國立臺灣大學管理學院財務金融學系
博士論文

Department of Finance
College of Management
National Taiwan University
Doctoral Dissertation

財務金融研究
Two Essays in Finance

王邦瑜
Pang-Yu Wang

指導教授：洪茂蔚 博士
Advisor: Mao-Wei Hung, PhD
共同指導教授：陳明賢 博士
Co-Advisor: Ming-shen Chen, PhD

中華民國 114 年 3 月
March 2025

doi:10.6342/NTU202500807

國立臺灣大學博士學位論文
口試委員會審定書
DOCTORAL DISSERTATION ACCEPTANCE CERTIFICATE
NATIONAL TAIWAN UNIVERSITY

（論文中文題目）財務金融研究

（論文英文題目）Two Essays in Finance

本論文係王邦瑜 (學號：D07723001) 在國立臺灣大學財務金融學系完成之博士學位論文，於民國 114 年 3 月 26 日承下列考試委員審查通過及口試及格，特此證明。

The undersigned, appointed by the Department of Finance on March 26, 2025 have examined a Doctoral Dissertation entitled above presented by WANG, PANG-YU (student ID: D07723001) candidate and hereby certify that it is worthy of acceptance.

口試委員 Oral examination committee:

陳明賢　　　　　　張瑄凌

（指導教授 Advisor）　（指導教授 Advisor）

鄭弘文

系（所、學位學程）主管 Director: 石百達

謝辭

本人衷心感謝我的指導教授洪茂蔚老師與陳明賢老師，在博士求學過程中給予我無私的指導、耐心的教誨與不斷的鼓勵。他們嚴謹的學術態度與深刻的洞見，深深影響了本論文的品質與研究方向。此外，亦感謝口試委員們所提供的寶貴建議與專業意見，使本論文在內容與論證上更加完善。我要特別感謝財金系與會計系博士班的所有同學與學長姐，在這段艱辛的旅程中給予我許多打氣、協助與陪伴。最後，誠摯感謝我的父母長期以來的支持與陪伴，是他們的愛與信念成就了我在學術道路上的堅持與成長。本文若有未盡之處，皆屬本人之責，與他人無涉。

財務金融研究

摘要

本論文由兩篇探討法規與科技如何影響財務決策與資產定價的研究組成。第一篇論文檢視臺灣於2014年實施企業社會責任（CSR）報告強制揭露制度對自願性財務揭露行為的影響，發現企業管理者在投資人與分析師關注度高的情況下，顯著增加盈餘預測的發布。此結果顯示，管理者進行自願性揭露主要是為了解決代理問題，而非為了降低專有資訊洩漏風險或資訊不對稱問題，突顯強制性非財務揭露與自願性財務揭露之間的互補關係。第二篇論文探討機器學習技術是否能提升盈餘貝他值（earnings beta）的預測效果，該指標為評估系統性風險的替代方案。研究比較傳統迴歸方法與套索（Lasso）與彈性網（Elastic Net）等機器學習模型的表現，發現儘管後者在預測準確度上略有提升，其在橫斷面資產定價的經濟解釋力仍然有限。綜合而言，這兩篇論文針對公司揭露行為與資產定價文獻做出貢獻，展示法規變遷與運算技術進步如何共同影響市場透明度、風險評估與資本市場效率。

關鍵詞：　非財務揭露；自願性揭露；環境、社會和公司治理；盈餘貝他值；機器學習；資產定價。

doi:10.6342/NTU202500807

Two Essays in Finance

Abstract

This dissertation comprises two essays that explore how regulation and technology shape financial decision-making and asset pricing. The first essay examines the impact of Taiwan's 2014 CSR reporting mandate on voluntary financial disclosures, finding that managers significantly increase the issuance of earnings forecasts, particularly under heightened investor and analyst scrutiny. These findings suggest that voluntary disclosures serve primarily to address agency concerns rather than proprietary costs or adverse selection, highlighting the complementary relationship between mandatory nonfinancial and voluntary financial reporting. The second essay investigates whether machine learning techniques enhance the estimation of earnings betas—an alternative to market betas in capturing systematic risk. While methods such as Lasso and Elastic Net slightly improve forecast accuracy over traditional regressions, their economic significance in pricing cross-sectional returns remains limited. Together, these essays contribute to the literature on corporate disclosure and asset pricing by illustrating how regulatory frameworks and computational advancements jointly influence transparency, risk assessment, and market efficiency.

Keywords: nonfinancial disclosures; voluntary disclosures; ESG; earnings beta; machine learning; asset pricing.

doi:10.6342/NTU202500807

Table of Contents

Chapter 1: Say More or Less? On Mandatory Nonfinancial Disclosures and Voluntary Financial Disclosures

doi:10.6342/NTU202500807

Chapter 2: Estimating Fundamental Betas via Linear Machine Learning

doi:10.6342/NTU202500807

Chapter 1:
Say More or Less? On Mandatory Nonfinancial Disclosures and Voluntary Financial Disclosures

1. Introduction

The accounting literature has revealed that managers convey signals to the capital market through mandatory and voluntary financial disclosures. The consequences of such combined disclosures, however, are mixed. While some studies find that these two kinds of disclosures are complements (Ball, Jayaraman, and Shivakumar 2012; X. Li and Yang 2016), others find that they are substitutes (Noh, So, and Weber 2019; Tuo, Rezaee, and Gao 2024). Besides financial disclosures, nonfinancial disclosures have been mandated in different countries in the past decade. For example, some countries and economies have required firms to disclose environmental, social, and corporate governance (ESG) information. They include the European Union (EU), China, India, Taiwan, etc. As ESG disclosures become popular, investors want managers to tell them the financial materiality of sustainable activities (J. Gelb et al. 2023). Moreover, managers communicate with investors more about their ESG activities in the financial reporting or conference calls (Eccles and Serafeim 2013; Tomlinson, Whelan, and Eckerle 2021). The quality of the communication, however, still has room for improvement, as some managers acknowledged that they did not sufficiently clarify the financial materiality of sustainable activities (Eccles and Serafeim 2013, p.9). Thus, some organizations, such as the International Sustainability Standards Board (ISSB), have endeavored to make ESG information useful for investors (IFRS 2024). Although many studies have investigated market and real effects of mandatory ESG disclosures (Y.-C. Chen, Hung, and Wang 2018; Manchiraju and Rajgopal 2017; Zhang et al. 2023), few studies have addressed interactions between mandatory nonfinancial disclosures and managers' voluntary financial disclosures,

doi:10.6342/NTU202500807

through which managers may convey the financial materiality of the sustainable activities. This study aims to explore such interactions.

We exploit the 2014 regulation on mandatory CSR reporting in Taiwan as a quasi-experiment to investigate the interplay between mandated and voluntary disclosures. We find three results. First, managers required to disclose CSR reports are 90% more likely to issue earnings forecasts than those whose firms are not. Second, managers whose firms are required to disclose CSR reports and under high investors' and analysts' scrutiny are 88% and 80% more likely to issue earnings forecasts, respectively, than the counterparts under low scrutiny. Both effects are robust to model triangulations and a propensity-score-matched model implementation.

This paper makes three contributions. First, this paper contributes to the literature on interactions among disclosures. Since Einhorn (2005) starts to address interactions between mandatory and voluntary disclosures from a theoretical perspective, subsequent studies have investigated this issue either theoretically or empirically (Noh, So, and Weber 2019; Ball, Jayaraman, and Shivakumar 2012; X. Li and Yang 2016; Hao 2023; Heinle, Samuels, and Taylor 2023). Most of these studies have focused on financial rather than nonfinancial disclosures. Tuo et al. (2024) address interactions between voluntary nonfinancial and financial disclosures. However, they do not investigate interactions between mandatory nonfinancial disclosures and voluntary financial disclosures. As mandatory ESG disclosures have been widely discussed and implemented around the globe, it is crucial to understand how investors and managers react to such disclosures. Our study is the first one to examine how managers change their voluntary disclosures to respond to mandatory nonfinancial disclosures.

Second, this paper contributes to the literature on the effects of mandatory ESG disclosures. Previous studies have found that compulsory ESG disclosures force firms to reduce pollution and

doi:10.6342/NTU202500807

increase ESG expenditures (Y.-C. Chen, Hung, and Wang 2018). Moreover, such disclosures also increase the quantity and quality of nonfinancial information, improve information environment, and enhance financial transparency and stock liquidity (Ioannou and Serafeim 2017; Xue et al. 2024; Shi et al. 2023; Nair et al. 2019; Deng et al. 2021). No wonder such disclosures lower cost of debt (Fonseka, Rajapakse, and Richardson 2019). On the other hand, such disclosures harm investors' interests by lowering profitability and widen ESG rating disagreement due to green bond issuance (Y.-C. Chen, Hung, and Wang 2018; Wang and Liu 2024). Unsurprisingly, such disclosures increase the cost of equity capital (Grewal, Riedl, and Serafeim 2019). These studies have addressed both market and real effects. However, they have not addressed how managers change their disclosure behavior to respond to ESG disclosure mandates. Our study is the first one to answer this question.

Third, this paper contributes to the literature on determinants of managerial voluntary disclosures and earnings forecasts. Many studies have investigated why managers voluntarily disclose or withhold information and to what extent. For example, managers may choose not to disclose certain information to lower proprietary costs or litigation risk (Verrecchia 1983). On the other hand, they may choose to disclose information even if they have to bear the costs and risks (Gigler 1994). According to the signaling theory, such disclosures may act as costly signals that help investors distinguish between good and bad firms (Spence 1973). Later studies have shown that institutional factors, such as legal environments, also affect managers' earnings forecasts, a form of voluntary disclosures (Guan et al. 2020; Baginski, Hassell, and Kimbrough 2002; W. Li et al. 2019; Xia, Zhang, and Guo 2024). Our study complements the findings in this literature, showing that managers are more likely to voluntarily disclose financial information because of the increase in investors’ demand for more information after the introduction of mandatory nonfinancial disclosures, a legal environment that encourages transparency.

2. Institutional Background

Between 2013 and 2014, a food safety crisis occurred in Taiwan. In 2013, Chang Chi Foodstuff Factory Co. was reported to have incorporated copper chlorophyllin, an illegal coloring agent for cooking oil, in its olive oil. The company adulterated its higher-end cooking oil with cheaper cottonseed oil. In 2014, Kaohsiung-based company Chang Guann Co. was found to have adulterated its Chuan Tung Fragrant Lard Oil cooking oil with recycled waste oil and animal feed oil. These two food safety frauds, among others, made the public in Taiwan concerned about food safety. Their concern was reflected in the significant spike in Google web searches for the Chinese term “food safety” (食安) during 2013 and 2014 (Figure 1).

To tackle the crisis and deter similar food safety scandals from recurring, the Financial Supervisory Commission (FEC), the SEC counterpart in Taiwan, passed on 20 September 2014 the first version of the “Taiwan Stock Exchange Corporation Rules Governing the Preparation and Filing of Corporate Social Responsibility Reports by TWSE Listed Companies.” These rules required certain listed firms to report on CSR activities. Specifically, listed firms satisfying one of the following conditions were required to file CSR reports:

1. At the end of the most recent fiscal year, the company falls into the food, chemical, and financial and insurance industries prescribed in the Taiwan Stock Exchange Corporation Key Points for Classifying and Adjusting Categories of Industries of Listed Companies.

doi:10.6342/NTU202500807

2. The financial report for the most recent fiscal year submitted under Article 36 of the Securities and Exchange Act indicates that no less than 50% of the company’s revenue is derived from food and beverage.

3. The financial report for the most recent fiscal year submitted under Article 36 of the Securities and Exchange Act indicates that the company’s capital stock has achieved no less than NT$5 billion[1] (Taiwan Stock Exchange Corporation Rules Governing the Preparation and Filing of Corporate Social Responsibility Reports by TWSE Listed Companies, 2019).

3. Literature Review and Hypothesis Development

The literature on voluntary disclosures has developed for decades. There have been two main explanations for why managers voluntarily disclose information. The first explanation is reducing information asymmetries between managers and shareholders, especially the adverse selection problem (Verrecchia 2001). The signaling theory suggests that, in an information-asymmetric game, the sender, the player with private information, is incentivized to convey costly signals to the receiver to reveal the sender’s type. By doing so, the sender of the good type can make the receiver tell good types from bad ones, achieving a separating rather than pooling equilibrium (Spence 1973). This signaling mechanism not only makes both the sender of the good type and the receiver better off but also prevents the market from collapsing (Akerlof 1970). Managers’ voluntary disclosures are like the signaling mechanism that helps investors distinguish between good and bad firms. Some empirical findings support this story. These studies find that, during

---

[1] The threshold was originally NT$10 billion when the rules were first announced in 2014. It was subsequently lowered to NT$5 billion in 2015.

quarterly earnings announcements, managers voluntarily disclose balance sheets when the information in earnings is unclear (S. Chen, DeFond, and Park 2002). They also find that managers provide voluntary disclosures to mitigate information asymmetries due to financial statement complexity, GAAP's limitations on reporting discretions, or terminations of analysts' coverage (Guay, Samuels, and Taylor 2016; Hribar et al. 2022; Balakrishnan et al. 2014).

Moreover, the literature has found that managers' disclosures in an information-asymmetric environment are usually partial rather than complete. This finding contradicts the implications of some of the classic information economic theories (Grossman 1981; Milgrom 1981; Grossman and Hart 1980). These theories argue that, even if managers have the information advantage, investors can expect that managers typically disclose good news and withhold bad news, thereby relentlessly discounting the share prices whenever managers disclose nothing. The punishment is so painful that the best response for managers is always to disclose everything, either good or bad. This full disclosure strategy, however, is rare in reality. One explanation is the presence of proprietary costs (Verrecchia 1983; Jovanovic 1982). Proprietary costs refer to the costs associated with the potential loss of competitive advantage that a company may face because of information disclosures. If there is proprietary cost, investors cannot completely make sure whether managers withhold information to avoid proprietary costs or hide bad news. Note that, even if they bear proprietary costs, managers may still provide voluntary disclosures to make their disclosures costly and thus credible to investors (Gigler 1994). Furthermore, the severity of the adverse selection problem and the amount of proprietary cost may be positively correlated, leading to the non-linearity property of the probability of voluntary disclosures (Kim, Taylor, and Verrecchia 2021). Another explanation is investors' uncertainty regarding managers' endowment of information (Dye 1985; Jung and Kwon 1988). In this world, investors do not know what information managers

doi:10.6342/NTU202500807

possess and whether managers have information. Thus, investors cannot tell whether managers withhold information to hide bad news or merely have no information.

The second explanation for managers' motivation to provide voluntary disclosures is reducing agency costs between managers and shareholders. Agency costs arise because of the separation of ownership and management. The separation leads to conflicts of interests between managers and shareholders, incentivizing managers to fulfill their benefits by sacrificing shareholders' interests (Jensen and Meckling 1976; Holmström 1979). Therefore, shareholders need to design contracts for and monitor managers to avoid agency costs. Asking managers to disclose more information can be a monitoring mechanism (D. S. Gelb 2000). Furthermore, designing contracts that contain stock-based incentives or longer pay durations can encourage managers to disclose more private information and terrible news (Nagar, Nanda, and Wysocki 2003; Cheng, Cho, and Kim 2021). The literature has also found that family firms, which usually suffer less from the agency problem than nonfamily ones, provide fewer earnings forecasts and conference calls than nonfamily firms (S. Chen, Chen, and Cheng 2008).

The literature has also provided other reasons to explain why managers provide voluntary disclosures. Some studies argue that, from the perspective of the upper echelons theory (Hambrick and Mason 1984), managers' characteristics can affect how managers disclose information (Bamber, Jiang, and Wang 2010). Other studies have found that firm-, industry-, and country-specific characteristics can also affect managers' voluntary disclosures (Meek, Roberts, and Gray 1995; Lang and Lundholm 1993). A recent study has shown that managers may consider the impact of disclosures on suppliers' perceptions when disclosing information (Arya and Ramanan 2021).

Since firms not only are required to provide mandatory disclosures but also can voluntarily disclose other information, there may be interactions between mandatory and voluntary disclosures.

doi:10.6342/NTU202500807

Einhorn (2005) is the first study to address such interactions from a theoretical perspective. In his model, Einhorn shows that, under different conditions, mandatory and voluntary disclosures are complementary, independent, or substitute. Subsequent studies have examined the relationship between mandatory and voluntary disclosures. They find mixed results. Some studies find that mandatory and voluntary disclosures are complements (X. Li and Yang 2016; Ball, Jayaraman, and Shivakumar 2012). Others find that the disclosures are substitutes (Noh, So, and Weber 2019; Tuo, Rezaee, and Gao 2024). Still others find that the relationships of the disclosures may not be monotonic (Arena, Bozzolan, and Imperatore 2023). Some recent studies have proposed new theoretical models to explain why mandatory and voluntary disclosures may be substitutes (Hao 2023; Heinle, Samuels, and Taylor 2023). Whether mandatory and voluntary disclosures share a complementary, substitute, or nonmonotonic relationship remains an open question.

Most mandatory disclosure regulations in the world require firms to reveal financial information. Recently, some countries have also required firms to disclose nonfinancial information, especially information regarding CSR or ESG. Christensen et al. (2021) provide a comprehensive literature review on the consequences of mandatory ESG disclosures. Specifically, the literature has exploited such regulations as quasi-experiments to find various results. For example, Ioannou and Serafeim (2017) utilize the mandatory ESG disclosure regulations in China, Denmark, Malaysia, and South Africa to conduct a difference-in-differences analysis. They find that, after the mandates, the firms required to disclose ESG information increase the quantity and quality of their nonfinancial information. The values of these firms also increase afterward. Chen et al. (2018) conduct a difference-in-differences test on the mandatory ESG disclosure regulations introduced in China in 2008. They find that the regulations increase ESG expenditures of the firms, harm profitability, and lower pollution. They suggest that the regulations mitigate external costs by sacrificing shareholders' interests. Further studies on disclosure regulations in China indicate that the regulations improve information environment and stock liquidity and lower cost of debt of energy firms (Xue et al. 2024; Shi et al. 2023; Deng et al. 2021; Fonseka, Rajapakse, and Richardson 2019) but widen ESG rating disagreement due to green bond issuance (Wang and Liu 2024). Even though mandatory disclosures harm profitability, Xue et al. (2023) find that voluntary ESG disclosures harm profitability greater than mandatory ones in China. Furthermore, Grewal et al. (2019) conduct a difference-in-differences test on the passage of European Union (EU) Directive 2014/95/EU on nonfinancial information disclosures. They find that the market reacts negatively to such disclosures. They suggest that the market perceives the cost of weak ESG disclosures as greater than the benefits. Nair et al. (2019) exploits the Companies Act in India in 2013 and finds that mandatory nonfinancial disclosures enhance financial transparency. They also find that retail ownership strengthens the enhancement.

The literature has yet to answer the relationship between mandatory nonfinancial disclosures and voluntary financial disclosures. Only Tuo et al. (2024) have addressed issues related to our research question. While they find that voluntary nonfinancial and financial disclosures are substitutes, their results may not be able to answer our question because the self-selection problem is much less severe in mandatory disclosures than in voluntary ones. Different theoretical stories can account for relationships between mandatory nonfinancial and voluntary financial disclosures. On the one hand, if disclosing mandatory nonfinancial information brings burdensome proprietary costs, managers may be less willing to provide voluntary disclosures (Verrecchia 1983). Furthermore, the more the disclosures, the lower the information asymmetry. Thus, the mandates of ESG disclosures may lead to a decrease in voluntary disclosures because the marginal benefit of providing voluntary disclosures may be less the marginal cost when ESG information is disclosed (Diamond and Verrecchia 1991). On the other hand, the introduction of immaterial nonfinancial disclosures may construct a more intricate information environment (Dhaliwal et al. 2012; Christensen, Hail, and Leuz 2021). In such a complicated information environment, the information asymmetry between managers and investors usually widens. If disclosing mandatory nonfinancial information worsens the adverse selection or agency problem, investors may pressure managers to offer more voluntary disclosures to fulfill the information needed from the capital market. For one thing, nonfinancial information is usually more diverse and less measurable than financial one (Christensen, Hail, and Leuz 2021 p. 1185), so investors may not easily understand the information content of nonfinancial disclosures, worsening the adverse selection problem. For another, since managers can exercise much more discretion over nonfinancial disclosures than financial ones, managers may have incentives to manipulate information in nonfinancial disclosures, such as greenwashing. In each case, managers may be incentivized or pressured to provide more voluntary disclosures to complement the information in nonfinancial disclosures. Thus, we form our first hypothesis in the null form:

Hypothesis 1: The mandate of ESG disclosures does not affect managers' voluntary disclosures.

The effect of ESG disclosures on voluntary disclosures may vary among different information environments. Typically, information environments with large institutional ownership and analyst coverage are better than the others. Thus, mandatory ESG disclosures may impact managers' voluntary disclosures through the channel of institutional investors' scrutiny. The literature has shown that institutional investors tend to push managers to commit more to ESG activities (T. Chen, Dong, and Lin 2020). This tendency is not confined to the US but prevails worldwide (Dyck et al. 2019). Even though institutional investors may push firms to do ESG

doi:10.6342/NTU202500807

activities, some research has also shown that forcing firms to disclose such activities may harm shareholders' interests (Y.-C. Chen, Hung, and Wang 2018). Moreover, institutional investors usually request managers to voluntarily disclose more information (B. Ajinkya, Bhojraj, and Sengupta 2005; Boone and White 2015). By doing so, the information environment can become more transparent. Since institutional investors care about pecuniary interests and ESG commitments, they may request managers to issue more earnings forecasts to understand better the cost and benefits of ESG projects in mandated ESG reports. Therefore, we form our first sub-hypothesis in the null form:

Hypothesis 1a: An increase in investors' scrutiny does not increase the manager's voluntary disclosures when the firm is required to disclose ESG information.

Furthermore, mandatory ESG disclosures may impact managers' voluntary disclosures through the channel of analysts' scrutiny. The literature has shown that analysts' coverage can enhance the positive association between voluntary ESG disclosures and firm values (Tsang et al. 2022). It is unclear whether this enhancement still holds in the case of mandatory ESG disclosures. Since managers' earnings forecasts are value-relevant (B. B. Ajinkya and Gift 1984; Waymire 1984; Pownall and Waymire 1989), analysts may require managers to provide more such forecasts after the adoption of mandatory ESG disclosures. Under high scrutiny, analysts may request managers to offer slightly more voluntary disclosures even if mandatory disclosures worsen the adverse selection problem. After all, the information environment under high analysts' scrutiny tends to be transparent, so the demand for more information on mandatory ESG disclosures may be mild. In comparison, if disclosing mandatory nonfinancial information worsens the agency problem, analysts may require managers to provide more voluntary disclosures to explain the information in the mandatory disclosures. Hence, we form our second sub-hypothesis in the null form. If the

doi:10.6342/NTU202500807

hypothesis is rejected, the result may imply that mandatory nonfinancial disclosures worsen the agency rather than the adverse selection problem.

Hypothesis 1b: An increase in analysts' scrutiny does not increase the manager's voluntary disclosures when the firm is required to disclose ESG information.

4. Research Design

4.1 Method

The passage of the Taiwanese CSR regulation provides a quasi-experimental setting that we can exploit to investigate the impact of mandatory CSR disclosures on managers' voluntary financial disclosures. Specifically, we test **Hypothesis 1** by running the following difference-in-differences (DiD) logit regression:

$$
\begin{aligned}
\mathrm{M}F_{it} = \mathrm{logit}(&\beta_0 + \beta_1 TREAT_i + \beta_2 POST_t + \beta_3 TREAT_i \times POST_t + \beta_4 SIZE_{it-1} \\
&+ \beta_5 MB_{it-1} + \beta_6 ROA_{it-1} + \beta_7 LOSS_{it-1} + \beta_8 LEV_{it-1} \\
&+ \beta_9 EARNVOL_{it-1} + \beta_{10} RETURN_{it-1} + \beta_{11} RETURNVOL_{it-1} \\
&+ \beta_{12} TURNOVER_{it-1} + \beta_{13} BETA_{it-1} + \beta_{14} INSTOWN_{it-1} \\
&+ \beta_{15} BOARD_{it-1} + \beta_{16} DUAL_{it-1} + \beta_{17} OPAQ_{it-1} + \beta_{18} OPAQ2_{it-1} \\
&+ \varepsilon_{it})
\end{aligned}
\tag{1}
$$

In (1), MF is an indicator equal to one if firm i issues at least one earnings forecast in fiscal year t and zero otherwise. $TREAT$ is an indicator equal to one if firm i is required to disclose a stand-alone CSR report in fiscal year 2014 and zero otherwise. POST is an indicator equal to one if fiscal year t equals or exceeds 2014 and zero otherwise. If β_3 is significantly different from zero, the null form of **Hypothesis 1** will be rejected. Furthermore, the issuance of mandatory CSR reports and the release of earnings forecasts will be complements if the coefficient is positive and substitutes

doi:10.6342/NTU202500807

if it is negative. We also include many common variables to mitigate the omitted variable bias. First, we control the fundamental characteristics of firm i by including $SIZE$, the natural logarithm of total assets; MB, the market-to-book ratio; ROA, the returns on assets; LOSS, an indicator equal to one if earnings are negative and zero otherwise; LEV, the financial leverage, defined as the ratio of long-term debt to total assets; and EARNVOL, the earnings volatility, defined as the standard deviation of quarterly earnings in fiscal year t. Second, we control the capital market characteristics by including RETURN, the annual average of monthly returns; RETURNVOL, the standard deviation of daily returns in fiscal year t; TURNOVER, the annual average of monthly turnovers; and BETA, the CAPM beta calculated using daily returns. Third, we control corporate governance by including INSTOWN, the institutional ownership; BOARD, the logarithm of the number of boards; DUAL, an indicator equal to one if the manager is on the board and zero otherwise. Finally, we control the audit quality by including OPAQ, the absolute value of discretionary accruals derived from the modified Jones model plus lagged returns on total assets as an additional control variable (Kothari, Leone, and Wasley 2005); and OPAQ2, the squared term of OPAQ.

