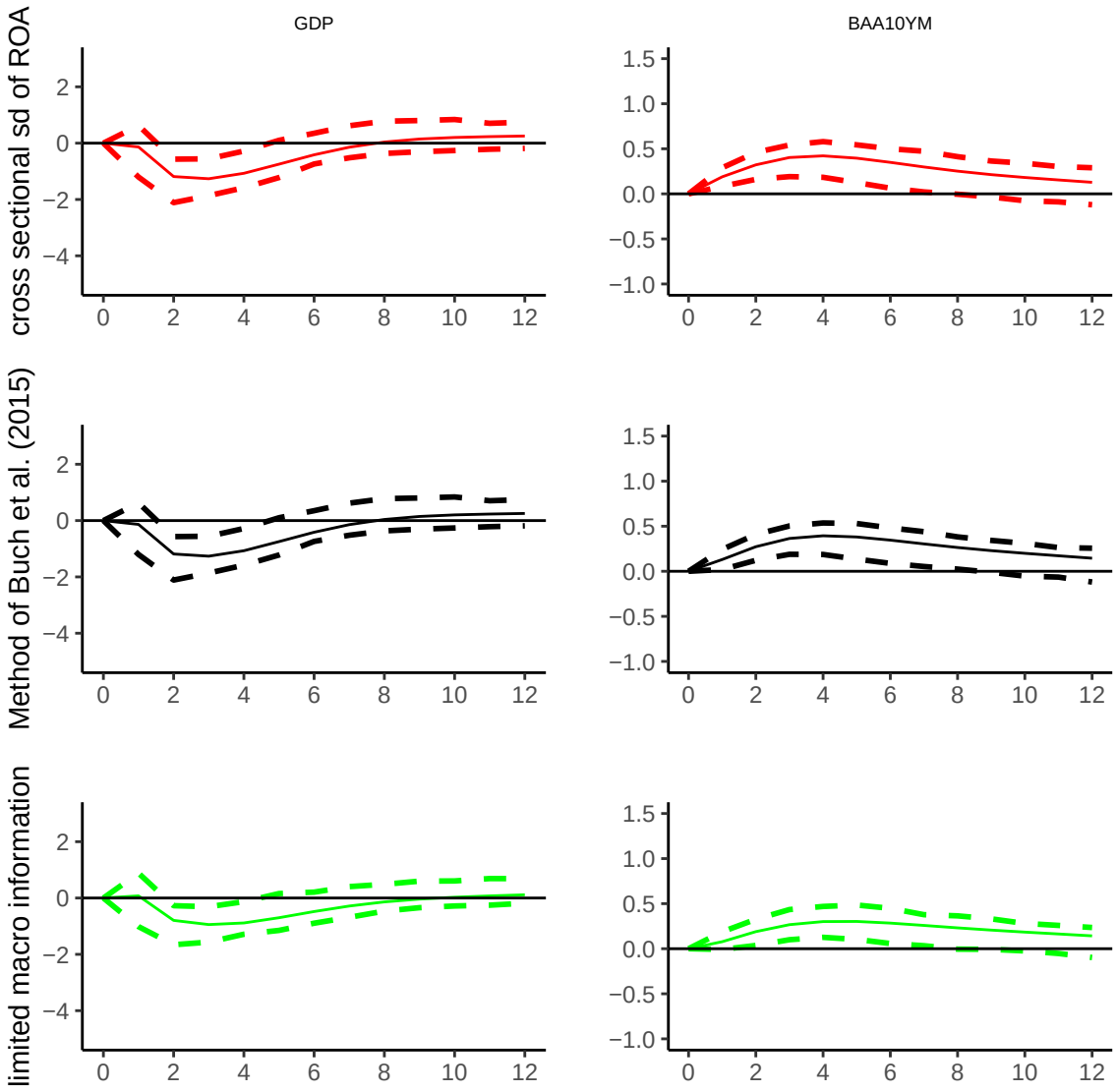
Note : The dots charts represent IRF lower and upper bounds and are constructed at 90% confidence level.

3.5.2 Comparing the bank level uncertainty to alternative dispersion based measures of uncertainty

The methodology presented in this paper, upon scrutiny, demands rigorous reconsideration. In pursuit of a more robust analysis, I replace the bank-level uncertainty, represented as U_t, with three alternative measures previously defined and derived from the dispersion of ROA or its delineated components: $U_t^{(1)}$, $U_t^{(2)}$ and $U_t^{(3)}$. The compelling results are showcased in Figure 3.5, which depicts the Impulse Response Functions (IRFs) of *GDP* and *BAA10YM* to these alternative measures. Figure 3.5 drives home an undeniable truth: the reactions of *GDP* and *BAA10YM* to perturbations in $U_t^{(1)}$, $U_t^{(2)}$ and $U_t^{(3)}$ are demonstra-

bly muted compared to their responses to the original bank-level uncertainty shock. This pattern persists unwaveringly across all forecast horizons at a 90% confidence interval. Even more telling is that despite harnessing the same bank-level uncertainty construction method that once evoked significant reactions for *GDP* and *BAA10YM*, the responses wane into near insignificance when limited to a truncated set of macroeconomic predictors. These results suggest and resoundingly affirm the vital need for comprehensive data integration. Furthermore, they emphasize the importance of precisely discerning between the non-forecastable and forecastable elements of ROA when architecting bank-level uncertainty.

Figure 3.5 IRFs of *GDP* and *BAA10YM* to a positive shock on alternative dispersion based measures of uncertainty $U_t^{(1)}$, $U_t^{(2)}$ and $U_t^{(3)}$



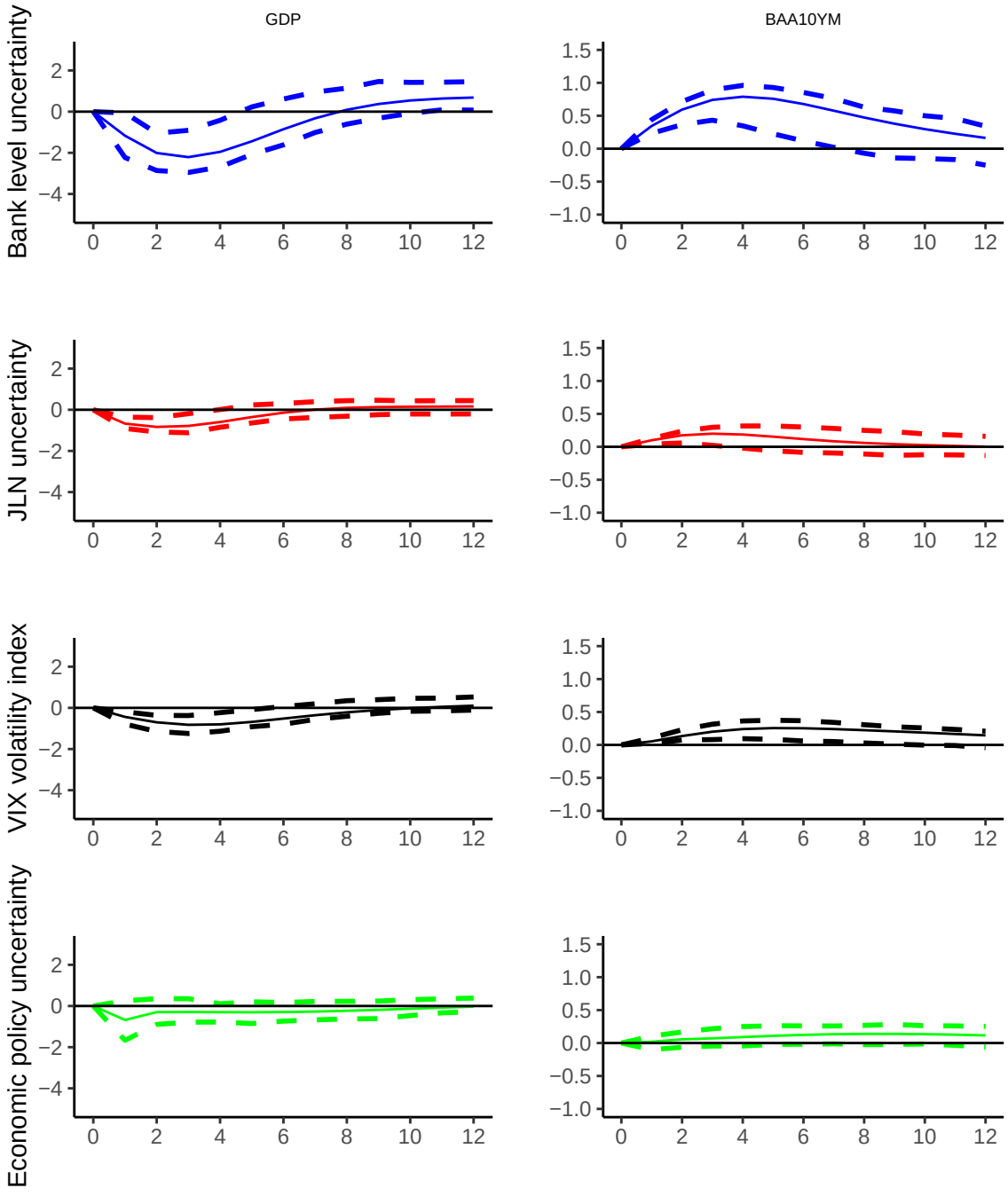
Note : The dots charts represent IRF lower and upper bounds and are constructed at 90% confidence level.

3.5.3 Comparing the bank level uncertainty to macroeconomic uncertainty measures

I replace the bank-level uncertainty with three measures of macroeconomic uncertainty sequentially: JLN, the VIX volatility index, and the economic policy uncertainty of Baker et al. (2016). Figure 3.6 plots the IRFs of *GDP* and *BAA10YM* to shocks on the bank-level uncertainty and the different macroeconomic uncertainty measures. According to this figure, the macroeconomic uncertainty measures impact much less the *GDP* and the *BAA10YM* than the bank-level uncertainty. The response function of *GDP* needs to be more neatly significant. These findings echo Ludvigson et al. (2021), who show that financial uncertainty is an impulse of economic activity, whereas macroeconomic uncertainty is an endogenous response to the business cycle.

Table 3.4 displays the decomposition of *GDP* and *BAA10YM* variance relative to the bank-level and macroeconomic uncertainty measures. This analytical framework corroborates our earlier observations and accentuates the divergence between the two types of uncertainty. Specifically, the results suggest that macroeconomic uncertainty measures are not pivotal determinants of future volatility in the credit market, in stark contrast to bank-level uncertainty. These findings also point to the necessity of differentiating between various forms of uncertainty when investigating their impact on economic cycles.

Figure 3.6 IRFs of GDP growth rate (*GDP*) and *BAA10YM* to a positive shock on the bank level uncertainty and on different measures of macroeconomic uncertainty



Note: The dots charts represent IRF lower and upper bounds and are constructed at 90% confidence level.

Table 3.4 Decomposition of variance of *GDP* and *BAA10YM* to bank-level uncertainty and macroeconomic proxies of uncertainty

Fraction of variance in GDP (percent)				
Explained by :	Bank level uncertainty	JLN uncertainty	VIX volatility index	Economic policy uncertainty
h=1	0.00	0.00	0.00	0.00
h=2	14.12	13.04	8.32	3.81
h=4	20.84	14.59	8.24	3.83
h=6	23.28	15.57	8.56	3.85
h=8	23.92	15.16	8.88	3.86
h=10	10.00	14.44	9.00	3.83
∞	24.03	14.60	9.04	3.67
Fraction of variation in BAA10YM (percent)				
Explained by :	Bank level uncertainty	JLN uncertainty	VIX volatility index	Economic policy uncertainty
h=1	0.00	0.00	0.00	0.00
h=2	17.59	3.08	0.61	1.31
h=4	24.76	7.16	3.60	2.16
h=6	27.04	8.50	4.98	2.38
h=8	27.62	8.76	5.69	2.46
h=10	27.58	8.58	6.18	2.48
∞	27.12	8.70	7.54	2.48

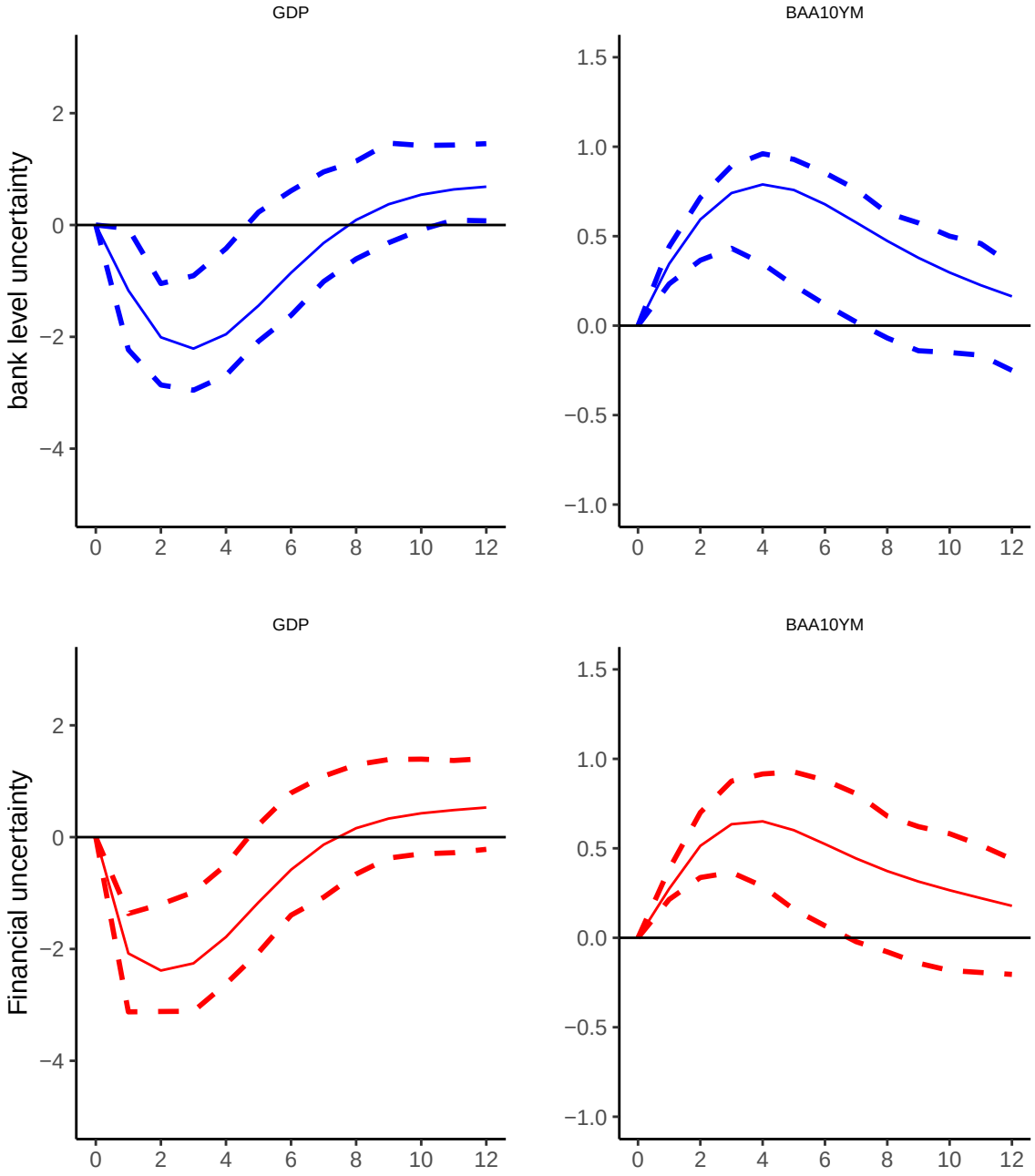
3.5.4 Comparing the bank-level uncertainty to financial uncertainty measure of Ludvigson et al. (2021)

Given that both bank-level uncertainty and the financial uncertainty measure proposed by Ludvigson et al. (2021) target the financial sector, a salient research question naturally arises: to what extent do these two measures differ in their influence on business cycle fluctuations? Addressing this query is more than just academic. It has practical implications for whether it is essential to consider a sector-specific measure of uncertainty for the banking industry. Therefore, the analysis comprises two dimensions. First, I assess the prospective impact of both uncertainty measures on key macro-financial variables, thereby projecting their potential role in shaping future economic conditions. Subsequently, I conduct an empirical analysis to scrutinize the historical influence exerted by these measures on the business cycle fluctuation.

