Does Better Information Favor Humans or Machines? Evidence from Global Earnings Forecasts

YUHAN YE

November 13, 2025

[Click here for the latest version]

Abstract

This paper investigates whether better information favors humans or machines. Using detailed analyst forecasts and self-built machine-learning predictions across 47 global stock markets from 1985 to 2024, I examine how information environments shape the relative forecasting performance of human analysts and algorithms. Controlling for year and industry effects, human forecasts remain relatively more accurate in stronger information environments, characterized by richer data availability, greater transparency, and better access to information. This pattern indicates that within the same year, moving from data-poor to data-rich environments amplifies humans' comparative advantage. Humans appear better able to extract and interpret qualitative and contextual information, or soft information that machines often miss, when credible data and transparent institutions are present. Over time, both humans and machines have become more accurate, but machines have improved more rapidly, narrowing the performance gap. These long-term trends are driven by technological progress and the expansion of publicly available data. Overall, richer information environments favor humans over machines in a cross-sectional sense, underscoring the central role of information quality in shaping human–machine complementarities in financial forecasting.

Keywords: Information Environment, Earnings Forecasts, Machine Learning, Transparency and Disclosure

JEL: G14, G17, C55, D83

YE is with the Swiss Finance Institute (SFI) at Università della Svizzera italiana (USI, Lugano). I am deeply grateful to Laurent Frésard for his guidance and support throughout this project and my doctoral studies. I thank François Derrien, Thierry Foucault, Loriano Mancini, Ernst Maug, Steven Ongena, Gordon Phillips, Marti G. Subrahmanyam, David Thesmar for helpful comments and suggestions. I also thank participants at various seminars and conferences for their feedback. Email: yuhan.ye@usi.ch. Preliminary draft. All errors are my own.

1. Introduction

Advances in machine learning have transformed financial forecasting, extending from earnings prediction to risk assessment. Algorithms now process vast datasets and uncover complex nonlinear relationships that often rival or even surpass human forecasts. Yet despite these advances, I find a striking pattern: as information environments improve, human analysts, rather than algorithms, benefit more. This finding challenges the view that technological progress merely substitutes for human judgment, suggesting instead that higher information quality enhances the value of contextual reasoning that algorithms cannot reproduce.

It is natural to expect that as information becomes richer and more transparent, algorithms designed to process large and structured data should benefit the most. Enhanced disclosure, standardized reporting, and digital accessibility indeed increase the volume of machine-readable information. Yet the evidence reveals the opposite pattern: humans benefit more from improved information. This apparent paradox motivates a closer examination of how the nature of information, ranging from hard to soft, shapes the comparative advantages of humans and algorithms.

Following Stein (2002) and Liberti and Petersen (2019), information can be viewed along a continuum from hard to soft. Hard information is quantitative and verifiable, while soft information is qualitative and contextual, requiring human interpretation. As modern data collection "hardens" information through automation, contextual richness may decline. Building on Black and Litterman (1990, 1992), I extend this idea to the information dimension: better information strengthens humans' ability to interpret context and extract meaning that algorithms cannot replicate.

This paper addresses a simple yet fundamental question: who benefits more from better information, humans or machines? Rather than asking whether machines outperform humans, I investigate how improvements in information quality amplify or reduce the relative advantage of human judgment across institutional settings. Specifically, I examine how the comparative forecasting performance of human analysts and machine learning models varies across countries that differ in information quality and transparency.

I develop a unified empirical framework that builds a machine analyst using state-of-the-art supervised learning models. The framework is implemented consistently across 47 markets, enabling a direct comparison of human and algorithmic forecasts under diverse institutional conditions. For each country, a separate machine analyst is trained exclusively on that market's historical and contemporaneous data, reflecting local information availability and institutional conditions. Each time a human analyst issues an earnings forecast, the corresponding machine analyst generates a parallel prediction based solely on information available up to that forecast date, ensuring identical information constraints and no look-ahead bias. To provide causal evidence, I exploit the 2018 MiFID II reform, which required asset managers to pay separately for research rather than bundle it with trading commissions. The reform sharply reduced analyst compensation and coverage, weakening incentives to produce qualitative research and soft information across Europe. To complement this soft-information shock, I also examine the staggered adoption of the *International Financial Reporting Standards* (*IFRS*) as a *hard-information shock*. While IFRS improves the structure and comparability of financial statements, the reform is followed by a decline in machine win rates, suggesting that analysts are better able than algorithms to translate richer hard information into value-relevant soft insights.

Over time, forecast errors for both humans and machines have declined worldwide, reflecting improvements in data accessibility and modeling techniques. Yet machines have improved more rapidly, gradually narrowing but not reversing their performance gap with human analysts within the shared short-horizon forecasts.

Using a global dataset covering 47 stock markets from 1985 to 2024, I show that richer information environments consistently enhance human relative performance in earnings forecasting, while machines play a complementary role where qualitative insights are scarce. I focus on short-horizon earnings forecasts ($\leq$ 1 year) because this horizon captures the most information-sensitive period relevant to investors' trading and valuation decisions. It is also where analyst coverage and forecast frequency are highest, providing the richest and most reliable data for a clean human–machine comparison.