We test Hypothesis 1a and1b by running the following DiD logit regression:

$$
\begin{aligned}
\mathrm{M}F_{it} = \mathrm{logit}(&\beta_0 + \beta_1 TREAT_i + \beta_2 POST_t + \beta_3 TREAT_i \times POST_t + \beta_{31} SCRUT_{it} \\
&+ \beta_{32} TREAT_i \times POST_t \times SCRUT_{it} + \beta_4 SIZE_{it-1} + \beta_5 MB_{it-1} \\
&+ \beta_6 ROA_{it-1} + \beta_7 LOSS_{it-1} + \beta_8 LEV_{it-1} + \beta_9 EARNVOL_{it-1} \\
&+ \beta_{10} RETURN_{it-1} + \beta_{11} RETURNVOL_{it-1} + \beta_{12} TURNOVER_{it-1} \\
&+ \beta_{13} BETA_{it-1} + \beta_{14} INSTOWN_{it-1} + \beta_{15} BOARD_{it-1} \\
&+ \beta_{16} DUAL_{it-1} + \beta_{17} OPAQ_{it-1} + \beta_{18} OPAQ2_{it-1} + \varepsilon_{it})
\end{aligned}
\quad (2)
$$

In (2), we include SCRUT and the product of SCRUT and TREAT $\times$ POST. SCRUT is either $INVSCRUT$ or $ANASCRUT$. INVSCRUT is an indicator equal to one if institutional ownership of firm i is larger than the cross-sectional median in year t and zero otherwise. $ANASCRUT$ is an indicator equal to one if the number of analysts following firm i is greater than the cross-sectional median in year t and zero otherwise. If β_{32} is significantly different from zero, the null form of Hypothesis 1a or 1b will be rejected. Economically speaking, a positive sign of the coefficient will indicate that managers are more likely to provide voluntary disclosures under higher scrutiny.

4.2 Sample Selection

We construct our sample by conducting the following procedures. Specifically, we retrieve all the data of financial and market variables from the Taiwan Economic Journal (TEJ). Moreover, we retrieve the data of managers' earnings forecasts and the number of analysts following from the Institutional Brokers' Estimate System (I/B/E/S). We join the data from the two sources by linking the observations with their stock market tickers. We set the sample period from 2011 to 2017. This period covers three years before or after 2014, when Taiwanese firms were required to disclose stand-alone CSR reports for the first time. The initial joined data contains 15,357 firm-year observations within the sample period. We exclude financial firms from our sample, yielding a sample of 14,792 observations. Following Chen et al. (2018), we exclude firms that voluntarily disclose CSR reports during our sample period. First, we exclude treated firms that voluntarily disclose CSR reports before 2014. Second, we exclude control firms that voluntarily disclose CSR reports in the sample period. These exclusions shrink the sample size to 10,005 observations. We further remove the observations that have missing values in the variables in (1) and (2). Therefore, the final sample consists of 4,244 firm-year observations that cover 793 distinct firms. The sample selection procedures are summarized in Table 2.

doi:10.6342/NTU202500807

5. Main Results

5.1 Descriptive Statistics

Table 3 provides the distributions of fiscal years, industries, and treatment and control observations in the sample. Panel A shows that the fiscal years are almost uniformly distributed. Panel B suggests that the treatment and control observations are uniformly distributed among the fiscal years. After we construct the sample, only 19 firms are labelled as treated firms. These firms did not provide any ESG reports before 2014 and were mandated to provide such reports after 2014. We adopt the strict requirement for being a treated firm because to make sure that the parallel trends assumption plausibly holds. By restricting our sample so, we can enhance the internal validity of the results, boosting the confidence in the causal inference on how mandatory nonfinancial disclosures affect managers' voluntary financial disclosures. Panel C shows that more than 80% of the firms in the sample belong to the electronics industry. The dominance of the electronics industry is consistent with the industry distribution of all the listed firms during this sample period.[2] Panel C also shows that most industries voluntarily disclosed ESG reports before 2014 or during our sample period. In our sample construction, such observations may contaminate the causal inference, so we exclude them. Panel D shows that the distribution of treatment observations among industries differs from that of control ones. While the control observations are distributed among the industries of textile, electric machinery, electronics, and culture and creativity, the treatment ones are distributed among only the textile industry and the electronics industry. To control the heterogeneity of industry distributions in the treatment sample and the control sample, we add fixed effects in the regression analyses.

---

[2] During the sample period, more than half of the listed firms in Taiwan belong to the electronics industry, according to the TEJ data.

doi:10.6342/NTU202500807

Table 4 provides summary statistics and correlations of the variables used in the regression analyses. Panel A of Table 4 provides the summary statistics of the variables. The mean, 25th percentile, median, and 75th percentile of MF are nearly zero, suggesting that the firms in our sample seldom issue earnings forecasts. Moreover, 3% of the sample are the observations of treated firms. Panel B provides the Pearson and Spearman correlations of the variables. The probability of managers' issuing earnings forecasts is positively correlated with the treatment indicator. This positive correlation is statistically significant, providing preliminary evidence of the interaction between mandatory nonfinancial disclosures and managers' voluntary financial disclosures. Furthermore, the probability is positively correlated with the analyst scrutiny indicator. The positive correlation is not only statistically significant but also larger than that between the probability and the treatment indicator. This robust correlation provides preliminary evidence regarding Hypothesis 1b.

5.2 Regression Analysis

Table 5 provides a strong case for rejecting Hypothesis 1. In column (1) of Table 5, the estimated coefficient associated with $TREAT \times POST$ is positive and statistically significant. The positive association between $TREAT \times POST$ and MF indicates that the managers of treated firms are more inclined to provide earnings forecasts after their firms are required to disclose CSR reports. This positive association remains intact even after industry and year fixed effects are controlled, as shown in columns (2) and (3). The average partial effect of the mandatory disclosure shock on the probability of managers' issuing earnings forecasts is 90%. Economically speaking, managers are 90% more likely to issue earnings forecasts if their firms are forced to disclose CSR reports. The economic significance is huge, since managers in Taiwan seldom issue earnings forecasts (see Table 4 Panel A). The results suggest that mandatory nonfinancial and voluntary financial

doi:10.6342/NTU202500807

disclosures are complements rather than substitutes. The results may imply that managers care less about proprietary costs than about the adverse selection or agency problem when voluntarily disclosing financial information.

Table 6 provides evidence supporting a rejection of Hypothesis 1a. In column (1) of Table 6, the coefficient associated with $TREAT \times POST$ remains positive, even though the magnitude is less than the counterparts in Table 5. The coefficient associated with $TREAT \times POST \times INVSCRUT$ is positive and statistically significant. The positive association between $TREAT \times POST \times INVSCRUT$ and MF indicates that, if the institutional ownerships of the treated firms are high, the managers will be more inclined to provide earnings forecasts than those whose institutional ownerships are low. This positive association remains intact even after industry and year fixed effects are controlled, as shown in columns (2) and (3). The average partial effect of the moderating effect of scrutiny is 88%. Economically speaking, if their firms are under high investors' scrutiny, managers of the treated firms are 88% more likely to issue earnings forecasts than those whose firms are under low scrutiny. The results indicate that institutional investors tend to require managers to provide more earnings forecasts to clarify whether the ESG activities revealed in the mandatory disclosures can harm their pecuniary interests.

Table 7 provides evidence supporting a rejection of Hypothesis 1b. In column (1) of Table 7, the coefficient associated with $TREAT \times POST$ remains positive and statistically significant, even though the magnitude is less than the counterparts in Table 5. The coefficient associated with TREAT $\times P$OST $\times ANAS$CRUT is also positive and statistically significant. The positive association between TREAT $\times P$OST $\times ANAS$CRUT and MF indicates that, if the numbers of analysts following the treated firms are high, the managers will be more inclined to provide earnings forecasts than those whose numbers of analysts following are low. This positive association remains intact even after industry and year fixed effects are controlled, as shown in columns (2) and (3). The average partial effect of the moderating effect of scrutiny is 80%. Economically speaking, if their firms are under high analysts' scrutiny, managers of the treated firms are 80% more likely to issue earnings forecasts than those whose firms are under low scrutiny. The results suggest that managers voluntarily use financial information to lower not the adverse selection problem but the agency problem. The information environment should be more transparent when the number of analysts following is high than when the number is low. Therefore, managers may not need to disclose further information to complement the disclosed nonfinancial information. On the other hand, if managers are suspected of manipulating the information presentation or content in nonfinancial disclosures, shareholders or analysts may require managers to disclose more financial information to help them understand the nonfinancial information. The results in Table 7 may support the latter explanation.

6. Robustness

To check the sensitivity of the results, we conduct a battery of robustness tests. Specifically, we exclude the observations in the years of the ESG disclosure mandates, change model specifications, and conduct a propensity score matching DiD (PSM-DiD). In general, the robustness tests buttress the results in the main analysis.

First, the regulation of mandatory CSR disclosures was implemented in 2014 and revised in 2015. Therefore, we exclude the observations in 2014 or 2015 to test the robustness. Table 8 presents the results. The results suggest that the exclusion of the observations does not affect the conclusions based on the results in Table 5, Table 6, and Table 7.

doi:10.6342/NTU202500807

Second, we run probit rather than logit regressions to examine whether the results are sensitive to model choices. The results in Table 9 indicate that the results are insensitive. We also run the logit regression with double-clustered standard errors to investigate whether the results are sensitive to different correlation assumptions on the regression residuals. Table 10 shows that the results are similar to those in the main analysis even if the standard errors are clustered by firms and years. Furthermore, we put the number of managers' earnings forecasts as a dependent variable and run Poisson regressions to examine whether the mandatory ESG disclosures lead to increases in the number of forecasts and whether the moderating effects of investors' and analysts' scrutiny also exist. The results in Table 11 indicate that the results are also insensitive to this model specification.

Third, we conduct a propensity score matching to match each treated firm with a control firm. Specifically, we run a logit regression of $TREAT$ on $SIZE$ and industry fixed effects, since $SIZE$ and industry fixed effects are the only variables relevant to whether a firm should disclose CSR reports according to the regulation. Moreover, running this parsimonious logit regression can ensure that the matched sample is large enough to run (1) and (2). We run the logit regression on all the sample observations before 2014. Panel A of Table 12 shows the results of the logit regression. The results indicate that the firm size is a good predictor of whether a firm will be mandated to disclose ESG reports. In the next step, we utilize the predicted probabilities from the logit regression to determine which firms in 2014 are treated or control firms. We conduct an one-to-one matching without replacement to pair each treated firm with only a control one. After the matching procedure, we collect all the observations of the matched firms during the sample period to run the regressions (1) and (2). Panel B presents the results of regressions (1) and (2) we run after conducting the propensity score matching. Even though the sample size shrinks dramatically

doi:10.6342/NTU202500807

after the matching is conducted, the conclusions based on the results in Table 12 are still the same as those based on the results in Table 5, Table 6, and Table 7.

7. Parallel and Placebo Tests

Table 13 shows the results of the parallel and placebo tests. The methods of the parallel and placebo tests are based on Bertrand and Mullainathan (2003) and Lee et al. (2021), respectively. Column (1) presents the results of the parallel regression of the voluntary disclosure indicator. The results suggest that the regulation of mandatory ESG disclosures does not impact managers' voluntary disclosure decisions before the implementation of the regulation. The regulation starts to impact voluntary disclosures only during and after the implementation. Column (2) presents the results of the placebo regression. Following Lee et al. (2021), we exclude (include) all firm-year observations that have received treatment (that have not received treatment, including those from treated firms before they receive treatment) to prevent the contamination of post-placebo treatment estimates. In fact, including treated observations in placebo tests can result in a significant estimate for the placebo treatment variable, as many of the post-placebo treatment observations are treated firm-year observations. Although the results in Column (2) still reveal some contamination of post-placebo treatment estimates, the economic significance of the pseudo-regulation effect is much smaller than that of the actual regulation effect. Specifically, the coefficient associated with TREAT × Pseudo-POST is only 10% statistically significant. Furthermore, the economic significance of this coefficient is only around one-eighth of that in column (3) of Table 5. Overall, the results in Table 13 tend to support the assumption of parallel trends in the DiD analysis.

Figure 2 illustrates two time series of the ratios of firms that issue earnings forecasts during the sample period. The dark line plots the time series of treated firms; the light one plots the time series of control firms. The time series of control firms remains steady: the ratios of control firms

doi:10.6342/NTU202500807

fluctuate around 0.15%. On the other hand, the ratios of treated firms remain at 0% before the initiation of mandatory CSR disclosures in 2014. The ratios skyrocket to 0.9% after the initiation. This figure reveals some clues that managers of treated firms increase voluntary financial disclosures after their firms are required to disclose nonfinancial information. These clues suggest that the parallel trend assumption should reasonably hold in our DiD analysis.

8. Conclusion

In this study, we examine how managers change their voluntary disclosures when their firms are required to disclose nonfinancial information. To test this question, we utilize the mandatory CSR reporting in Taiwan in 2014 as a quasi-experiment. We find that managers are 90% more likely to issue earnings forecasts if their firms are forced to disclose CSR reports. Furthermore, if their firms are under high scrutiny, the managers are 80%-90% more likely to issue earnings forecasts than those whose firms are under low scrutiny. The findings suggest that managers voluntarily disclose financial information more to ease the agency problem than to solve the adverse selection problem or prevent proprietary costs. The findings can contribute to the literature on voluntary disclosures, mandatory nonfinancial disclosures, and interactions among mandatory and voluntary disclosures.

doi:10.6342/NTU202500807

Chapter 2:
Estimating Fundamental Betas via Linear Machine Learning

1. Introduction

In recent years, machine learning and artificial intelligence have revolutionized various fields, from healthcare and social sciences to finance and accounting. The ability of machine learning algorithms to analyze vast amounts of data, identify complex patterns, and improve predictive accuracy has made them an indispensable tool in financial research. In asset pricing, where traditional models often struggle with high-dimensional data and non-linear relationships, machine learning offers a promising alternative for refining risk estimation and enhancing return predictability (Gu, Kelly, and Xiu 2020). Given its success in improving market beta forecasts (Drobetz et al. 2024), a natural question arises: Can machine learning also enhance the estimation of fundamental betas, providing a more robust measure of systematic risk? This paper explores this question by investigating the potential of machine learning to improve fundamental beta forecasts, ultimately bridging the gap between fundamental firm-level information and asset pricing models.

The fundamental beta, which measures the association between firm fundamentals and market systematic risk, has been proposed as an alternative to the traditional market beta in explaining cross-sectional return variations. Such betas include cash-flow betas and earnings betas (e.g., Campbell and Vuolteenaho 2004; Ellahie 2021). However, traditional estimation methods, such as variance decomposition and vector autoregressions (VARs), face challenges related to implementation complexity, specification errors, and instability (L. Chen and Zhao 2009; Penman 2016). To address these issues, Ellahie (2021) introduced a simpler approach to estimate earnings betas using different earnings series and found that investors priced six out of eleven constructed

doi:10.6342/NTU202500807

betas. Despite these advancements, the reliance on analyst forecasts limits sample coverage and raises concerns about selection bias (Ball, Sadka, and Tseng 2022). Recent asset pricing studies (e.g., Gu, Kelly, and Xiu 2020; Drobetz et al. 2024) suggest that machine learning could enhance fundamental beta estimation by handling high-dimensional data, capturing non-linear interactions, and reducing biases inherent in traditional regression models. Given its demonstrated success in improving market beta forecasts, applying machine learning to fundamental beta estimation may yield more accurate, stable, and economically meaningful insights, advancing its role in empirical asset pricing.

Improving fundamental beta forecasts via machine learning is important to not only the academia but also the financial industry. Advancing the precision of fundamental beta estimates enriches the literature on empirical asset pricing and financial accounting, addressing the drawbacks in traditional methodologies and approaching classic research questions from a new perspective. The implications of improved fundamental beta forecasts extend beyond the academia. For investors and fund managers, accurate fundamental beta predictions are essential for constructing risk-adjusted portfolios and developing market-neutral investment strategies. Machine learning's ability to improve these predictions can lead to better resource allocation and more effective risk management.

By investigating whether machine learning improves fundamental beta forecasts, we seek to bridge the theoretical appeal of fundamental beta with the practical capabilities of machine learning. This study aims to provide evidence on how computational advancements can enhance our understanding of systematic risk and contribute to the development of more robust and actionable asset pricing models.

doi:10.6342/NTU202500807

We find that machine learning models—especially the Lasso model and the Elastic Net model—mostly outperform traditional rolling regressions in forecasting fundamental betas, achieving lower RMSEs and more stable performance across firm types, beta portfolios, and periods. These improvements are particularly pronounced for EBETA and TBETA and are most evident in firms with moderate systematic risk exposure. However, despite their slightly superior statistical accuracy, the economic value of these forecasts remains limited. RMSHEs show that machine learning models often match or underperform the rolling regression in hedging applications, especially post-2000. Furthermore, Fama-MacBeth regressions reveal that the estimated betas—regardless of model—carry little pricing power, with coefficients that are generally small, statistically weak, and economically insignificant. These results underscore a key finding: while machine learning enhances fundamental beta estimation, it does not necessarily improve its usefulness in explaining cross-sectional stock returns.

This research makes several key contributions to the literature on fundamental beta estimation, machine learning, and beta pricing. First, it extends the empirical asset pricing literature by proposing a novel machine learning-based framework for estimating fundamental betas, addressing the limitations of traditional estimation methods and enhancing their applicability in financial research. Second, it contributes to the growing body of work on machine learning in finance by demonstrating its effectiveness in capturing complex, non-linear relationships in firm-level and market-level earnings data, ultimately improving beta estimation accuracy. Lastly, this study provides new insights into beta pricing by evaluating whether machine learning-generated fundamental betas exhibit stronger explanatory power in cross-sectional asset pricing tests. By integrating machine learning techniques into fundamental beta estimation, this research not only refines our understanding of systematic risk but also offers practical insights for investors, financial analysts, and policymakers seeking to improve risk assessment and investment decision-making.

2. Literature Review

Understanding whether and how adopting machine learning improves fundamental beta forecasts is a vital inquiry in the field of empirical asset pricing and financial accounting. The failure of the capital asset pricing model (CAPM) to account for cross-sectional return variations has motivated the literature to seek other betas that have better explanatory power. One of them is the fundamental beta, known as the cash-flow or earnings beta in the literature (e.g., Campbell and Vuolteenaho 2004; Bansal, Dittmar, and Lundblad 2005; L. P. Hansen, Heaton, and Li 2008; Campbell, Polk, and Vuolteenaho 2010; Botshekan, Kraeussl, and Lucas 2012; Maio 2013). The fundamental beta, which measures the association between firm-level fundamentals and market-level systematic risk, provides a critical link between accounting data and systematic risk factors. As highlighted by Ellahie (2021), such betas are theoretically appealing and practical for explaining cross-sectional variations in returns, offering an alternative to traditional market betas, which have often failed to account for the complexities of systematic risk effectively.

Despite the potential explanatory power of the fundamental beta, traditional methods of estimating fundamental beta face several challenges. Specifically, the traditional approach relies on the variance decomposition based on financial and accounting identity log-linearization and vector autoregressions (VARs) (see Campbell and Shiller 1988; Campbell 1991; Vuolteenaho 2002). This approach is notoriously difficult to implement and often suffers from specification errors, sub-sample instability, and sensitivity to model assumptions (L. Chen and Zhao 2009; Penman 2016; Penman and Zhang 2020). These limitations hinder the broader application of the fundamental beta in asset pricing models and investment practices. To solve the methodological

doi:10.6342/NTU202500807

problem, Ellahie (2021) adopted a different approach to develop several simple and easy-to-construct earnings betas and examine their relative ability to explain cross-sectional variation in returns. Specifically, he estimated earnings betas using 11 different earnings series, including (1) historical realizations of earnings over the past 12 months or analysts' expectations of earnings for the next 12 months; (2) earnings levels or earnings changes (i.e., growth); (3) three alternatives for scaling earnings, including prior period earnings, book value of equity, or market value of equity; and (4) analysts' expectations of long-term growth in earnings. He regressed each earnings series on a corresponding aggregate-level earnings series using backward-rolling five-year window estimations to generate time-varying earnings betas. By doing so, Ellahie prevented the notorious measurement error problem of the traditional approach. By conducting the cross-sectional asset pricing test, he found that 6 of the 11 earnings betas he constructed were correctly priced in risk premiums.

Even though it solves the methodological problem of the earnings beta estimation and yields convincing results on the pricing of earnings betas, the limited coverage of analyst earnings forecasts may constrain Ellahie (2021)'s approach, restricting the sample to firms with analyst coverage and the IBES data availability period (Ball, Sadka, and Tseng 2022, p. 610). This limitation raises concerns about potential sample selection bias and the generalizability of the findings across a broader set of firms. Furthermore, Ellahie (2021)'s approach may suffer from the model misspecification problem of regressions exposed and addressed with machine learning by recent asset pricing studies (e.g., Gu, Kelly, and Xiu 2020; Akyildirim, Goncu, and Sensoy 2021; Bianchi, Büchner, and Tamoni 2021; Leippold, Wang, and Zhou 2022; Bali et al. 2023; Drobetz et al. 2024; J. H. Hansen and Siggaard 2024). In estimating stock market betas, Drobetz et al. (2024) showed that machine learning can provide significant advantages in handling high-dimensional

doi:10.6342/NTU202500807

data, identifying non-linear relationships, and capturing interactions between predictors. These features not only enhance predictive accuracy but also reduce biases associated with traditional regression-based approaches. Applying these machine learning techniques to fundamental beta estimation may address persistent empirical challenges, such as capturing time-varying relationships and accounting for the interaction of firm-level and aggregate fundamentals, thereby improving the robustness and applicability of fundamental beta in explaining systematic risk. Thus, using machine learning may provide a more comprehensive and reliable framework for understanding the determinants and the role of fundamental betas in asset pricing.

Moreover, the empirical evidence from asset pricing studies on machine learning applications highlights the potential economic benefits of adopting machine learning. For example, Drobetz et al. (2024) showed that machine learning-based beta estimators consistently outperformed traditional approaches in predictive accuracy, particularly during periods of economic distress. These estimators reduced forecast errors, mitigated biases, and generated more stable and reliable estimates. Translating these advantages to fundamental beta forecasts may revolutionize our understanding of the fundamental drivers of systematic risk and enhance their integration into asset pricing models.

3. Research Design

3.1 Method

3.1.1 Performance Evaluation

The analysis aims to investigate whether machine learning methods to estimate fundamental betas have better predictive performance than traditional ones. Specifically, we conduct traditional and machine learning methods to estimate time-varying fundamental betas and compare the estimation performance of the two kinds of methods. The traditional method is the rolling regression approach

doi:10.6342/NTU202500807

commonly used in beta estimation. The machine learning methods include a linear regression with all the predictors and an elastic net. Building on the methodologies of the studies on machine learning and asset pricing (Cosemans et al. 2016; Hollstein and Prokopczuk 2016; Gu, Kelly, and Xiu 2020; Drobetz et al. 2024), we conduct our estimation and forecast evaluation at the individual stock level. The general framework of our forecasting involves generating out-of-sample beta estimates at the firm level every quarter and employing an iterative approach. In the first step, we utilize all or part of the data available up to the end of quarter *t* to generate forecasts for the beta of each stock *i* over the out-of-sample forecast period, which spans from quarter *t+1* to quarter *t+4*. We denote these beta forecasts as $\beta^{\mathrm{F}}_{\mathrm{i,t}}$. In the second step, we extend the data set to include observations up to the end of quarter *t+1* and compute beta forecasts for the next out-of-sample period, covering months *t+2* through *t+5*. By continuously iterating through the data, we construct a time series of overlapping out-of-sample beta estimates, which we compare with their realized counterparts.

We follow the approaches of Andersen et al. (2006) and Drobetz et al. (2024) to calculate realized fundamental betas. Specifically, we define the future realized fundamental beta over the 4-quarter forecast horizon as

$$\beta^{R}_{i,t+4} \equiv \frac{COV^{R}_{iM,t+4}}{VAR^{R}_{M,t+4}}, \tag{3}$$

where $COV^{R}_{iM,t+4}$ represents the realized covariance between the earnings growth of stock *i* and one of the following aggregate measures: the aggregate earnings growth, the household wealth growth, or the total factor productivity growth. The $VAR^{R}_{M,t+4}$ represents the realized variance of the aggregate measures. To compute the realized fundamental beta, we run the regression at quarter $t\, X_{i,t} = \beta_{i,0} + \beta_{i,t} X_{MKT,t} + \varepsilon_{i,t}$ on the observations between quarter *t+1* and quarter *t+4*, , where

$X_{i,t}$ represents the earnings measure for a firm, and $X_{MKT,t}$ represents the market-level earnings measure or macroeconomic variables.