3.5.4.1 How different are the bank level and financial uncertainty in affecting future financial and business cycle fluctuations?

Figure 3.7 delineates the Impulse Response Functions (IRFs) of *GDP* and *BAA10YM* following bank-level and broader financial uncertainty measure shocks. According to this figure, the *GDP* response to fluctuations in these two types of uncertainty is virtually identical. However, the response of $BAA10YM$ is markedly more pronounced when influenced by bank-level uncertainty. Table 3.5 offers further insight and presents the forecast variance decomposition of *GDP* and *BAA10YM* for bank-level and financial uncertainty. These results corroborate our earlier observations. While both types of uncertainty explain a comparable proportion of future fluctuations in *GDP*, bank-level uncertainty exerts approximately twice the influence on *BAA10YM* fluctuations, as does financial uncertainty. This finding lends credence to the argument that credit market dynamics are more acutely affected by perturbations specific to the banking sector than the broader financial sector.

Figure 3.7 IRFs of *GDP* and *BAA10YM* to a positive shock on the bank level uncertainty and financial uncertainty of Ludvigson et al. (2021)



Note : *The dots charts represent IRF lower and upper bounds and are constructed at 90% confidence leve*

Table 3.5 Decomposition of variance of *GDP* and *BAA10YM* to bank level uncertainty and financial uncertainty(%)

Fraction of variance in GDP (percent)		
Explained by :	Bank level uncertainty	Financial uncertainty
h=1	0.00	0.00
h=2	14.12	40.53
h=4	20.84	31.22
h=6	23.28	28.13
h=8	23.92	27.56
h=10	23.84	27.41
∞	24.03	26.39
Fraction of variation in BAA10YM (percent)		
Explained by :	Bank level uncertainty	Financial uncertainty
h=1	0.00	0.00
h=2	17.59	15.36
h=4	24.76	15.15
h=6	27.04	14.43
h=8	27.62	14.54
h=10	27.58	14.97
∞	27.12	16.55

3.5.4.2 How important is the bank-level uncertainty for historic fluctuation of the business cycle?

Historical decomposition is the standard tool used in the literature to ascertain the relative importance of structural shocks during historical episodes. Within a VAR framework, the historical decomposition at date t provides the cumulative effect of each structural shock on every endogenous variable up to that point in time. Kilian and Lee (2014) propose two formats to present the information conveyed by the historical decomposition. The first format expresses the historical decomposition as counterfactuals. In contrast, the second determines the contribution of each structural shock to the cumulative change of a specific variable of interest between two specific points in time. This paper considers the first approach.

In an n-variables Vector Autoregression (VAR) model, let y_{kt} denote the value of the k^{th}

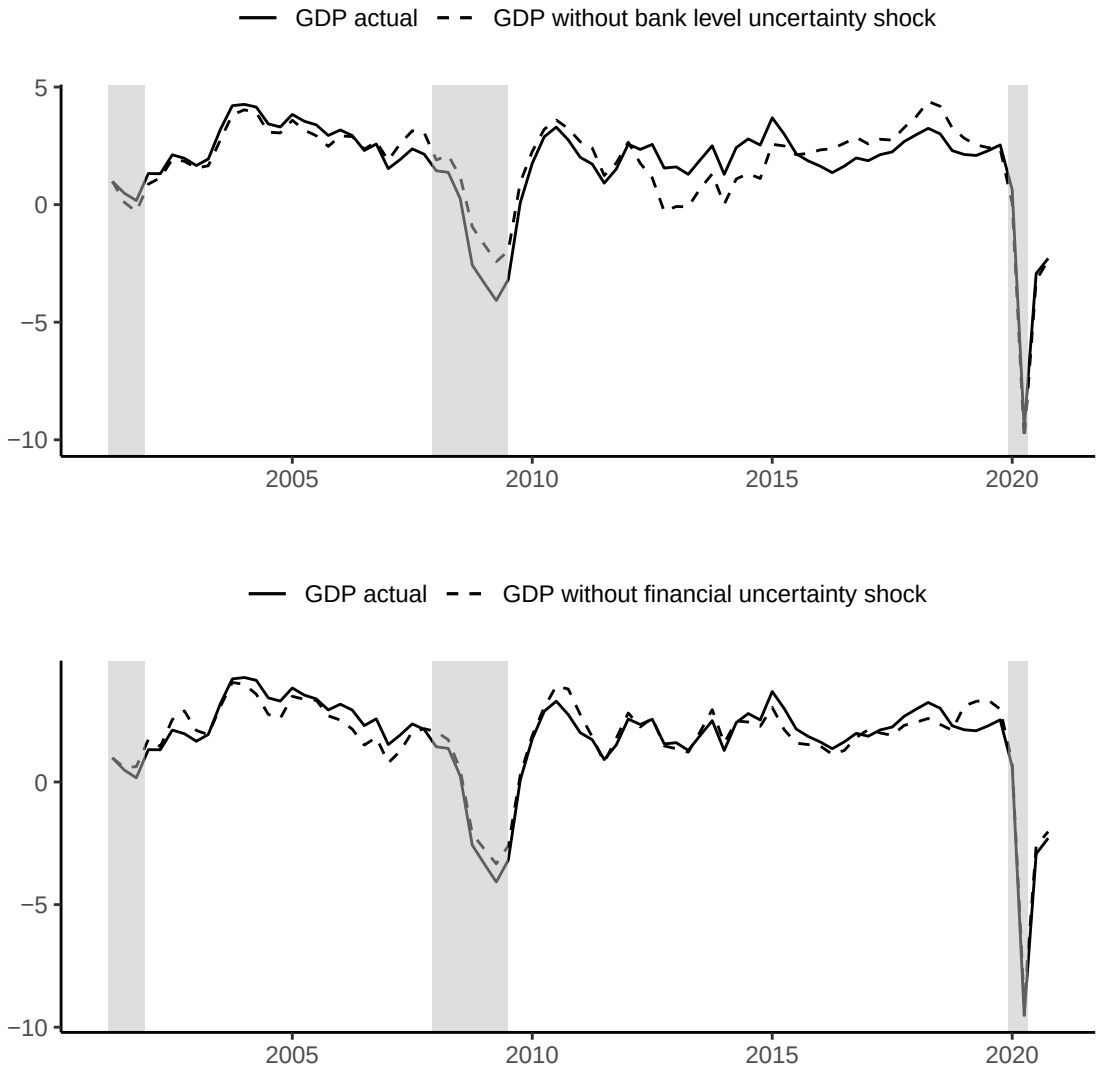
variable at time t, where $1 \leqslant k \leqslant n$. Additionally, let $\hat{y}_{kt}^{(j)}$ $(1 \leqslant j \leqslant n)$ represent the cumulative contribution of shock j to the variable y_{kt}. Following the work of Kilian and Lee (2014), counterfactuals are defined as:

$$y_{kt} - \hat{y}_{kt}^{(j)}.$$

This counterfactual estimate reveals how the variable of interest, k, might have evolved if all realizations of shock j up to time t had been nullified. If a counterfactual falls below the observed value y_{kt}, then the shock j has positively influenced the value of the variable of interest. Conversely, if a counterfactual is above y_{kt}, it indicates that the shock has negatively impacted the value of the variable of interest. Fig. 3.8 portrays the historical bank-level and financial uncertainty counterfactuals. Notably, it indicates that bank-level uncertainty played a substantial role in the onset and magnitude of the Great Recession. If this specific uncertainty had been absent, GDP growth would have exceeded its actual level during the crisis.

In contrast, while financial uncertainty contributed to the economic downturn, its effect was significantly less pronounced than bank-level uncertainty. Furthermore, it is worth noting that between 2013 and 2016, bank-level uncertainty positively affected GDP growth. During this period, the counterfactuals for GDP growth stemming from financial shocks showed no appreciable deviation from the observed values. These insights emphasize the nuanced impact of bank-level versus financial uncertainty on economic fluctuations over time. Moreover, considering banks' pivotal role in the economic landscape—particularly during the Great Recession—these findings advocate a more focused examination of bank-level uncertainty. This is imperative, as a generalized financial uncertainty measure may not capture the unique attributes and influence of the banking sector.

Figure 3.8 Historical conterfactuals of the bank-level uncertainty and financial uncertainty



3.6 Robustness analysis

I evaluate the robustness of the results based on three criteria: (1) varying the VAR specification, (2) considering alternative measures of financial condition, and (3) considering alternative definitions for bank-level uncertainty. Through these three criteria, I aim to ensure the reliability and stability of results and draw robust conclusions from the analysis.

3.6.1 Macroeconomic and financial dynamics of bank-level shock : Alternatives VAR specifications

In the literature, the SVAR framework, particularly the Cholesky decomposition, is widely used to analyze the dynamic behavior of *GDP* and *BAA10YM* in response to uncertainty shocks. However, consensus has not yet been reached on the appropriate composition of the VAR model and the convenient ordering of uncertainty measures. In an influential paper, JLN uses an 11-variable VAR model (VAR-11) and places the uncertainty measure as the last variable. In contrast, Bloom (2009) adopts an 8-variables VAR model (VAR-8) and positions the VIX volatility index, its measure of uncertainty, as the second variable. Similarly, Baker et al. (2016) employ a 5-variables VAR model (VAR-5) and place the measure of uncertainty, economic policy uncertainty, as the first variable.

To check the robustness of findings to VAR specification, I replicate the identification of shocks to bank-level uncertainty (U_t) using the VAR-5, VAR-8, and VAR-11 frameworks. I replace in each specification the uncertainty measure by U_t. The VAR-11 model comprises the following variables: (1) *GDP*, (2) *UNRATE*, (3) *CONSUMPTION*, (4) *INFLATION*, (5) *ORDERS*, (6) *WAGE*, (7) *HOURS*, (8) *FEDFUNDS*, (9) *BAA10YM*, (10) *M2REAL*, (11) U_t.

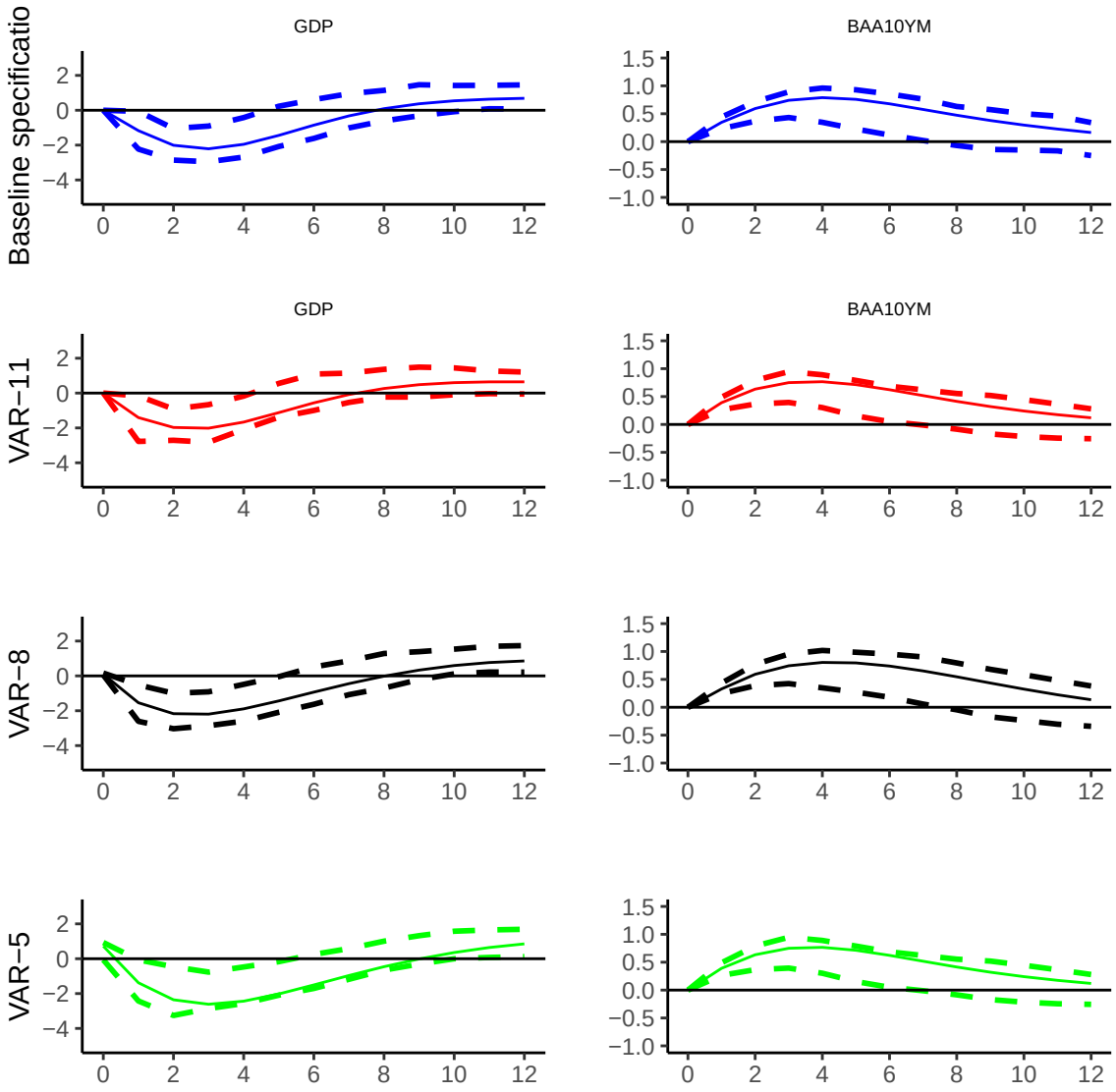
The VAR-8 model consists of the following variables: (1) *BAA10YM*, (2) U_t, (3) *FED-*

FUNDS, (4) *WAGE*, (5) *INFLATION*, (6) *HOURS*, (7) *UNRATE*, and (8) *GDP*.