In essence, humans outperform machines in richer information environments not because they process more data, but because they reason differently about its meaning. Analysts interpret contextual cues to infer which economic forces are at work, such as whether a shock is transitory or structural or whether a policy change signals a regime shift, whereas algorithms depend on a single estimated mapping between inputs and outcomes. This capacity to flexibly update and switch among competing models of the world, as described by Kurz (1994b), becomes increasingly valuable when information is credible, abundant, and interpretable.

This paper contributes in three ways. First, it introduces a new perspective on human–machine comparison by focusing on the marginal effects of information improvements rather than absolute performance. Second, it provides the first global evidence that institutional transparency systematically amplifies human comparative advantage in forecasting. Third, it identifies a causal link between information shocks such as MiFID II and the erosion of human advantage, showing that soft-information scarcity benefits machines.

The objective is not to search for the most powerful algorithm, but to compare humans and machines under a consistent and transparent modeling framework applied uniformly across countries and time.

The global earnings-forecast setting provides an ideal laboratory for this comparison, as it offers standardized expectations data across 47 markets where both analysts and algorithms operate under comparable information constraints, making it possible to test how transparency affects human–machine complementarity.

The distinction between publicly available and soft information is becoming increasingly blurred in the age of AI. Improvements in information environments may reflect not only greater availability of structured, public data, but also enhanced transparency and accessibility of contextual or qualitative information. To capture variation in information environments, I construct country-level measures that reflect both the quantity and the interpretability of available information. These indicators combine aspects of data accessibility, institutional transparency, and media freedom, providing a broad view of how public and contextual information evolve across countries and over time. I focus on country-level rather than firm-level disclosure measures because firm transparency is partly endogenous to analyst activity and thus unsuitable for testing how information environments shape the relative performance of humans and machines.

Together, the construction of an original machine analyst, the broad global coverage, and the strict control of information timing provide a clean and externally valid test of how information environments shape the relative advantage of humans and algorithms. These design features also distinguish this study from prior work focused on a single, already transparent market.

The remainder of the paper is organized as follows. Section 2 describes the global dataset, the construction of forecasting targets, and the measurement of information environments. Section 3 presents the hypotheses to be tested. Section 4 documents the main results on how information quality shapes the relative performance of humans and machines. Section 5 provides causal evidence from the MiFID II reform and from IFRS adoption. Section 6 concludes and discusses implications for market efficiency and human–AI interaction.

Related Literature

This paper contributes to the growing literature comparing human and machine forecasting. Early studies such as Ball and Ghysels (2018) use mixed data sampling (MIDAS) regressions to generate automated earnings forecasts and show that these models contain information that complements analyst forecasts. Subsequent research including Van Binsbergen, Han, and Lopez-Lira (2023), Cao et al. (2024), and De Silva and Thesmar (2024) employs modern machine-learning techniques to examine when algorithms outperform or complement human analysts. While most existing evidence focuses on U.S. data, I provide the first global analysis showing that the relative advantage of humans and machines systematically varies with the richness and credibility of national information environments. Consistent with earlier findings, machines perform better in weak institutional settings where structured data are scarce and stable signals are valuable, whereas in richer and more transparent environments humans outperform by leveraging contextual and qualitative insights.

This paper also connects to the literature on forecast informativeness and horizon effects. Dessaint, Foucault, and Frésard (2021) show that the use of alternative data can reshape analysts' attention allocation, leading them to concentrate more effort on short-horizon forecasts where data are timelier and easier to exploit. My analysis focuses precisely on short-horizon forecasts within one year, consistent with this behavioral mechanism. By comparing human and machine performance across global information environments, I test whether improvements in data richness and transparency amplify the human advantage at short horizons.

Building on this comparative perspective, my paper also contributes to the theoretical literature on the nature of information. Following Stein (2002), information can be conceptualized along a continuum from hard to soft. Hard information refers to data that are quantitative, verifiable, and easily transmitted within organizations, whereas soft information is qualitative, contextual, and often embedded in human judgment. In hierarchical settings, soft information cannot be credibly communicated, which favors decentralized decision-making where agents rely on their own interpretation rather than codified data. Liberti and Petersen (2019) extend this distinction to modern financial markets, showing that advances in data collection and automation have increasingly hardened information by emphasizing quantification and standardization, often at the cost of contextual richness. This shift toward hard information benefits algorithmic decision-making but can reduce the scope for human judgment based on qualitative insight. Building on these ideas, my study examines whether improvements in information environments through better disclosure, transparency, and data accessibility strengthen or weaken the role of human interpretation. By relating Stein (2002) and Liberti and Petersen (2019) to the global forecasting context, I test whether richer and more credible information environments amplify humans' comparative advantage in processing contextual cues.

This perspective echoes Black and Litterman (1990, 1992), who emphasized that quantitative optimization should be conditioned on subjective human beliefs, suggesting that even in a data-rich world interpretation remains indispensable.