To compare the predictive performance of each model, we follow Drobetz et al. (2024)'s approach to calculate the value-weighted mean squared error (MSE) at the end of each quarter t:

$$MSE_{t+4|t} \equiv \sum_{t=1}^{N_t} w_{i,t}(\beta_{\mathrm{i,t}}^{\mathrm{F}} - \beta_{i,t+4}^{R})^2, \tag{4}$$

where N_t represents the number of stocks included in the sample at the end of quarter *t*, while $w_{i,t}$ denotes the weight of stock *i* based on its market capitalization. It is essential to recognize that realized betas are also estimates. Nonetheless, using future realized betas as a benchmark for evaluating forecasts has been shown to be statistically effective (P. R. Hansen and Lunde 2006). Furthermore, in the context of volatility forecasting, Patton (2011) demonstrates that the mean squared error (MSE) criterion remains robust even when the evaluation proxy contains mean-zero noise. To facilitate economic interpretations of forecast errors, we take the square root of the MSE to get the root mean squared error (RMSE), which equals $\sqrt{MSE}$.

Unlike many traditional models, machine learning models may naturally excel when assessed using the mean squared error (MSE), as they are explicitly optimized to minimize this criterion. To address this potential advantage, we follow Drobetz et al. (2024) to complement the MSE analysis with an alternative metric—the mean squared hedging error (MSHE). Specifically, for each stock and each quarter, we calculate the hedging error as the squared difference between the actual return and the return predicted by the market model using the forecasted beta. As is the case with the MSE, we calculate the cross-sectionally value-weighted average of the MSHEs in each quarter. We summarize the calculation in the following equation:

$$MSHE_{t+4|t} \equiv \sum_{t=1}^{N_t} w_{i,t}(R_{i,t+4} - \beta_{i,t}^{F} \times R_{M,t+4})^2, \tag{5}$$

where $R_{i,t+4}$ is the realized return four quarters later, and $R_{M,t+4}$ is the realized market return, proxied using the CRSP value-weighted portfolio return. By evaluating the models according to the alternative measure, we can obtain an additional perspective on out-of-sample predictive performance. To facilitate economic interpretations of forecast errors, we take the square root of the MSHE to get the root mean squared hedging error (RMSHE), which equals $\sqrt{MSHE}$.

3.1.2 Traditional Method

We employ the rolling regression approach to estimate the betas using the quarterly data. First, following Ball et al. (2022), we calculate in each quarter the firm-level annual change in earnings, defined as operating incomes minus preferred stock cash dividends or net incomes before extraordinary items if operating incomes are missing, scaled by the beginning market value. Second, we calculate the annual change in aggregate earnings and growth rates of household wealth and total factor productivity. To calculate the aggregate earnings, we compute firm-level earnings and aggregate them across firms to construct a market-level earnings series. The aggregation ensures that the same set of firms contributes to both the numerator (earnings measures) and denominator (prior-period scalars) to maintain consistency (Ellahie 2021, p. 90). Third, following Ellahie (2021), we estimate the regressions of the firm-level changes in earnings on the aggregate measures using the five-year rolling window. Using the constructed measures, we estimate the regression model $X_{i,t} = \beta_{i,0} + \beta_{i,t}X_{MKT,t} + \varepsilon_{i,t}$, where $X_{i,t}$ represents the earnings measure for a firm, and $X_{MKT,t}$ represents the market-level earnings measure or macroeconomic variables. These regressions are performed in backward-rolling five-year windows, with a minimum of three years of data required when full five-year data is unavailable. The rolling

doi:10.6342/NTU202500807

regression approach provides time-varying estimates of fundamental betas, capturing the dynamic relationship regarding the systematic risk on earnings. This methodology ensures that the estimates reflect the evolving nature of earnings sensitivity to market-wide earnings fluctuations or macroeconomic dynamics. At the end of each quarter, we use all the data from the previous five years to predict the fundamental betas in the next four quarters using the fundamental beta estimates in Ball et al. (2022).

3.1.3 Machine Learning Methods

Machine learning-based methods adopt a distinct and more structured approach to distill cross-sectional variations in future fundamental betas. In fact, machine learning techniques are designed specifically to optimize predictive performance, making such techniques suitable for beta prediction. In these models, the realized fundamental beta serves as the outcome variable, while sample beta estimates, firm characteristics, and other relevant factors act as predictors. By structuring the model around the forecasting objective from the outset and incorporating diverse sources of information, this approach has the potential to enhance predictive accuracy. To model the fundamental beta, we adapt the additive prediction error framework proposed by Gu et al. (2020):

$$\beta^{R}_{i,t+4} = E_t\left(\beta^{R}_{i,t+4}\right) + \varepsilon_{i,t+4}. \tag{6}$$

In this framework, the expectation function E_t indicates the best prediction based on all the available information at quarter t. The variable $\varepsilon_{i,t+4}$ represents the noise term. This identity provides an intuition that the realized fundamental beta equals the expected beta plus some noise. Furthermore, the expected beta can be modeled as:

$$E_t\left(\beta^R_{i,t+4}\right) = \mathrm{f}^*\left(\Omega_{i,t}\right), \tag{7}$$

where f^* represents the true model of the expectation, and $\Omega_{i,t}$ represents all the predictors utilized in the expectation. While our machine learning-based forecasting models encompass various methodologies—such as linear regression, tree-based approaches, and neural networks—they share a common objective: approximating the underlying prediction model by minimizing out-of-sample mean squared error (MSE). The true function f^* can take either a linear or a nonlinear form and may be parametric or nonparametric, shattering the limitations of the traditional regression approach.

To estimate the betas using machine learning techniques, we follow the standard procedures in the literature on asset pricing and machine learning (Drobetz et al. 2024, Section IV) without cross-validation because of the machine limits. Specifically, the dataset is divided into training and test samples, with a rolling-window approach to account for the temporal structure of the data. The training sample, which covers 40 quarters, is used to fit models according to Drobetz et al. (2024)'s hyperparameter configurations. The test sample, representing the four quarters following the training period, remains entirely out-of-sample and serves to evaluate predictive accuracy. This process is repeated quarterly to incorporate new data dynamically, refitting models each quarter. We employ four models: (i) the linear pooled OLS regression (OLS), (ii) the Lasso regression (Lasso), (iii) the Ridge regression (Ridge), and (iv) the Elastic Net model. The latter three of the four models are regularized regressions. Reviewing Drobetz et al. (2024)'s hyperparameter settings, we set the penalization strength in our regularized regressions as one. Furthermore, we set the penalty mixture of the Elastic Net model as 0.5. This setting makes the Elastic Net model behave as a linear combination of half of the Lasso model and half of the Ridge model. We summarize the machine learning and traditional models in Table 14.

3.1.4 Cross-Sectional Asset Pricing Tests

After deriving the various beta estimates based on the traditional or machine learning methods, we conduct cross-sectional asset pricing tests to evaluate whether the estimated fundamental betas are priced and to what extent. Following Ball et al. (2022), we run the following regressions:

$$Ret_{i,m} = \alpha_0 + \alpha_1 \hat{\beta}_{EARN,i,t} + \varepsilon_{\alpha,i,m} \quad (8)$$

$$Ret_{i,m} = \beta_0 + \beta_1 \hat{\beta}_{WEALTH,i,t} + \beta_2 \hat{\beta}_{TFP,i,t} + \varepsilon_{\beta,i,m},$$

where $Ret_{i,m}$ is firm i's stock return in month *m*. $\hat{\beta}_{EARN,i,t}$, $\hat{\beta}_{WEALTH,i,t}$, and $\hat{\beta}_{TFP,i,t}$ are the estimates of the fundamental betas at quarter *t*, measured as the correlations between firm-level earnings and aggregate earnings, household wealth growth, and total factor productivity growth, respectively. If investors price $\hat{\beta}_{EARN,i,t}$ in their risk premiums, the estimated α_1 should be significantly positive. Similarly, if investors price $\hat{\beta}_{WEALTH,i,t}$ and $\hat{\beta}_{TFP,i,t}$, the estimated β_1 and β_2 should be significantly positive. In fact, the coefficients α_1, β_1, and β_2 capture the realized relationship between risk and return for a given month *m*. The average values of these coefficients over time are commonly regarded as estimates of risk premiums. We also run a regression of returns on all the three betas. To estimate the coefficients associated with the betas, we use the Fama-MacBeth regression (Fama and MacBeth 1973). This regression involves two steps. First, we run cross-sectional regressions of returns on firm-level betas in each period to obtain a time series of slope coefficients. Second, we average these estimated slopes over time to infer risk premia. This approach accounts for time variation in factor loadings and allows for valid inference by adjusting standard errors for cross-sectional correlation.

3.2 Data

The data sources include the CRSP, Compustat, Fama-French Data Library, and Fred. We construct all the earnings measures and macroeconomic variables used to calculate fundamental betas in Ball et al. (2022) and most of the predictors found in Gu et al. (2020), Ellahie (2021), Ball et al. (2022), and Drobetz et al. (2024). The descriptions and definitions of the predictors are in Table 15.

During the data processing, we restrict our sample to normal firms and keep as many observations as possible. Specifically, we confine our sample to ordinary equity securities, whose CRSP share codes are 10 or 11, listed on the New York Stock Exchange (NYSE), the American Stock Exchange (AMEX), or the National Association of Securities Dealers Automated Quotations (NASDAQ) (Ball, Sadka, and Tseng 2022; Drobetz et al. 2024). We also confine our sample to the firms whose end months of fiscal years are December to ensure the time consistency of announcements of macroeconomic series and earnings (Ball, Sadka, and Tseng 2022, p. 640). After constructing the predictors, we cross-sectionally winsorize each continuous predictor by 5% and 95% (Drobetz et al. 2024) and replace the missing values with the cross-sectional medians (Gu, Kelly, and Xiu 2020). Afterward, we conduct the cross-sectional mean normalization on each firm-level predictor in each quarter, subtracting the cross-sectional mean of each predictor in each quarter and dividing the difference by the cross-sectional maximum minus the cross-sectional minimum. After conducting the selection procedures, we obtain a sample that contains 615,946 firm-quarter observations in 1972Q2 – 2022Q4. We summarize the sample selection in Table 16.

doi:10.6342/NTU202500807

4. Results

4.1 Descriptive Statistics

Table 17 shows the summary statistics of the outcome variables and predictors. Panel A reports the summary statistics of the three fundamental beta measures—EBETA, TBETA, and WBETA—across 615,946 firm-quarter observations. EBETA has the highest mean (1.322) and standard deviation (19.671), indicating substantial heterogeneity and extreme values in firms' sensitivity to aggregate earnings. In contrast, TBETA and WBETA have means near zero (0.025 and –0.075, respectively) and lower dispersion (SDs of 0.973 and 4.797), suggesting weaker and more stable relationships with total factor productivity and household wealth growth. These differences highlight the varying economic channels through which firm fundamentals respond to systematic risk.

Panel B presents the summary statistics of the predictors utilized in the machine learning models to forecast EBETA, TBETA, and WBETA across 615,946 firm-quarter observations. After the mean-normalization, all the firm-level predictors are bounded within -1 and 1. Additionally, all the means are near zero. The standard deviations range from 0.1 to 0.3, indicating moderate variability. The distributions are generally symmetric, though a few variables (e.g., EPVAR, SGA, FXC) show mild skewness or longer tails, as reflected in wider percentile spreads. The relatively tight interquartile ranges suggest that most predictors are well-behaved, facilitating stable training of machine learning models.

Panel C summarizes the time-series and macroeconomic predictors used in the machine learning models to forecast EBETA, TBETA, and WBETA. Unlike the firm-level ones, these predictors do not vary across firms within a given period. The mean values indicate modest average growth rates across economic indicators, with aggregate earnings growth (dAEARN) averaging

doi:10.6342/NTU202500807

0.2%, consumption growth (dCON) 2.1%, and real GDP growth (dGDP) 2.8%. Household wealth growth (dWEALTH) is the highest, averaging 7.1%, while total factor productivity growth (dTFP) averages 0.8% but exhibits the largest dispersion (standard deviation = 0.033), suggesting higher volatility. Overall, the interquartile ranges suggest relatively stable macroeconomic conditions, though occasional extreme values—especially in dTFP—reflect sensitivity to broader economic fluctuations. In general, the distributions of the predictors are smooth enough to mitigate the outlier problem or the scale problem in conducting machine learning models.

Table 18 shows the distributions of industries and years in the sample. Panel A reports the industry composition of the sample based on the Fama-French 48-industry classification. The sample covers a wide range of sectors, with Banking being the most represented industry, accounting for 11.5% of all firm-quarter observations, followed by Business Services (9.53%), Pharmaceutical Products (6.80%), and Trading (5.67%). Other notable sectors include Petroleum and Natural Gas (4.56%), Electronic Equipment (4.00%), and Utilities (3.83%). In contrast, industries such as Shipbuilding and Railroad Equipment, Tobacco, and Defense are minimally represented, each comprising less than 0.2% of the sample. Overall, the distribution reflects a relatively broad industry coverage, with a concentration in finance, healthcare, technology, and industrial sectors.

Panel B presents the distribution of firm-quarter observations by year within the sample period. The sample shows a steady increase in coverage over time, with earlier years like 1972–1981 each contributing around 1% of the sample annually. The number of observations rises notably from the mid-1980s, peaking in 1996 at 3.04%, and remains consistently high through the early 2000s. From 2001 onward, the annual share stabilizes between 1.88% and 2.61%, reflecting a mature and stable dataset in later decades. Overall, the sample provides broad temporal coverage

doi:10.6342/NTU202500807

across five decades, supporting analyses of long-run trends and time-varying dynamics in fundamental beta estimation.

4.2 General Evaluation of Model Performance

Table 19 shows the evaluation results of the forecast performance of five models—Rolling Regression, OLS, Lasso, Ridge, and Elastic Net—in estimating three types of fundamental betas: EBETA (based on aggregate earnings growth), TBETA (based on total factor productivity growth), and WBETA (based on household wealth growth). The models are assessed using two metrics: the root mean squared error (RMSE) in Panel A, which measures the accuracy of beta forecasts, and the root mean squared hedging error (RMSHE) in Panel B, which evaluates the economic usefulness of those forecasts in predicting future returns.

In Panel A, we observe that the OLS model performs poorly, producing implausibly large RMSE values across all beta types, indicating overfitting or instability due to high-dimensional predictors without regularization. In contrast, the Lasso, Ridge, and Elastic Net models all achieve RMSEs comparable to or better than the Rolling Regression benchmark, particularly for TBETA and EBETA. For example, the Elastic Net achieves the lowest RMSE for EBETA (8.042), slightly outperforming the Rolling Regression (8.279) and Lasso (8.311). Similarly, for TBETA, both Lasso and Elastic Net achieve the lowest RMSE (0.368), suggesting these regularized models are more effective at handling noise and multicollinearity in the predictor set. While the Ridge model underperforms slightly compared to Lasso and Elastic Net, it still performs much better than the unregularized OLS.

Panel B reports the RMSHEs, which provide insight into the economic value of the beta forecasts. Again, the OLS model fails to deliver meaningful results, with extremely large hedging errors. Among the remaining models, the Elastic Net model achieves the lowest RMSHE for

doi:10.6342/NTU202500807

EBETA (0.106), indicating that its beta forecasts lead to the most accurate return predictions in a market model context. For TBETA and WBETA, all models—including Rolling Regression, Lasso, Ridge, and Elastic Net—deliver similar RMSHEs, suggesting that the gains from machine learning methods are more pronounced when forecasting EBETA, which may involve more complex firm-level and macro interactions.

Overall, the results demonstrate that machine learning models—especially Lasso and Elastic Net—slightly outperform traditional rolling regressions and unregularized OLS in both statistical accuracy and practical hedging performance. These findings underscore the value of incorporating regularization techniques when systematic risk components are forecasted using high-dimensional firm and macroeconomic information.

4.3 Time Series of Model Performance

Figure 3, Figure 4, and Figure 5 present the time series of relative forecast errors (RMSEs) for three types of fundamental betas—EBETA, TBETA, and WBETA—estimated using Lasso, Ridge, and Elastic Net models, benchmarked against the traditional Rolling Regression. The relative errors are expressed as the percentage difference between each machine learning model's RMSE and that of the benchmark, where positive values indicate worse performance than the Rolling Regression, and negative values indicate better performance. Shaded grey areas represent NBER-defined recession periods.

Across all three figures, Lasso (blue) and Elastic Net (green) generally outperform the benchmark, with relative forecast errors often hovering around or below the zero line, indicating improved predictive accuracy. This pattern is most evident in EBETA and TBETA, where Lasso achieves the lowest error rates in most of the quarters, while Elastic Net delivers stable and modest

doi:10.6342/NTU202500807

improvements. For WBETA, the differences across models narrow over time, with all three approaches converging toward similar performance after the early 2000s.

In contrast, the Ridge model (red) displays significantly greater volatility and frequent underperformance, particularly in earlier periods and during recessions. Its relative error spikes are especially pronounced in WBETA during the 1980s and in EBETA during downturns. These findings suggest that Ridge is more sensitive to macroeconomic instability and sample noise. Taken together, the results highlight the superior robustness and stability of Lasso and Elastic Net, making them more reliable alternatives to traditional rolling regressions for forecasting fundamental betas.

Figure 6, Figure 7, and Figure 8 present the time series of relative root mean squared hedging errors (RMSHEs) for the three types of fundamental betas—EBETA, TBETA, and WBETA—estimated using Lasso, Ridge, and Elastic Net models, with the traditional Rolling Regression serving as the benchmark. The relative errors are expressed as the percentage difference between each model's RMSHE and that of the benchmark, where values above zero indicate worse hedging performance than the rolling regression. Across all three figures, Ridge (red) consistently exhibits the highest volatility and frequent large deviations, particularly during the 1980s and early 1990s, and often around NBER-defined recession periods (shaded in grey). Although its performance stabilizes somewhat in recent decades, it generally underperforms across all beta types.

In contrast, the Lasso (blue) and Elastic Net (green) models offer more stable and comparable hedging performance, especially for TBETA and WBETA. In the case of TBETA, the two models perform nearly identically and hover close to the zero line, occasionally outperforming the rolling regression before 2000, but generally underperforming after 2000. For WBETA, both

doi:10.6342/NTU202500807

models show more positive relative errors throughout the sample, suggesting limited economic improvement over the benchmark, despite better statistical accuracy in RMSEs.

For EBETA, the hedging performance of all machine learning models tends to lag behind the rolling regression, with most relative RMSHEs exceeding zero. Although Elastic Net remains the most stable of the three, even it struggles to deliver consistent economic value in return forecasting. These findings underscore a key insight: while machine learning techniques often improve statistical forecasts of fundamental betas, their advantages do not always translate into better hedging performance. In particular, the rolling regression remains a competitive and often superior benchmark in terms of economic relevance, especially in earlier decades.

4.4 Model Performance among Different Beta Portfolios

Figure 9, Figure 10, and Figure 11 report the time-series averages of quarterly market-capitalization-weighted root mean squared errors (RMSEs) across decile portfolios sorted by estimated EBETA, TBETA, and WBETA values, respectively, for four models: Rolling Regression, Lasso, Ridge, and Elastic Net. Across all three figures, a consistent U-shaped error pattern emerges: forecast errors are highest in the lowest (Decile 1) and highest (Decile 10) beta portfolios and lowest in the middle deciles (especially Deciles 4 to 6). This pattern suggests that firms with extreme beta exposures—whether to aggregate earnings, total factor productivity, or household wealth—are harder to forecast, likely due to heightened volatility or nonlinear sensitivities to macroeconomic risk.

Among the models, Lasso and Elastic Net consistently outperform Rolling Regression and Ridge, particularly in the middle deciles and, in some cases, even in the tails. These regularized models demonstrate superior accuracy in capturing firm-level earnings sensitivities, benefiting from their ability to handle high-dimensional predictors and reduce overfitting. In contrast, Ridge

doi:10.6342/NTU202500807

exhibits the highest forecast errors, especially in the outer deciles, indicating its limited effectiveness when applied to firms with extreme beta exposures.

Overall, the results show that while all models perform best for firms with moderate risk exposure, Lasso and Elastic Net may offer clear advantages across the beta spectrum, delivering more accurate forecasts and demonstrating the potential of machine learning in refining beta estimation across diverse macroeconomic dimensions.

4.5 Asset Pricing Tests

Table 20 reports the Fama-MacBeth regression results that assess whether estimated fundamental betas (EBETA, TBETA, and WBETA) are priced in the cross-section of stock returns. Across all models—Rolling Regression, OLS, Lasso, Ridge, and Elastic Net—the results reveal very weak statistical and economic significance of the estimated risk premia. In most cases, the coefficients on the betas are economically small (close to zero) and statistically insignificant, with few exceptions. Across all panels, the adjusted R-squared values are extremely low, typically under 1%, indicating that the explanatory power of these models in explaining cross-sectional return variation is minimal, regardless of the estimation method.

The Rolling Regression in Panel A shows marginal evidence that EBETA and WBETA carry weak pricing signals. Specifically, EBETA and WBETA are marginally significant at the 10% level in specifications (1) and (3), respectively. However, the magnitudes of these coefficients are essentially zero, suggesting negligible economic impact, despite statistical significance. Similarly, TBETA fails to demonstrate any pricing power in any specification.

Machine learning-based models provide mixed signals. In Panel C (Lasso), EBETA is weakly priced at the 10% level, but again with economically trivial coefficients. Ridge (Panel D)

doi:10.6342/NTU202500807

shows slightly stronger signals: EBETA is significant at the 5% level in the model (1), and WBETA is marginally significant (5% or 10%) in models (2) and (3). Yet, the small magnitudes again suggest that even when statistically significant, the economic relevance remains limited. Elastic Net (Panel E) fails to detect any statistically significant pricing across all beta estimates. The OLS model (Panel B) shows no statistically significant relationship between returns and any of the betas. This underperformance is likely due to the overfitting and instability associated with unregularized high-dimensional models, which underscores the advantage of using regularized approaches like Lasso and Ridge in beta estimation.

The lack of meaningful and consistent pricing of fundamental betas across methods suggests that the systematic risk captured by EBETA, TBETA, and WBETA—at least as estimated in this study—is not strongly priced in the cross-section of monthly stock returns. This may reflect several possibilities: fundamental betas might not effectively capture sources of risk priced by the market in our sample period; the risks associated with firms' fundamental earnings exposures could be diversifiable and thus not rewarded with a risk premium; and while machine learning techniques can enhance forecasting accuracy, they do not necessarily translate into more economically informative betas for explaining return variation.

In general, the Fama-MacBeth regressions demonstrate that while some models, particularly Ridge and Lasso, yield marginally significant pricing estimates for certain betas, the overall economic significance is weak. These findings caution against assuming that improvements in beta estimation accuracy via machine learning necessarily lead to better cross-sectional return predictability or stronger asset pricing implications.

doi:10.6342/NTU202500807

5. Conclusion

This study examines whether machine learning can improve the estimation and asset pricing relevance of fundamental betas—systematic risk measures that link firm fundamentals to aggregate economic indicators. Leveraging a comprehensive sample of U.S. firms from 1972 to 2022 and comparing traditional rolling regressions with machine learning approaches including Lasso, Ridge, and Elastic Net, we find that machine learning models consistently outperform traditional methods in predictive accuracy. In particular, Lasso and Elastic Net models deliver lower root mean squared errors (RMSEs) across the board and demonstrate robustness during periods of macroeconomic stress. However, the gains in statistical forecast performance do not fully translate into stronger economic value. The root mean squared hedging errors (RMSHEs) indicate limited improvements in return predictability, and Fama-MacBeth regressions show that fundamental betas—regardless of estimation method—carry weak and inconsistent pricing power in the cross-section of stock returns. These findings suggest that while machine learning enhances the modeling of firm-level earnings sensitivity to macroeconomic risk, it does not necessarily elevate the fundamental beta to a priced factor in asset markets. The results highlight a fundamental distinction between statistical performance and economic significance, offering a cautionary note for researchers and practitioners aiming to extract asset pricing insights from machine learning-generated fundamentals.

While this study demonstrates the potential of machine learning in improving fundamental beta forecasts, it is not without limitations. Because of computational constraints, the machine learning models employed in this paper—OLS, Lasso, Ridge, and Elastic Net—did not undergo rigorous hyperparameter tuning or cross-validation, thereby limiting the precision and generalizability of the forecasts. Moreover, this study exclusively focuses on linear models,

doi:10.6342/NTU202500807

omitting powerful nonlinear methods such as random forests, gradient boosting machines, or neural networks. These models, known for their ability to capture complex interactions and nonlinearities, may yield further improvements in both statistical and economic performance. Future research could extend this framework by incorporating advanced machine learning techniques, performing robust hyperparameter optimization, and utilizing high-performance computing environments to fully exploit the predictive potential of fundamental betas in asset pricing.

doi:10.6342/NTU202500807

References

Ajinkya, Bipin B., and Michael J. Gift. 1984. "Corporate Managers' Earnings Forecasts and Symmetrical Adjustments of Market Expectations." *Journal of Accounting Research* 22 (2): 425–44. https://doi.org/10.2307/2490657.