On the other hand, the VAR-5 model includes (1) U_t, (2) *BAA10YM*, (3) *FEDFUNDS*, (4) *UNRATE*, and (5) *GDP*.

Fig 3.9 plots the IRFs of *GDP* and *BAA10YM* to U_t identified by Cholesky decomposition successively in the baseline specification, VAR-11 of JLN, VAR-8 of Bloom (2009) and VAR-5 of Baker et al. (2016). Based on Fig 3.9, the impact of bank-level uncertainty shock on GDP and $BAA10YM$ is consistent across different specifications. The response of *GDP* and *BAA10YM* to the shock is similar and has the same sign, indicating significant economic and financial consequences. These results suggest that the bank-level uncertainty shock has a substantial macroeconomic and financial effect, regardless of the ordering of the uncertainty measure and the VAR specification.

Figure 3.9 IRFs of GDP growth rate (*GDP*) and *BAA10YM* to a positive shock on the bank level uncertainty identified in different VAR specifications

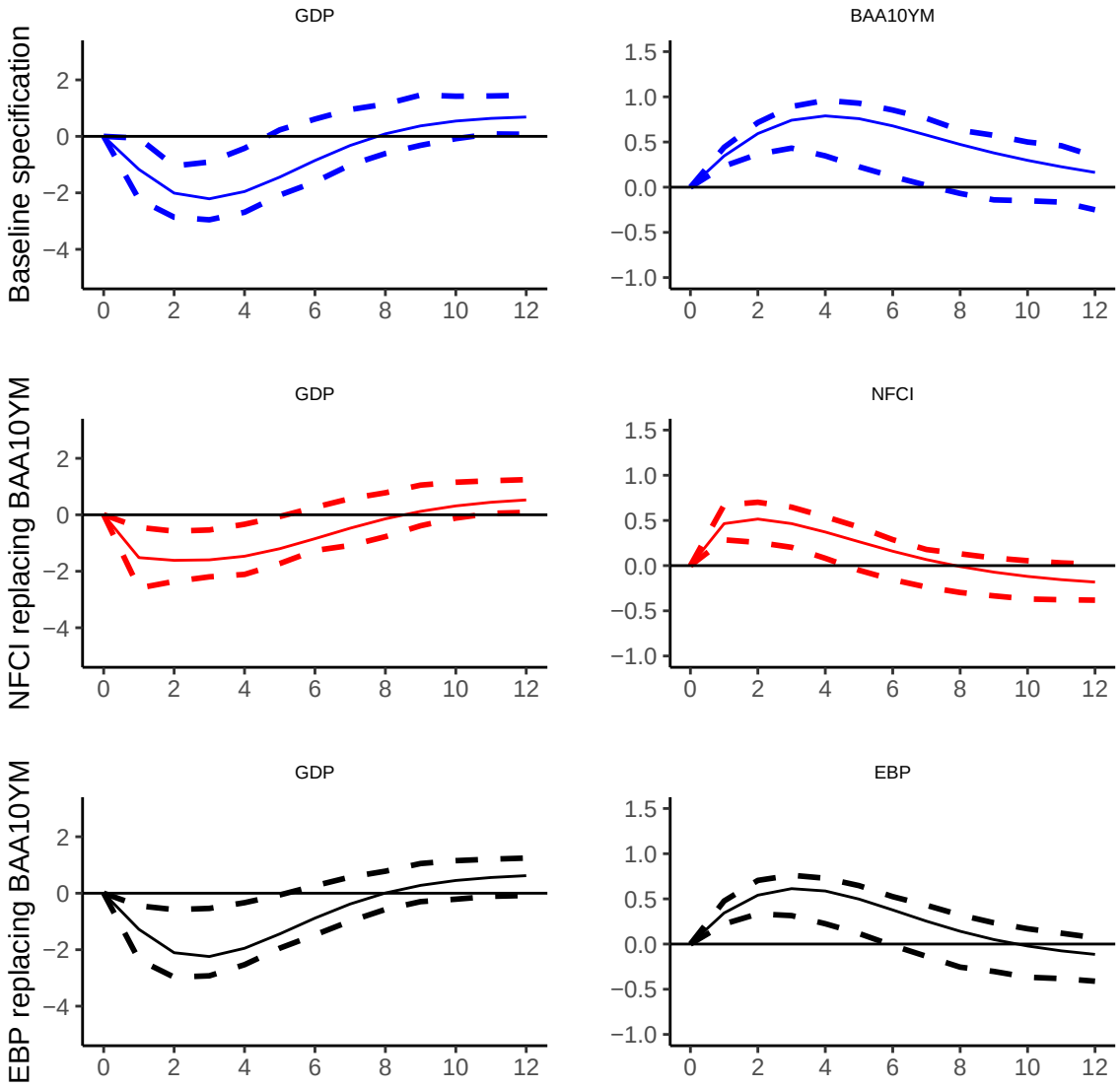


Note : The dots charts represent IRF lower and upper bounds and are constructed at 90% confidence level

3.6.2 Macroeconomic and financial dynamics of bank-level uncertainty : Alternatives measures of financial condition

I have replaced *BAA10YM* with two widely used measures for characterizing financial health: the National Financial Conditions Index (*NFCI*) and the excedent bond premium (*EBP*). The purpose is to assess the consistency and robustness of the results when employing different measures of financial health. Fig. 3.10 compares the IRFs based on the baseline specification with the IRFs obtained sequentially by replacing *BAA10YM* with *NFCI* and *EBP*, respectively. The findings reveal that *GDP* exhibits significant reactions to a bank-level uncertainty shock when *NFCI* and *EBP* are used as replacements for *BAA10YM*. The direction of the responses remains consistent with the baseline specification. However, the magnitudes of these reactions are lower than the baseline specification. Furthermore, following a positive shock on the bank-level uncertainty, *NFCI* and *EBP* significantly increase, similar to *BAA10YM*. Therefore, findings are robust to the measure of financial health used.

Figure 3.10 Comparing baseline IRFs with sequentially *NFCI* and *EBP* replacing *BAA10YM*

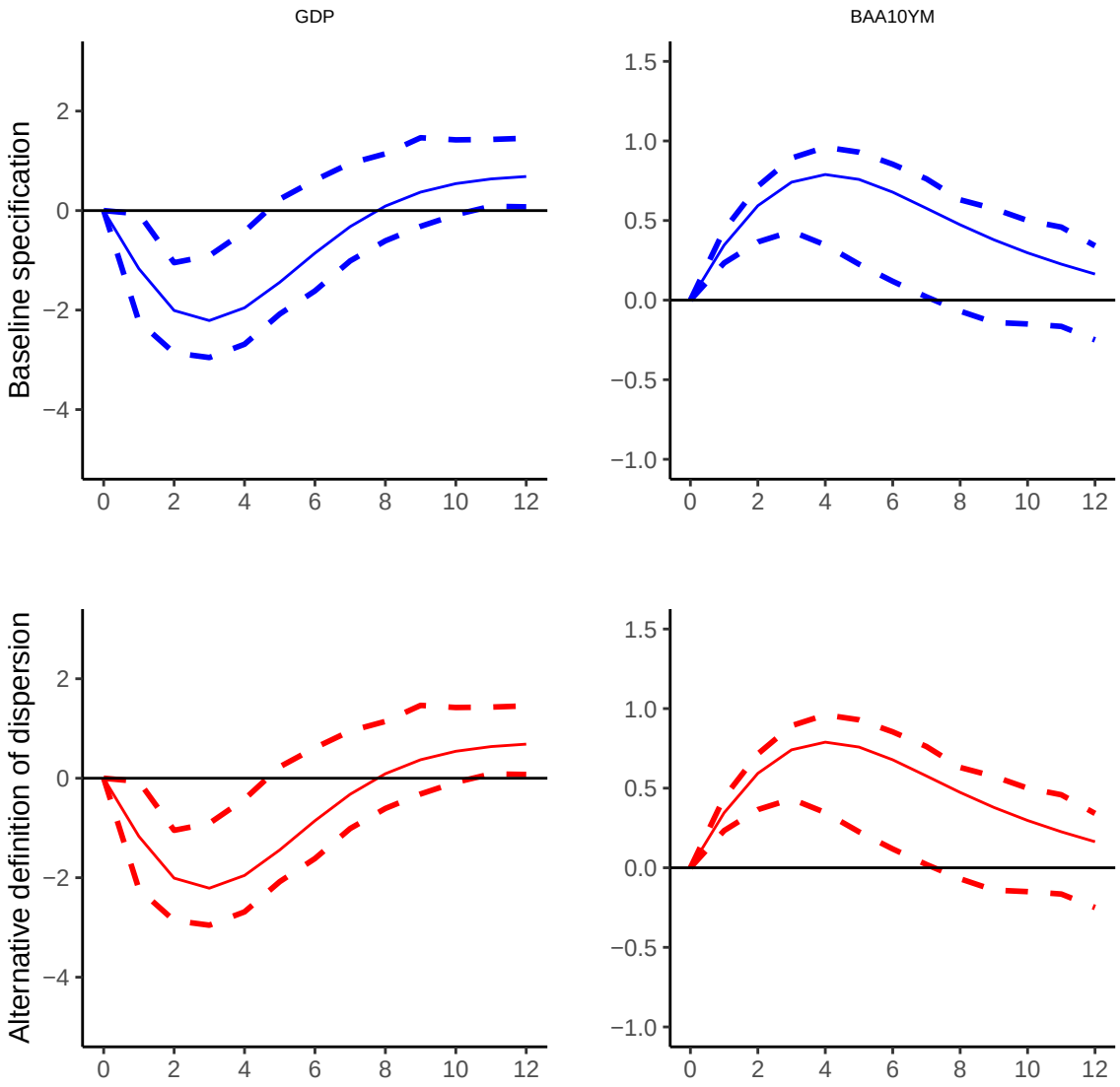


Note : The dots charts represent IRF lower and upper bounds and are constructed at 90% confidence level.

3.6.3 Macroeconomic and financial dynamics of bank-level shock: using an alternative measure of the dispersion of ROA non-forecastable component

To evaluate the impact of the standard deviation versus variance as measures of dispersion for the non-forecastable component of ROA, I replace the standard deviation with variance as an alternative measure and plot the two IRFs in Fig 3.11. This figure shows that the responses of *GDP* and *BAA10YM* are similar in both specifications, indicating the robustness of the approach.

Figure 3.11 Comparing baseline VAR to VAR based on alternative measure of dispersion



Note : The dots charts represent IRF lower and upper bounds and are constructed at 90% confidence level.

3.7 Conclusion

This paper presents a pioneering methodology for constructing a bank-level uncertainty measure employing a data-rich environment and machine learning techniques. The study unfolds in two primary phases. First, I develop an optimal forecast for the bank's return on assets (ROA) by averaging forecasts across various machine learning forecasting models. Second, this optimal forecast is the foundation for the uncertainty measure, defined as the standard deviation of the ROA's non-forecastable component. Implementing this bank-level uncertainty measure in a Vector Autoregression (VAR) framework reveals its significant implications for financial markets and the business cycle. A positive shock to this uncertainty measure triggers a pronounced decline in *GDP* and exacerbates credit market conditions. Compared to alternative dispersion-based measures, the indicator stands out in its unique impact on *GDP* and credit market conditions.

This finding underscores the value of incorporating broader macroeconomic information and distinguishing between forecastable and non-forecastable components when constructing uncertainty measures. Furthermore, results highlight the different effects on the business cycle and the financial sector between, on the one hand, the bank-level uncertainty and the macroeconomic uncertainty and, on the other hand, between the bank-level uncertainty and financial uncertainty. These results suggest that the bank-level uncertainty differs from the macroeconomic and financial uncertainty.

These findings are crucial for financial stability, as they highlight that uncertainty originating in the banking sector can lead to recessions and financial market turmoil. Therefore, policymakers and regulators should monitor developments in the banking sector closely and consider more targeted indicators, as broader financial uncertainty measures may not adequately capture the nuances specific to banks. Future studies could extend this work

by investigating the applicability of bank-level uncertainty measures in different economic contexts and their interplay with other types of uncertainty.

CONCLUSION

The Great Recession of 2007-2009 highlighted the pivotal role of the financial sector in general and the banking sector in particular in maintaining macroeconomic stability, emphasizing its intricate connection with the broader economy. Macroeconomic stability is characterized by an economy that consistently tracks its potential growth, exhibiting steady economic growth and low inflation. Recognizing the importance of this macro-financial relationship, Quadrini (2011) recommends integrating the financial sector into macroeconomic models. Machine learning techniques, known for their adaptability and aptitude for detecting complex non-linearities and managing large datasets, offer promise for constructing these models. Their effectiveness is evident in studies such as Goulet Coulombe et al. 2022 and Fraisse and Laporte 2022.

This thesis aims to assess the effectiveness of machine learning (ML) techniques when combined with granular bank-level data and macroeconomic insights to enhance our understanding of the macro-financial relationship.

In Chapter 1, we employ machine learning techniques within a comprehensive data framework to pinpoint a bank credit supply shock. Our metric — derived from averaging the forecast errors of a predictive model for the bank's capital-to-asset ratio — effectively captures the supply dynamics of bank credit. Indeed, an unfavorable metric fluctuation moves the credit price and volume in the opposite direction, showing that our shock fluctuation translates into supply shock. Intriguingly, a negative shift in this measure results in a pronounced GDP downturn, reduced inflation, and a marked decrease in the volume of bank

credit. It is worth highlighting that the impact of this shock is more sustained and distinct on bank credit than GDP. Comparing our measure to alternative measures of bank credit supply highlights the importance of using more information and finding a suitable model to identify the bank credit supply shock. This evidence backs our agnostic approach.