This study further contributes to the literature on expectation formation (e.g., Sims 2003; Woodford 2003; Landier, Ma, and Thesmar 2017; De Silva and Thesmar 2024) and the role of noise expectation in forecasting and human decision-making across various domains such as medicine, finance, hiring, and judicial decisions (Kahneman, Sibony, and Sunstein 2021). Subjective forecast noise has been extensively discussed in the literature on noisy information and behavioral economics.

Recent work by Fedyk et al. (2024) examines the behavioral alignment between human investors and large language models. Using experimental surveys comparing GPT-4 with human respondents, they find that AI-generated investment judgments mirror human reasoning about risk and return, achieving roughly 70% correlation with human preference rankings. However, AI responses tend to reproduce demographic biases, overrepresenting young, high-income males. This evidence highlights that while AI can imitate human judgment structures, it may also inherit systematic biases, reinforcing the need to distinguish algorithmic learning from contextual human-specific interpretation of information.

This study also resonates with the governance-based view of home bias proposed by Pinkowitz, Stulz, and Williamson (2001) and Dahlquist et al. (2003), extending its logic from the domain of asset ownership to that of forecast composition. Specifically, I show that analysts are more likely to generate value when they possess privileged access to local, non-public information. This result also connects to the literature on local analyst advantage, particularly Bae, Stulz, and Tan (2008). In doing so, my paper contributes to both strands of literature by introducing direct, quantitative measures of local informational advantage in the context of financial forecasting. These measures allow for a systematic evaluation of when and how locally embedded analysts outperform, thereby offering new empirical content to theories of home bias and local expertise.

Prior literature extensively documents the prevalence of analyst optimism, whereby analysts tend to overestimate firm earnings and stock values due to various incentives. These include the motivation to maintain access to management (Lim 2001), the desire to secure trading commissions and retain clients (Cowen, Groysberg, and Healy 2006), as well as career concerns and institutional affiliations (Hong and Kubik 2003a; Mola and Guidolin 2009), among a broader literature on analyst behavior. This optimistic bias has been identified as a form of analyst activism influencing market expectations. Extending this understanding, I leverage international datasets to uncover a complementary pattern in noisy forecasting environments. In these markets, characterized by less transparent information and greater uncertainty, analysts exhibit systematic pessimism, consistently underestimating firm earnings and stock prices. This finding suggests that analyst biases are context-dependent and that pessimism may dominate when informational noise impedes accurate forecasting.

Finally, this paper builds on the idea that forecast heterogeneity may arise not from irrationality but from rational differences in beliefs. As argued by Kurz (1994b,a), individuals may hold rational yet distinct models of how the world works, what he terms rational beliefs rather than rational expectations. Even when agents share the same public information, small uncertainty about the signal distribution can lead to persistent belief differences (Acemoglu, Chernozhukov, and Yildiz 2016). In this sense, soft information can be viewed as analysts' contextual interpretation of data under different perceived data-generating processes (DGPs). Humans' ability to select and update among competing DGPs provides a rational foundation for their advantage in richer information environments (see also Kandel and Pearson 1995).

2. Data and Methodology

2.1. Empirical Framework

The empirical design is implemented separately for each country. I train one individual *machine analyst* per market using that country's historical firm-level, financial, and macroeconomic data, and retrain the model annually in a rolling five-year window. This setup captures cross-country heterogeneity in information environments: markets with richer and more transparent data enable the machine analyst to learn more complex patterns, whereas data-scarce markets naturally constrain algorithmic performance.

The analysis is conducted at the detailed individual-forecast level. Whenever a human analyst j issues a forecast for firm i's annual earnings (EPS) on date t, the corresponding machine analyst generates a forecast on the same firm, for the same target variable, and with the same forecast horizon (measured in calendar days). This one-to-one matching ensures that both forecasts are made under identical information sets and temporal constraints, allowing a direct comparison of forecasting ability across 47 stock markets from 1985 to 2024.

This design provides a global yet locally grounded framework to evaluate whether richer information environments systematically favor humans or machines. It effectively simulates a team of country-specific machine analysts operating under the same data availability and timing constraints as their human counterparts.

Construction of Human–Machine Forecast Pairs

[Insert Figure 1 here]

Figure 1 illustrates the matching process. For each analyst forecast of next-year EPS, a corresponding machine forecast is generated on the exact same forecast-issue day to ensure perfect temporal alignment. Both the forecasted and realized EPS are scaled by the most recent stock price available before the forecast date, ensuring comparability across firms, years, and countries. Repeating this procedure across all firms, analysts, and markets yields a comprehensive global panel of human–machine forecast pairs that forms the empirical foundation of the study.

2.2. Data Sources, Processing, and Coverage

This study integrates three main categories of publicly available data: firm fundamentals, analyst forecasts, and market data.

Analyst Forecasts

Analyst earnings forecasts are obtained from the *I/B/E/S Global* database for 1985–2023, encompassing more than 20 million observations across 47 countries. To enhance comparability and minimize seasonal effects, I restrict the sample to firms with fiscal years ending on December 31. Both forecasted and realized EPS are scaled by the most recent stock price available prior to the forecast date, ensuring comparability across firms, years, and countries. Within each country and year, EPS forecasts are winsorized at the 5% level to mitigate the influence of outliers. Forecast horizons are measured precisely as the number of calendar days between the forecast issue date and the corresponding earnings announcement date. The analysis focuses on short-horizon forecasts (within one year), which are most relevant to investors and financial analysts.