Ajinkya, Bipin, Sanjeev Bhojraj, and Partha Sengupta. 2005. "The Association between Outside Directors, Institutional Investors and the Properties of Management Earnings Forecasts." *Journal of Accounting Research* 43 (3): 343–76. https://doi.org/10.1111/j.1475-679x.2005.00174.x.

Akerlof, George A. 1970. "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism." *The Quarterly Journal of Economics* 84 (3): 488–500. https://doi.org/10.2307/1879431.

Akyildirim, Erdinc, Ahmet Goncu, and Ahmet Sensoy. 2021. "Prediction of Cryptocurrency Returns Using Machine Learning." *Annals of Operations Research* 297 (1): 3–36. https://doi.org/10.1007/s10479-020-03575-y.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Ginger Wu. 2006. "Realized Beta: Persistence and Predictability." In *Econometric Analysis of Financial and Economic Time Series*, 20 Part 2:1–39. Emerald Group Publishing Limited. https://doi.org/10.1016/S0731-9053(05)20020-8.

Arena, Claudia, Saverio Bozzolan, and Claudia Imperatore. 2023. "The Trade-Off Between Mandatory and Voluntary Disclosure: Evidence From Oil Companies' Risk Reporting." *Journal of Accounting, Auditing & Finance* 38 (4): 986–1008. https://doi.org/10.1177/0148558X211025250.

Arya, Anil, and Ram N.V. Ramanan. 2021. "Voluntary Disclosure in Light of Control Concerns." *Contemporary Accounting Research* 38 (4): 2824–50. https://doi.org/10.1111/1911-3846.12717.

Baginski, Stephen P., John M. Hassell, and Michael D. Kimbrough. 2002. "The Effect of Legal Environment on Voluntary Disclosure: Evidence from Management Earnings Forecasts Issued in U.S. and Canadian Markets." *The Accounting Review* 77 (1): 25–50. https://doi.org/10.2308/accr.2002.77.1.25.

Balakrishnan, Karthik, Mary Brooke Billings, Bryan Kelly, and Alexander Ljungqvist. 2014. "Shaping Liquidity: On the Causal Effects of Voluntary Disclosure." *The Journal of Finance* 69 (5): 2237–78. https://doi.org/10.1111/jofi.12180.

Bali, Turan G, Heiner Beckmeyer, Mathis Mörke, and Florian Weigert. 2023. "Option Return Predictability with Machine Learning and Big Data." *The Review of Financial Studies* 36 (9): 3548–3602. https://doi.org/10.1093/rfs/hhad017.

Ball, Ray, Sudarshan Jayaraman, and Lakshmanan Shivakumar. 2012. "Audited Financial Reporting and Voluntary Disclosure as Complements: A Test of the Confirmation Hypothesis." *Journal of Accounting and Economics* 53 (1): 136–66. https://doi.org/10.1016/j.jacceco.2011.11.005.

Ball, Ray, Gil Sadka, and Ayung Tseng. 2022. "Using Accounting Earnings and Aggregate Economic Indicators to Estimate Firm-Level Systematic Risk." *Review of Accounting Studies* 27 (2): 607–46. https://doi.org/10.1007/s11142-021-09594-9.

Bamber, Linda Smith, John (Xuefeng) Jiang, and Isabel Yanyan Wang. 2010. "What's My Style? The Influence of Top Managers on Voluntary Corporate Financial Disclosure." *The Accounting Review* 85 (4): 1131–62. https://doi.org/10.2308/accr.2010.85.4.1131.

Bansal, Ravi, Robert F. Dittmar, and Christian T. Lundblad. 2005. “Consumption, Dividends, and the Cross Section of Equity Returns.” *The Journal of Finance* 60 (4): 1639–72. https://doi.org/10.1111/j.1540-6261.2005.00776.x.

Bertrand, Marianne, and Sendhil Mullainathan. 2003. “Enjoying the Quiet Life? Corporate Governance and Managerial Preferences.” *Journal of Political Economy* 111 (5): 1043–75. https://doi.org/10.1086/376950.

Bianchi, Daniele, Matthias Büchner, and Andrea Tamoni. 2021. “Bond Risk Premiums with Machine Learning.” *The Review of Financial Studies* 34 (2): 1046–89. https://doi.org/10.1093/rfs/hhaa062.

Boone, Audra L., and Joshua T. White. 2015. “The Effect of Institutional Ownership on Firm Transparency and Information Production.” *Journal of Financial Economics* 117 (3): 508–33. https://doi.org/10.1016/j.jfineco.2015.05.008.

Botshekan, Mahmoud, Roman Kraeussl, and Andre Lucas. 2012. “Cash Flow and Discount Rate Risk in Up and Down Markets: What Is Actually Priced?” *Journal of Financial and Quantitative Analysis* 47 (6): 1279–1301. https://doi.org/10.1017/S0022109012000567.

Campbell, John Y. 1991. “A Variance Decomposition for Stock Returns.” *The Economic Journal* 101 (405): 157–79. https://doi.org/10.2307/2233809.

Campbell, John Y., Christopher Polk, and Tuomo Vuolteenaho. 2010. “Growth or Glamour? Fundamentals and Systematic Risk in Stock Returns.” *The Review of Financial Studies* 23 (1): 305–44. https://doi.org/10.1093/rfs/hhp029.

Campbell, John Y., and Robert J. Shiller. 1988. “The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors.” *The Review of Financial Studies* 1 (3): 195–228. https://doi.org/10.1093/rfs/1.3.195.

Campbell, John Y., and Tuomo Vuolteenaho. 2004. “Bad Beta, Good Beta.” *American Economic Review* 94 (5): 1249–75. https://doi.org/10.1257/0002828043052240.

Chen, Long, and Xinlei Zhao. 2009. “Return Decomposition.” *The Review of Financial Studies* 22 (12): 5213–49. https://doi.org/10.1093/rfs/hhp017.

Chen, Shuping, Xia Chen, and Qiang Cheng. 2008. “Do Family Firms Provide More or Less Voluntary Disclosure?” *Journal of Accounting Research* 46 (3): 499–536. https://doi.org/10.1111/j.1475-679X.2008.00288.x.

Chen, Shuping, Mark L. DeFond, and Chul W. Park. 2002. “Voluntary Disclosure of Balance Sheet Information in Quarterly Earnings Announcements.” *Journal of Accounting and Economics* 33 (2): 229–51. https://doi.org/10.1016/S0165-4101(02)00043-5.

Chen, Tao, Hui Dong, and Chen Lin. 2020. “Institutional Shareholders and Corporate Social Responsibility.” *Journal of Financial Economics* 135 (2): 483–504. https://doi.org/10.1016/j.jfineco.2019.06.007.

Chen, Yi-Chun, Mingyi Hung, and Yongxiang Wang. 2018. “The Effect of Mandatory CSR Disclosure on Firm Profitability and Social Externalities: Evidence from China.” *Journal of Accounting and Economics* 65:169–90.

Cheng, Qiang, Young Jun Cho, and Jae B. Kim. 2021. “Managers’ Pay Duration and Voluntary Disclosures.” *Journal of Business Finance & Accounting* 48 (7–8): 1332–67. https://doi.org/10.1111/jbfa.12516.

Christensen, Hans B., Luzi Hail, and Christian Leuz. 2021. “Mandatory CSR and Sustainability Reporting: Economic Analysis and Literature Review.” *Review of Accounting Studies* 26 (3): 1176–1248. https://doi.org/10.1007/s11142-021-09609-5.

Cosemans, Mathijs, Rik Frehen, Peter C. Schotman, and Rob Bauer. 2016. "Estimating Security Betas Using Prior Information Based on Firm Fundamentals." *The Review of Financial Studies* 29 (4): 1072–1112. https://doi.org/10.1093/rfs/hhv131.

Deng, Lu, Mingqing Liao, Rui Luo, Jianfei Sun, and Chen Xu. 2021. "Does Corporate Social Responsibility Reduce Share Price Premium? Evidence from China's A- and H-Shares." *Pacific-Basin Finance Journal* 67 (June):101569. https://doi.org/10.1016/j.pacfin.2021.101569.

Dhaliwal, Dan S., Suresh Radhakrishnan, Albert Tsang, and Yong George Yang. 2012. "Nonfinancial Disclosure and Analyst Forecast Accuracy: International Evidence on Corporate Social Responsibility Disclosure." *The Accounting Review* 87 (3): 723–59. https://doi.org/10.2308/accr-10218.

Diamond, Douglas W., and Robert E. Verrecchia. 1991. "Disclosure, Liquidity, and the Cost of Capital." *The Journal of Finance* 46 (4): 1325–59. https://doi.org/10.1111/j.1540-6261.1991.tb04620.x.

Drobetz, Wolfgang, Fabian Hollstein, Tizian Otto, and Marcel Prokopczuk. 2024. "Estimating Stock Market Betas via Machine Learning." *Journal of Financial and Quantitative Analysis*, February, 1–37. https://doi.org/10.1017/S0022109024000036.

Dyck, Alexander, Karl V. Lins, Lukas Roth, and Hannes F. Wagner. 2019. "Do Institutional Investors Drive Corporate Social Responsibility? International Evidence." *Journal of Financial Economics* 131 (3): 693–714. https://doi.org/10.1016/j.jfineco.2018.08.013.

Dye, Ronald A. 1985. "Disclosure of Nonproprietary Information." *Journal of Accounting Research* 23 (1): 123–45. https://doi.org/10.2307/2490910.

Eccles, Robert G., and George Serafeim. 2013. "A Tale of Two Stories: Sustainability and the Quarterly Earnings Call." *Journal of Applied Corporate Finance* 25 (3): 8–19. https://doi.org/10.1111/jacf.12023.

Einhorn, Eti. 2005. "The Nature of the Interaction between Mandatory and Voluntary Disclosures." *Journal of Accounting Research* 43 (4): 593–621. https://doi.org/10.1111/j.1475-679X.2005.00183.x.

Ellahie, Atif. 2021. "Earnings Beta." *Review of Accounting Studies* 26 (1): 81–122. https://doi.org/10.1007/s11142-020-09561-w.

Fama, Eugene F., and James D. MacBeth. 1973. "Risk, Return, and Equilibrium: Empirical Tests." *Journal of Political Economy* 81 (3): 607–36. https://doi.org/10.1086/260061.

Fonseka, Mohan, Theja Rajapakse, and Grant Richardson. 2019. "The Effect of Environmental Information Disclosure and Energy Product Type on the Cost of Debt: Evidence from Energy Firms in China." *Pacific-Basin Finance Journal* 54 (April):159–82. https://doi.org/10.1016/j.pacfin.2018.05.001.

Gelb, David S. 2000. "Managerial Ownership and Accounting Disclosures: An Empirical Study." *Review of Quantitative Finance and Accounting* 15 (2): 169–85. https://doi.org/10.1023/A:1008321230900.

Gelb, Jay, Rob McCarthy, Werner Rehm, and Andrey Voronin. 2023. "Investors Want to Hear from Companies about the Value of Sustainability." McKinsey & Company. September 15, 2023. https://www.mckinsey.com/capabilities/strategy-and-corporate-finance/our-insights/investors-want-to-hear-from-companies-about-the-value-of-sustainability.

Gigler, Frank. 1994. "Self-Enforcing Voluntary Disclosures." *Journal of Accounting Research* 32 (2): 224–40. https://doi.org/10.2307/2491283.

Grewal, Jody, Edward J. Riedl, and George Serafeim. 2019. “Market Reaction to Mandatory Nonfinancial Disclosure.” *Management Science* 65 (7): 3061–84. https://doi.org/10.1287/mnsc.2018.3099.

Grossman, Sanford J. 1981. “The Informational Role of Warranties and Private Disclosure about Product Quality.” *The Journal of Law and Economics* 24 (3): 461–83. https://doi.org/10.1086/466995.

Grossman, Sanford J., and Oliver D. Hart. 1980. “Disclosure Laws and Takeover Bids.” *The Journal of Finance* 35 (2): 323–34. https://doi.org/10.2307/2327390.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu. 2020. “Empirical Asset Pricing via Machine Learning.” *The Review of Financial Studies* 33 (5): 2223–73. https://doi.org/10.1093/rfs/hhaa009.

Guan, Yuyan, Gerald J. Lobo, Albert Tsang, and Xiangang Xin. 2020. “Societal Trust and Management Earnings Forecasts.” *The Accounting Review* 95 (5): 149–84. https://doi.org/10.2308/tar-2017-0023.

Guay, Wayne, Delphine Samuels, and Daniel Taylor. 2016. “Guiding through the Fog: Financial Statement Complexity and Voluntary Disclosure.” *Journal of Accounting and Economics*, Conference papers 2015, 62 (2): 234–69. https://doi.org/10.1016/j.jacceco.2016.09.001.

Hambrick, Donald C., and Phyllis A. Mason. 1984. “Upper Echelons: The Organization as a Reflection of Its Top Managers.” *The Academy of Management Review* 9 (2): 193–206. https://doi.org/10.2307/258434.

Hansen, Jacob H., and Mathias V. Siggaard. 2024. “Double Machine Learning: Explaining the Post-Earnings Announcement Drift.” *Journal of Financial and Quantitative Analysis* 59 (3): 1003–30. https://doi.org/10.1017/S0022109023000133.

Hansen, Lars Peter, John C. Heaton, and Nan Li. 2008. “Consumption Strikes Back? Measuring Long-Run Risk.” *Journal of Political Economy* 116 (2): 260–302. https://doi.org/10.1086/588200.

Hansen, Peter Reinhard, and Asger Lunde. 2006. “Consistent Ranking of Volatility Models.” *Journal of Econometrics* 131 (1): 97–121. https://doi.org/10.1016/j.jeconom.2005.01.005.

Hao, Jinji. 2023. “Disclosure Regulation, Cost of Capital, and Firm Values.” *Journal of Accounting and Economics*, May, 101605. https://doi.org/10.1016/j.jacceco.2023.101605.

Heinle, Mirko, Delphine Samuels, and Daniel Taylor. 2023. “Disclosure Substitution.” *Management Science* 69 (8): 4774–89. https://doi.org/10.1287/mnsc.2022.4549.

Hollstein, Fabian, and Marcel Prokopczuk. 2016. “Estimating Beta.” *Journal of Financial and Quantitative Analysis* 51 (4): 1437–66. https://doi.org/10.1017/S0022109016000508.

Holmström, Bengt. 1979. “Moral Hazard and Observability.” *The Bell Journal of Economics* 10 (1): 74–91. https://doi.org/10.2307/3003320.

Hribar, Paul, Richard Mergenthaler, Aaron Roeschley, Spencer Young, and Chris X. Zhao. 2022. “Do Managers Issue More Voluntary Disclosure When GAAP Limits Their Reporting Discretion in Financial Statements?” *Journal of Accounting Research* 60 (1): 299–351. https://doi.org/10.1111/1475-679X.12401.

IFRS. 2024. “About the International Sustainability Standards Board.” IFRS - International Sustainability Standards Board. 2024. https://www.ifrs.org/groups/international-sustainability-standards-board/.

Ioannou, Ioannis, and George Serafeim. 2017. “The Consequences of Mandatory Corporate Sustainability Reporting.” SSRN Scholarly Paper. Rochester, NY. https://papers.ssrn.com/abstract=1799589.

doi:10.6342/NTU202500807

Jensen, Michael C., and William H. Meckling. 1976. “Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure.” *Journal of Financial Economics* 3 (4): 305–60. https://doi.org/10.1016/0304-405X(76)90026-X.

Jovanovic, Boyan. 1982. “Truthful Disclosure of Information.” *The Bell Journal of Economics* 13 (1): 36–44. https://doi.org/10.2307/3003428.

Jung, Woon-Oh, and Young K. Kwon. 1988. “Disclosure When the Market Is Unsure of Information Endowment of Managers.” *Journal of Accounting Research* 26 (1): 146–53. https://doi.org/10.2307/2491118.

Kim, Jung Min, Daniel J. Taylor, and Robert E. Verrecchia. 2021. “Voluntary Disclosure When Private Information and Disclosure Costs Are Jointly Determined.” *Review of Accounting Studies* 26 (3): 971–1001. https://doi.org/10.1007/s11142-021-09601-z.

Kothari, S. P., Andrew J. Leone, and Charles E. Wasley. 2005. “Performance Matched Discretionary Accrual Measures.” *Journal of Accounting and Economics* 39 (1): 163–97. https://doi.org/10.1016/j.jacceco.2004.11.002.

Lang, Mark, and Russell Lundholm. 1993. “Cross-Sectional Determinants of Analyst Ratings of Corporate Disclosures.” *Journal of Accounting Research* 31 (2): 246–71. https://doi.org/10.2307/2491273.

Lee, Yoojin, Shaphan Ng, Terry Shevlin, and Aruhn Venkat. 2021. “The Effects of Tax Avoidance News on Employee Perceptions of Managers and Firms: Evidence from Glassdoor.Com Ratings.” *The Accounting Review* 96 (3): 343–72. https://doi.org/10.2308/TAR-2019-0148.

Leippold, Markus, Qian Wang, and Wenyu Zhou. 2022. “Machine Learning in the Chinese Stock Market.” *Journal of Financial Economics* 145 (2, Part A): 64–82. https://doi.org/10.1016/j.jfineco.2021.08.017.

Li, Wenjing, Jeff Ng, Albert Tsang, and Oktay Urcan. 2019. “Country-Level Institutions and Management Earnings Forecasts.” *Journal of International Business Studies* 50 (1): 48–82. https://doi.org/10.1057/s41267-018-0190-4.

Li, Xi, and Holly I. Yang. 2016. “Mandatory Financial Reporting and Voluntary Disclosure: The Effect of Mandatory IFRS Adoption on Management Forecasts.” *The Accounting Review* 91 (3): 933–53. https://doi.org/10.2308/accr-51296.

Maio, Paulo. 2013. “Intertemporal CAPM with Conditioning Variables.” *Management Science* 59 (1): 122–41. https://doi.org/10.1287/mnsc.1120.1557.

Manchiraju, Hariom, and Shivaram Rajgopal. 2017. “Does Corporate Social Responsibility (CSR) Create Shareholder Value? Evidence from the Indian Companies Act 2013.” *Journal of Accounting Research* 55 (5): 1257–1300. https://doi.org/10.1111/1475-679X.12174.

Meek, Gary K., Clare B. Roberts, and Sidney J. Gray. 1995. “Factors Influencing Voluntary Annual Report Disclosures By U.S., U.K. and Continental European Multinational Corporations.” *Journal of International Business Studies* 26 (3): 555–72. https://doi.org/10.1057/palgrave.jibs.8490186.

Milgrom, Paul R. 1981. “Good News and Bad News: Representation Theorems and Applications.” *The Bell Journal of Economics* 12 (2): 380–91. https://doi.org/10.2307/3003562.

Nagar, Venky, Dhananjay Nanda, and Peter Wysocki. 2003. “Discretionary Disclosure and Stock-Based Incentives.” *Journal of Accounting and Economics* 34 (1): 283–309. https://doi.org/10.1016/S0165-4101(02)00075-7.

Nair, Rajiv, Mohammad Muttakin, Arifur Khan, Nava Subramaniam, and V. S. Somanath. 2019. “Corporate Social Responsibility Disclosure and Financial Transparency: Evidence from

India." *Pacific-Basin Finance Journal* 56 (September):330–51. https://doi.org/10.1016/j.pacfin.2019.06.015.

Noh, Suzie, Eric C. So, and Joseph P. Weber. 2019. "Voluntary and Mandatory Disclosures: Do Managers View Them as Substitutes?" *Journal of Accounting and Economics* 68 (1): 101243. https://doi.org/10.1016/j.jacceco.2019.101243.

Patton, Andrew J. 2011. "Volatility Forecast Comparison Using Imperfect Volatility Proxies." *Journal of Econometrics*, Realized Volatility, 160 (1): 246–56. https://doi.org/10.1016/j.jeconom.2010.03.034.

Penman, Stephen. 2016. "Valuation: Accounting for Risk and the Expected Return." *Abacus* 52 (1): 106–30. https://doi.org/10.1111/abac.12067.

Penman, Stephen, and Xiao-Jun Zhang. 2020. "A Theoretical Analysis Connecting Conservative Accounting to the Cost of Capital." *Journal of Accounting and Economics* 69 (1): 101236. https://doi.org/10.1016/j.jacceco.2019.101236.

Pownall, Grace, and Gregory Waymire. 1989. "Voluntary Disclosure Credibility and Securities Prices: Evidence from Management Earnings Forecasts, 1969-73." *Journal of Accounting Research* 27 (2): 227–45. https://doi.org/10.2307/2491233.

Shi, Haina, Byron Y. Song, Huifeng Xu, and Xiaodong Xu. 2023. "Mandatory CSR Disclosure and Analyst Forecast Properties: Evidence from a Quasi-Natural Experiment in China." *China Journal of Accounting Research* 16 (2): 100301. https://doi.org/10.1016/j.cjar.2023.100301.

Spence, Michael. 1973. "Job Market Signaling." *The Quarterly Journal of Economics* 87 (3): 355–74. https://doi.org/10.2307/1882010.

Tomlinson, Brian, Tensie Whelan, and Kevin Eckerle. 2021. "How to Bring ESG Into the Quarterly Earnings Call." *MIT Sloan Management Review*, May 10, 2021. https://sloanreview.mit.edu/article/how-to-bring-esg-into-the-quarterly-earnings-call/.

Tsang, Albert, Kun Tracy Wang, Yue Wu, and Jeff Lee. 2022. "Nonfinancial Corporate Social Responsibility Reporting and Firm Value: International Evidence on the Role of Financial Analysts." *European Accounting Review* 0 (0): 1–36. https://doi.org/10.1080/09638180.2022.2094435.

Tuo, Ling, Zabihollah Rezaee, and Lei Gao. 2024. "Is There a Tradeoff between Management Earnings Forecasts and Sustainability Reporting?" *Global Finance Journal* 59 (March):100908. https://doi.org/10.1016/j.gfj.2023.100908.

Verrecchia, Robert E. 1983. "Discretionary Disclosure." *Journal of Accounting and Economics* 5 (January):179–94. https://doi.org/10.1016/0165-4101(83)90011-3.

———. 2001. "Essays on Disclosure." *Journal of Accounting and Economics* 32 (1): 97–180. https://doi.org/10.1016/S0165-4101(01)00025-8.

Vuolteenaho, Tuomo. 2002. "What Drives Firm-Level Stock Returns?" *The Journal of Finance* 57 (1): 233–64. https://doi.org/10.1111/1540-6261.00421.

Wang, Xue, and Qingyuan Liu. 2024. "Information Disclosure and ESG Rating Disagreement: Evidence from Green Bond Issuance in China." *Pacific-Basin Finance Journal* 85 (June):102350. https://doi.org/10.1016/j.pacfin.2024.102350.

Waymire, Gregory. 1984. "Additional Evidence on the Information Content of Management Earnings Forecasts." *Journal of Accounting Research* 22 (2): 703–18. https://doi.org/10.2307/2490672.

Xia, Yanchun, Huan Zhang, and Shijun Guo. 2024. "Mandatory Internal Control Audits and Management Earnings Forecast." *Pacific-Basin Finance Journal* 85 (June):102362. https://doi.org/10.1016/j.pacfin.2024.102362.

Xue, Shuyu, Qi Chang, and Jingwen Xu. 2023. "The Effect of Voluntary and Mandatory Corporate Social Responsibility Disclosure on Firm Profitability: Evidence from China." *Pacific-Basin Finance Journal* 77 (February):101919. https://doi.org/10.1016/j.pacfin.2022.101919.

Xue, Shuyu, Huilin Wu, Yishu Ling, and Ye Lu. 2024. "Mandatory CSR Disclosure and Stock Liquidity: Evidence from Chinese Listed Firms." *Finance Research Letters* 59 (January):104817. https://doi.org/10.1016/j.frl.2023.104817.

Zhang, Qiyu, Rong Ding, Ding Chen, and Xiaoxiang Zhang. 2023. "The Effects of Mandatory ESG Disclosure on Price Discovery Efficiency around the World." *International Review of Financial Analysis* 89 (October):102811. https://doi.org/10.1016/j.irfa.2023.102811.