In Chapter 2, we assess if machine learning techniques can improve the bank's stress test risk analysis. Furthermore, we investigate the role that using more micro and macro information as usual could play in ameliorating the forecast accuracy of key banking variables and, therefore, affecting the stress test. We show that machine learning techniques can improve the stress test indirectly and directly. Indirectly, ML techniques relative to linear models enhance the predictive accuracy of two banking variables: pre-provision net revenue (PPNR) and net charge-off (NCO). Moreover, combining ML models with extensive macroeconomic information further improves the forecast accuracy of PPNR and NCO. Directly, machine learning techniques provide a closer approximation of risk under adverse macroeconomic conditions compared to linear models. As a result, these models offer a more accurate representation of bank vulnerabilities and potential risks during economic downturns.

In Chapter 3, we construct a novel measure of bank-level uncertainty using machine learning techniques combined with granular bank data and big macroeconomic data. Our measure derived as the standard deviation of the non-forecastable component of the bank's return on asset (ROA) follows Jurado et al. (2015), who link uncertainty to difficulty in predicting. We include the measure in a VAR, finding that a positive fluctuation triggers GDP growth slowdown and credit deterioration more critically than with standard macroeconomic and financial measures of uncertainty. This finding highlights the difference between bank-level uncertainty and macroeconomic uncertainty. Furthermore, it recalls that financial uncertainty does not capture all idiosyncratic bank development. It is, therefore,

essential, given the bank's role in the financial system and the economy, to consider more precisely a measure of uncertainty tailored to the bank sector.

This thesis has three significant contributions. First, ML techniques combined with comprehensive data can improve the structural analysis of the economic impact of the bank credit shock by allowing its effective identification. This issue is essential for macroeconomics as wrong identification of the bank credit supply may mislead the real impact of such shock for the broader economy. Second, this thesis shows that ML techniques can enhance risk analysis in a stress test relative to linear models. This issue is essential for financial stability since a wrong picture of risk leads to inaccurate estimation of the systemic risk and corollary to inadequate policy decisions. Third, this thesis suggests using flexible tools of ML techniques and big data to help build a more robust measure of bank-level uncertainty.

BIBLIOGRAPHY

Adrian, T., Boyarchenko, N. and Giannone, D. (2019). Vulnerable growth. *American Economic Review*, *109*(4), 1263–1289.

Adrian, T. and Shin, H. S. (2010). Financial intermediaries and monetary economics. In *Handbook of monetary economics*, volume 3 pp. 601–650. Elsevier.

Ahrens, A., Hansen, C. B. and Schaffer, M. E. (2020). lassopack: Model selection and prediction with regularized regression in stata. *The Stata Journal*, *20*(1), 176–235.

Amiti, M., McGuire, P. and Weinstein, D. E. (2017). *Supply-and demand-side factors in global banking*. Technical report, National Bureau of Economic Research.

Athey, S., Tibshirani, J. and Wager, S. (2019). Generalized random forests.

Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *65*(2), 367–389.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, *70*(1), 191–221.

Bai, J. and Ng, S. (2007). Determining the number of primitive shocks in factor models. *Journal of Business & Economic Statistics*, *25*(1), 52–60.

Baker, S. R., Bloom, N. and Davis, S. J. (2016). Measuring economic policy uncertainty. *The quarterly journal of economics*, *131*(4), 1593–1636.

Barth, M., Emrich, E. and Güllich, A. (2019). A machine learning approach to "revisit" specialization and sampling in institutionalized practice. *Sage Open*, *9*(2), 2158244019840554.

Bassett, W. F., Chosak, M. B., Driscoll, J. C. and Zakrajšek, E. (2014). Changes in bank lending standards and the macroeconomy. *Journal of Monetary Economics*, *62*, 23–40.

Bernanke, B. S., Lown, C. S. and Friedman, B. M. (1991). The credit crunch. *Brookings papers on economic activity*, *1991*(2), 205–247.

Berrospide, J. M. and Edge, R. M. (2010). The effects of bank capital on lending: What do we know, and what does it mean?

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

Bloom, N., Floetotto, M., Jaimovich, N., Saporta-Eksten, I. and Terry, S. J. (2018). Really uncertain business cycles. *Econometrica*, *86*(3), 1031–1065.

Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5–32.

Breiman, L., Friedman, J., Olshen, R. and Stone, C. (1984). Cart. *Classification and regression trees*.

Buch, C. M., Buchholz, M. and Tonzer, L. (2015). Uncertainty, bank lending, and bank-level heterogeneity. *IMF Economic Review*, *63*(4), 919–954.

Carriero, A., Clark, T. E. and Marcellino, M. (2018). Measuring uncertainty and its impact on the economy. *Review of Economics and Statistics*, *100*(5), 799–815.

Chevillon, G. and Hendry, D. F. (2005). Non-parametric direct multi-step estimation for forecasting economic processes. *International Journal of Forecasting*, *21*(2), 201–218.

Covas, F. B., Rump, B. and Zakrajšek, E. (2014). Stress-testing us bank holding companies: A dynamic panel quantile regression approach. *International Journal of Forecasting*, *30*(3), 691–713.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, *2*(4), 303–314.

Diebold, F. X. and Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & economic statistics*, *20*(1), 134–144.

Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004). Least angle regression.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the econometric society*, pp. 987–1007.

Fraisse, H. and Laporte, M. (2022). Return on investment on artificial intelligence: The case of bank capital requirement. *Journal of Banking & Finance*, *138*, 106401.

Friedman, J., Hastie, T., Höfling, H. and Tibshirani, R. (2007). Pathwise coordinate optimization.

Friedman, J., Hastie, T. and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, *33*(1), 1.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pp. 1189–1232.

Fu, W. J. (1998). Penalized regressions: the bridge versus the lasso. *Journal of computational and graphical statistics*, *7*(3), 397–416.

Galvao Jr, A. F. (2011). Quantile regression for dynamic panel data with fixed effects. *Journal of Econometrics*, *164*(1), 142–157.

Gilchrist, S., Sim, J. W. and Zakrajšek, E. (2014). *Uncertainty, financial frictions, and investment dynamics*. Technical report, National Bureau of Economic Research.

Gilchrist, S. and Zakrajšek, E. (2012). Credit spreads and business cycle fluctuations. *American economic review*, *102*(4), 1692–1720.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, *102*(477), 359–378.

Goulet Coulombe, P., Leroux, M., Stevanovic, D. and Surprenant, S. (2022). How is machine learning useful for macroeconomic forecasting? *Journal of Applied Econometrics*, *37*(5), 920–964.

Grover, S. and McCracken, M. W. (2014). Factor-based prediction of industry-wide bank stress. *Sean Grover & Michael W. McCracken." Factor-Based Prediction of Industry-Wide Bank Stress." Review*, *96*(2), 173–193.

Gu, S., Kelly, B. and Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, *33*(5), 2223–2273.

Guerrieri, L. and Welch, M. (2012). Can macro variables used in stress testing forecast the performance of banks?

Hirtle, B., Kovner, A., Vickery, J. and Bhanot, M. (2016). Assessing financial stability: The capital and loss assessment under stress scenarios (class) model. *Journal of Banking & Finance*, *69*, S35–S55.

Hornik, K., Stinchcombe, M. and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, *2*(5), 359–366.

Iacopini, M., Ravazzolo, F. and Rossini, L. (2023). Proper scoring rules for evaluating density forecasts with asymmetric loss functions. *Journal of Business & Economic Statistics*, *41*(2), 482–496.

James, G., Witten, D., Hastie, T., Tibshirani, R. et al. (2013). *An introduction to statistical learning*, volume 112. Springer.

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

Kapinos, P. and Mitnik, O. A. (2016). A top-down approach to stress-testing banks. *Journal of Financial Services Research*, *49*, 229–264.

Kilian, L. and Lee, T. K. (2014). Quantifying the speculative component in the real price of oil: The role of global oil inventories. *Journal of International Money and Finance*, *42*, 71–87.

Liu, L., Moon, H. R. and Schorfheide, F. (2020). Forecasting with dynamic panel data models. *Econometrica*, *88*(1), 171–201.

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

Marcellino, M., Stock, J. H. and Watson, M. W. (2006). A comparison of direct and iterated multistep ar methods for forecasting macroeconomic time series. *Journal of econometrics*, *135*(1-2), 499–526.

Masters, T. (1993). *Practical neural network recipes in C++*. Morgan Kaufmann.

McCracken, M. and Ng, S. (2020). *FRED-QD: A quarterly database for macroeconomic research*. Technical report, National Bureau of Economic Research.

Mésonnier, J.-S. and Stevanovic, D. (2017). The macroeconomic effects of shocks to large banks' capital. *Oxford Bulletin of Economics and Statistics*, *79*(4), 546–569.

Plagborg-Møller, M., Reichlin, L., Ricco, G. and Hasenzagl, T. (2020). When is growth at risk? *Brookings Papers on Economic Activity*, *2020*(1), 167–229.

Quadrini, V. (2011). Financial frictions in macroeconomic fluctuations. *FRB Richmond Economic Quarterly*, *97*(3), 209–254.

Romer, C. D. and Romer, D. H. (2004). A new measure of monetary shocks: Derivation and implications. *American economic review*, *94*(4), 1055–1084.

Schorfheide, F. (2005). Var forecasting under misspecification. *Journal of Econometrics*, *128*(1), 99–136.

Shevade, S. K. and Keerthi, S. S. (2003). A simple and efficient algorithm for gene selection using sparse logistic regression. *Bioinformatics*, *19*(17), 2246–2253.

Soto, P. E. (2021). Breaking the word bank: Measurement and effects of bank level uncertainty. *Journal of Financial Services Research*, *59*, 1–45.

Stock, J. H. and Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of forecasting*, *23*(6), 405–430.

Stock, J. H. and Watson, M. W. (2005). Implications of dynamic factor models for var analysis.

Watkins, G. P. (1922). Knight's risk, uncertainty and profit. *The Quarterly Journal of Economics*, *36*(4), 682–690.

Winkler, R. L. and Murphy, A. H. (1979). Use of probabilities in forecasts of maximum and minimum temperatures. *Meteorological Magazine*, *108*(1288), 317–329.

APPENDIX

Appendix A. Description of different machine learning techniques

In this appendix, we provide detailed description of machine learning techniques, we employ in this thesis.

Appendix A.1 Penalized linear regression methods

I have considered the main penalized linear regression methods, which encompass the "lasso" regression and its variant, the "adaptive lasso" regression, as well as the "elastic net" regression and the "ridge" regression.

Model

These methods are based on the linear assumption, where the relationship between variables can be represented as $g(Z_{i,t}, \theta)$=$Z'_{i,t}\theta$.

The objective function is defined as follows:

$$Q(\theta; .) \quad = L(\theta; .) + \Phi(\theta; .) \; . \tag{A.1}$$

where $L(\theta; .)$ represents the mean square error (MSE), and $\Phi(\theta; .)$ is the penalty term that varies across different methods. Generally, the penalty function takes the following form :

$$\Phi(\theta, \lambda, \alpha, \Psi) \quad = \tfrac{\lambda}{NT}[\alpha \textstyle\sum_{j=1}^{K} \Psi_j |\theta_j| + (1-\alpha) \sum_{j=1}^{K} \Psi_j^2 \theta_j^2 \,] \; . \tag{A.2}$$

Here, K denotes the number of potential predictors, while λ and α are hyperparameters. Ψ_j (j=1...K) represents the weights assigned to the predictors. Additionally, N stands for the number of banks, and T represents the number of quarters in the estimation sample. The optimization problem can be formulated as:

$$\hat{\theta} \quad = argmin_{\theta}(L(\theta;.) + \Phi(\theta;.)) \,, \tag{A.3}$$

where $\hat{\theta}$ denotes the estimate of the vector of parameters θ obtained using the selected method.

Several cases arise:

$\alpha = 1$, $\Psi_j = 1$, K=1,...p corresponds to the "lasso" regression. The "lasso" regression uses the L1 norm[12] to penalize the coefficients. For the high values of λ , the "lasso" regression sets certain coefficients to zero. It simultaneously estimates the forecasting model and selects relevant predictors. To determine the optimal value of λ, multiple forecasting models are estimated using different values of λ from a grid, and the optimal value is chosen through cross-validation[13].

$\alpha = 0$, $\Psi_j = 1$, K=1,...p corresponds to the "ridge" regression. The "ridge" regression

[12]Let $\theta = (\theta_1, ..., \theta_p)$ a vector, $\|\theta\|_{L1}$=$\sum_{j=1}^{K}|\theta_j|$

[13]It is the standard approach to calculating the performance of a forecasting model, which consists of : (1) dividing the sample randomly into k groups of approximately equal size, (2) considering group 1 as the validation sample and estimating the forecast model on the remaining k-1 groups, (3) calculating the mean squared error of the forecast model on this validation sample (without harming the generality, it can be noted CV1), (4) redoing the same exercise k times by treating each of the remaining k-1 groups as a validation sample, (5) computing the performance of the forecasting model by determining the mean of CVi i = 1, ... k. By noting this performance by CV, we can write CV=$\frac{CV1 + CV2 + ... + CVk}{k}$.

employs the L2 norm[14] to penalize the coefficients. Unlike the L1 penalty, the L2 penalty shrinks some coefficients towards zero but does not set them exactly to zero. Therefore, the "ridge" regression does not perform predictor selection. Similar to the "lasso" regression, the optimal value of λ is determined by minimizing the model mean squared error (MSE) on a validation set.

$\alpha \in (0,1)$, $\Psi_j = 1$, j=1,...K corresponds to the "elastic net" regression. The "elastic net" regression combines the L1 and L2 norms to penalize the coefficients. The hyperparameter α determines the relative weight between the two norms. Grids of values are constructed for both α and λ, and multiple models are built for all possible combinations. The optimal combination is selected through cross-validation.

$\alpha = 1$, Ψ_j j=1...K varying as a function of j, corresponds to the "adaptive lasso" regression. In the "adaptive lasso" regression, the weights Ψ_j of the predictors are determined from the data. Zou (2006) suggests a two-step approach to determine these weights.In the first step, a weight allocation technique is employed to assign weights to different variables. Zou (2006) suggests regressing the variable of interest on the potential predictors using OLS (or any other estimation method such as "ridge" regression). The estimated coefficient for variable j, denoted as $\hat{\theta}_{jOLS}$, along with a chosen parameter γ>0 is used to define the weight attached to variable z_{jt} as $\Psi_j = \frac{1}{|\hat{\theta}_{jOLS}|^{\gamma}}$. The value of γ is determined through cross-validation.