Firm Fundamentals & Returns

Over 200 firm-level variables are sourced from Compustat Global and Datastream, including balance sheet items, profitability metrics, and market prices. For stock price and market capitalization data, I use CRSP for firms in the United States, and Eikon Datastream for other countries, ensuring consistent global coverage. The variable selection ensures contemporaneously observable items to avoid look-ahead bias. To control for industry-level heterogeneity, 2-digit SIC industry dummies are included as additional features. To avoid look-ahead bias, I rely on raw accounting items rather than financial ratios.[1] The variable selection therefore focuses on contemporaneously observable fundamentals available to analysts and investors at the time of forecasting. To account for sectoral heterogeneity, I include a set of 2-digit SIC industry dummies as additional features in the training data. A detailed list of all accounting, market, and return variables, as well as the subset required to be non-missing for inclusion, is provided in Appendix Note D.1.

[1] Many financial ratios were introduced in studies published after the sample training period, and their retrospective inclusion could implicitly incorporate future information.

Macroeconomic Variables

A set of 28 country-level macroeconomic indicators is obtained primarily from *Trading Economics*, covering key dimensions of real activity, inflation, and financial conditions. The variables include industrial production (IP), consumer price index (CPI), oil prices, short-term interest rates (3-month Treasury bill), long-term interest rates (10-year government bond), and credit spreads, among others. These indicators are collectively referred to as the macro dataset throughout the paper.

These macro variables provide time-varying information on aggregate economic and financial conditions, allowing the machine analyst to detect regime shifts such as expansions, recessions, and monetary tightening cycles in real time. All macro variables are aligned at the monthly frequency and lagged so that only information available prior to each forecast date is used in model training. This design preserves strict out-of-sample validity and mirrors the information set realistically accessible to both analysts and algorithms at the time forecasts are issued. The full list of macroeconomic indicators used in the model is provided in Section D.2.

Global Coverage

The final dataset spans 47 stock markets, 23 developed and 24 emerging economies, covering all MSCI ACWI constituent countries and approximately 85% of the global investable equity market (see Appendix Section C.1). This cross-country variation in transparency, disclosure regulation, and data infrastructure enables a powerful test of how national information environments shape the comparative advantage of humans and algorithms.

2.3. Machine Forecast Construction

To construct machine forecasts, I implement supervised learning algorithms trained exclusively on publicly observable *hard information*, including firm fundamentals, market indicators, and macroeconomic variables. These inputs reflect the structured data that algorithms can efficiently process, in contrast to the qualitative insights that human analysts may access through soft information channels. All models are estimated using a rolling five-year window and trained only on data available prior to the forecast date, ensuring strict out-of-sample validity and eliminating look-ahead bias.

For each firm i and analyst j, let $F^{\text{Human}}_{i,j,t,h}$ denote the human forecast issued at time t for horizon h, predicting the realized outcome $Y_{i,t+h}$. A corresponding machine forecast $F^{\text{Machine}}_{i,k,t,h}$ is generated by algorithm k, trained only on information observable before t. The detailed forecast unit (i,j,t,h) thus represents a one-to-one pairing of human and machine forecasts for the same firm, forecast date, and horizon.

$$(1) \qquad |FE^{\text{Human}}_{i,j,t,h}| = \left|F^{\text{Human}}_{i,j,t,h} - Y_{i,t+h}\right|, \qquad |FE^{\text{Machine}}_{i,k,t,h}| = \left|F^{\text{Machine}}_{i,k,t,h} - Y_{i,t+h}\right|.$$

I define the *Relative Performance Index* comparing absolute forecast errors:

$$(2) \qquad \text{RelPerf}_{i,j,t,h} = \frac{|FE^{\text{Machine}}_{i,k,t,h}| - |FE^{\text{Human}}_{i,j,t,h}|}{\max\left\{|FE^{\text{Machine}}_{i,k,t,h}|,\ |FE^{\text{Human}}_{i,j,t,h}|\right\}}.$$

A positive value indicates that the human forecast is more accurate.

The binary indicator of human dominance is:

$$(3) \qquad \text{Beat}_{i,j,t,h} = \begin{cases} 1, & \text{if } |FE^{\text{Human}}_{i,j,t,h}| < |FE^{\text{Machine}}_{i,k,t,h}|, \\ 0, & \text{otherwise.} \end{cases}$$

At the aggregated level (e.g., firm–year or country–year), I define the *Beat Ratio* as the fraction of forecasts in which humans outperform machines:

$$(4) \qquad \text{BeatRatio}_g = \frac{1}{N_g} \sum_{(i,j,t,h)\in g} 1\left(|FE^{\text{Human}}_{i,j,t,h}| < |FE^{\text{Machine}}_{i,k,t,h}|\right).$$

2.4. Benchmark Forecasting Models

Building on Van Binsbergen, Han, and Lopez-Lira (2020), Cao et al. (2021), and De Silva and Thesmar (2024), I implement a dual-model forecasting framework combining quasi-linear and non-linear algorithms. These algorithms are trained identically across countries and periods to ensure comparability, with hyperparameters selected through nested cross-validation within each rolling window.