Figures and Tables

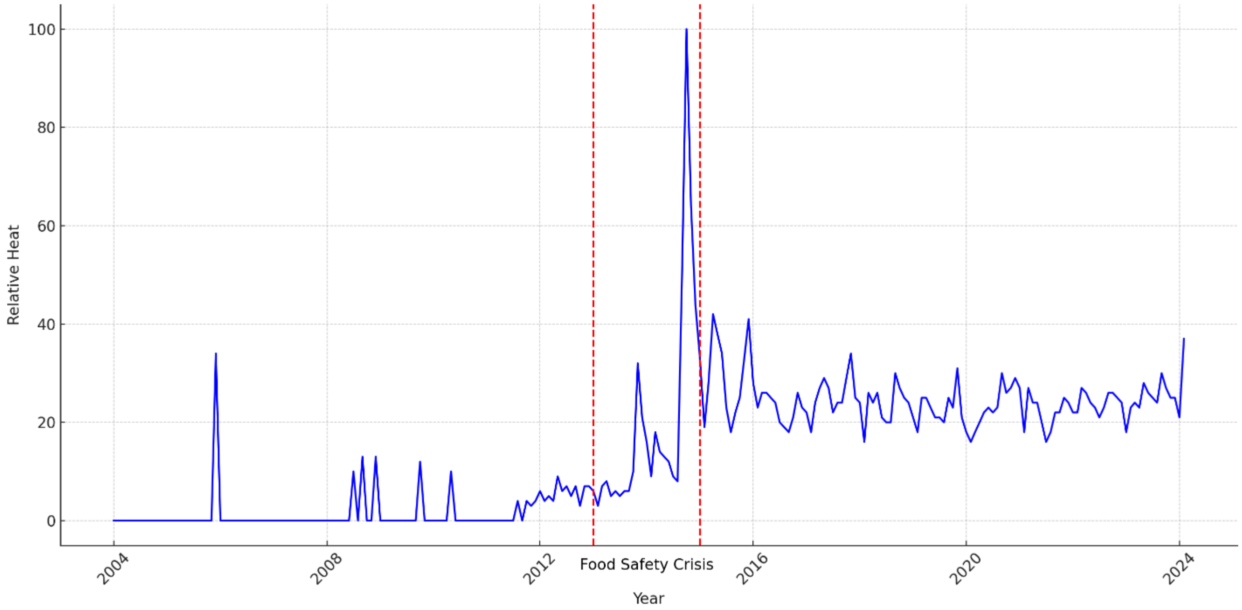

Figure 1: Time Series of Relative Heat on the Keyword “食安” in Google Searches

This figure illustrates the time series of the relative heat of the keyword “食安” on the Google search engine over the years. The highest heat is standardized as 100. The left red dashed line indicates the beginning of 2013. The right one indicates the end of 2015. Between 2013 and 2015, the food safety crisis occurred in Taiwan.

doi:10.6342/NTU202500807

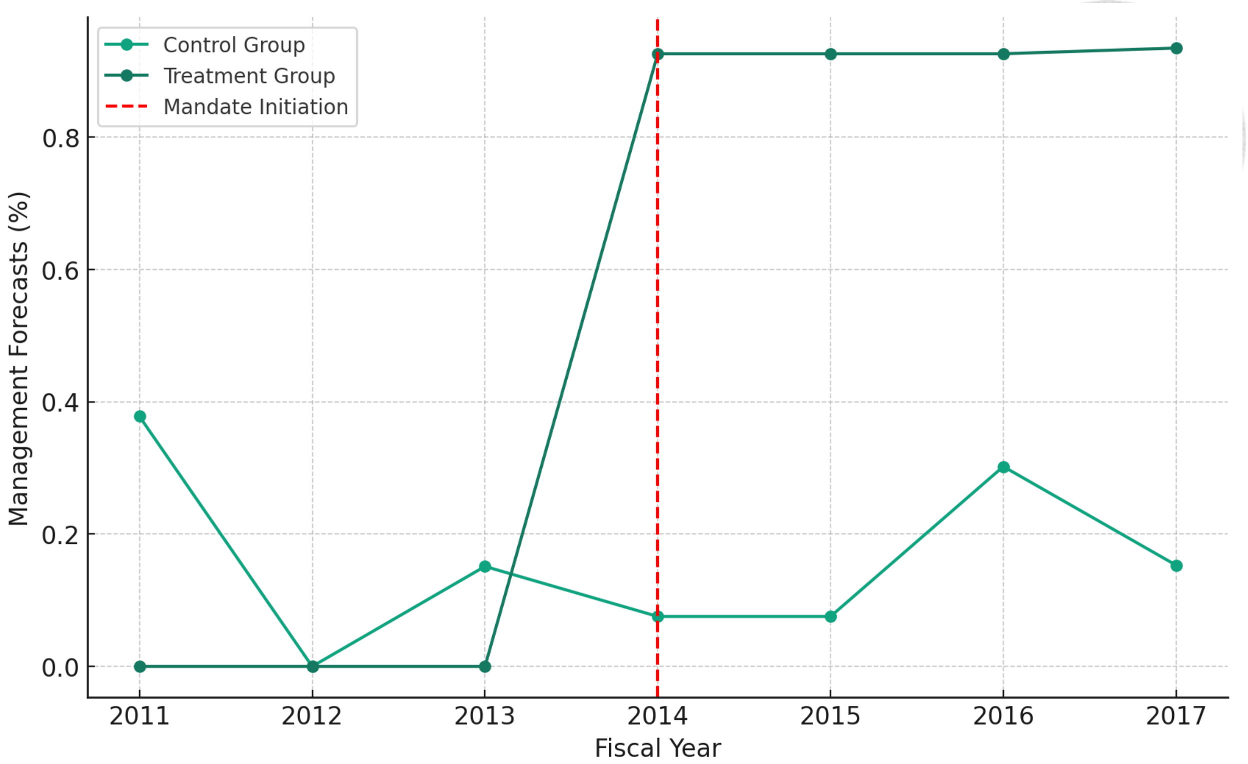

Figure 2: Time Series of the Ratios of Firms that Issue Management Forecasts Every Year

This figure illustrates the time series of the ratios of firms that voluntarily disclose financial information over the years. The dark line plots the time series of the treated firms. The light line plots the time series of the control firms. The red dashed line indicates the fiscal year 2014, when firms are mandated to disclose CSR reports in Taiwan. The sample period is 2011-2017.

doi:10.6342/NTU202500807

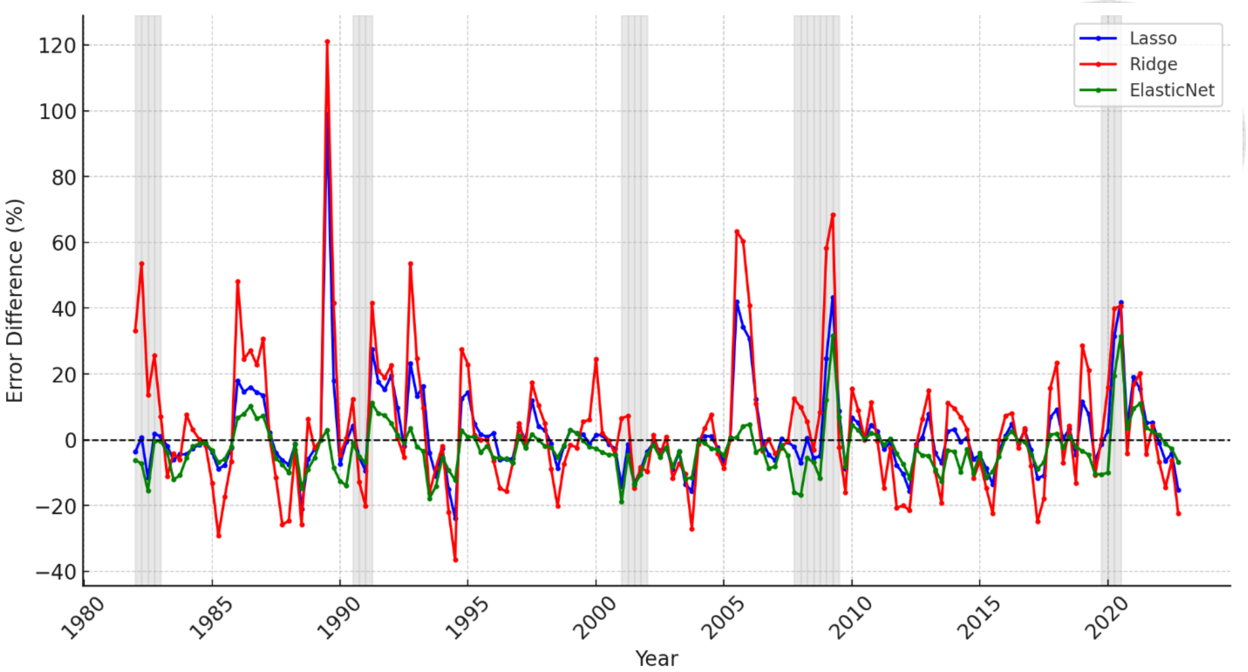

Figure 3: Relative Forecast Errors (RMSEs) of EBETA Over Time

This figure shows the time series of the relative quarterly market-capitalization-weighted root mean squared errors (RMSEs) of the EBETA estimated from the Lasso, Ridge, or Elastic Net models. The relative errors are calculated as the difference between a machine learning model's RMSEs and the Rolling Regression's RMSEs divided by the Rolling Regression's RMSEs. The grey areas indicate the recession periods defined by the National Bureau of Economic Research (NBER). The sample period is 1972Q1 – 2022Q4. The descriptions of the models are in Table 14. The variable definitions are in Table 15.

doi:10.6342/NTU202500807

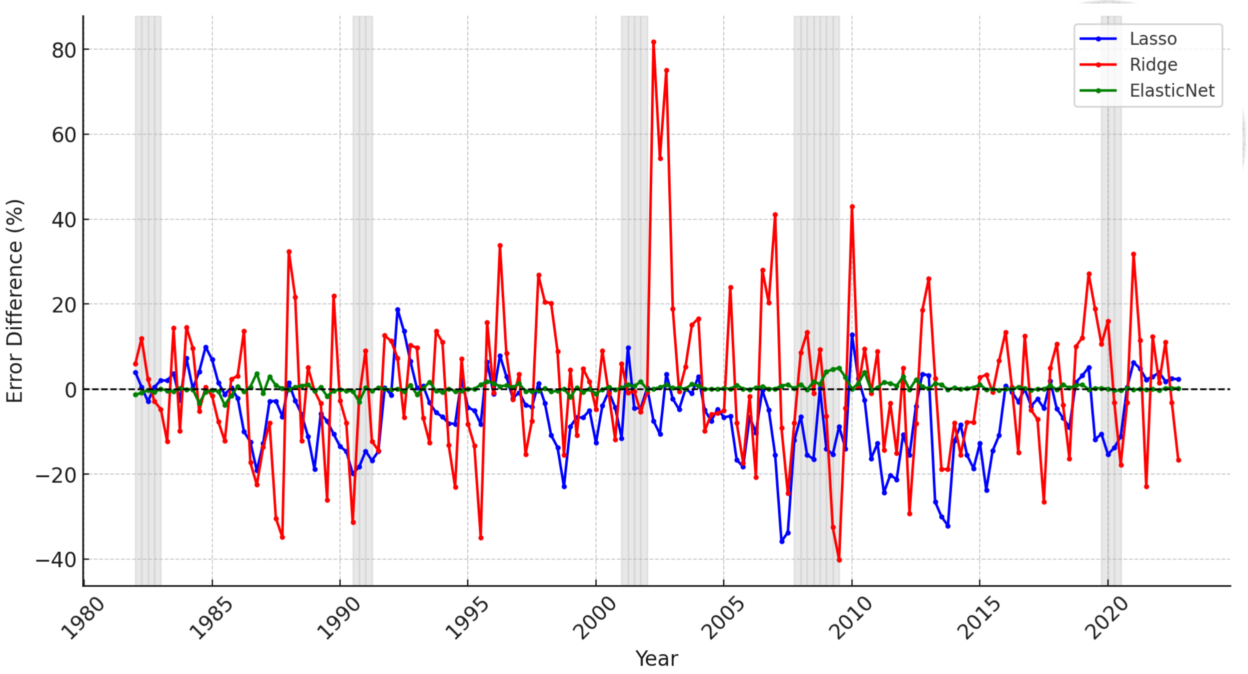

Figure 4: Relative Forecast Errors (RMSEs) of TBETA Over Time

This figure shows the time series of the relative quarterly market-capitalization-weighted root mean squared errors (RMSEs) of the TBETA estimated from the Lasso, Ridge, or Elastic Net models. The relative errors are calculated as the difference between a machine learning model's RMSEs and the Rolling Regression's RMSEs divided by the Rolling Regression's RMSEs. The grey areas indicate the recession periods defined by the National Bureau of Economic Research (NBER). The sample period is 1972Q1 – 2022Q4. The descriptions of the models are in Table 14. The variable definitions are in Table 15.

doi:10.6342/NTU202500807

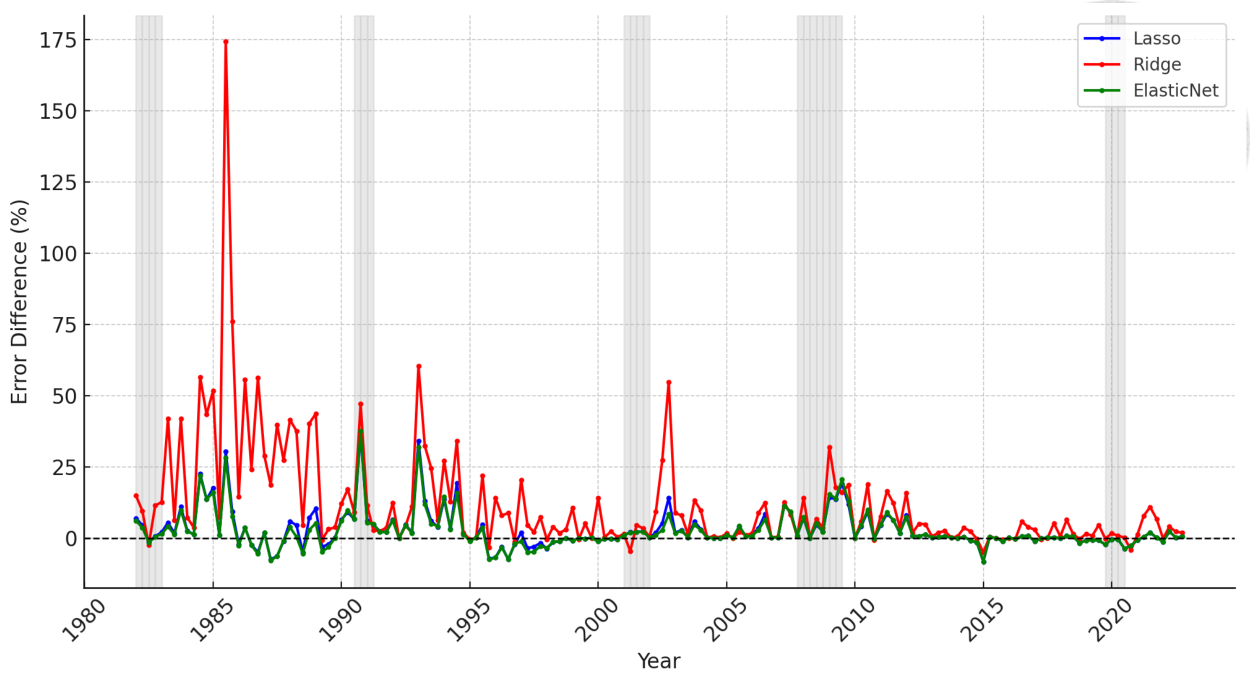

Figure 5: Relative Forecast Errors (RMSEs) of WBETA Over Time

This figure shows the time series of the relative quarterly market-capitalization-weighted root mean squared errors (RMSEs) of the WBETA estimated from the Lasso, Ridge, or Elastic Net models. The relative errors are calculated as the difference between a machine learning model's RMSEs and the Rolling Regression's RMSEs divided by the Rolling Regression's RMSEs. The grey areas indicate the recession periods defined by the National Bureau of Economic Research (NBER). The sample period is 1972Q1 – 2022Q4. The descriptions of the models are in Table 14. The variable definitions are in Table 15.

doi:10.6342/NTU202500807

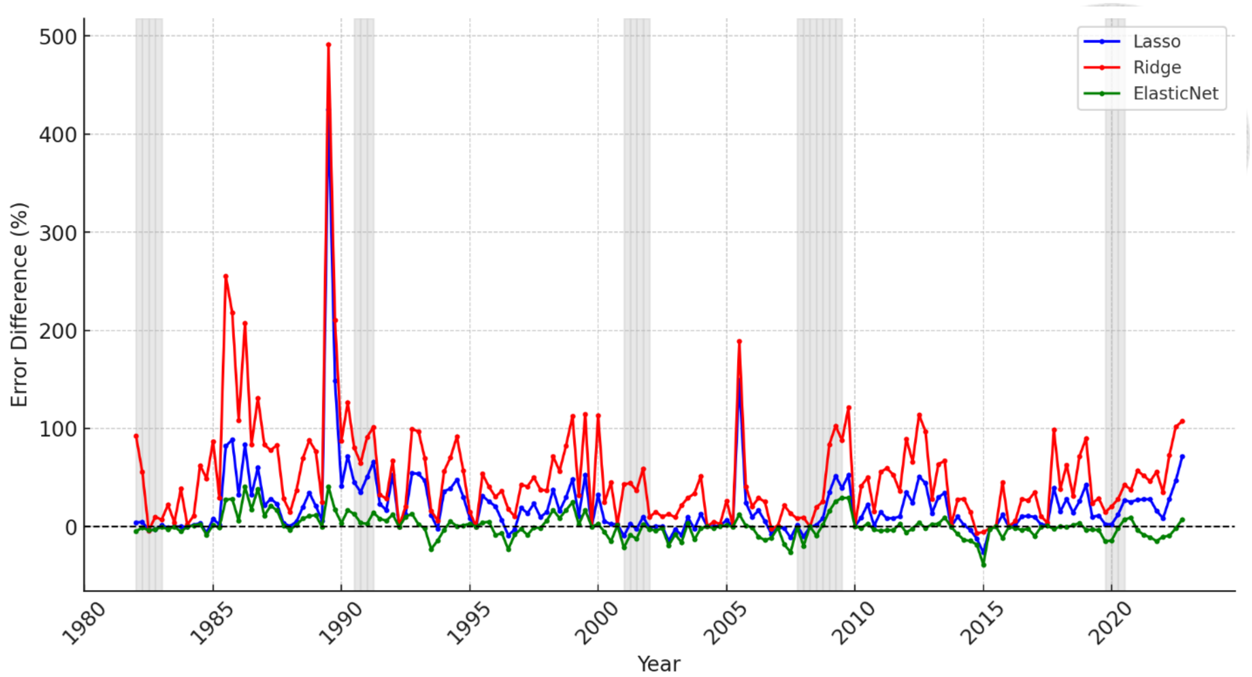

Figure 6: Relative Forecast Errors (RMSHEs) of EBETA Over Time

This figure shows the time series of the relative quarterly market-capitalization-weighted root mean squared hedging errors (RMSHEs) of the EBETA estimated from the Lasso, Ridge, or Elastic Net models. The relative errors are calculated as the difference between a machine learning model's RMSHEs and the Rolling Regression's RMSHEs divided by the Rolling Regression's RMSHEs. The grey areas indicate the recession periods defined by the National Bureau of Economic Research (NBER). The sample period is 1972Q1 – 2022Q4. The descriptions of the models are in Table 14. The variable definitions are in Table 15.

doi:10.6342/NTU202500807

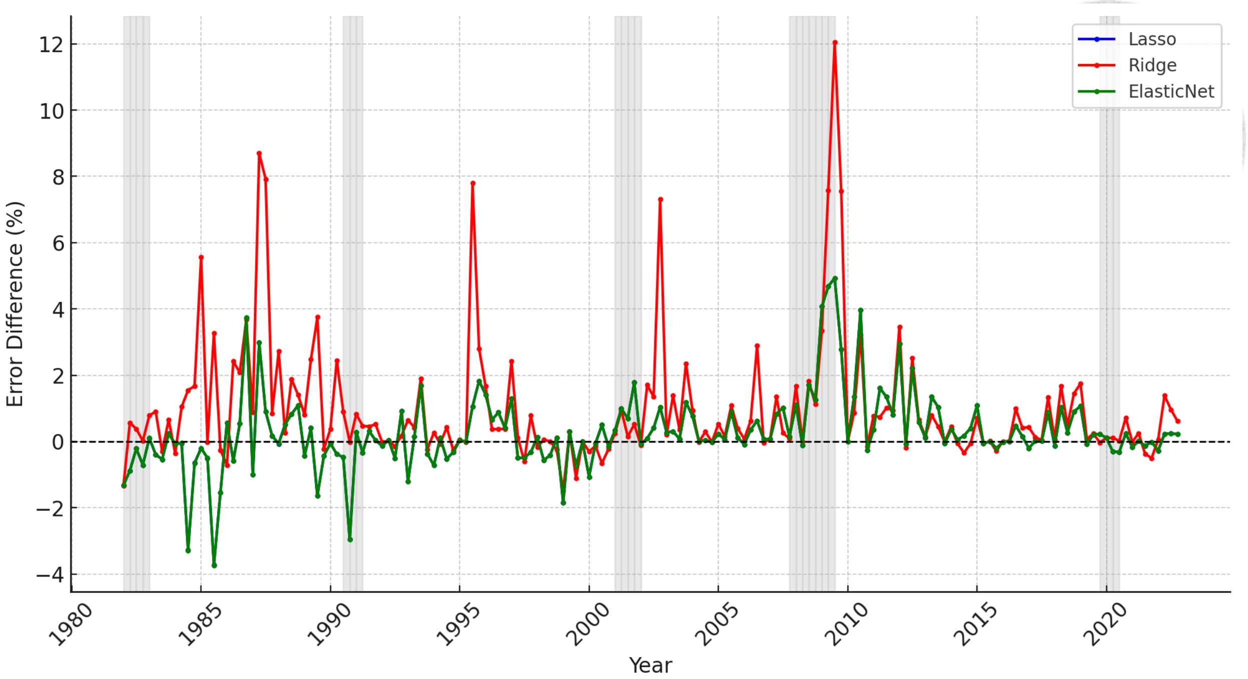

Figure 7: Relative Forecast Errors (RMSHEs) of TBETA Over Time

This figure shows the time-series of the relative quarterly market-capitalization-weighted root mean squared hedging errors (RMSHEs) of the TBETA estimated from the Lasso, Ridge, or Elastic Net models. The relative errors are calculated as the difference between a machine learning model's RMSHEs and the Rolling Regression's RMSHEs divided by the Rolling Regression's RMSHEs. The grey areas indicate the recession periods defined by the National Bureau of Economic Research (NBER). Note that the time series of the relative forecast errors of the Lasso model is so close to the series of the errors of the Elastic model that they are indistinguishable in the figure. The sample period is 1972Q1 – 2022Q4. The descriptions of the models are in Table 14. The variable definitions are in Table 15.