In the second step, the weights Ψ_j obtained in the first step are used in the "lasso" regression. The optimal value of λ is also determined using cross-validation.

To solve the optimization problem for the different estimation techniques, I employ the

---

[14]Let $\theta = (\theta_1, ..., \theta_p)$ a vector, $\|\theta\|_{L2}$=$\sum_{j=1}^{p} \theta_j^2$

coordinated rapid descent algorithm, which is outlined below.

Algorithm

The coordinated rapid descent algorithm was first proposed by Fu (1998) for estimating the "lasso" regression. Subsequently, it has been employed in various empirical studies, including Shevade and Keerthi (2003). In a study by Friedman et al. (2007), it was demonstrated that this algorithm is as efficient as the "Least Angle Regression" algorithm (LARS) developed by Efron et al. (2004). LARS is widely used in the literature for estimating the "lasso" regression.

Now, let's establish the framework of a linear regression model. In this model, N denotes the sample size, where y_i i=1...N represents the variable of interest and x_{ij} i=1...N, $1 \leqslant j \leqslant K$ represents the K potential predictors. For simplicity, let's assume that the predictors are centered and standardized, meaning that $\sum_{i=1}^{N} x_{ij}$=0 et $\frac{1}{N}\sum_{i=1}^{N} x_{ij}^2$=1, j=1...K . The "elastic net" regression solves the following problem:

$$min_{(\beta_0,\beta)\in\mathbb{R}^{K+1}} R_\lambda(\beta_0,\beta) \quad = min_{(\beta_0,\beta)\in\mathbb{R}^{K+1}} [\tfrac{1}{2N}\sum_{i=1}^{N}(y_i-\beta_0-x_i'\beta)^2+\lambda\Phi_\alpha(\beta)] \tag{A.4}$$

where

$$\Phi_\alpha(\beta) \quad = \sum_{j=1}^{K}[\tfrac{1}{2}(1-\alpha)\beta_j^2+\alpha|\beta_j|] \, . \tag{A.5}$$

Here, λ and α are the two hyperparameters of the "elastic net" regression. The hyperparameter α controls a trade-off between "lasso" regression (α =1) and the "ridge" regression (α =0), both of which are special cases of the "elastic net" regression.

Without loss of generality, let $x_{i,j}$, j=1...K represent a prédictor. The conjectures $\tilde{\beta}_0$ and $\tilde{\beta}_l$ for l$\#j$serve as the initial values in the coordinated descent algorithm. In the second step, the objective function is optimized solely with respect to β_j, requiring the computation of

the gradient at $\beta = \tilde{\beta}$. This procedure updates $\tilde{\beta}_j$, $1 \leqslant j \leqslant p$. Two cases arise:

First case: $\tilde{\beta}_j > 0$

$$\frac{\partial R_\lambda}{\partial \beta_j} \mid \beta = \tilde{\beta} \;\; = -\tfrac{1}{N} \textstyle\sum_{i=1}^{N} x_{ij}(y_i - \beta_0 - x_i'\beta) + \lambda(1-\alpha)\beta_j + \lambda\alpha \; . \qquad \text{(A.6)}$$

Second case: $\tilde{\beta}_j < 0$

$$\frac{\partial R_\lambda}{\partial \beta_j} \mid \beta = \tilde{\beta} \;\; = -\tfrac{1}{N} \textstyle\sum_{i=1}^{N} x_{ij}(y_i - \beta_0 - x_i'\beta) + \lambda(1-\alpha)\beta_j - \lambda\alpha \; . \qquad \text{(A.7)}$$

The optimization of R_λ with respect to β_j implies that $\frac{\partial R_\lambda}{\partial \beta_j} \mid \tilde{\beta} = 0$.
After simplification, I get:

$$\begin{cases} \tilde{\beta}_j &= \frac{\frac{1}{N}\sum_{i=1}^{N}(y_i - \tilde{y_i}^{(j)}) - \lambda\alpha}{1+\lambda(1-\alpha)} \; si \; \frac{1}{N}\sum_{i=1}^{N}(y_i - \tilde{y_i}^{(j)}) > 0 \; et \; |\frac{1}{N}\sum_{i=1}^{N}(y_i - \tilde{y_i}^{(j)})| > \lambda\alpha \; ; \\ \tilde{\beta}_j &= \frac{\frac{1}{N}\sum_{i=1}^{N}(y_i - \tilde{y_i}^{(j)}) + \lambda\alpha}{1+\lambda(1-\alpha)} \; si \; \frac{1}{N}\sum_{i=1}^{N}(y_i - \tilde{y_i}^{(j)}) < 0 \; et \; |\frac{1}{N}\sum_{i=1}^{N}(y_i - \tilde{y_i}^{(j)})| > \lambda\alpha \; ; \\ \tilde{\beta}_j &= 0 \; si \; |\frac{1}{N}\sum_{i=1}^{N}(y_i - \tilde{y_i}^{(j)})| \leqslant \lambda\alpha \; . \end{cases}$$

$y_i^{(j)} = \tilde{\beta}_0 + \sum_{l \neq j} x_{il}\tilde{\beta}_l$ is the predicted value of the variable of interest excluding the contribution of the variable x_{ij}. If I set Z=$\frac{1}{N}\sum_{i=1}^{N}(y_i - \tilde{y_i}^{(j)})$ and $\gamma = \lambda\alpha$, then the expression can be simplified as follows:

$$\begin{cases} \tilde{\beta}_j &= \frac{Z-\gamma}{1+\lambda(1-\alpha)} \; si \; Z > 0 \; et \; |Z| > \gamma \; ; \\ \tilde{\beta}_j &= \frac{Z+\gamma}{1+\lambda(1-\alpha)} \; si \; Z < 0 \; et \; |Z| > \gamma \; ; \\ \tilde{\beta}_j &= 0 \; si \; |Z| \leqslant \gamma \; . \end{cases}$$

Formally, the algorithm is broken down into three stages:

Step 1:

Conjecture estimates $\tilde{\beta}_j$, $0 \leqslant j \leqslant K$.

Step 2:

Update $\tilde{\beta}_j$, $1 \leqslant j \leqslant p$ by the formulas listed above.

Step 3:

Repeat step 2 until convergence.

Friedman et al. (2007) demonstrated that the estimates $\tilde{\beta}_j$ converge to the estimates of penalized linear regression, irrespective of the initial conjectures. To facilitate the estimation of these regressions, various computer routines have been developed, particularly in R and STATA. In R, Friedman et al. (2010) created routines that are available in the "glmnet" package. These routines provide the capability to estimate these regressions using the coordinated descent algorithm. Similarly, in STATA, Ahrens et al. (2019) developed routines for estimating these regressions. These STATA routines offer the functionality to apply the coordinated descent algorithm for estimation.

Appendix A.2 Principal component regression method

The principal component regression is a data machine learning technique that can be applied to linear models. However, it differs from previous approaches in not involving variable selection or coefficient reduction. This method is structured in two stages. In the first stage, given that $Z_{i,t}$ represents a vector composed of K potential predictors, M linear combinations of the $Z_{i,t}$ vectors are constructed to explain the most significant part of the data's variance. These linear combinations are known as principal components. In the second stage, these principal component combinations replace the p potential regressors, and the forecast model is estimated using ordinary least squares ("OLS"). The principle of the method is to summarize the information contained in the potential regressors by a

reduced number of principal components, which substitute the initial predictors and effectively reduce the problem's dimensionality.

Model

The linear model $Y_{i,t+h} = Z'_{i,t}\theta + \epsilon_{i,t+h}$ can be expressed in matrix form as:

$$Y \quad = Z\theta + \varepsilon \,, \tag{A.8}$$

which can be further written as :

$$Y \quad = (Z\Omega_M)\theta_M + \varepsilon' \,. \tag{A.9}$$

where Ω_M is a matrix of type K $\times$M consisting of columns $\omega_1, \omega_2, ...\omega_M$, which represent the weight vectors assigned to the predictors in the construction of M principal components. The matrix $Z\Omega_M$ is an NT$\times$ M matrix composed of the M principal components. θ_M is an M$\times$ 1 vector that contains M parameters to be estimated instead of K. The model in equation (31) represents the reduced version of the initial model obtained through principal component analysis. The hyperparameter in this method is denoted as M, which represents the number of principal components. Similar to the previous methods, a grid of potential values for M is constructed. Different models corresponding to different potential values of the hyperparameter are then estimated using the training sample. The optimal value of M is determined through cross-validation on the estimation sample. The principal component regression method selects the weights, ω_j $(1 \leq j \leq M)$, recursively. The jth linear combination solves the following optimization problem:

$$w_j \quad = argmax_w Variance(Zw)\ s.c\ \ w'_j w_j = 1,\ cov(Zw_j, Zw_l) = 0\ for\ l \neq j \tag{A.10}$$

Appendix A.3 Methods based on decision trees

The previous methods assume that the relationship between the predictors and parameters, denoted as $g(Z_{i,t}, \theta)$, can be approximated by a linear function. However, this linearity assumption may be too restrictive in cases where there is an evident nonlinearity between the variable of interest and the predictors and/or parameters. Furthermore, linear models ignore any interactions between predictors. To this end, intending to find the best model for forecasting PPNR or NCO, it is also necessary to consider the nonlinearity hypothesis.

Tree regression is indeed a widely used nonlinear machine learning technique that offers the advantage of capturing potential interactions between predictors. This method consists of: (1) stratifying, segmenting the space of predictors into a number of simple regions, (2) calculating the mean of the variable of interest for the observations of the estimation sample belonging to each region, (3) classify each observation of the validation sample in one of the regions, (4) predict the variable of interest for an observation of the validation sample by the mean calculated in (2) of his region. The subdivision of the predictor space into regions, guided by the mean squared error minimization criterion, results in the appearance of a tree-like structure. This is why the term "tree regression" is used to describe this method.

By way of illustration, I present an example of Tibshirani et al. (2013). The purpose of the exercise is to predict the log of baseball players' salaries based on two characteristics: (1) the number of years competing in the big leagues and (2) the number of successes in the league. the previous year. The tree regression in this example is summarized by the following subdivision of the predictor space:

1) if "number of years of competition in the big leagues" <4.5 then log (salary) = 5.11;

2) if "number of years of competition in the big leagues"> 4.5 and "number of successes

of the previous year" <117.5 then log (salary) = 6.00;
3) if "number of years of competition in the big leagues"> 4.5 and "number of successes of the previous year"> 117.5 then log (salary) = 6.74.

In the provided example, the predictor space is divided into three simple regions, and the projected salaries on a logarithmic scale are assigned based on the region. It's important to note that in real-world applications, the subdivisions of the predictor space can be more complex and involve multiple conditions or thresholds. One of the advantages of tree regression is its interpretability. The resulting tree structure allows straightforward interpretation and understanding of the relationship between predictors and the target variable. Each region represents a specific combination of predictor values and provides a precise prediction. However, tree regression can suffer from high variability, negatively impacting prediction performance on the validation sample. Two commonly used methods in the literature that address this issue and reduce variance are random forest and "boosting" regression.

A.3.1 Random Forest

The random forest consists of (1) estimating several regressions of decorrelated trees on samples modified by "bootstrap"[15], (2) constructing the prediction of the variable of interest, and (3) computing the prediction's average on the different trees. Averaging predictions allows for reducing and stabilizing the forecast's variance, unlike the simple regression of trees. To proceed to the decorrelation of the trees, a number m of predictors drawn

[15]The "bootstrap" is a statistical technique that consists of constructing several samples from an initial one by random selection with or without replacement to estimate the moment of an estimator or to improve its precision.

randomly from among K potential predictors (m <K) are used at each tree subdivision. This exercise prevents the case of a very predominant predictor from being found at the first subdivision of all the trees. It thus creates a correlation between them, as is the case in the "bagging" method, which is similar in every way to the technique studied with the only difference that m = K.

Model

Suppose that I construct B trees on the bootstrap estimation sample. Suppose further that the subdivision of each tree randomly uses m predictors with m <K.

The regression on a tree j, (j = 1 .... B) supposes a model of the form $f_j(Z_{i,t}, \theta) = \sum_{k=1}^{n_j} \theta_k . 1_{(Z_{i,t} \in R_k)}$ where $R_1, R_2, ...R_{n_j}$ are the regions of tree j; $Z_{i,t}$ is the vector comprising the characteristics of the banks, the dichotomous variables and the macroeconomic variables. θ is the vector of the parameters to be estimated.

The random forest thus assumes the following model:

$$g(Z_{i,t}, \theta) = \frac{1}{B} \sum_{j=1}^{B} f_j(Z_{i,t}, \theta) \, . \tag{A.11}$$

B represents the number of trees constructed, and m is the number of predictors drawn randomly among K potential predictors and used for the subdivision of trees. They are the two hyperparameters of this method. The question of determining optimal hyperparameter values is central to the success of any machine learning technique, and this method is no exception. For this purpose, I construct two grids of values for these two parameters, and the optimal values of B and m are those that minimize the criterion "out of bag error" (OOB) [16] provided by the estimation of the model.

---

[16] It is shown that the construction of each tree uses on average two-thirds of the observations. The remaining third can be used as a validation sample. For each observation i, about B / 3 trees was not built with this observation; the latter is used to construct predictions outside the estimation sample, which the

The regression of the tree j (j = 1 ... B) is solved thanks to the following problem:

$$argmin_{R,\theta} \sum_{i=1}^{N} \sum_{t=1}^{T} (Y_{i,t+1} - g(Z_{i,t}, \theta))^2 \,. \tag{A.12}$$

Indeed, the subdivision of the tree and the parameters are chosen to minimize the mean square error while ensuring a decorrelation of the trees.

Algorithm

If we know the objective of tree regression, which is to determine for a tree j the regions $R_1, R_2, ..., R_{n_j}$ in a way to minimize the RMSE, we have not yet specified the procedure for achieving this objective. The most common procedure in the literature was highlighted by Breiman (2001). This top-down approach, better known as "recursive binary splitting," consists of looking for the optimal subdivision at each step. In more detail, it starts at the top of the tree, where all observations belong to the same region.