- **Quasi-linear Models:** Lasso, Ridge, and Elastic Net regression for sparse high-dimensional data.
- **Non-linear Models:** Random Forests, Gradient-Boosted Trees, and Feedforward Neural Networks to capture nonlinearities and higher-order interactions.

The purpose is not to identify the most powerful predictive algorithm, but to evaluate humans and machines within a consistent and transparent modeling framework. Because the same procedures and inputs are applied uniformly across markets, any performance differences reflect variation in information environments rather than model capacity. A detailed discussion of the algorithms is provided in Appendix A.

2.5. Forecast Horizons

Following Dessaint, Foucault, and Frésard (2021), forecast horizons are computed as the number of calendar days between the forecast date and the earnings release date:

$$\text{Horizon} = \text{Earnings Release Date - Forecast Date}$$

For example, if an analyst issues an earnings forecast for Tesla on January 1st, and the actual earnings are released on March 31st, the forecast horizon is 89 calendar days.

This continuous measure provides a more precise indication of the time remaining until the realization of the target variable, compared to the traditional use of the `FPI` (Forecast Period Indicator) in I/B/E/S datasets, which only reflects the fiscal period being forecasted. It better captures the actual temporal distance faced by forecasters, either human or machine, at the time of prediction.

This paper focuses on short-term horizons, defined as forecasts with a horizon of less than one year (365 days). Short-term forecasts represent the most relevant horizon for both investors and equity analysts, as they directly relate to the upcoming fiscal year's earnings expectations and are closely tied to market valuation and trading decisions. To ensure comparability across firms and over time, the sample is further restricted to companies whose fiscal year ends on December 31, which aligns the forecast window across firms and avoids confounding variation arising from different fiscal calendars.

2.6. Summary Statistics

Table 1 reports summary statistics of human and machine forecast errors, their differences, and the resulting beat ratios across countries.

[Insert Table 1 here]

2.7. Information Environment Measures

To quantify cross-country differences in information environments, I draw on three complementary indicators from the *Quality of Government (QoG) Standard Dataset* (Teorell et al. 2025), covering 1980–2022. These include: (i) the *Data Services Score* (2016–2022), capturing data accessibility and openness; (ii) the *Transparency Index* (1980–2010), combining information and accountability subcomponents; and (iii) the *Access to Information Index* (2013–2021), measuring media freedom and public access to government information. Together, these measures capture both the quantity and quality of available data and the openness of national information systems.

3. Hypotheses

In this section, I outline the main hypotheses that guide the empirical analysis.

Conceptual Framework

Before stating the formal hypotheses, it is useful to outline the underlying conceptual framework. The framework is based on the idea that national information environments shape how humans and machines process information. Better data infrastructure, higher transparency, and broader access to information change not only the amount but also the nature of information available in capital markets.

Humans and algorithms use fundamentally different mechanisms to interpret this information. Humans are capable of contextual reasoning, qualitative assessment, and narrative understanding, which become increasingly valuable as soft information is released. Algorithms, in contrast, rely on structured and quantifiable data patterns. When information environments improve, more qualitative and contextual signals are available, benefiting human analysts disproportionately. When information becomes scarce or opaque, structured data becomes relatively more important, favoring machine learning models.

Conceptually, this distinction can be understood through the idea of rational heterogeneity: humans may interpret the same signals under different yet internally consistent models of the economy (Kurz 1994b) (Kurz 1994a). Even when information is public, small uncertainty about how signals are generated can lead to persistent differences in interpretation (Acemoglu, Chernozhukov, and Yildiz 2016). In this sense, soft information reflects analysts' contextual understanding of data under alternative perceived data-generating processes, rather than mere noise.

This conceptual distinction motivates the first hypothesis on how the information environment affects relative forecast performance.

3.1. Information Environment and Relative Performance

Humans and algorithms rely on fundamentally different mechanisms for processing information. Human analysts can integrate qualitative narratives, contextual cues, and cross-firm patterns into their judgments, whereas algorithms primarily extract patterns from structured numerical data. When information environments improve, through richer disclosures, greater transparency, or broader data access, more contextual and qualitative signals become available. Such improvements should enhance human forecasts disproportionately, as humans can interpret and contextualize these signals more effectively than algorithms.

Hypothesis 1 (Information Environment). Across countries, within the same year, richer information environments increase the *relative advantage of human analysts* over machine forecasts.

Formally,

$$\left.\frac{\partial(\text{Relative Performance}_{c,t})}{\partial(\text{Information Environment}_{c,t})}\right|_{t} > 0, \tag{5}$$

where c indexes countries and t indexes years. The derivative is taken within the same year, holding time effects constant, so that the comparison reflects cross-country differences in information environments rather than global or temporal trends.