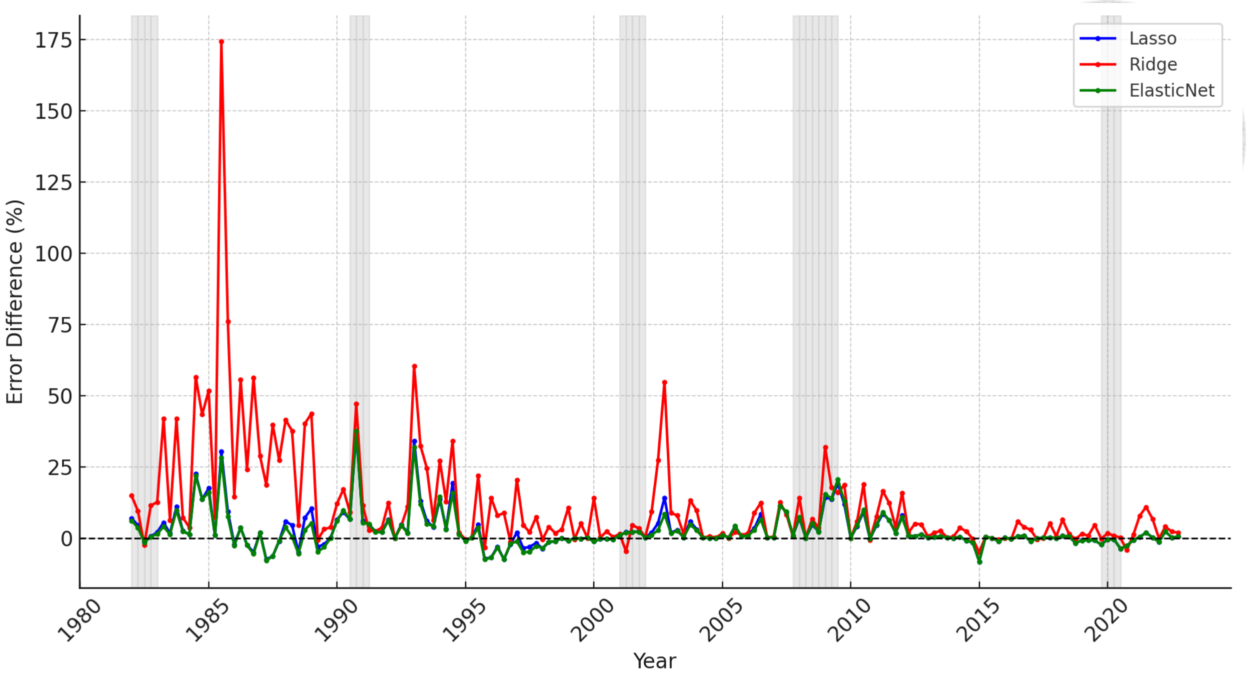

Figure 8: Relative Forecast Errors (RMSHEs) of WBETA Over Time

This figure shows the time-series of the relative quarterly market-capitalization-weighted root mean squared hedging errors (RMSHEs) of the WBETA estimated from the Lasso, Ridge, or Elastic Net models. The relative errors are calculated as the difference between a machine learning model's RMSHEs and the Rolling Regression's RMSHEs divided by the Rolling Regression's RMSHEs. The grey areas indicate the recession periods defined by the National Bureau of Economic Research (NBER). The sample period is 1972Q1 – 2022Q4. The descriptions of the models are in Table 14. The variable definitions are in Table 15.

doi:10.6342/NTU202500807

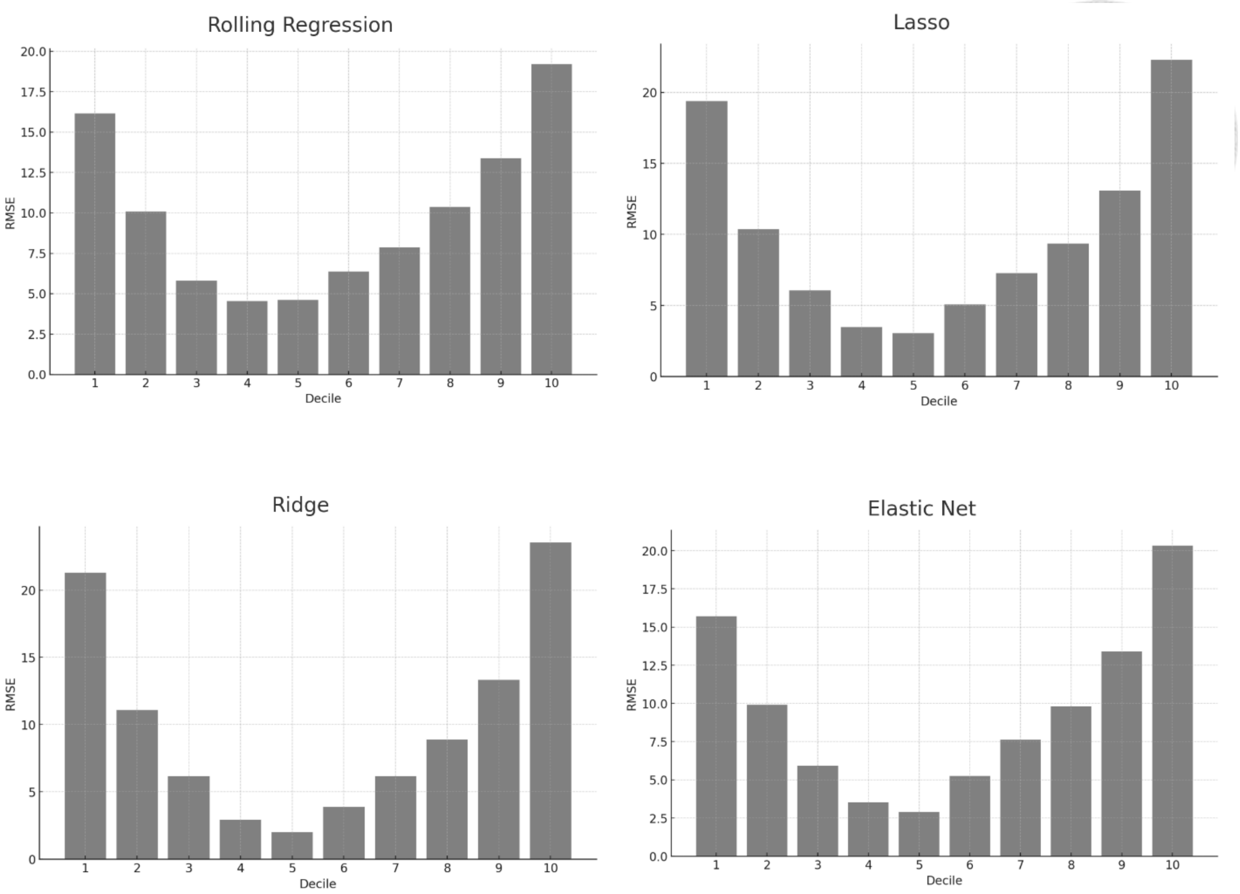

Figure 9: Average Forecast Errors (RMSEs) of Portfolio Sorts Based on EBETA Estimates

This figure shows the time-series average of the quarterly market-capitalization-weighted root mean squared errors (RMSEs) in each portfolio. In each quarter, the firms are allocated into decile portfolios according to the estimated EBETAs. The sample period is 1972Q1 – 2022Q4. The descriptions of the models are in Table 14. The variable definitions are in Table 15.

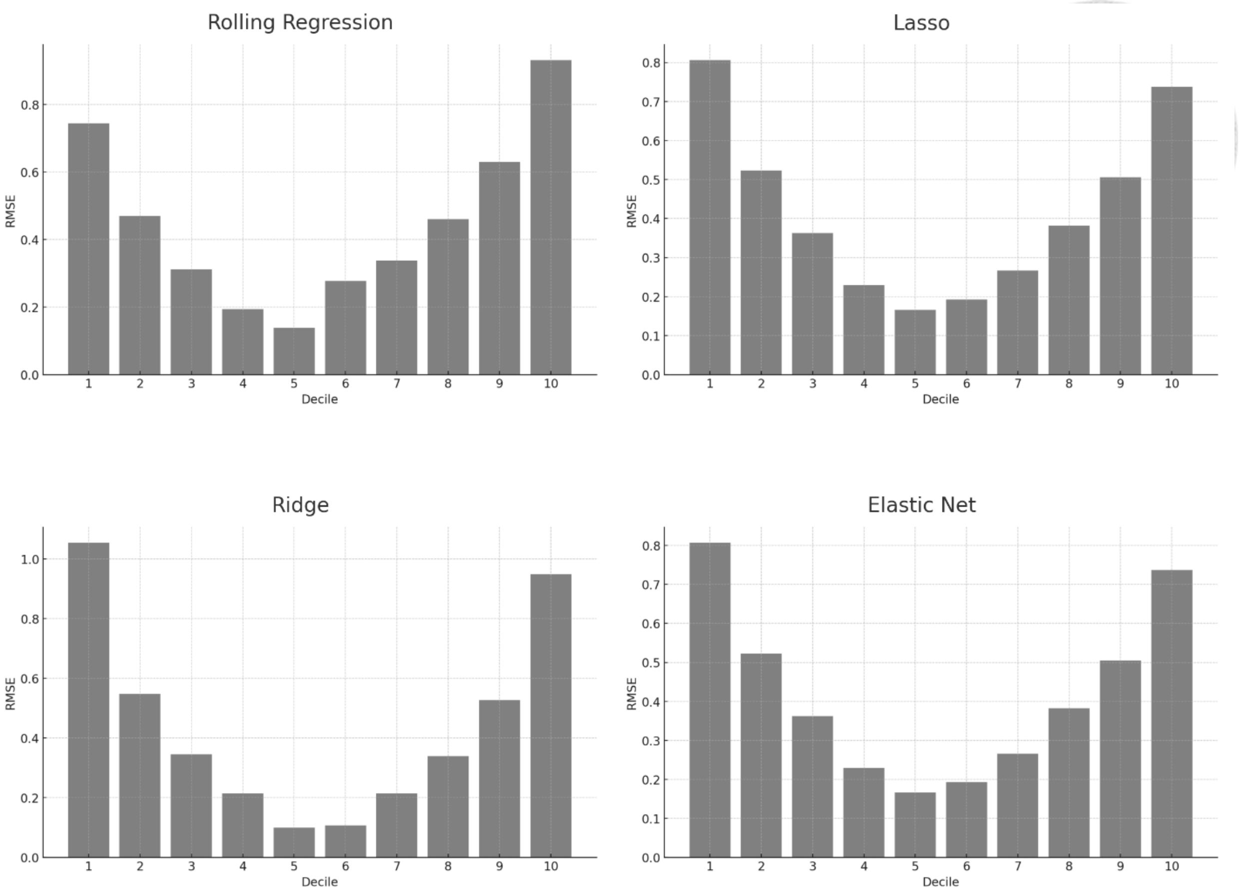

Figure 10: Average Forecast Errors (RMSEs) of Portfolio Sorts Based on TBETA Estimates

This figure shows the time-series average of the quarterly market-capitalization-weighted root mean squared errors (RMSEs) in each portfolio. In each quarter, the firms are allocated into decile portfolios according to the estimated TBETAs. The sample period is 1972Q1 – 2022Q4. The descriptions of the models are in Table 14. The variable definitions are in Table 15.

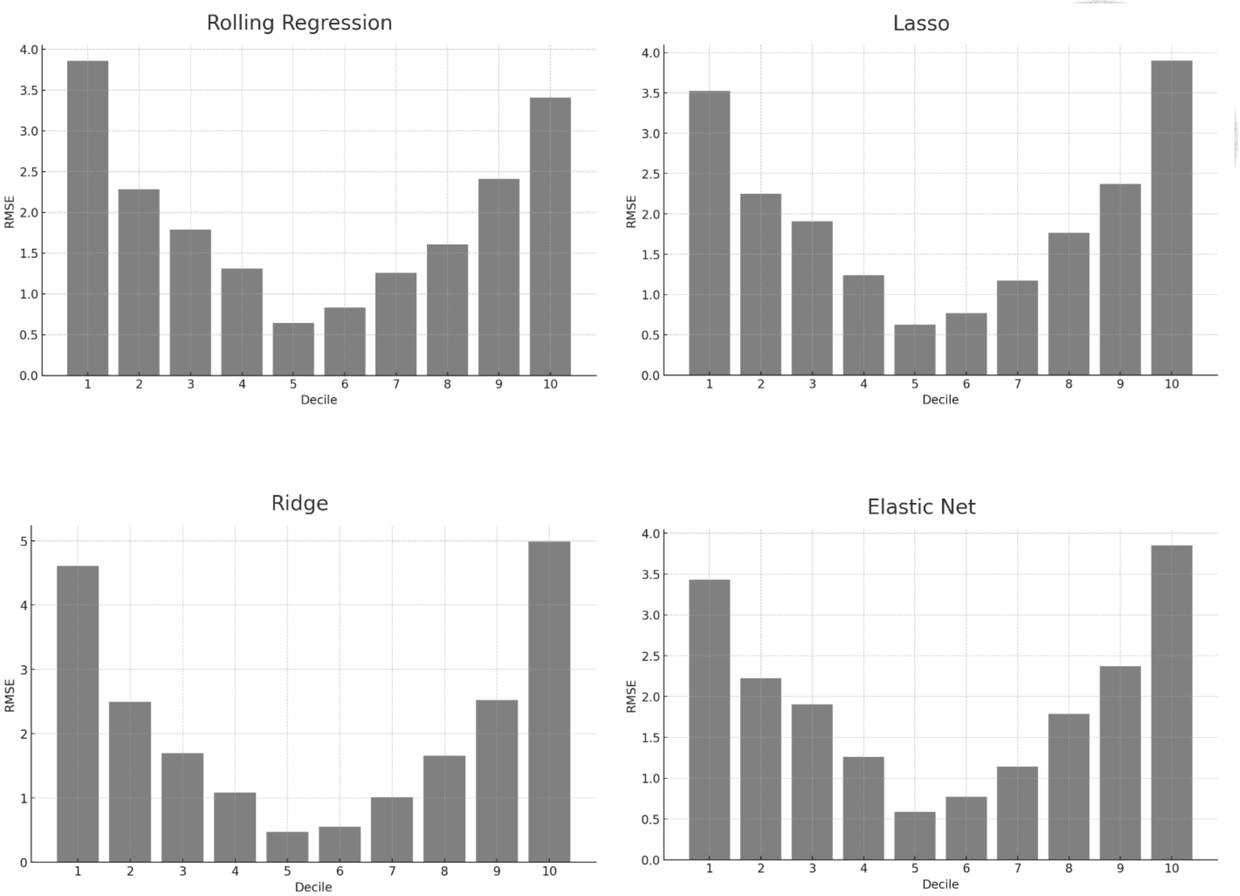

Figure 11: Average Forecast Errors (RMSEs) of Portfolio Sorts Based on WBETA Estimates

This figure shows the time-series average of the quarterly market-capitalization-weighted root mean squared errors (RMSEs) in each portfolio. In each quarter, the firms are allocated into decile portfolios according to the estimated WBETAs. The sample period is 1972Q1 – 2022Q4. The descriptions of the models are in Table 14. The variable definitions are in Table 15.

Table 1: Variable Definitions in Chapter 1

This table shows the names, definitions, and sources of the variables used in Chapter 1.

Variable Name	Definition	Source
$\mathrm{MF}_{\mathrm{it}}$	An indicator equal to one if the manager of firm i provides at least one earnings forecast in fiscal year t and zero otherwise	I/B/E/S
NMF_{it}	The number of earnings forecasts the manager of firm i provides in fiscal year t	I/B/E/S
$\mathrm{TREA}T_i$	An indicator equal to one if the firm i in fiscal year 2014 is required to disclose a stand-alone CSR report and zero otherwise	TEJ
$POST_t$	An indicator equal to one if the fiscal year is in or after 2014 and zero otherwise	TEJ
$INVSCRUT_{it}$	An indicator equal to one if the institutional ownership of firm i in year t is greater than the cross-sectional median of the institutional ownerships in year t	TEJ
$\mathrm{ANASCRU}T_{it}$	An indicator equal to one if the number of analysts following firm i in year t is greater than the cross-sectional median of the numbers of analysts following in year t	I/B/E/S
$SIZE_{it}$	The natural logarithm of total assets of firm i in year t	TEJ
MB_{it}	The market-to-book ratio of firm i in year t	TEJ
ROA_{it}	The returns on total assets of firm i in year t	TEJ
$LOSS_{it}$	An indicator equal to one if the earnings of firm i in year t is negative and zero otherwise	TEJ
LEV_{it}	The ratio of long-term debts to total assets	TEJ

doi:10.6342/NTU202500807

$EARNVOL_{it}$	The standard deviation of quarterly earnings of firm i in year t	TEJ
$RETURN_{it}$	The annual average of monthly returns of firm i in year t	TEJ
$RETURNVOL_{it}$	The standard deviation of daily returns of firm i in year t	TEJ
$TURNOVER_{it}$	The annual average of monthly stock turnovers of firm i in year t	TEJ
$BETA_{it}$	The market beta of the firm i in year t calculated using daily stock returns	TEJ
$INSTOWN_{it}$	Institutional ownership of firm i in year t	TEJ
$BOARD_{it}$	The natural logarithm of the number of boards of firm i in year t	TEJ
$DUAL_{it}$	An indicator equal to one if the manager of firm i in year t is in the board and zero otherwise	TEJ
$OPAQ_{it}$	The absolute value of discretionary accruals of firm i in year t derived from the modified Jones model plus returns on total assets as an additional variable (Kothari et al. 2005)	TEJ
$\mathrm{OPAQ2}_{\mathrm{it}}$	The squared term of $OPAQ_{it}$	TEJ

doi:10.6342/NTU202500807

Table 2: Sample Selection Process

This table summarizes the number of observations and the number of distinct firms the sample contains after each procedure. The sample period is 2011-2017.

Procedure	Number of Observations	Number of Distinct Firms
Initial joined data	15,357	2,462
Exclude financial firms	14,792	2,381
Exclude the treated firms that voluntarily disclosed CSR reports before 2014 and the control firms that voluntarily disclosed CSR reports during the sample period	10,005	1,434
Exclude missing values of the variables in (1) and (2)	4,244	793

doi:10.6342/NTU202500807

Table 3: Distribution of Fiscal Years, Industries, and Treatment and Control Observations

Panel A: Distribution of Fiscal Years

This table shows the distribution of fiscal years during the sample period 2011-2017. The Freq. Column shows the number of observations in each fiscal year. The Percent column shows the ratio, expressed in percentage, of the number of observations in each year to the number of all the observations in the sample.

Fiscal Year	Freq.	Percent
2011	634	14.94
2013	683	16.09
2014	705	16.61
2015	725	17.08
2016	748	17.62
2017	749	17.65
Total	4244	100.00

Panel B: Distribution of Treatment and Control Observations among the Fiscal Years

This table shows the distribution of treatment and control observations among the fiscal years during the sample period 2011-2017. The False, True, and Total sub-columns of the TREAT (number) column show the number of control, treatment, and total observations in each fiscal year, respectively. The False, True, and Total sub-columns of the TREAT (%) column show the ratio, expressed in percentage, of the number of control, treatment, and total observations in each year to the number of all the control, treatment, and total observations in the sample, respectively.

Fiscal Year	TREAT (number)			TREAT (%)		
	False	True	Total	False	True	Total
2011	615	19	634	14.89	16.67	14.94
2013	664	19	683	16.08	16.67	16.09
2014	686	19	705	16.61	16.67	16.61
2015	706	19	725	17.09	16.67	17.08
2016	729	19	748	17.65	16.67	17.62
2017	730	19	749	17.68	16.67	17.65
Total	4130	114	4244	100.00	100.00	100.00

Panel C: Distribution of Industries

This table shows the distribution of industries during the sample period 2011-2017. The Freq. column shows the number of observations in each industry. The Percent column shows the ratio, expressed in percentage, of the number of observations in each industry to the number of all the observations in the sample.

Industry	Freq.	Percent
Textile	264	6.22
Electric Machinery	359	8.46
Electronics	3545	83.53
Cultural and Creativity	76	1.79
Total	4244	100.00

Panel D: Distribution of Treatment and Control Observations among the Industries

This table shows the distribution of treatment and control observations among the industries during the sample period 2011-2017. The False, True, and Total sub-columns of the TREAT (number) column show the number of control, treatment, and total observations in each fiscal year, respectively. The False, True, and Total sub-columns of the TREAT (%) column show the ratio, expressed in percentage, of the number of control, treatment, and total observations in each year to the number of all the control, treatment, and total observations in the sample, respectively.

Industry	TREAT (number)			TREAT (%)		
	False	True	Total	False	True	Total
Textile	246	18	264	5.96	15.79	6.22
Electric Machinery	359	0	359	8.69	0.00	8.46
Electronics	3449	96	3545	83.51	84.21	83.53
Cultural and Creative	76	0	76	1.84	0.00	1.79
Total	4130	114	4244	100.00	100.00	100.00

doi:10.6342/NTU202500807

Table 4: Descriptive Statistics

Panel A: Summary Statistics

This table presents the number of observations (N), the mean (Mean), the standard deviation (SD), the minimum (Min), the 25th percentile (P25), the 50th percentile (P50), the 75th percentile (P75), and the maximum (Max) of each variable. The sample period is 2011-2017. The variable definitions are in Table 1.

VARIABLES	N	Mean	SD	Min	P25	P50	P75	Max
MF	4,244	0.00	0.07	0.00	0.00	0.00	0.00	1.00
TREAT	4,244	0.03	0.16	0.00	0.00	0.00	0.00	1.00
POST	4,244	0.69	0.46	0.00	0.00	1.00	1.00	1.00
INVSCRUT	4,242	0.29	0.45	0.00	0.00	0.00	1.00	1.00
ANASCRUT	4,244	0.12	0.32	0.00	0.00	0.00	0.00	1.00
SIZE	4,244	14.83	1.17	11.16	14.10	14.77	15.48	19.56
MB	4,244	1.67	1.49	0.14	0.83	1.36	2.05	24.18
ROA (%)	4,244	2.50	9.83	-50.44	-1.89	2.84	7.43	37.28
LOSS	4,244	0.32	0.47	0.00	0.00	0.00	1.00	1.00
LEV	4,244	0.08	0.09	0.00	0.01	0.04	0.11	0.50
EARNVOL	4,244	0.01	0.02	0.00	0.00	0.01	0.01	0.14
RETURN (%)	4,244	1.15	3.72	-8.88	-1.03	0.66	2.78	18.70
RETURNVOL (%)	4,244	2.41	0.88	0.69	1.82	2.33	2.84	8.57
TURNOVER (%)	4,244	14.35	14.42	0.12	4.00	9.57	19.71	66.81
BETA	4,244	0.85	0.39	-0.06	0.57	0.85	1.13	1.77
INSTOWN (%)	4,244	32.03	20.40	0.94	15.81	28.55	45.78	94.79
NANA	4,244	0.43	0.73	0.00	0.00	0.00	0.69	3.71
BOARD	4,244	2.02	0.19	1.61	1.79	2.08	2.08	2.77
DUAL	4,244	0.45	0.50	0.00	0.00	0.00	1.00	1.00
OPAQ	4,244	0.07	0.08	0.00	0.02	0.05	0.09	0.60
OPAQ2	4,244	0.01	0.03	0.00	0.00	0.00	0.01	0.36

doi:10.6342/NTU202500807

Panel B: Pearson-Spearman Correlations

This table presents the Pearson and Spearman correlations of the variables. The Pearson correlations are in the lower triangle. The Spearman correlations are in the upper triangle. The sample period is 2011-2017. The variable definitions are in Table 1.

VARIABLES	MF	TREAT	POST	INVSCRUT	ANASCRUT	SIZE	MB	ROA	LOSS	LEV
MF	1.00***	0.08***	-0.01	0.05***	0.12***	0.07***	-0.04**	0.04***	-0.02	-0.03*
TREAT	0.08***	1.00***	-0.01	0.14***	0.21***	0.28***	0.10***	0.01	-0.03*	0.08***
POST	-0.01	-0.01	1.00***	0.01	-0.01	0.01	0.01	-0.00	-0.02	-0.01
INVSCRUT	0.05***	0.14***	0.01	1.00***	0.25***	0.18***	-0.14***	0.14***	-0.09***	-0.01
ANASCRUT	0.12***	0.21***	-0.01	0.25***	1.00***	0.33***	-0.23***	0.34***	-0.22***	-0.05***
SIZE	0.10***	0.47***	0.01	0.21***	0.37***	1.00***	0.27***	0.22***	-0.25***	0.22***
MB	-0.03*	0.13***	-0.02	-0.07***	-0.16***	0.27***	1.00***	-0.48***	0.27***	0.31***
ROA	0.03**	0.01	0.00	0.12***	0.31***	0.26***	-0.29***	1.00***	-0.81***	-0.12***
LOSS	-0.02	-0.03*	-0.02	-0.09***	-0.22***	-0.25***	0.27***	-0.70***	1.00***	0.09***
LEV	-0.02	0.06***	-0.00	-0.00	-0.06***	0.16***	0.32***	-0.16***	0.16***	1.00***
EARNVOL	-0.03*	-0.04***	-0.06***	0.03*	-0.04***	-0.26***	-0.12***	-0.14***	0.17***	-0.01
RETURN	-0.00	-0.02	0.02	0.05***	0.11***	-0.03**	-0.22***	0.27***	-0.19***	-0.02
RETURNVOL	-0.01	-0.06***	-0.08***	0.02	-0.01	-0.32***	-0.03	-0.17***	0.23***	0.08***
TURNOVER	0.02	-0.04***	-0.08***	-0.11***	0.20***	0.12***	-0.25***	0.28***	-0.19***	-0.05***
BETA	0.04**	0.09***	-0.19***	-0.03*	0.21***	0.29***	-0.10***	0.20***	-0.12***	-0.03**
INSTOWN	0.08***	0.16***	0.06***	0.76***	0.27***	0.25***	-0.11***	0.17***	-0.13***	-0.02
NANA	0.13***	0.29***	0.08***	0.26***	0.71***	0.52***	-0.17***	0.36***	-0.28***	-0.05***
BOARD	0.03*	0.19***	0.06***	0.10***	0.15***	0.26***	0.03	0.07***	-0.09***	0.04***
DUAL	-0.02	0.01	-0.03*	-0.09***	0.07***	0.05***	-0.07***	0.07***	-0.07***	-0.06***
OPAQ	-0.02	-0.04**	-0.03*	0.07***	-0.02	-0.18***	-0.10***	-0.10***	0.07***	-0.00
OPAQ2	-0.01	-0.03**	-0.02	0.07***	-0.02	-0.16***	-0.09***	-0.08***	0.05***	0.00

Panel B (continued)

doi:10.6342/NTU202500807

VARIABLES	EARNVOL	RETURN	RETURN VOL	TURNOVER	BETA	INSTOWN	NANA	BOARD	DUAL	OPAQ	OPAQ2
MF	-0.03**	-0.00	-0.01	0.03**	0.04**	0.07***	0.08***	0.03*	-0.02	-0.02	-0.02
TREAT	-0.08***	-0.01	-0.06***	-0.02	0.10***	0.14***	0.21***	0.14***	0.01	-0.03*	-0.03*
POST	-0.10***	0.02	-0.10***	-0.08***	-0.20***	0.06***	0.09***	0.06***	-0.03*	-0.03**	-0.03**
INVSCRUT	-0.02	0.05***	0.01	-0.13***	-0.02	0.72***	0.20***	0.11***	-0.09***	0.05***	0.05***
ANASCRUT	-0.04**	0.12***	-0.02	0.22***	0.21***	0.25***	0.60***	0.15***	0.07***	-0.01	-0.01
SIZE	-0.30***	0.01	-0.30***	0.19***	0.28***	0.23***	0.47***	0.19***	0.05***	-0.13***	-0.13***
MB	-0.26***	-0.27***	-0.19***	-0.32***	-0.13***	-0.15***	-0.24***	-0.03**	-0.06***	-0.11***	-0.11***
ROA	-0.06***	0.30***	-0.19***	0.33***	0.18***	0.18***	0.40***	0.09***	0.08***	-0.06***	-0.06***
LOSS	0.20***	-0.22***	0.25***	-0.23***	-0.11***	-0.13***	-0.30***	-0.09***	-0.07***	0.07***	0.07***
LEV	-0.11***	0.00	-0.02	-0.05***	-0.02	-0.03*	-0.01	0.04***	-0.05***	-0.05***	-0.05***
EARNVOL	1.00***	0.03**	0.39***	0.11***	0.05***	-0.04**	-0.07***	-0.04***	-0.00	0.22***	0.22***
RETURN	0.12***	1.00***	0.17***	0.25***	-0.00	0.02	0.11***	0.01	0.00	0.00	0.00
RETURNVOL	0.34***	0.28***	1.00***	0.36***	0.37***	0.01	-0.06***	-0.02	-0.04***	0.16***	0.16***
TURNOVER	0.06***	0.30***	0.33***	1.00***	0.66***	-0.11***	0.32***	0.11***	0.13***	0.02	0.02
BETA	-0.05***	-0.01	0.25***	0.56***	1.00***	0.00	0.28***	0.14***	0.11***	-0.01	-0.01
INSTOWN	0.02	0.03*	0.03**	-0.10***	-0.01	1.00***	0.25***	0.13***	-0.11***	0.03*	0.03*
NANA	-0.08***	0.08***	-0.05***	0.25***	0.28***	0.32***	1.00***	0.19***	0.09***	-0.04***	-0.04***
BOARD	-0.06***	-0.00	-0.04**	0.07***	0.13***	0.14***	0.21***	1.00***	0.12***	-0.05***	-0.05***
DUAL	-0.04**	-0.01	-0.04***	0.10***	0.11***	-0.11***	0.09***	0.11***	1.00***	-0.01	-0.01
OPAQ	0.32***	0.11***	0.24***	0.02	-0.06***	0.07***	-0.05***	-0.06***	-0.02	1.00***	1.00***
OPAQ2	0.34***	0.15***	0.22***	-0.01	-0.09***	0.07***	-0.04***	-0.05***	-0.04**	0.89***	1.00***

Table 5: Difference-in-Differences Logit Regressions for Hypothesis 1

This table presents the difference-in-differences logit regressions of the indicator of managerial earnings forecasts. The sample period is 2011-2017. The variable definitions are in Table 1. Ind FE and Year FE indicate industry fixed effects and year fixed effects, respectively. The standard errors are clustered by firms. The *, **, and *** indicate the statistical significance levels of 10%, 5%, and 1%, respectively.