In the first step, I consider all of the predictors $Z_1, Z_2, ..., Z_p$ and all of the possible values of the subdivision points for each predictor. Then, for any predictor Z_j and any subdivision point s, I define the subdivision of the following predictor space:R1(j,s)= $\{Z/Z_j < s\}$ and R2(j,s)=$\{Z/Z_j \geqslant s\}$. I select the values j and s for which the subdivision allow for the minimization of the equation:

$$\sum_{i:\, Z_{ij} \in R1(j,s)} (y_i - \hat{y}_{R_1})^2 + \sum_{i:\, Z_{ij} \in R2(j,s)} (y_i - \hat{y}_{R_2})^2$$

where $y_{i,t}$ is the variable of interest, $\hat{y}_{R_1}$ is the mean of the variable of interest and for the observations of the estimation sample belonging to R1(j,s) and $\hat{y}_{R_2}$. After this step, I

---

mean will synthesize. "OOB" is then the mean square error resulting from these forecasts.

have two regions. I continue the process of subdivision of each of these regions always to minimize the above equation until the stopping point dictated by a previously defined criterion.[17]. Since the minimization of the mean squared error relates to the estimation sample, the risk of over-training the data is real. Breiman (1984) recommends pruning optimal tree branches for this purpose. It is a trade-off similar to the cases of penalized linear regressions made between the model's fit to the data and its complexity. The tree regression algorithm of Breiman (1984) synthesized below is taken from James et al. (2013).

Step 1:

Use the binary splitting described above to build an optimal large tree T_0 based on the estimation sample. The criterion for stopping the subdivision of the tree is the imposition of a minimum number of observations in each region.

Step 2:

For each potential value of α that is the coefficient penalizing the complexity of the tree, construct a subtree T$\subset T_0$ such that:

$$\sum_{m=1}^{N_T} \sum_{x_i \in R_m} (y_i - \hat{y}_{R_m})^2 + \alpha N_T \, ,$$

has the lowest possible value. N_T is the number of terminal nodes of subtree T and, α is a coefficient that makes it possible to ensure the trade-off between the model's suitability to the data and its complexity. The higher α is, the less terminal the subtree has. $\hat{y}_{R_m}$ is the average of the variable of interest y_i for the observations of the estimation sample belonging to the region R_m.

Step 3:

---

[17]The criterion defined in general is the imposition on the regions $R_1, R_2, ..., R_{n_j}$ to each contain a minimum number of observations.

Use cross-validation on the estimation sample to determine the optimal value of α that minimizes RMSE.

Step 4:

Go back to step 2 to build the subtree corresponding to the optimal value of α. This optimal subtree is used for forecasting.

Since the random forest is obtained from the regression of trees, the algorithm is also deduced from it. This algorithm is structured in three stages:

Stage 1:

Build a number B of samples from the bootstrap estimation sample and estimate the tree regression on each sample according to the above algorithm by randomly selecting from each subdivision m predictors among K (m <K).

Stage 2:

Build the$\hat{y}_{ij}$ forecasts for $1 \leqslant j \leqslant B$ obtained from the different tree regressions.

Stage 3:

Construct the forecast obtained by the random forest as $\hat{y}_i = \frac{1}{B} \sum_{j=1}^{B} \hat{y}_j$

A.3.2 Boosting regression

The boosting regression approach is similar to the previous method, which involves constructing multiple trees. However, there is a fundamental difference. While the random decision tree drill method relies on regressing noncorrelated trees using independent 'bootstrap' samples, boosting regression constructs trees sequentially. Each tree is built using

information from the preceding trees. The underlying principle of this technique is slow and progressive learning. In each iteration, a decision tree is estimated by replacing the variable of interest with the residuals from the model. This new tree is then added to the prediction from the previous iteration, updating both the prediction and the residuals. The process is governed by a regularization parameter denoted as λ. This incremental learning approach gradually improves prediction performance at each iteration and continues until the final tree is constructed. In the literature, there are several variants of boosting regression. The most widely used variant is known as 'gradient boosting,' introduced by Friedman (2001).

Model

Let B be the number of trees built sequentially according to a regularization parameter λ, each tree having d subdivisions. At iteration j, where $1 \leqslant j$ B, let $f_j(Z_{i,t}, \theta)$ denote the prediction obtained:

$$g(Z_{i,t}, \theta) = \lambda \sum_{j=1}^{B} f_j(Z_{i,t}, \theta). \tag{A.13}$$

To determine the optimal combination of these hyperparameters, I construct value grids for each parameter and employ cross-validation. This process helps identify the best hyperparameter settings. Optimization is performed for each tree regression, and the final predictor is obtained by combining the predictions from all iterations using Equation B.13.

Algorithm

Let y_i be the variable of interest, $\hat{f}(x)$ the prediction of y, x the vector of predictors, and r_i the residue of the regression. The algorithm comes in three steps.

Step 1:

I set $\hat{f}(x)$=0 and consequently $r_i = y_i$ for any observation i belonging to the estimation

sample.

Step 2:

Let B be the number of trees to build, d the number of subdivisions of each tree, λ the regularization parameter, and m the fraction of the estimation sample chosen at random. For j = 1,2, ... B repeat:

- Build a regression tree whose dependent variable is r_i and has d subdivisions (d + 1 terminal nodes) on the estimation sample and extract the prediction from it $\hat{f}^j(x)$.
- Update $\hat{f}(x)$ by the formula: $\hat{f}(x) \leftarrow \hat{f}(x)+\lambda\ \hat{f}^j(x)$.
- Update r_i by the formula : $r_i \leftarrow r_i - \lambda\hat{f}^j(x_i)$.

Step 3:

The forecast model is given by $\hat{f}(x) =\lambda \sum_{j=1}^{B} \hat{f}^j(x)$.

Appendix A.4 Neural network method

The neural network method is the last nonlinear technique studied. It is by far the most complex and powerful technique for machine learning. Its complexity is due to its hyper parametrization, making it not very transparent and difficult to interpret. Its power comes from its ability to approximate most nonlinear functions (see Hornik (1989), Cybenko (1989)).

Model

I opted for the neural network method with forward tuning, which is widely used in the literature. This technique consists of an input layer composed of the initial predictors- one

or more hidden layers that interact and transform the predictors nonlinearly and an output layer that aggregates the hidden layers into a final result, the prediction.

Synapses link the layers together. Each layer is made up of a group of neurons[18].The neurons of the first layer directly receive the signals of the initial predictors via the synapses and transform them nonlinearly using a function called the activation function. The signals thus constructed at the level of each neuron of the first hidden layer are transmitted to neurons of the second layer, which transform them nonlinearly by the same activation function. This process continues until the final aggregation of the signals is built at the level of the neurons of the last layer which provides the prediction.

Suppose that K is the number of potential predictors and k is the number of hidden layers. Suppose for this purpose that n_1 is the number of neurons for the first hidden layer, n_2, the number of neurons for the second hidden layer, ...n_k, the number of neurons for the kth hidden layer. Let us denote by $x_1^{(1)}$, $x_2^{(1)}$, ...$x_{n_1}^{(1)}$ the signals constructed at the level of neurons of the first hidden layer. Analogously, $x_1^{(j)}$, $x_2^{(j)}$, ...$x_{n_j}^{(j)}$ represent the signals constructed at the level of the neurons of the hidden layer j ($1 \leqslant j \leqslant k$). f being the activation function, I can write:

$x_i^{(1)}$=f($\theta_{i,0}^{(1)} + \sum_{h=1}^{p} \theta_{i,h}^{(1)} z_h$), $1 \leqslant i \leqslant n_1$. In general, the signal constructed at the level of the neuron i of the hidden layer j can be written: $x_i^{(j)}$=f($\theta_{i,0}^{(j)} + \sum_{h=1}^{n_{j-1}} \theta_{i,h}^{(j)} x_h^{(j-1)}$) ($1 \leqslant i \leqslant n_j$, $1 \leqslant j \leqslant k$). The choice of the activation function f is an essential prerequisite in implementing this technique. For this purpose, I have chosen one of the most popular forms in the literature: the rectified linear function, denoted LR. It is defined by:

---

[18]These are the points where the signals from the variables transformed at the lower layer are constructed.

$$LR(x) = \begin{cases} x \; si \; x \geqslant 0 \, , \\ 0 \; si \; x < 0 \, . \end{cases}$$

This function is practical because it is less expensive in terms of the running time of the computer. So, the signal built at the level of each neuron of the last layer k can be written:

$$x_i^{(k)} \quad = \quad LR(\theta_{i,0}^{(k)} + \sum_{h=1}^{n_{k-1}} \theta_{i,h}^{(k)} x_h^{(k-1)}) \;\; i = 1, ...n_k \, , \tag{A.14}$$

and the forecast model that is obtained by aggregating the signals produced by the neurons of the last layer is written:

$$g(Z_{i,t}, \theta) \quad = \quad \theta_0 + \sum_{i=1}^{n_k} \theta_i x_i^{(k)} \, . \tag{A.15}$$

Without a hidden layer, this model is reduced to the linear model.

The estimation of the parameters of the forecast model involves solving the following problem:

$$argmin_\theta R(\theta) = \sum_{i=1}^{N} \sum_{t=1}^{T} (Y_{i,t+1} - g(Z_{i,t}, \theta))^2 \, . \tag{A.16}$$

The objective function R(θ) is RMSE. n_j ($1 \leqslant j \leqslant k$). The hyper parametrization of the model and the multitude of potential architectures make it extremely difficult to find the optimal values of these hyperparameters by cross-validation. I follow the methodology developed by Gu et al. (2020), which consists of considering a finite number of potential architectures and determining the number of neurons per hidden layer following the geometric pyramid rule (see Masters (1993)). For this purpose, I build three forecast models I can denote by NN1, NN2, and NN3. NN1 has a hidden layer that has 32 neurons. NN2

has two hidden layers with 32 and 16 neurons, respectively. NN3 has three hidden layers with 32, 16, and 8 neurons, respectively.

Appendix B. Method of Mésonnier and Stevanovic (2013)

This method, which is also a dimension reduction technique, was developed by Mesonnier and Stevanovic (2013) in an inferential approach. We will adapt it to the forecast, which will not fundamentally affect the methodology. First of all, it consists of constructing the dynamic factors, which are components justifying a large part of the co-movement of macroeconomic predictors. In the second step, innovations of the factors are extracted and substituted for the potential macroeconomic predictors. The forecast model obtained is then estimated by "OLS." In contrast to principal components, dynamic factors are generated by an autoregressive structure.

Model

The linear model $CapitalRatio_{i,t+1} = Z'_{i,t}\theta + \epsilon_{i,t+1}$ is transformed by the method into :

$$CapitalRatio_{i,t+1} \quad = Y'_{i,t}\theta_x + \eta'_{i,t}\theta_q + \varepsilon_{i,t} \; . \tag{B.1}$$

$Z_{i,t}$ is the vector grouping together all the predictors made up in this case of (1) a limited number of microeconomic variables, (2) dichotomous variables and, (3) macroeconomic variables. $Y_{i,t}$ is a subvector of $Z_{i,t}$ consisting only of microeconomic predictors, dichotomous variables and macroeconomic expectations. $\eta_{i,t}$ is a vector made up of q innovations with dynamic factors. θ_x et θ_k are the vectors of parameters to be estimated.
The number of dynamic factors q is the hyperparameter of this technique. The determination of the optimal value of q is detailed later in this text.

Optimization problem

Let us denote by X_t the subvector of $Z_{i,t}$ consisting only of macroeconomic predictors with the exception of macroeconomic expectations. If we admit that these predictors have a dynamic factor structure, then we can write:

$$X_t = \lambda(L)f_t + u_t = \lambda_0 f_t + \lambda_1 f_{t-1} + ... + \lambda_s f_{t-s} + u_t \; ; \tag{B.2}$$

$$u_t = D(L)u_{t-1} + v_{Y_t} \; ; \tag{B.3}$$

$$f_t = \Theta(L)f_{t-1} + \eta_t \; . \tag{B.4}$$

f_t is the vector with q common dynamic factors. $\lambda(L)$ is a polynomial of lagged operator L[19] of order s with matrix coefficients λ_i ($0 \leqslant i \leqslant s$). u_t is the vector made up of idiosyncratic shocks associated with the time series. D (L) is a polynomial of lagged operator L of order q* (q * $\geqslant 0$) and whose coefficients are diagonal matrices. $\Theta(L)$ is a polynomial of order p (p $\geqslant 1$) of lagged operator L. η_t is the vector consisting of q innovations with dynamic factors. v_{Y_t} is a vector made up of white noise.

This representation admits for a given time series, the autocorrelation of the idiosyncratic shock while excluding a correlation of idiosyncrastic shocks between two time series.

By multiplying both sides of equality (8) by A(L) = $I - D(L)L$ [20], and using equation (9) we get the following expression:

$$A(L)X_t = A(L)\lambda(L)f_t + v_{Y_t} \; . \tag{B.5}$$

---

[19] A lagged operator noted L is a function which transforms a time series y_t into a series $L^j(y_t) = y_{t-j}$, j being a non-zero natural integer.

[20] Equation (9) implies that $(I - D(L)L)u_t = v_{Y_t}$.

If we define $\tilde{X}_t = A(L)X_t$ et $\Lambda(L) = A(L)\lambda(L)$ then we can rewrite the equation (11) :

$$\tilde{X}_t \;=\; \Lambda(L)f_t + v_{Y_t} = \Lambda_0 f_t + \Lambda_1 f_{t-1} + ... + \Lambda_m f_{t-m} + v_{Y_t}\ . \tag{B.6}$$

$\Lambda(L)$ is a polynomial of lagged operator L of degree m $\leqslant q^* s$ which has matrix coefficients Λ_i (0$\leqslant i \leqslant m$).

By defining $F_t = (f'_t, f'_{t-1}, ..., f'_{t-s})'$ and $\Gamma = (\Lambda_0, \Lambda_1, ..., \Lambda_m)$ and further assuming that m$\leqslant p$, the dynamic factor model can be written

$$\tilde{X}_t \;=\; \Gamma F_t + v_{Y_t}\ ; \tag{B.7}$$

$$F_t \;=\; \Phi_F F_{t-1} + \epsilon_{F_t}\ . \tag{B.8}$$

This writing is the static form of the dynamic factor model. F_t is the vector consisting of R static factors and ϵ_{F_t} the vector comprising the shocks to the static factors. Note that ϵ_{F_t} = G ϵ_{f_t} where G is a matrix of type R $\times$ q which creates a correspondence between shocks to fixed factors and shocks or innovations to dynamic factors.

The optimization problem successively consists of: (1) estimating F_t and Γ by the principal components method, (2) estimating Φ_F in a VAR model of order 1, (3) construct the vector $\epsilon_t = \tilde{X}_{it} - \Gamma'_i \Phi_F F_{t-1}$ and (4) finally construct the q principal components of this vector which constitute the innovations with dynamic factors. The extraction of innovations from dynamic factors first requires the estimation of static factors. The econometer must first determine the number of static factors. Let T be the number of observation periods, N the number of time series, R_{max} the maximum number of factors defined by economics; Bai and Ng (2002) propose the following criterion to determine the number of static factors R: R = argmin $_{R=1,....,R_{max}}$ log (V (R)) + R $\frac{N+T}{NT} ln \frac{NT}{N+T}$ where V (R) = $(NT)^{-1} \sum_{t=1}^{T} (A(L)X_t - \Gamma F_t)'(A(L)X_t - \Gamma F_t)$. V (R) represents the variance explained

by idiosyncratic shocks.