As defined in Equation (**??**), the relative performance index compares human and machine forecast errors at the detailed-forecast level. Empirically, I measure relative performance using multiple complementary indicators introduced in Section **??**, including the *Beat Ratio* and the continuous *Relative Performance Index*. A higher value of either measure indicates superior human forecast accuracy.

Empirically, this prediction will be tested using three complementary measures of information quality: the *Data Services Pillar Overall Score*, the *Transparency Index*, and *Access to Information*. All three measures are expected to have positive coefficients when regressed on the relative performance of human over machine forecasts.

3.2. Temporal Trend in Forecast Accuracy

Hypothesis 2 (Temporal Catch-Up). Over time, improvements in data availability and computational technology enable machines to *catch up with humans faster* in forecasting accuracy. That is, forecast errors of both humans and machines decline over time, but the reduction is stronger for machines.

Formally,

$$\frac{\partial(\text{Relative Performance})}{\partial t} < 0, \tag{6}$$

Empirically, this pattern is reflected in a negative time-trend coefficient when regressing forecast errors on time, controlling for industry fixed effects. The finding suggests that technological advances and better data processing allow machine forecasts to improve at a faster rate than human forecasts.

The second hypothesis examines how forecast accuracy evolves over time. Advances in data infrastructure, disclosure standards, and computational technology are expected to improve the predictive precision of both humans and machines. Thus, while their relative performance may depend on institutional differences, the overall level of forecast accuracy should improve over time for both groups.

Taken together, these two hypotheses highlight that while institutional information environments continue to favor humans through richer soft-information channels, technological progress over time allows machines to close the performance gap more rapidly.

4. Empirical Findings

Section 4 presents the main empirical results. Across all specifications and measures of information environments, richer information infrastructures systematically enhance the relative performance of human analysts. In contrast, in weaker information environments, where qualitative insights are scarce and hard information is less reliable, machine forecasts play a complementary role by providing consistent and data-driven signals. I document consistent directional results across three alternative measures of national information environments: the *Data Service Score*, the *Transparency Index*, and the *Access to Information* indicator. The following subsections present these results in turn.

4.1. Forecast Errors and Beat Ratios over Time: Developed vs. Emerging Economies

Before turning to the regression analysis, it is useful to visualize how the forecasting accuracy of humans and machines has evolved over time. Figures 11 and 12 together provide descriptive evidence on these dynamics for representative developed and emerging economies.

Cross-country evolution in human dominance.

Figure 11 plots the average *beat ratio* over time for four countries (the United States, France, Indonesia, and Thailand). The beat ratio measures the fraction of firm–analyst–machine forecast pairs in which human analysts outperform the machine. Across all countries, human dominance remains evident but shows a gradual downward trend, especially in emerging markets. In developed economies such as the United States, the beat ratio remains high (around 0.85–0.9) and relatively stable, reflecting consistently strong analyst performance in well-developed data environments. In contrast, emerging economies such as Indonesia and Thailand display more volatility and a more pronounced decline, indicating that as digital infrastructures and data accessibility improve, algorithmic forecasts are catching up faster where information quality had previously been low.

[Insert Figure 11]

Convergence in forecast accuracy.

Figure 12 illustrates the evolution of forecast errors for human analysts and machine models in the United States and Thailand over the past two decades. Both countries display a clear downward trend in forecast errors for both humans and machines, reflecting substantial improvements in data availability, computational capacity, and institutional transparency. The continuous expansion of digital reporting platforms, the rise of big data analytics, and the global diffusion of accounting standards have collectively enhanced the overall efficiency of information processing in financial markets.

[Insert Figure 12 here]

Despite these parallel declines, the figure reveals that **machine forecasts have improved more rapidly than human forecasts**. Although machines still exhibit higher short-horizon forecast errors (within one year) than human analysts on average, which is the primary focus of this paper, the gap has been narrowing steadily over time. This convergence highlights the technological progress that enables algorithms to better capture nonlinear patterns and adapt to evolving information environments, even within fundamentally judgment-intensive tasks such as earnings forecasting.

Developed versus emerging economies.

The cross-country comparison further underscores the role of institutional development. In the United States, both human and machine forecast errors are substantially lower, and the convergence between the two has progressed more smoothly, suggesting that richer and more stable data ecosystems facilitate faster machine learning gains. In contrast, Thailand and other emerging economies exhibit greater volatility and a wider error gap, especially in earlier years, consistent with the view that less standardized disclosure and weaker data infrastructure still favor human analysts' contextual judgment. Overall, these dynamics confirm that improvements in data quality and accessibility benefit both humans and machines, yet the pace of convergence depends critically on the maturity of a country's information environment.

4.2. Empirical Design: Cross-Country Variation

To examine how national information environments shape the relative forecasting performance of humans and machines, I estimate the following panel regression model:

$$Y_{i,j,t,h} = \alpha + \beta\, \text{InfoEnvironment}_{c,t} + \gamma X_{i,t} + \mu + \epsilon_{i,j,t,h}, \tag{7}$$

where the dependent variable $Y_{i,j,t,h}$ is either the binary indicator $\text{Beat}_{i,j,t,h}$, equal to one if the analyst's forecast is more accurate than the machine's, or the continuous measure $\text{RelPerf}_{i,j,t,h}$, defined as the relative performance difference between human and machine forecast errors.