	(1)	(2)	(3)
VARIABLES	MF	MF	MF
TREAT	-17.414***	-17.527***	-17.797***
	(-9.395)	(-8.960)	(-9.346)
POST	-0.922*	-1.121**	-1.775***
	(-1.938)	(-2.365)	(-2.642)
TREAT × POST	16.688***	16.865***	16.905***
	(17.306)	(16.927)	(16.508)
SIZE	1.260***	1.347***	1.264***
	(3.118)	(3.099)	(2.630)
MB	-1.507***	-1.422***	-1.290***
	(-3.500)	(-3.102)	(-2.893)
ROA	-0.013	-0.007	-0.017
	(-0.247)	(-0.124)	(-0.284)
LOSS	1.349	1.458	1.558
	(1.462)	(1.510)	(1.564)
LEV	-5.743	-7.693	-7.629
	(-0.981)	(-1.462)	(-1.401)
EARNVOL	-134.608***	-133.048***	-142.158***
	(-3.692)	(-3.296)	(-3.379)
RETURN	-0.043	-0.050	0.025
	(-0.464)	(-0.544)	(0.211)
RETURNVOL	0.400	0.409	0.058
	(1.495)	(1.452)	(0.134)
TURNOVER	0.010	0.012	0.011
	(0.458)	(0.525)	(0.445)
BETA	-0.234	-0.296	0.230
	(-0.254)	(-0.320)	(0.226)
INSTOWN	0.007	0.005	0.008
	(0.625)	(0.483)	(0.586)
NANA	0.265	0.277	0.430
	(0.613)	(0.662)	(0.900)
BOARD	0.441	0.350	0.513
	(0.356)	(0.292)	(0.392)
DUAL	-1.107**	-1.194*	-1.243*
	(-1.975)	(-1.955)	(-1.912)
OPAQ	-3.036	-3.142	-2.265
	(-0.309)	(-0.313)	(-0.211)

OPAQ2	-1.070	-1.269	-1.232
	(-0.047)	(-0.056)	(-0.054)
Constant	-22.995***	-23.536***	-22.381***
	(-4.332)	(-4.166)	(-3.490)
Observations	4,244	4,244	4,244
Ind FE	No	Yes	Yes
Year FE	No	No	Yes
Pseudo R^2	0.329	0.341	0.360

Table 6: Difference-in-Differences Logit Regressions for Hypothesis 1a

This table presents the difference-in-differences logit regressions of the indicator of managerial earnings forecasts. The sample period is 2011-2017. The variable definitions are in Table 1. Ind FE and Year FE indicate industry fixed effects and year fixed effects, respectively. The standard errors are clustered by firms. The *, **, and *** indicate the statistical significance levels of 10%, 5%, and 1%, respectively.

VARIABLES	(1) MF	(2) MF	(3) MF
TREAT	-17.596***	-17.700***	-18.034***
	(-8.978)	(-8.662)	(-8.942)
POST	-0.921*	-1.116**	-1.864***
	(-1.867)	(-2.326)	(-2.799)
TREAT × POST	3.090	2.999	3.206
	(1.373)	(1.111)	(1.552)
INVSCRUT	-1.097	-1.020	-1.243
	(-1.179)	(-1.033)	(-1.211)
TREAT × POST × INVSCRUT	13.716***	14.009***	13.827***
	(6.389)	(5.295)	(6.232)
SIZE	1.320***	1.400***	1.324***
	(3.015)	(2.988)	(2.581)
MB	-1.521***	-1.443***	-1.300***
	(-3.372)	(-3.091)	(-2.850)
ROA	-0.016	-0.010	-0.019
	(-0.296)	(-0.179)	(-0.318)
LOSS	1.357	1.451	1.569
	(1.488)	(1.544)	(1.614)
LEV	-5.307	-7.241	-7.276
	(-0.882)	(-1.335)	(-1.307)
EARNVOL	-132.042***	-128.873***	-137.801***
	(-3.498)	(-3.106)	(-3.151)
RETURN	-0.034	-0.044	0.038
	(-0.380)	(-0.514)	(0.347)
RETURNVOL	0.414	0.422	0.063
	(1.479)	(1.448)	(0.147)
TURNOVER	0.006	0.008	0.005
	(0.274)	(0.335)	(0.193)
BETA	-0.234	-0.296	0.295
	(-0.253)	(-0.319)	(0.295)
INSTOWN	0.022	0.019	0.025
	(1.269)	(1.080)	(1.277)
NANA	0.275	0.285	0.454
	(0.624)	(0.656)	(0.931)
BOARD	0.489	0.476	0.620
	(0.354)	(0.340)	(0.425)

DUAL	-1.097**	-1.193**	-1.208*
	(-1.990)	(-2.002)	(-1.883)
OPAQ	-2.201	-2.430	-1.314
	(-0.230)	(-0.244)	(-0.124)
OPAQ2	-3.970	-3.822	-4.202
	(-0.179)	(-0.168)	(-0.186)
Constant	-24.193***	-24.603***	-23.625***
	(-4.060)	(-3.914)	(-3.350)
Observations	4,242	4,242	4,242
Ind FE	No	Yes	Yes
Year FE	No	No	Yes
Pseudo R^2	0.336	0.347	0.368

Table 7: Difference-in-Differences Logit Regressions for Hypothesis 1b

This table presents the difference-in-differences logit regressions of the indicator of managerial earnings forecasts. The sample period is 2011-2017. The variable definitions are in Table 1. Ind FE and Year FE indicate industry fixed effects and year fixed effects, respectively. The standard errors are clustered by firms. The *, **, and *** indicate the statistical significance levels of 10%, 5%, and 1%, respectively.

	(1)	(2)	(3)
VARIABLES	MF	MF	MF
TREAT	-17.140***	-17.267***	-17.291***
	(-9.661)	(-9.053)	(-9.067)
POST	-0.827*	-1.030**	-1.677**
	(-1.795)	(-2.187)	(-2.416)
TREAT × POST	4.218**	4.165*	4.676**
	(2.217)	(1.860)	(2.508)
ANASCRUT	1.147	1.084	0.902
	(1.452)	(1.400)	(1.101)
TREAT × POST × ANASCRUT	12.248***	12.492***	11.747***
	(6.456)	(5.497)	(5.730)
SIZE	1.300***	1.387***	1.296***
	(3.137)	(3.133)	(2.659)
MB	-1.576***	-1.501***	-1.346***
	(-3.612)	(-3.180)	(-2.922)
ROA	-0.022	-0.013	-0.020
	(-0.392)	(-0.218)	(-0.321)
LOSS	1.371	1.473	1.518
	(1.479)	(1.513)	(1.509)
LEV	-5.384	-7.147	-7.195
	(-0.931)	(-1.393)	(-1.369)
EARNVOL	-127.591***	-127.892***	-138.485***
	(-3.590)	(-3.272)	(-3.305)
RETURN	-0.062	-0.071	0.001
	(-0.624)	(-0.710)	(0.007)
RETURNVOL	0.408*	0.435*	0.147
	(1.659)	(1.712)	(0.373)
TURNOVER	0.010	0.011	0.009
	(0.431)	(0.490)	(0.361)
BETA	-0.226	-0.296	0.183
	(-0.239)	(-0.316)	(0.182)
INSTOWN	0.006	0.004	0.006
	(0.456)	(0.316)	(0.463)
NANA	-0.108	-0.083	0.126
	(-0.233)	(-0.176)	(0.226)
BOARD	0.447	0.389	0.507
	(0.333)	(0.294)	(0.365)

doi:10.6342/NTU202500807

DUAL	-1.125**	-1.209**	-1.236*
	(-1.994)	(-1.987)	(-1.905)
OPAQ	-3.183	-3.002	-2.030
	(-0.332)	(-0.299)	(-0.186)
OPAQ2	-1.203	-2.027	-2.816
	(-0.054)	(-0.086)	(-0.115)
Constant	-23.609***	-24.166***	-22.901***
	(-4.428)	(-4.225)	(-3.538)
Observations	4,244	4,244	4,244
Ind FE	No	Yes	Yes
Year FE	No	No	Yes
Pseudo R^2	0.338	0.350	0.365

Table 8: Difference-in-Differences Logit Regressions that Exclude Adoption Years

This table presents the difference-in-differences logit regressions of the indicator of managerial earnings forecasts. The observations in the fiscal year 2014 or 2015 are excluded in the sample. The sample period is 2011-2017. The variable definitions are in Table 1. Ind FE and Year FE indicate industry fixed effects and year fixed effects, respectively. The standard errors are clustered by firms. The *, **, and *** indicate the statistical significance levels of 10%, 5%, and 1%, respectively.

VARIABLES	(1) MF	(2) MF	(3) MF	(4) MF	(5) MF	(6) MF
TREAT	-16.971***	-16.134***	-17.225***	-16.907***	-17.215***	-16.671***
	(-9.443)	(-8.929)	(-9.132)	(-9.140)	(-9.613)	(-9.036)
POST	-1.715***	-1.501***	-1.809***	-1.614***	-1.605**	-1.407**
	(-2.657)	(-2.612)	(-2.847)	(-2.809)	(-2.384)	(-2.331)
TREAT × POST	16.229***	15.174***	2.588	2.386	4.456***	4.964***
	(16.334)	(13.346)	(1.341)	(1.270)	(2.647)	(3.129)
INVSCRUT			-1.343	-1.709		
			(-1.198)	(-1.380)		
ANASCRUT					1.005	1.006
					(1.145)	(1.072)
TREAT × POST × INVSCRUT			13.801***	13.431***		
			(6.303)	(6.246)		
TREAT × POST × ANASCRUT					12.019***	10.721***
					(6.289)	(6.547)
SIZE	1.150**	1.225**	1.214**	1.286**	1.175**	1.251**
	(2.271)	(2.089)	(2.180)	(2.068)	(2.273)	(2.124)
MB	-1.123***	-1.389***	-1.122***	-1.406***	-1.154***	-1.414***
	(-2.795)	(-3.797)	(-2.738)	(-3.865)	(-2.763)	(-3.650)
ROA	-0.011	-0.012	-0.013	-0.011	-0.015	-0.015
	(-0.187)	(-0.245)	(-0.211)	(-0.224)	(-0.223)	(-0.292)
LOSS	1.545	1.607	1.542	1.646	1.512	1.563
	(1.547)	(1.536)	(1.583)	(1.591)	(1.489)	(1.459)
LEV	-6.473	-6.709	-6.109	-6.195	-6.314	-6.794
	(-1.281)	(-1.158)	(-1.196)	(-1.101)	(-1.268)	(-1.190)

doi:10.6342/NTU202500807

EARNVOL	-151.809***	-149.062***	-148.796***	-146.172***	-147.839***	-145.570***
	(-3.515)	(-3.143)	(-3.369)	(-2.970)	(-3.399)	(-3.090)
RETURN	0.029	-0.017	0.040	0.001	0.003	-0.038
	(0.221)	(-0.127)	(0.312)	(0.007)	(0.020)	(-0.269)
RETURNVOL	0.119	0.302	0.131	0.297	0.189	0.355
	(0.300)	(0.923)	(0.324)	(0.838)	(0.510)	(1.154)
TURNOVER	0.012	0.018	0.006	0.008	0.010	0.015
	(0.490)	(0.712)	(0.249)	(0.321)	(0.425)	(0.613)
BETA	0.250	0.315	0.297	0.473	0.217	0.346
	(0.207)	(0.233)	(0.242)	(0.338)	(0.179)	(0.257)
INSTOWN	0.012	0.012	0.031	0.038*	0.011	0.011
	(0.833)	(0.834)	(1.452)	(1.783)	(0.742)	(0.754)
NANA	0.295	0.166	0.307	0.175	-0.045	-0.171
	(0.646)	(0.390)	(0.655)	(0.404)	(-0.081)	(-0.317)
BOARD	0.777	0.442	0.896	0.603	0.785	0.450
	(0.541)	(0.285)	(0.573)	(0.369)	(0.549)	(0.309)
DUAL	-1.221*	-1.343*	-1.191*	-1.271*	-1.203*	-1.333*
	(-1.943)	(-1.943)	(-1.900)	(-1.841)	(-1.911)	(-1.937)
OPAQ	-6.546	-9.996	-5.763	-8.958	-6.234	-9.606
	(-0.502)	(-0.839)	(-0.438)	(-0.740)	(-0.468)	(-0.807)
OPAQ2	8.148	13.810	5.696	10.800	6.443	12.122
	(0.312)	(0.648)	(0.215)	(0.489)	(0.228)	(0.526)
Constant	-21.488***	-21.961***	-22.841***	-23.504***	-22.017***	-22.465***
	(-3.185)	(-2.761)	(-3.005)	(-2.644)	(-3.158)	(-2.758)
Observations	3,539	2,814	3,537	2,813	3,539	2,814
Excluded Years	2014	2014, 2015	2014	2014, 2015	2014	2014, 2015
Ind FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R^2	0.330	0.329	0.340	0.344	0.337	0.336

Table 9: Difference-in-Differences Probit Regressions

This table presents the difference-in-differences probit regressions of the indicator of managerial earnings forecasts. The sample period is 2011-2017. The variable definitions are in Table 1. Ind FE and Year FE indicate industry fixed effects and year fixed effects, respectively. The standard errors are clustered by firms. The *, **, and *** indicate the statistical significance levels of 10%, 5%, and 1%, respectively.

VARIABLES	(1) MF	(2) MF	(3) MF
TREAT	-5.288***	-5.313***	-5.511***
	(-5.566)	(-5.360)	(-5.639)
POST	-0.787***	-0.834***	-0.745***
	(-3.126)	(-3.321)	(-2.889)
TREAT × POST	5.187***	1.300*	1.962***
	(9.520)	(1.812)	(2.741)
INVSCRUT		-0.495	
		(-1.398)	
ANASCRUT			0.441
			(1.395)
TREAT × POST × INVSCRUT		3.952***	
		(6.099)	
TREAT × POST × ANASCRUT			3.477***
			(4.992)
SIZE	0.454**	0.462**	0.472**
	(2.441)	(2.329)	(2.484)
MB	-0.498***	-0.501***	-0.524***
	(-3.165)	(-3.164)	(-3.170)
ROA	-0.009	-0.011	-0.011
	(-0.506)	(-0.601)	(-0.580)
LOSS	0.501	0.511	0.517
	(1.479)	(1.553)	(1.515)
LEV	-2.391	-2.352	-2.269
	(-1.472)	(-1.492)	(-1.421)
EARNVOL	-57.569***	-56.727***	-57.071***
	(-3.512)	(-3.451)	(-3.577)
RETURN	0.013	0.019	0.007
	(0.322)	(0.515)	(0.161)
RETURNVOL	0.052	0.051	0.073
	(0.411)	(0.401)	(0.617)
TURNOVER	0.001	-0.001	0.001
	(0.137)	(-0.153)	(0.068)
BETA	0.211	0.227	0.202
	(0.578)	(0.650)	(0.550)
INSTOWN	0.002	0.009	0.001
	(0.430)	(1.230)	(0.271)

NANA	0.187	0.209	0.030
	(1.129)	(1.289)	(0.148)
BOARD	0.322	0.407	0.349
	(0.792)	(0.917)	(0.804)
DUAL	-0.524**	-0.514**	-0.547**
	(-2.038)	(-2.089)	(-2.142)
OPAQ	0.543	0.736	0.611
	(0.157)	(0.221)	(0.178)
OPAQ2	-4.597	-5.515	-5.119
	(-0.523)	(-0.642)	(-0.555)
Constant	-8.938***	-9.257***	-9.250***
	(-3.542)	(-3.364)	(-3.615)
Observations	4,244	4,242	4,244
Ind FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Pseudo R^2	0.348	0.358	0.356

Table 10: Difference-in-Differences Logit Regressions with Double-Clustered Standard Errors

This table presents the difference-in-differences logit regressions of the indicator of managerial earnings forecasts. The sample period is 2011-2017. The variable definitions are in Table 1. Ind FE and Year FE indicate industry fixed effects and year fixed effects, respectively. The standard errors are clustered by firms and fiscal years. The *, **, and *** indicate the statistical significance levels of 10%, 5%, and 1%, respectively.

VARIABLES	(1) MF	(2) MF	(3) MF
TREAT	-17.797***	-18.034***	-17.291***
	(-10.657)	(-10.295)	(-10.301)
POST	-0.392	-0.359	-0.369
	(.)	(.)	(.)
TREAT × POST	16.905***	3.206*	4.676***
	(13.936)	(1.821)	(4.052)
INVSCRUT		-1.243**	
		(-2.223)	
TREAT × POST × INVSCRUT		13.827***	
		(7.448)	
ANASCRUT			0.902
			(1.457)
TREAT × POST × ANASCRUT			11.747***
			(7.655)
SIZE	1.264**	1.324***	1.296**
	(2.441)	(2.757)	(2.469)
MB	-1.290***	-1.300***	-1.346***
	(-3.430)	(-3.181)	(-3.507)
ROA	-0.017	-0.019	-0.020
	(-0.313)	(-0.348)	(-0.345)
LOSS	1.558**	1.569***	1.518**
	(2.408)	(2.592)	(2.381)
LEV	-7.629	-7.276	-7.195
	(-1.624)	(-1.530)	(-1.599)
EARNVOL	-142.158***	-137.801***	-138.485***
	(-2.940)	(-2.718)	(-2.840)
RETURN	0.025	0.038	0.001
	(0.338)	(0.535)	(0.011)
RETURNVOL	0.058	0.063	0.147
	(0.128)	(0.153)	(0.358)
TURNOVER	0.011	0.005	0.009
	(0.650)	(0.283)	(0.502)
BETA	0.230	0.295	0.183
	(0.501)	(0.687)	(0.492)
INSTOWN	0.008	0.025	0.006

	(0.769)	(1.567)	(0.610)
NANA	0.430	0.454	0.126
	(1.019)	(1.052)	(0.310)
BOARD	0.513	0.620	0.507
	(0.476)	(0.538)	(0.448)
DUAL	-1.243**	-1.208**	-1.236**
	(-1.994)	(-2.073)	(-1.985)
OPAQ	-2.265	-1.314	-2.030
	(-0.308)	(-0.177)	(-0.280)
OPAQ2	-1.232	-4.202	-2.816
	(-0.071)	(-0.250)	(-0.147)
Constant	-22.165***	-23.838***	-22.617***
	(-4.434)	(-5.108)	(-4.611)
Observations	4,244	4,242	4,244
Ind FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes

Table 11: Difference-in-Differences Poisson Regressions

This table presents the difference-in-differences Poisson regressions of the number of managerial earnings forecasts. The sample is refined after the propensity score matching based on a logit regression of the treatment indicator on the firm size and industry fixed effects. The sample period is 2011-2017. The variable definitions are in Table 1. Ind FE and Year FE indicate industry fixed effects and year fixed effects, respectively. The standard errors are clustered by firms. The *, **, and *** indicate the statistical significance levels of 10%, 5%, and 1%, respectively.

VARIABLES	(1) NMF	(2) NMF	(3) NMF
TREAT	-16.896***	-14.834***	-14.992***
	(-11.934)	(-10.655)	(-10.489)
POST	-1.422**	-1.570**	-1.345**
	(-2.302)	(-2.556)	(-2.174)
TREAT × POST	16.555***	0.420	1.148
	(19.329)	(0.245)	(0.827)
INVSCRUT		-2.368*	
		(-1.752)	
ANASCRUT			0.901
			(0.991)
TREAT × POST × INVSCRUT		13.863***	
		(7.401)	
TREAT × POST × ANASCRUT			13.547***
			(9.241)
SIZE	1.213***	1.390***	1.244***
	(3.545)	(3.408)	(3.583)
MB	-1.441***	-1.459***	-1.495***
	(-3.945)	(-3.760)	(-3.925)
ROA	-0.012	-0.016	-0.017
	(-0.281)	(-0.353)	(-0.350)
LOSS	1.607	1.581	1.567
	(1.594)	(1.628)	(1.514)
LEV	-3.677	-3.299	-3.438
	(-0.730)	(-0.665)	(-0.710)
EARNVOL	-116.618***	-106.368***	-113.138***
	(-3.340)	(-2.846)	(-3.192)
RETURN	-0.042	-0.015	-0.067
	(-0.398)	(-0.169)	(-0.572)
RETURNVOL	0.127	0.120	0.196
	(0.348)	(0.319)	(0.593)
TURNOVER	0.016	0.005	0.014
	(0.724)	(0.208)	(0.627)
BETA	-0.110	0.021	-0.129
	(-0.119)	(0.022)	(-0.139)
INSTOWN	0.006	0.040*	0.005

doi:10.6342/NTU202500807

	(0.605)	(1.795)	(0.435)
NANA	0.098	0.083	-0.206
	(0.220)	(0.175)	(-0.360)
BOARD	0.781	0.879	0.809
	(0.640)	(0.694)	(0.636)
DUAL	-1.396**	-1.390**	-1.406**
	(-2.356)	(-2.180)	(-2.323)
OPAQ	-5.454	-4.144	-4.894
	(-0.563)	(-0.461)	(-0.496)
OPAQ2	4.366	-0.654	1.807
	(0.201)	(-0.032)	(0.076)
Constant	-22.150***	-25.252***	-22.686***
	(-4.604)	(-4.070)	(-4.747)
Observations	4,244	4,242	4,244
Ind FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Pseudo R^2	0.328	0.354	0.334

Table 12: Propensity-Score-Matched Difference-in-Differences Logit and Poisson Regressions

Panel A: Logit Regression Used to Calculate Propensity Scores

This table presents the logit regression results used to calculate the propensity scores. All the observations in the sample whose fiscal years are before 2014, when the mandatory ESG disclosures in Taiwan were implemented, are used in the logit regression. The sample period is 2011-2017. The variable definitions are in Table 1. Ind FE indicates industry fixed effects. The standard errors are clustered by firms. The *, **, and *** indicate the statistical significance levels of 10%, 5%, and 1%, respectively.

VARIABLES	(1) TREAT
SIZE	1.443***
	(18.855)
Constant	-19.722***
	(-16.996)
Observations	3,886
Ind FE	Yes
Pseudo R^2	0.533

Panel B: Difference-in-Differences Logit and Poisson Regressions

This table presents the difference-in-differences logit regression results of the indicator of managerial earnings forecasts and Poisson regression results of the number of managerial earnings forecasts. The sample is refined after the propensity score matching based on a logit regression of the treatment indicator on the firm size and industry fixed effects, whose results are in Panel A of this table. The sample period is 2011-2017. The variable definitions are in Table 1. Ind FE and Year FE indicate industry fixed effects and year fixed effects, respectively. The standard errors are clustered by firms. The *, **, and *** indicate the statistical significance levels of 10%, 5%, and 1%, respectively.