Intuition

The first log term (V (R)) decreases as the number of factors increases, because an additional factor increases the variance explained by the factors and thereby decreases the variance explained by idiosyncratic shocks. The reduction of this term indicates a greater correspondence between the movements of the factors and the data. The second term R $\frac{N+T}{NT} ln \frac{NT}{N+T}$ which is increasing with R translates the penalization of the addition of an additional factor. newline The criterion of Bai and Ng (2002) therefore makes it possible to best summarize the data by a reasonable number of factors.

The number of static factors being determined, the next step consists in implementing an algorithm based on the iterative principal components method following the approach of Stock and Watson (2005) to estimate the static factors.

Algorithm

The algorithm proceeds in four steps:

Step 1:

We initialize A(L) by estimating the autoregressive model of order q^* footnote The order of the autoregressive model is determined by the economist. where X_{it} (component of rank i of vector X_t, $1 \leqslant i \leqslant K$) is the dependent variable.

step 2:

Estimate by the principal components method Γ and F_t according to the following program:

$min_{F_t,\Gamma} SSR = \ sum_{t=1}^{T} (A(L)X_t - \Gamma F_t)'(A(L)X_t - \Gamma F_t)$.

Let Vect be the vector made up of R eigenvectors[21] corresponding to the largest eigenval-

[21] Let A be a square matrix of order n, a scalar λ is an eigenvalue of A if there is a non-zero vector x of $\mathbb{R}^n$ such that $Ax = \lambda x$. x is an eigenvector associated with the eigenvalue λ.

ues of the matrix $\tilde{X}'\tilde{X}$ where $\tilde{X}$ = $(\tilde{X}_{1t}, \tilde{X}_{2t}, ..., \tilde{X}_{Kt})$ is the matrix made up of the K time series transformed by A(L)[22]. $F_t = \sqrt{T}Vect$.

Step 3:

Estimate A (L) X_t = $\Gamma F_t + v_{X_t}$ using F_t obtained in step 2 to update A (L).

Step 4:

Iterate steps 2 and 3 until F_t converges. Using the normalization $F_t'F_t/T = I_R$ we obtain at the convergence $\Gamma' = F_t'x_t/T$

The economist at this stage must first determine the number of dynamic factors and equivalently the number of innovations with dynamic factors.

Bai and Ng (2007) propose a procedure for determining the number of dynamic factors. The starting point of this procedure is the relation ϵ_{F_t} = Gϵ_{f_t} where ϵ_{f_t} is the constituted vector from innovations to dynamic factors and ϵ_{F_t} the vector comprising innovations to static factors. G is a correspondence matrix between innovations with dynamic factors and innovations with static factors. This relation suggests that the variance-covariance matrix of ϵ_{F_t} is Σ=G$\Sigma_{\epsilon_{f_t}}$G' where $\Sigma_{\epsilon_{f_t}}$ is the variance-covariance matrix of the vector comprising q innovations with common dynamic factors or macroeconomic shocks. The common macroeconomic shocks by definition being uncorrelated, the matrix $\Sigma_{\epsilon_{f_t}}$ has full rank q which makes Σ a square matrix of order R and rank q . The procedure can be summarized in three steps:

Step 1:

Estimate the VAR in F_t, $F_t = \Phi_F F_{t-1} + \epsilon_{F_t}$ and deduce the vector of residuals ϵ_{F_t}.

Step 2:

Classify the R eigenvalues of the estimated variance-covariance matrix of the vector $epsilon_{F_t}$, Σ = $T^{-1}\sum_{t=1}^{T}\epsilon_{F_t}\epsilon'_{F_t}$ in descending order $\rho_1 \geqslant \rho_2 \geqslant ... \geqslant \rho_R$

---

[22] $\tilde{X_{it}} = A(L)X_{it}$

Step 3:

Define $D_1(r)$=$\sqrt{\frac{\rho_{r+1}^2}{\sum_{i=1}^{R}\rho_i^2}}$ et $D_2(r)$=$\sqrt{\frac{\sum_{i=r+1}^{R}\rho_i^2}{\sum_{i=1}^{R}\rho_i^2}}$. Based on these two quantities, Bai and Ng (2007) propose to estimate the number of dynamic factors or common macroeconomic shocks by

q =min r$\in$ $\{r \mid D_1(r) < \frac{1}{T^{1/2-\delta}}\}$ where q =min r$\in$ $\{r \mid D_2(r) < \frac{1}{T^{1/2-\delta}}\}$. δ is a number between 0 and 1/2. In the simulations, Bai and Ng (2007) choose the value $delta = 1/4$ which seems to give them reasonable results.

Intuition

This criterion is based on the principle according to which a variance-covariance matrix in this case the matrix Σ semi-definite positive[23] of order R and of rank q, has q non zro eigenvalues among the R eigenvalues; the R-q other eigenvalues all being zero. The eigenvalues of this matrix being positive due to the fact that it is positive semi definite. By classifying them in decreasing order, $D_1(r) = D_2(r) = 0$ for r $\geqslant q$ and for r<q, $D_1(r)$ and $D_2(r)$ are strictly positive. In most empirical studies, including ours, T is a large number which makes $\frac{1}{T^{1/2-\delta}}$ tend towards 0. Therefore, the number of common macroeconomic shocks q appears as the minimum defined above. The number of common macroeconomic shocks being known, the last point is their estimation. This last point comes in two stages.

Step 1 :

We construct the vector $\epsilon_t = \tilde{X}_{it} - \Gamma_i'\Phi_F F_{t-1}$. Γ_i et F_t were obtained when estimating the static factors. Φ_F was obtained in step 1 above.

Step 2 :

ϵ_{f_t} is estimated as the q first principal components of ϵ_t.

---

[23] A square matrix A of order nxn is semi-definite positive if for any non-zero vector x of $\mathbb{R}^n$ x'Ax $\geqslant 0$.

Appendix C. Variables computation

Table 1: Variables computation

Variable name	Variable label	Source
Capital ratio	Ratio of capital to total assets	(BHCK3210/BHCK2170)*100
Size	Bank size	log(BHCK2170)
ROA	Return on assets	(BHCK4340/BHCK2170)*100
NetChargeoffRatio	Net allocation to provisions	((BHCK4635-BHCK4605)/BHCK2170)*100
MortgageLoanRatio	Real estate loan volume	(BHCK1410/BHCK2170)*100
CIloanRatio	Volume of commercial and industrial credit	(BHCK1766/BHCK2170)*100

Other database variables excluded from the above transformations are normalized by dividing their values by the corresponding values of the variable BHCK2170. We then harmonize the variables on the same scale and thus avoid the effects of levels. Table 2 highlights the links between equity capital and, therefore, $CapitalRatio$ and the other essential bank variables. It is noteworthy to reveal that $NetChargeoffRatio$ and ROA approximate the provisions for loan losses and the pre-tax net income, impacting the dynamic of $CapitalRatio$.

Table 2 : Relationship between the Capital ratio and other accounting items

Net interest income+ non interest income- non interest expense= **pre provision net revenue (PPNR)**
PPNR + other revenue – provisions – available for sale (AFS)/held to maturity (HTM) securities losses – held for sale (HFS)/fair value option (FVO) loan losses – trading and counterparty losses = **pre tax net income**
Pre tax net income – taxes – income attributable to minority interest – change in the valuation allowance = **after tax net income**
After tax net income – net distributions to common and preferred shareholders and other net reductions to shareholder's equity + other comprehensive income = **change in equity capital**
Change in equity capital – change in adjustments and deductions from regulatory capital + other additions to regulatory capital = **change in regulatory capital**

Appendix D. List of selected BHC

No	ID	Name	minimum asset	maximum asset	Observations	Start date	End date
1	1020603	Continental Bank corporation	2,16e+07	3,34e+07	32	1986q3	1994q2
2	1021075	Republic New york corporation	1.71e+07	5.99e+07	53	1986q3	1999q3
3	1021758	Natwest Holdinds INC	1.12+07	3.37+07	39	1986q3	1996q1
4	1022362	Lasalle National Corporation	3431221	2.90e+07	38	1989q4	1999q1
5	1023060	Meridian Bancorp. INC.	6511743	1.53e+07	39	1986q3	1996q1
6	1023314	Citicorp Holding INC	1.54e+07	2.89+07	39	1990q1	1999q3
7	1023453	Suntrust bank of georgia INC	9459810	2.89e+07	40	1990q1	1999q4
8	1023538	First Interstate Bancorp	4.89e+07	5.91e+07	39	1986q3	1996q1
9	1024058	First Security Corporation	4887095	2.33e+07	68	1986q3	2003q2
10	1025309	Bank of Hawaii Corporation	5053417	1.81e+07	134	1986q3	2019q4
11	1025608	Bancwest Corporation	3346544	9.62e+07	134	1986q3	2019q4
12	1025701	U.S Bancorp	9007465	3.40e+07	44	1986q3	1997q2
13	1026016	BankAmerica corporation	9.28e+07	2.65e+08	48	1986q3	1998q2
14	1027095	Wells Fargo et Company	4.27e+07	1.09e+08	49	1986q3	1998q3
15	1028739	Bank one Arizona Corp	1,01e+07	1,58e+07	46	1986q3	1997q4
16	1028953	West one Bancorp	3264714	9244503	37	1986q3	1995q3
17	1033470	Bank of New York Company	1.84e+07	1.26e+08	84	1986q3	2007q2
18	1033872	Summit Bancorp	5765299	3.97e+07	58	1986q3	2007q2
19	1033993	First Fidelity incorporated	1.43e+07	3.35e+07	38	1986q3	1997q4
20	1034888	Chemical New Jersey Holding	3256963	6217134	33	1986q3	1995q3
21	1035166	Bancorp Nj	3703851	4915584	6	1986q3	1987q4
22	1037115	JP Morgan et company INC	7.47e+07	2.99e+08	57	1986q3	2000q3
23	1039502	JP Morgan chase and company	5.60e+07	2.76e+09	134	1986q3	2019q4
24	1040795	Chase Manhattan Corporation	9.00e+07	1.21e+08	38	1986q3	1995q4
25	1042351	Citicorp	1.86e+08	9.71e+08	76	1986q3	2005q2

No	ID	Name	minimum asset	maximum asset	Observations	Start date	End date
26	1049341	Commerce Bancshares	4917787	2.61e+07	134	1986q3	2019q4
27	1068025	Keycorp	8633737	1.47e+08	134	1986q3	2019q4
28	1068191	Huntington Bancshares INC	7016598	1.09e+08	134	1986q3	2019q4
29	1068294	Bank One corporation	1.87e+07	3.27e+08	72	1986q3	2004q2
30	1068762	Mellon Financial Corporation	2.87e+07	5.10e+07	83	1986q3	2007q2
31	1069125	National City Corporation	1.31e+07	1.55e+08	89	1986q3	2008q3
32	1069778	PNC Financial Services INC	2.09e+07	4.10e+08	134	1986q3	2019q4
33	1070251	Star Bank Corporation	3989648	1.73e+07	49	1986q3	1998q3
34	1070345	Fifth Third Bancorp	3207136	1.71e+08	134	1986q3	2019q4
35	1071968	First Virginia Banks INC	3670960	1.13e+07	68	1986q3	2003q2
36	1072107	Signet Banking Corporation	9214776	1.30e+07	45	1986q3	1997q3
37	1072237	Crestar Financial Corporation	8703206	2.78e+07	54	1986q3	1999q4
38	1072291	First Union corporation of Virginia	5377889	2.03e+07	44	1986q3	1995q3
39	1072554	South Carolina NC	4522295	7424140	37	1986q3	1995q3
40	1073551	Wachovia Corporation	2.10e+07	8.12e+08	89	1986q3	2008q3
41	1073757	Bank of America Corporation	2.46e+07	2.43e+09	134	1986q3	2019q4
42	1074660	Allfirst Financial INC	5016877	1.89e+07	67	1986q3	2003q1
43	1074875	Central Fidelity Banks INC	3701715	1.08e+07	45	1986q3	1997q3
44	1075126	Riggs National Corporation	4426221	7637650	75	1986q3	2005q1
45	1076776	Barnett Banks INC	1,64e+07	4.65e+07	46	1986q3	1997q4
46	1078332	Regional Financial Corporation	3941334	4.99e+07	72	1986q3	2004q2
47	1078426	First american corporation	5322423	2.22e+07	53	1986q3	1999q3
48	1078529	BBVA USA Bancshares INC	3494388	9.38e+07	134	1986q3	2019q4
49	1078604	Amsouth Bancorporation	5549050	5.43e+07	81	1986q3	2006q3
50	1078921	Hibernia Corporation	4113161	2.32e+07	77	1986q3	2005q3

No	ID	Name	minimum asset	maximum asset	Observations	Start date	End date
51	1079441	Southtrust Corporation	4746757	5.39e+07	73	1986q3	2004q3
52	1079638	Bank South Corporation	3142295	7684650	38	1986q3	1995q4
53	1080148	Suntrust Banks of Tenessee INC	5047345	9114637	41	1986q3	1999q4
54	1080371	Louisiana Bank One Corporation	3421048	9507297	49	1986q3	1998q3
55	1093586	Boatmen Bancshares INC	9228286	4.12e+07	42	1986q3	1996q4
56	1094211	Mercantile Bancorporation	6287766	3.60e+07	52	1986q3	1999q2
57	1094640	First Horizon National corporation	5352901	4.37e+07	134	1986q3	2019q4
58	1102367	Cullen-Frost Bankers INC	3042869	3.41e+07	134	1986q3	2019q4
59	1111435	State street Corporation	6172269	2.95e+08	134	1986q3	2019q4
60	1112076	Bankboston corporation	3.05+07	7.76e+07	53	1986q3	1999q3
61	1113514	Fleetboston financial corporation	1.07e+07	2.12e+08	71	1986q3	2004q1
62	1116300	Corestates Financial Corp	1.23e+07	4.85e+07	47	1986q3	1998q1
63	1119794	US Bancorp	1.70e+07	4.64e+08	134	1986q3	2019q4
64	1120754	Wells Fargo	1.96e+07	1.95e+09	134	1986q3	2019q4
65	1130892	Premier Bancorp INC	3760037	5511975	37	1986q3	1995q3
66	1131787	Suntrust Banks INC	1.99e+07	2.28e+08	134	1986q3	2019q4
67	1136157	Wachovia Corporation	1.75e+07	7.56e+07	60	1986q3	2001q3
68	1140743	Midatlantic corporation	1.33e+07	2.43e+07	36	1986q3	1995q3
69	1199479	Bank One Corp	6589462	3.87e+07	53	1986q3	1999q3
70	1199488	Bank One Wiscosin Corp	3846284	8755366	34	1986q3	1996q3
71	1199497	MI LLC	5037646	6.08e+07	85	1986q3	2007q3
72	1199611	Northern Trust Corporation	7824700	1.39e+08	134	1986q3	2019q4
73	1199648	First of America Bank Corporation	5578498	2.46e+07	46	1986q3	1997q4
74	1199705	Old Kent Financial Corporation	5400488	2.45e+07	59	1986q3	2001q1
75	1199714	Michigan National Corporation	7826412	1.20e+07	59	1986q3	2001q1
76	1199778	First chicago NBD corp	1.91e+07	1.22e+08	49	1986q3	1998q3
77	1199844	Comerica INC	9336328	7.44e+07	134	1986q3	2019q4
78	1200432	Bank One Indiana corp	4284241	9400736	37	1986q3	1998q3
79	1201028	First Chicago Corporation	3.91e+07	7.57e+07	37	1986q3	1995q3
80	1255415	Harris financial corp	9880000	1.32e+08	134	1986q3	2019q4