The key explanatory variable, $\text{InfoEnvironment}_{c,t}$, represents a country-level measure of the information environment, captured by one of three complementary indicators from the *World Economic Forum*: the *Data Services Pillar – Overall Score* (2016–2022), the *Transparency Index* (1980–2010), and the *Access to Information Index* (2013–2021). These measures vary both in their coverage periods and in the specific dimensions of information quality they capture. Together, they provide a comprehensive and complementary view of national information environments, ranging from institutional transparency reforms to digital data accessibility, with each measure also covering a distinct set of countries.

Accordingly, I estimate six regression specifications, combining two alternative outcome variables (Beat and RelPerf) with three information-environment proxies (Data Services, Transparency, and Access to Information). The vector $X_{i,t}$ includes firm- and macro-level control variables, and μ denotes a set of fixed effects (e.g., industry and year).

Across all specifications, the estimated coefficient β is consistently positive, indicating that richer information environments favor human analysts over machine forecasters. Detailed results and interpretations for each measure are presented in Sections 4.3, 4.4, and 4.5, corresponding to Tables 2, 3, and 4, respectively. In each table, Panel A reports results using the continuous outcome variable RelPerf, while Panel B reports results using the binary indicator Beat.

4.3. Data Services Score: Modern Digital Infrastructure (2016–2022)

The *Data Services Pillar – Overall Score* measures the quality of a country's digital data infrastructure, including the availability of financial data services, open data platforms, and real-time analytics accessibility. Covering the period 2016 to 2022, it captures the most recent phase of technological and institutional development that enables both humans and algorithms to access and process data more efficiently.

To examine how national information infrastructures influence the relative forecasting performance of human analysts and machine learning models, I estimate the following panel regression at the firm i, analyst j, date t, and forecast horizon h levels:

$$Y_{i,j,t,h} = \alpha + \beta \,\text{DataServicesScore}_{c,t} + \gamma X_{i,t} + \mu_f + \epsilon_{i,j,t,h}, \tag{8}$$

where $Y_{i,j,t,h}$ is the outcome variable. In **Panel A**, $Y_{i,j,t,h} = \text{RelPerf}_{i,j,t,h}$, defined as the difference between the machine's and the analyst's forecast errors, normalized by the larger of their absolute values, so that positive values indicate greater human accuracy. In **Panel B**, $Y_{i,j,t,h} = \text{Beat}_{i,j,t,h}$, an indicator equal to one if the analyst's forecast error is smaller than the machine's and zero otherwise. The key explanatory variable, $\text{DataServicesScore}_{c,t}$, measures the quality and accessibility of national data infrastructures for country c in year t. The control vector $X_{i,t}$ includes firm and macro-level characteristics as discussed in Section 2, and μ_f denotes industry and year fixed effects.

[Insert Table 2 here]

Table 2 reports the regression results. Column (2), which includes industry and year fixed effects, provides the cleanest cross-country identification and directly tests Hypothesis 1. Consistent with this hypothesis, the coefficient on *Data Service Score* is positive and highly significant, indicating that analysts outperform machines in countries with more developed data services. The estimated coefficient of 0.005 indicates that a one-point increase in the *Data Service Score* (ranging from 0 to 100) is associated with a 0.005 increase in relative forecast performance, equivalent to 0.5 percentage points, as shown in Panel A. Consistently, the probability that analysts outperform machines (*Beat* ratio) rises by 0.3%, as reported in Panel B. Both dimensions thus point to the same conclusion: countries with more reliable and accessible data infrastructures exhibit higher analyst accuracy relative to machine forecasts and a higher frequency of human wins.

Column (1), in contrast, includes industry fixed effects and a linear time trend, which allows me to examine the evolution of the human–machine gap over time, corresponding to Hypothesis 2. The estimated coefficient on the time-trend term is -0.014 in Panel A and -0.005 in Panel B, both significant at the 1% level. These estimates indicate that, on average, each additional year is associated with a 1.4% decline in relative forecast performance and a 0.5% decrease in the probability that humans outperform machines. This pattern shows that machine forecasts have improved more rapidly than human forecasts, gradually narrowing the accuracy gap as digital data availability and algorithmic techniques advanced. The decline reflects the increasing sophistication of machine-learning models and the growing access to structured financial data, which together have enhanced algorithmic forecasting accuracy at a faster pace than human learning, even though analysts still tend to outperform machines in short-horizon predictions.

Economically, these results suggest that richer data infrastructures strengthen analysts' ability to transform hard information into soft insights, while technological progress over time accelerates the pace at which machines catch up. In well-developed information environments, human analysts can contextualize standardized data and incorporate qualitative judgments, whereas algorithms remain constrained by structured numerical inputs.

Discussion of Significant Accounting Variables

In this subsection, I examine the key firm-level accounting variables that significantly correlate with the relative forecasting performance of human analysts and machine learning models. These variables capture underlying balance-sheet strength, earnings quality, and managerial policies that may systematically influence analysts' and algorithms' predictive accuracy. By interpreting their economic implications, I aim to highlight the plausible firm-level mechanisms that drive the observed performance gap between human and machine forecasts.