VARIABLES	(1) MF	(2) MF	(3) MF	(4) NMF	(5) NMF	(6) NMF
TREAT	-17.141***	-18.053***	-18.034***	-18.746***	-18.156***	-15.571***
	(-10.289)	(-7.507)	(-8.155)	(-10.604)	(-10.771)	(-8.702)
POST	-4.536***	-5.045***	-4.776***	-3.530***	-3.675***	-3.779***
	(-4.983)	(-4.468)	(-5.038)	(-3.819)	(-3.214)	(-3.278)
TREAT × POST	19.000***	4.368	9.140***	19.750***	3.754*	7.294***
	(10.501)	(1.294)	(2.609)	(12.253)	(1.915)	(3.528)
INVSCRUT		1.078			0.114	
		(0.613)			(0.095)	
ANASCRUT			4.039**			4.504***

			(2.019)			(3.558)
TREAT × POST × INVSCRUT		16.531***			15.780***	
		(6.640)			(5.119)	
TREAT × POST × ANASCRUT			10.410***			9.028***
			(3.827)			(4.528)
SIZE	0.011	-0.142	0.285	0.360	0.298	0.526
	(0.010)	(-0.124)	(0.303)	(0.526)	(0.423)	(0.822)
MB	-0.690	-0.403	-1.086	-0.621	-0.419	-0.969
	(-0.970)	(-0.507)	(-1.463)	(-0.794)	(-0.482)	(-1.539)
ROA	-0.108	-0.092	-0.090*	-0.062	-0.050	-0.071
	(-1.575)	(-1.509)	(-1.653)	(-1.035)	(-0.961)	(-0.935)
LOSS	0.888	1.715	2.256	1.275	1.752	2.310
	(0.700)	(1.551)	(1.324)	(0.787)	(1.367)	(1.087)
LEV	-32.103	-32.503*	-32.173	-23.702**	-22.943**	-26.942
	(-1.579)	(-1.844)	(-1.470)	(-2.085)	(-2.252)	(-1.615)
EARNVOL	-30.552	-21.143	-46.393	-32.193	-30.659	-42.204
	(-0.768)	(-0.616)	(-0.996)	(-0.771)	(-0.676)	(-0.919)
RETURN	0.161	0.124	0.166	0.035	0.017	0.099
	(0.555)	(0.423)	(0.710)	(0.139)	(0.067)	(0.522)
RETURNVOL	-0.450	-0.493	-0.165	-0.207	-0.208	0.053
	(-0.469)	(-0.427)	(-0.166)	(-0.252)	(-0.236)	(0.059)
TURNOVER	-0.007	0.010	-0.061*	0.020	0.026	-0.054
	(-0.180)	(0.374)	(-1.702)	(0.597)	(1.044)	(-1.277)
BETA	2.241***	2.853**	2.272***	1.902**	2.082**	2.133*
	(3.557)	(2.512)	(2.803)	(2.335)	(2.030)	(1.653)
INSTOWN	0.011	-0.007	-0.002	0.033	0.029	0.012
	(0.257)	(-0.111)	(-0.059)	(1.047)	(0.730)	(0.548)
NANA	1.354	1.463	0.196	0.609	0.652	-0.404
	(1.053)	(1.116)	(0.133)	(0.780)	(0.938)	(-0.560)
BOARD	0.952	1.543	1.394	-0.687	-0.407	0.374
	(0.395)	(0.621)	(0.959)	(-0.397)	(-0.225)	(0.246)
DUAL	0.336	0.374	0.800	0.083	0.063	0.596
	(0.405)	(0.424)	(0.863)	(0.114)	(0.083)	(0.748)
OPAQ	-16.284	-17.060	-23.559	-25.806	-25.318	-33.430*
	(-0.621)	(-0.653)	(-0.860)	(-1.205)	(-1.152)	(-1.678)
OPAQ2	47.198	53.430	78.204	83.326	84.086	117.668*
	(0.525)	(0.647)	(0.806)	(1.281)	(1.340)	(1.878)
Constant	-5.447	-5.733	-10.404	-21.862**	-20.229**	-23.416**
	(-0.367)	(-0.394)	(-0.672)	(-2.322)	(-2.292)	(-2.172)
Observations	425	424	425	580	579	580
Ind FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R^2	0.421	0.441	0.453	0.474	0.479	0.517

Table 13: Parallel and Placebo Tests

This table presents in Column (1) the parallel test results (Bertrand and Mullainathan 2003) of the difference-in-differences logit regression of the indicator of managerial earnings forecasts. This table also presents in Column (2) the placebo test results (Lee et al. 2021) of the difference-in-differences logit regression of the indicator of managerial earnings forecasts. PRE^{-1} is the fiscal year 2013 indicator. POST^{0} is the fiscal year 2014 indicator. POST^{1} is the fiscal year 2015 indicator. POST^{2+} is an indicator that equals one if the fiscal year is in or after 2016 and 0 otherwise. Pseudo-POST equals one if the fiscal year is in or after 2013 and 0 otherwise. The sample period is 2011-2017. The variable definitions are in Table 1modes. Ind FE and Year FE indicate industry fixed effects and year fixed effects, respectively. The standard errors are clustered by firms. The *, **, and *** indicate the statistical significance levels of 10%, 5%, and 1%, respectively.

VARIABLES	(1) MF	(2) MF
TREAT	-18.309***	-17.354***
	(-9.084)	(-7.523)
PRE^{-1}	-1.413	
	(-1.493)	
POST^{0}	-2.363***	
	(-3.465)	
POST^{1}	-2.778***	
	(-4.012)	
POST^{2+}	-1.598**	
	(-2.518)	
Pseudo-Post		-1.759***
		(-2.759)
TREAT × PRE^{-1}	1.853	
	(1.643)	
TREAT × POST^{0}	18.285***	
	(15.106)	
TREAT × POST^{1}	18.703***	
	(10.694)	
TREAT × POST^{2+}	16.672***	
	(13.804)	
TREAT × Pseudo-POST		2.170*
		(1.822)
SIZE	1.285***	1.139**
	(2.694)	(2.131)
MB	-1.285***	-1.224***
	(-2.950)	(-3.489)
ROA	-0.024	-0.025
	(-0.430)	(-0.383)
LOSS	1.494	1.463
	(1.517)	(1.403)
LEV	-8.011	-7.605

	(-1.491)	(-1.235)
EARNVOL	-145.496***	-142.978***
	(-3.441)	(-3.324)
RETURN	0.027	0.031
	(0.231)	(0.241)
RETURNVOL	0.062	0.085
	(0.148)	(0.229)
TURNOVER	0.012	0.014
	(0.471)	(0.562)
BETA	0.271	0.223
	(0.268)	(0.208)
INSTOWN	0.008	0.014
	(0.600)	(1.001)
NANA	0.437	0.415
	(0.917)	(0.780)
BOARD	0.393	0.560
	(0.304)	(0.290)
DUAL	-1.356**	-2.110**
	(-2.023)	(-2.183)
OPAQ	-3.360	-8.554
	(-0.314)	(-0.789)
OPAQ2	0.912	9.710
	(0.041)	1.139**
Constant	-22.436***	(2.131)
	(-3.521)	-1.224***
Observations	4,244	4,168
Ind FE	Yes	Yes
Year FE	Yes	Yes
Pseudo R^2	0.366	0.340

doi:10.6342/NTU202500807

Table 14: Summary of the Models

This table summarizes the models we use to estimate fundamental betas. The table contains the name, description, and parameters of each of the models.

Name	Description	Parameters
Rolling Regression	A five-year (20-quarter) rolling regression of firm-level earnings growth on aggregate earnings growth, total factor productivity growth, or household wealth growth	NA
OLS	A linear regression of realized fundamental betas on all the predictors described in Table 15	NA
Lasso	A Lasso regression, which adopts the L1 regularization, of realized fundamental betas on all the predictors described in Table 15	Penalization Strength = 1
Ridge	A Ridge regression, which adopts the L2 regularization, of realized fundamental betas on all the predictors described in Table 15	Penalization Strength = 1
Elastic Net	A linear mixture of the Lasso regression and the Ridge regression of realized fundamental betas on all the predictors described in Table 15	Penalization Strength = 1; Penalty Mixture = 0.5

doi:10.6342/NTU202500807

Table 15: Variable and Predictor Descriptions and Definitions in Chapter 2

This table alphabetically lists the outcome variables and predictors used in the research and their descriptions and definitions. Panel A shows the outcome variables. Panel B shows the firm-level predictors. Panel C shows the time-series predictors.

Panel A: Outcome Variables

Variable	Description	Definition
EBETA	Earnings Beta	OLS coefficient of firm-level earnings growth on aggregate earnings growth in the last 4 quarters
TBETA	TFP Beta	OLS coefficient of firm-level earnings growth on total factor productivity growth in the last 4 quarters
WBETA	Wealth Beta	OLS coefficient of firm-level earnings growth on household wealth growth in the last 4 quarters

Panel B: Firm-Level Predictors

Predictor	Description	Definition
AG	Asset Growth	Annual change in total assets divided by the total assets 4 quarters ago
AGE	Age	Natural logarithm of years since first inclusion in CRSP
BETA	Market Beta	Coefficient of firm-level excess returns on market excess returns of the 5-year rolling regression
BM	Book-to-Market Ratio	Ratio of common share equity to market capitalizations
DIVPAY	Dividend Payout Ratio	Ratio of dividends paid during the last fiscal year to net income
DP	Dividend Yield	The dividend yield derived from gross returns with dividends divided by gross returns without dividends minus one
EP	Earnings-to-Price Ratio	Ratio of earnings divided by market capitalization
EPCOV	Covariability in Earnings	Coefficient estimate in the time series ordinary least squares regression of monthly earnings-to-price ratios on the market's monthly earnings-to-price ratio over the last 3 years
EPVAR	Variability in Earnings	Natural logarithm of the standard deviation of monthly earnings-to-price ratios over the last 3 years
EV	Earnings Volatility	Standard deviation of returns on equity in the last 5 years
FLEV	Financial Leverage	Ratio of total assets to total equity

doi:10.6342/NTU202500807

FXC	Fixed Cost	Ratio of selling, general, and administrative expenses plus research and development expenses plus advertising expenses to sales
ILLIQ	Illiquidity	Ratio of monthly absolute return to monthly dollar trading volume at the end of the quarter
IND	Industry Classification	Fama and French (1997)'s 48 industry classifications
lnAT	Book Size	Natural logarithm of total assets
lnMC	Market Size	Natural logarithm of market capitalizations
lnVOL	Trading Volume	Natural logarithm of trading volumes
MARGIN	Sales Margin	Cumulative earnings divided by cumulative sales in the past 4 quarters
MOM[-12;-2]	Momentum	Cumulative returns from month -12 to month -2
MOM[-12;-7]	Intermediate Momentum	Cumulative returns from month -12 to month -7
MOM[-36;-13]	Long-Term Reversal	Cumulative returns from month -36 to month -13
NOA	Net Operating Assets	Ratio of operating assets minus operating liabilities to book value of total assets
OLEV	Operating Leverage	Ratio of the annual change in unlevered earnings to the annual change in sales, where unlevered earnings are defined as earnings plus tax-adjusted interest expenses minus tax-adjusted interest expenses
PPE	PPE Change-to-Assets Ratio	Ratio of changes in property, plants, and equipment (PPE) to lagged book value of total assets
RET	Return	Market return
ROA	Return on Assets	Cumulative earnings in the last 4 quarters divided by total assets 4 quarters ago
ROE	Return on Equity	Cumulative earnings in the last 4 quarters divided by total equity 4 quarters ago
RON	Return on Net Operating Assets	Ratio of operating income after depreciation to lagged net operating assets
SALE	Asset Turnover	Cumulative sales in the last 4 quarters divided by total assets 4 quarters ago
SGA	SGA-to-Sales Ratio	Ratio of selling, general, and administrative (SGA) expenses to net sales
SP	Sales-to-Price Ratio	Ratio of sales divided by market capitalization
SV	Asset Turnover Volatility	Standard deviation of asset turnovers in the last 5 years
TTURN	Turnover	Stock trading volumes divided by market capitalizations

Panel C: Time-Series Predictors

Predictor	Description	Definition
dAEARN	Aggregate Earnings Growth	Annual change in aggregate earnings divided by the aggregate market capitalizations 4 quarters ago
dCON	Consumption Growth	Annual growth rate of real personal consumption expenditures per capita
dGDP	GDP Growth	Annual growth rate of real domestic products
dTFP	Total Factor Productivity Growth	Growth rate of total factor productivity in the production function
dWEALTH	Wealth Growth	Growth rate of household wealth

Table 16: Sample Selection

This table shows the sample selection procedures and the changes in observations and distinct firms in each procedure. Obs. is the abbreviation for observations. The sample period is 1972Q1 – 2022Q4.

Procedure	Obs.	Distinct Firms	Obs. (%)	Distinct Firms (%)
Initial	858,373	19,001	100.00%	100.00%
Less: securities that are not common shares, that don't have December fiscal ends, or that are not listed in NYSE, AMEX, or NASDAQ	-164,142	-6,048	-19.12%	-31.83%
Common shares with December fiscal ends listed in NYSE, AMEX, and NASDAQ	694,231	12,953	80.88%	68.17%
Less: observations with missing values	-78,285	-327	-9.12%	-1.72%
Observations without missing values	615,946	12,626	71.76%	66.45%

doi:10.6342/NTU202500807

Table 17: Summary Statistics

This table shows the number of observations (N), the mean (Mean), the SD (standard deviation), the 5th percentile (P5), the 25th percentile (P25), the median (P50), the 75th percentile (P75), and the 95th percentile (P95) of each variable and predictor. Panels A, B, and C show the summary statistics of the outcome variables, firm-level predictors, and time-series predictors, respectively. The sample period is 1972Q1 – 2022Q4. The variable definitions are in Table 15.

Panel A: Outcome Variables

Variable	N	Mean	SD	P5	P25	P50	P75	P95
EBETA	615,946	1.322	19.671	-21.383	-1.574	0.554	3.566	26.741
TBETA	615,946	0.025	0.973	-1.038	-0.099	0.006	0.130	1.127
WBETA	615,946	-0.075	4.797	-4.684	-0.465	-0.001	0.406	4.437

Panel B: Firm-Level Predictors

Predictor	N	Mean	SD	P5	P25	P50	P75	P95
AG	615,946	0.003	0.197	-0.239	-0.076	-0.037	0.016	0.484
AGE	615,946	0.007	0.203	-0.187	-0.117	-0.053	0.054	0.448
BETA	615,946	0.000	0.215	-0.351	-0.086	-0.014	0.050	0.467
BM	615,946	-0.001	0.223	-0.319	-0.126	-0.032	0.066	0.508
dEARN	615,946	0.001	0.179	-0.323	-0.040	-0.006	0.031	0.371
DIVPAY	615,946	0.007	0.268	-0.170	-0.134	-0.119	0.040	0.657
DP	615,946	0.006	0.254	-0.128	-0.082	-0.076	-0.069	0.800
EP	615,946	0.004	0.203	-0.416	-0.043	0.022	0.093	0.310
EPCOV	615,946	0.001	0.192	-0.267	-0.068	-0.022	0.024	0.418
EPVAR	615,946	-0.003	0.225	-0.419	-0.115	-0.006	0.092	0.447
EV	615,946	0.000	0.170	-0.147	-0.052	-0.031	-0.023	0.334
FLEV	615,946	0.000	0.252	-0.191	-0.120	-0.091	-0.034	0.693
FXC	615,946	-0.002	0.211	-0.238	-0.094	-0.043	0.001	0.484
ILLIQ	615,946	0.001	0.220	-0.095	-0.083	-0.074	-0.050	0.580
lnAT	615,946	0.005	0.240	-0.432	-0.123	-0.007	0.123	0.484
lnMC	615,946	0.007	0.248	-0.425	-0.141	-0.012	0.147	0.502
lnVOL	615,946	0.001	0.244	-0.491	-0.122	0.012	0.145	0.423
MARGIN	615,946	0.001	0.195	-0.370	-0.002	0.031	0.066	0.190
MOM[-12;-2]	615,946	0.002	0.216	-0.360	-0.096	-0.011	0.076	0.464
MOM[-12;-7]	615,946	0.002	0.216	-0.374	-0.093	-0.007	0.080	0.450
MOM[-36;-13]	615,946	0.003	0.202	-0.315	-0.067	-0.017	0.032	0.450
NOA	615,946	0.003	0.256	-0.528	-0.114	0.040	0.160	0.385
OLEV	615,946	0.000	0.167	-0.303	-0.026	-0.005	0.026	0.317
PPE	615,946	0.002	0.154	-0.155	-0.039	-0.021	-0.011	0.302

RET	615,946	0.001	0.217	-0.397	-0.099	-0.007	0.092	0.432
ROA	615,946	0.003	0.197	-0.424	-0.021	0.016	0.077	0.282
ROE	615,946	0.003	0.190	-0.398	-0.030	0.019	0.069	0.315
RON	615,946	0.002	0.171	-0.331	-0.022	0.007	0.031	0.311
SALE	615,946	0.002	0.237	-0.290	-0.142	-0.031	0.072	0.552
SGA	615,946	-0.002	0.194	-0.245	-0.067	-0.032	-0.004	0.430
SP	615,946	-0.000	0.222	-0.198	-0.118	-0.069	0.008	0.574
SV	615,946	0.001	0.184	-0.209	-0.054	-0.030	-0.014	0.414
TTURN	615,946	0.001	0.234	-0.241	-0.137	-0.062	0.042	0.602

Panel C: Time-Series Predictors

Predictor	N	Mean	SD	P5	P25	P50	P75	P95
dAEARN	204	0.002	0.005	-0.003	-0.000	0.002	0.005	0.012
dCON	204	0.021	0.023	-0.018	0.011	0.021	0.032	0.051
dGDP	204	0.028	0.024	-0.016	0.017	0.029	0.042	0.062
dTFP	204	0.008	0.033	-0.044	-0.010	0.008	0.027	0.057
dWEALTH	204	0.071	0.037	0.024	0.048	0.069	0.097	0.128

Table 18: Distribution of Industries and Years

This table shows the distribution of industries defined in Fama and French (1997) and years. Panel A shows the distribution of industries. Panel B shows the distribution of years. The sample period is 1972Q1 – 2022Q4. The variable definitions are in Table 15.

Panel A: Industry Distribution

Fama-French industry code (48 industries)	Freq.	Percent
Agriculture	1498	0.24
Food Products	7051	1.14
Candy & Soda	1555	0.25
Beer & Liquor	1844	0.30
Tobacco Products	1152	0.19
Recreation	5206	0.85
Entertainment	7430	1.21
Printing and Publishing	4973	0.81
Consumer Goods	9989	1.62
Apparel	6134	1.00
Healthcare	10645	1.73
Medical Equipment	16448	2.67
Pharmaceutical Products	41899	6.80
Chemicals	11732	1.90
Rubber and Plastic Products	4959	0.81
Textiles	3703	0.60
Construction Materials	13687	2.22
Construction	7428	1.21
Steel Works Etc	8088	1.31
Fabricated Products	2357	0.38
Machinery	18767	3.05
Electrical Equipment	7722	1.25
Automobiles and Trucks	9402	1.53
Aircraft	3048	0.49
Shipbuilding, Railroad Equipment	1130	0.18
Defense	1156	0.19
Precious Metals	2624	0.43
Non-Metallic and Industrial Metal Mining	3311	0.54
Coal	1438	0.23
Petroleum and Natural Gas	28070	4.56
Utilities	23566	3.83
Communication	16815	2.73
Personal Services	5445	0.88
Business Services	58700	9.53
Computers	16207	2.63
Electronic Equipment	24652	4.00
Measuring and Control Equipment	9943	1.61
Business Supplies	6882	1.12

doi:10.6342/NTU202500807

Shipping Containers	2930	0.48
Transportation	16204	2.63
Wholesale	16294	2.65
Retail	14522	2.36
Restaraunts, Hotels, Motels	10026	1.63
Banking	70848	11.50
Insurance	24287	3.94
Real Estate	6285	1.02
Trading	34951	5.67
Almost Nothing	12943	2.10
Total	615946	100.00

Panel B: Year Distribution

Year	Freq.	Percent
1972	6232	1.01
1973	6365	1.03
1974	6383	1.04
1975	6351	1.03
1976	6347	1.03
1977	6160	1.00
1978	5953	0.97
1979	5837	0.95
1980	5816	0.94
1981	5724	0.93
1982	8861	1.44
1983	10270	1.67
1984	10588	1.72
1985	10770	1.75
1986	11679	1.90
1987	12369	2.01
1988	12551	2.04
1989	12654	2.05
1990	12930	2.10
1991	13492	2.19
1992	14274	2.32
1993	15481	2.51
1994	16238	2.64
1995	16766	2.72
1996	18743	3.04
1997	18390	2.99
1998	17428	2.83
1999	18034	2.93
2000	17259	2.80
2001	16087	2.61
2002	15385	2.50

2003	15047	2.44
2004	14877	2.42
2005	14514	2.36
2006	14243	2.31
2007	13717	2.23
2008	13329	2.16
2009	12936	2.10
2010	12483	2.03
2011	12060	1.96
2012	11994	1.95
2013	12271	1.99
2014	12287	1.99
2015	11898	1.93
2016	11654	1.89
2017	11616	1.89
2018	11591	1.88
2019	11606	1.88
2020	12172	1.98
2021	12256	1.99
2022	11978	1.94
Total	615946	100.00

doi:10.6342/NTU202500807

Table 19: Forecast Errors

This table shows the forecast errors of the rolling regression, the OLS model with all the predictors, the Lasso regression, the Ridge regression, and the Elastic Net. Panel A shows the time-series average of the quarterly market-capitalization-weighted root mean squared errors (RMSEs). Panel B shows the time-series average of the quarterly market-capitalization-weighted root mean squared hedging errors (RMSHEs). The sample period is 1972Q1 – 2022Q4. The variable definitions are in Table 15.

Panel A: Value-Weighted RMSEs

Model	Rolling Regression	OLS	Lasso	Ridge	Elastic Net
EBETA	8.279	652.546	8.311	8.314	8.042
TBETA	0.387	29.014	0.368	0.383	0.368
WBETA	1.729	177.607	1.784	1.896	1.772

Panel B: Value-Weighted RMSHEs

Model	Rolling Regression	OLS	Lasso	Ridge	Elastic Net
EBETA	0.107	20.588	0.132	0.168	0.106
TBETA	0.070	0.837	0.070	0.071	0.070
WBETA	0.073	4.922	0.075	0.081	0.074

Table 20: Asset Pricing Tests

This table shows the Fama-McBeth regression results of monthly returns on the fundamental betas estimated using different models. Panel A, B, C, D, and E show the results of Rolling Regression, OLS, Lasso, Ridge, and Elastic Net, respectively. The *, **, and *** stand for 10%, 5%, and 1% statistical significances, respectively. The abbreviation cons stands for the constant term or the intercept of the regression. Adj. is the abbreviation for Adjusted. The sample period is 1972Q1 – 2022Q4. The variable definitions are in Table 15.

Panel A: Rolling Regression

	(1)	(2)	(3)
VARIABLES	RET	RET	RET
EBETA	0.000		0.000*
	(0.116)		(1.873)
TBETA		0.003	0.002
		(1.161)	(0.783)
WBETA		-0.001	-0.004*
		(-0.307)	(-1.879)
cons	0.012***	0.012***	0.012***
	(5.594)	(5.575)	(5.587)
Observations	707,734	707,734	707,734
R-squared	0.005	0.007	0.010
Adj. R-squared	0.00388	0.00514	0.00789

Panel B: OLS

	(1)	(2)	(3)
VARIABLES	RET	RET	RET
EBETA	0.000		0.000
	(0.448)		(0.860)
TBETA		0.000	0.000
		(1.051)	(1.156)
WBETA		0.000	-0.000
		(0.217)	(-0.018)
cons	0.012***	0.012***	0.012***
	(5.453)	(5.450)	(5.452)
Observations	707,734	707,734	707,734
R-squared	0.002	0.003	0.005
Adj. R-squared	0.000795	0.00156	0.00237

Panel C: Lasso

VARIABLES	(1) RET	(2) RET	(3) RET
EBETA	0.000*		0.000*
	(1.695)		(1.657)
TBETA		-0.000	-0.000
		(-0.264)	(-0.328)
WBETA		0.000	0.000
		(0.129)	(0.299)
cons	0.012***	0.012***	0.012***
	(5.425)	(5.442)	(5.421)
Observations	707,734	707,734	707,734
R-squared	0.003	0.004	0.007
Adj. R-squared	0.00257	0.00240	0.00476

Panel D: Ridge

VARIABLES	(1) RET	(2) RET	(3) RET
EBETA	0.000**		0.000
	(1.990)		(1.461)
TBETA		-0.000	0.000
		(-0.209)	(0.502)
WBETA		-0.000**	-0.000*
		(-2.019)	(-1.810)
cons	0.012***	0.012***	0.012***
	(5.393)	(5.411)	(5.378)
Observations	707,734	707,734	707,734
R-squared	0.004	0.006	0.009
Adj. R-squared	0.00307	0.00407	0.00695

Panel E: Elastic Net

VARIABLES	(1) RET	(2) RET	(3) RET
EBETA	0.000		0.000
	(0.908)		(0.678)
TBETA		-0.000	-0.000
		(-0.123)	(-0.351)
WBETA		0.000	0.000
		(0.077)	(0.390)

doi:10.6342/NTU202500807

cons	0.012***	0.012***	0.012***
	(5.476)	(5.433)	(5.454)
Observations	707,734	707,734	707,734
R-squared	0.004	0.004	0.007
Adj. R-squared	0.00277	0.00240	0.00496

doi:10.6342/NTU202500807