No	ID	Name	minimum asset	maximum asset	Observations	Start date	End date
81	1246216	Franklin resources INC	5900311	1.66e+07	53	2001q2	2014q2
82	1248612	First fidelity bancorporation	2.82e+07	3.62e+07	32	1988q1	1995q4
83	1250932	Bank One Ohio Corporation	1.50e+07	2.83e+07	32	1988q1	1997q4
84	1378434	Unionbancal corp	4091351	1.73e+08	134	1986q3	2019q4
85	1416774	NationsBank Texas Bancorp INC	3.04e+07	6.24e+07	33	1990q1	1998q1
86	1473562	Banc One Texas Corporation	1.28e+07	3.13e+07	43	1990q1	2000q3
87	1871159	MBNA corporation	5146165	6.30e+07	60	1991q1	2005q4
88	1888193	Wilmington Trust Corp	3969244	1.26e+07	79	1991q3	2011q1
89	1951350	Citigroup INC	6.69e+08	2.36e+09	85	1998q4	2019q4
90	2081124	Greenpoint financial corp	6955179	2.7e+07	43	1994q1	2004q3
91	2389941	TCF financial corp	7429623	2.46e+07	91	1997q2	2019q4
92	2477754	Investor Bancorp MHC	3172821	2.71e+07	84	1997q1	2019q4
93	2744894	First Bancorp	4017352	2.06e+07	85	1998q4	2019q4
94	2801546	W. Holding company INC	3374588	1.80e+07	42	1999q4	2010q1
95	2816906	Taunus corp	1.48e+08	7.6e+08	57	1999q2	2011q4
96	2847115	Santander Bancorp	5396415	9288663	65	2000q2	2016q2
97	2894230	Discount Bancorp INC	4610604	9983916	80	2000q1	2019q4
98	2945824	Metlife INC	2.52e+08	8.46e+08	47	2001q1	2012q3
99	2945824	BBVA PR Holding corp	4724153	7045588	48	2000q4	2012q3
100	3005332	FNB corp	4062977	3.46e+07	75	2001q2	2019q4

Total assets are expressed in thousands of US dollars.

Appendix E. VAR model

E.1 VAR method

Let y_t be a vector which includes K time series (K $\geqslant$ 2) observed over T periods. Let p be a non-zero natural number. A structural VAR of order p is a set of K linear regression equations (equation 15)

$$B_0 y_t \quad = \quad B_1 y_{t-1} + ... + B_p y_{t-p} + \omega_t \, . \tag{E.1}$$

The matrices B_i i = 0, ... p have the dimension K x K. p is the maximum lag fixed by the economist. ω_t is the vector with K structural shocks. The variance-covariance matrix of the vector ω_t is

E ($\omega_t \omega_t'$) = I_K (the variance of structural shocks is normalized to 1). The structural shocks corresponding to the K equations are uncorrelated by definition. We also make the hypothesis that the shocks of the structural form are white noise[24].

The matrix B_0 translates the contemporary relation between the time series while the matrices B_i

$1 \leqslant$ i $\leqslant$ p, translate the relation between each time series, its lagged values and lagged of other time series. An estimate of these equations by ordinary least squares (OLS) is affected by an endogeneity bias caused by the contemporary relationship between the time series. This problem is solved by writing the VAR in reduced form.

So, by multiplying equation (15) by B_0^{-1} we get equation (16).

$$y_t = B_0^{-1} B_1 y_{t-1} + ... + B_0^{-1} B_p y_{t-p} + B_0^{-1} \omega_t \, . \tag{E.2}$$

Equation (16) can be rewritten in the form of equation (17) which is the reduced form of the VAR

$$y_t = A_1 y_{t-1} + ... + A_p y_{t-p} + \mu_t \, . \tag{E.3}$$

where $A_i = B_0^{-1} B_i$ $1 \leqslant i \leqslant p$ et $\mu_t = B_0^{-1} \omega_t$. μ_t is the vector of reduced form shocks which, unlike structural shocks, can be correlated. The estimate of the coefficients of the VAR in reduced form by OLS is convergent because E($y_{t-i}' u_t$)=0 ($1 \leqslant i \leqslant p$). Beyond the estimation of the VAR in reduced form, the interest of the use of this model in empirical

[24] A time series ϵ_t is white noise if E (ϵ_t) = 0 for all t and cov ($\epsilon_t, \epsilon_{ti}$) = 0 for all i.

work lies in the construction of the response functions of the variables contained in the vector y_t to the K structural shocks, that is to say a total of K^2 response functions. For this purpose, if y_t is stationary[25] then y_t can be written as a weighted average of the shocks in reduced form μ_t. So,

$$y_t = \sum_{i=0}^{\infty} \Phi_i \mu_{t-i} \, , \tag{E.4}$$

knowing that Φ_i is a square matrix of order K $\times$ K obtained from the matrices A_i $1 \leqslant i \leqslant p$ and which summarizes the impact of shocks in reduced form over the different macroeconomic series at horizon i. A typical element of the Φ_i matrix is $\Phi_i^{(j,k)} = \frac{\partial y_{jt+i}}{\partial \mu_{kt}}$ ($1 \leqslant j \leqslant K$ and $1 \leqslant k \leqslant K$). $\Phi_i^{(j,k)}$ represents the effect of the deviation of one unit of the kth shock of the vector mu_t on the jth macroeconomic series of the vector y_t at horizon i.

Using equations (18) and (19) we get:

$$y_t = \sum_{i=0}^{\infty} \Phi_i B_0^{-1} \omega_{t-i} = \sum_{i=0}^{\infty} \Theta_i \omega_{t-i} \, . \tag{E.5}$$

So, $\Theta_i = \Phi_i B_0^{-1}$ is a square matrix of order K $times$ K which represents the impact of structural shocks on time series at horizon i. A typical element of the matrix Θ_i is $\Theta_i^{(j,k)}$=$\frac{\partial y_{jt+i}}{\partial \omega_{kt}}$ ($1 \leqslant j \leqslant K$ et $1 \leqslant k \leqslant K$). $\Theta_i^{(j,k)}$ represents the effect of the deviation of one unit of the k-th shock of the vector $omega_t$ on the j-th macroeconomic series of the vector y_t at the horizon i.

To deduce the response function of the time series to structural shocks from the estimates

---

[25] y_t is stationary if E (y_t) = μ for all t (the unconditional expectation is constant) and cov (y_t, y_{tk}) = γ_k (the covariance between two points of a series depends only on the step k which separates these two points). In general, in empirical work, variables which are not stationary are made stationary.

of the reduced form, it suffices to recover the matrix B_0^{-1} which reflects the cyclical impact of structural shocks on time series. By reusing the relationship μ_t=$B_0^{-1}\omega_t$, we obtain E($\mu_t\mu_t'$) $=$ Σ_μ $=$ $B_0^{-1}B_0^{-1'}$. The matrix Σ_μ due to its symmetry[26] has K (K + 1) / 2 elements while $B_0^{-1}B_0^{-1'}$ consists of K^2 elements. The deduction of B_0^{-1} requires the imposition of K (K-1) / 2 restrictions to B_0^{-1} called short-term restrictions[27].

The method of identifying shocks (or restriction) commonly used in the literature is the recursive method. According to this method, no restriction is imposed on the first column of the matrix B_0^{-1}, an exclusion restriction[28] is imposed on the first element of the second column, ... K-1 exclusion restrictions are imposed on the first K-1 elements of the Kth column of B_0^{-1} . The identification is based on the economic facts that the first shock has a contemporary impact on all variables in y_t, the second shock has no contemporary impact on the first variable in y_t, ... the Kth structural shock has no contemporary impact on the K-1 first variables in y_t. Without harming the generality, the application of the recursive method of identification to the case where K = 3 allows to obtain the following matrix

B_0^{-1}= $\begin{pmatrix} b11 & 0 & 0 \\ b21 & b22 & 0 \\ b31 & b23 & b33 \end{pmatrix}$ dssssssz

---

[26]It is a variance covariance matrix which by definition is symmetric

[27]The identification of structural shocks can also be done through other restrictions such as the long term restriction which is used when the variables are not stationary. It consists in imposing restrictions on the long-term impact matrix of structural shocks which induce restrictions on the matrix B_0^{-1}. Short-term restrictions are widely used in empirical studies.

[28]An exclusion restriction is a constraint that we impose on an element of the matrix B_0^{-1} to be equal to 0.

E.2 Variables included in VAR

	mean	standard deviation	minimum	maximum
GDP	0.49	0.60	-2.19	1.81
Inflation	0.53	0.56	-2.29	1.54
TFED	1.71	1.98	0.07	6.52
Volume of credit	4.36	4.89	-8.80	17.20
Aggregate credit shock	0.17	0.76	-3.14	2.80
Risk premium	0.12	0.74	-0.80	3.04
Yield gap	2.69	0.75	1.62	5.59

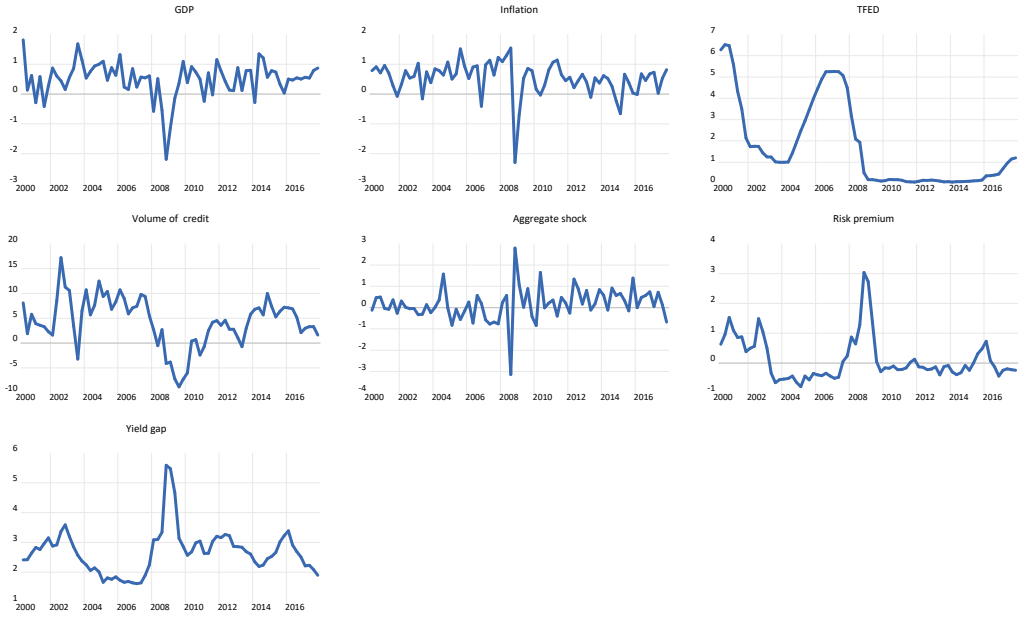


Figure. VAR variables

Appendix F. Details on Diebold Mariano test

Suppose I denote by $e_{i,t}^{(k)}$ and $e_{i,t}^{(l)}$ the forecast errors [29] of the two models to compare, by H, the loss function associated with the forecast[30]. I can formulate the test as follows:

$$H_0 : \quad E(H(e_{i,t}^{(k)})) = E(H(e_{i,t}^{(l)}))$$

$$H_1 : \quad E(H(e_{i,t}^{(k)})) < E(H(e_{i,t}^{(l)}))$$

$$H_1 : \quad E(H(e_{i,t}^{(k)})) > E(H(e_{i,t}^{(l)}))$$

I test the hypothesis of equality of performance between the two models against one of the following two alternatives: (1) the predictive performance of model K is higher than that of model L, (2) the predictive performance of model K is lower than that of model L.
The test statistic is

$$\bar{d} = \frac{1}{NT} \sum_{i,t} H(e_{i,t}^{(k)}) - H(e_{i,t}^{(l)}) \, .$$

By setting $d_{i,t} = H(e_{i,t}^{(k)}) - H(e_{i,t}^{(l)})$, the statistics can be simplified as:

$$\bar{d} = \frac{1}{NT} \sum_{i,t} d_{i,t}$$

---

[29] The forecast error of a model is the difference between the value of $Y_{i,t+h}$ of the dependent variable and its predicted value. Diebold and Mariano's test (2002) is effective even if the means of these forecast errors are non-zero, non-gaussian, and autocorrelated, and the two series of forecast errors are contemporaneously correlated.

[30] The Diebold Mariano test admits several possible loss functions, including the quadratic loss function that I have chosen. If $e_{i,t}$ represents a forecast error, H the quadratic loss function, then H ($e_{i,t}$) = $e_{i,t}^2$. The loss function is a measure of the performance of the forecasting model. The performance of a model is all the better as the average loss is low.

Diebold and Mariano (2002) show that if series $d_{i,t}$ is stationary[31], then for large value of T and under H_0, $\sqrt{T}(\bar{d} - \mu) \rightarrow N(0, \sigma)$ where μ and σ are respectively the expectation and the standard deviation of $\bar{d}$[32]. Diebold and Mariano (2002) calculate σ by integrating the autocovariance of series $d_{i,t}$, which depends on the forecast horizon.

[31] A time series is stationary if it has a constant expectation and an autocovariance that depends only on the number of periods and not on the origin.

[32] The standard deviation of $d_{i,t}$ depends on the number of non-null autocovariance terms linked to the forecast horizon h.