Among the firm-level controls,[2] Table 2 reports the coefficients on the key accounting variables. These controls are included to absorb firm-level heterogeneity in financial structure and reporting practices, rather than to explain cross-country differences in information environments. They serve to isolate the effects of national information infrastructures from firm-specific accounting characteristics.

Their signs provide intuitive checks consistent with the information-structure interpretation: human analysts perform better when interpretation and qualitative judgment matter more (e.g., high cash holdings or repurchases), and their advantage diminishes when firms are easier to value using standardized data (e.g., high current assets). Similar patterns are observed in the other information environment regressions in Table 3 and 4.

[2] About 150 variables, not all reported due to space constraints.

Firms with higher *current assets* (*ACT*) exhibit a smaller human advantage, as they are easier to value and involve less information asymmetry. When balance sheets are dominated by liquid and standardized assets, financial statements already convey most relevant information in a form that algorithms can efficiently process. Consequently, the relative value of analysts' qualitative judgment and soft information diminishes.

Greater *cash holdings* (*CHE*) are associated with stronger analyst outperformance, consistent with analysts' superior ability to interpret firms' strategic flexibility. High cash reserves provide firms with multiple strategic options such as acquisitions, share repurchases, or precautionary savings, whose implications depend on qualitative assessments of management intent and market conditions. Such contextual interpretation is difficult for algorithms that rely purely on quantitative inputs, but it is precisely where analysts' soft information and forward-looking judgment add value.

Acquisition expenditures (*AQC*) show no significant effect, suggesting that merger events pose similar challenges for both humans and machines.

More *repurchase activity* (*PRSTKC*) is linked to stronger analyst dominance, reflecting their qualitative understanding of signaling motives. Repurchases can convey diverse messages, ranging from managerial confidence and perceived undervaluation to short-term earnings management, whose interpretation depends on contextual and qualitative assessments. Analysts can differentiate between genuine value signals and cosmetic buybacks by incorporating information from management communication, industry dynamics, and governance cues, while such nuance is difficult for algorithms relying purely on numerical inputs.

Higher *special items* (*SPI*) weaken the human edge, as one-off accounting charges add noise and reduce predictability. These nonrecurring items often reflect temporary shocks such as impairments, restructuring costs, or legal settlements that obscure the underlying earnings process. When transitory accounting adjustments dominate reported figures, both human and machine forecasts become less accurate, and analysts' qualitative advantage diminishes because soft information offers limited guidance in noisy environments.

Finally, stronger *recent returns* (*RET*) correspond to greater human dominance, indicating that analysts can better interpret market sentiment and contextual cues beyond raw price data. Machines can recognize that prices have increased, but they cannot distinguish whether such movements reflect genuine improvements in fundamentals or temporary investor optimism. Analysts can incorporate qualitative information from earnings calls, news coverage, and macroeconomic developments to interpret the nature of price changes, which enhances their forecasting accuracy when markets are sentiment-driven.

Taken together, Column (2) supports Hypothesis 1 that cross-country information quality strengthens human performance relative to machines, while Column (1) supports Hypothesis 2, revealing a gradual, time-trend-driven improvement in human comparative advantage. These findings highlight that richer information environments enhance the role of human interpretation rather than mechanical data processing, underscoring the enduring value of contextual reasoning even in the age of digital information.

4.4. Transparency Index: Institutional Disclosure Reforms (1980–2010)

The *Transparency Index* captures the early phase of institutional disclosure reforms between 1980 and 2010, a period marked by significant improvements in financial reporting standards, corporate governance, and public access to fiscal and political information. This index combines two subcomponents: the *Information Transparency Index*, which measures media freedom, fiscal openness, and political constraints, and the *Accountability Transparency Index*, which captures the strength of accountability mechanisms and checks on political power. Together, they provide a comprehensive measure of how institutional transparency and disclosure credibility shape national information environments.

To examine how institutional transparency reforms shaped the relative forecasting performance of human analysts and machine learning models, I estimate the following panel regression at the firm i, analyst j, date t, and forecast horizon h levels:

$$Y_{i,j,t,h} = \alpha + \beta \, \text{TransparencyIndex}_{c,t} + \gamma X_{i,t} + \mu_f + \epsilon_{i,j,t,h}, \tag{9}$$

where $Y_{i,j,t,h}$ is the outcome variable. In **Panel A**, $Y_{i,j,t,h} = \text{RelPerf}_{i,j,t,h}$, defined as the difference between the machine's and the analyst's forecast errors, normalized by the larger of their absolute values, such that positive values indicate greater human accuracy. In **Panel B**, $Y_{i,j,t,h} = \text{Beat}_{i,j,t,h}$, an indicator equal to one if the analyst's forecast error is smaller than the machine's and zero otherwise. The key explanatory variable, $\text{TransparencyIndex}_{c,t}$, reflects the quality of institutional disclosure for country c in year t. The control vector $X_{i,t}$ includes firm and macro-level variables as described in Section 2, and μ_f

represents industry and year fixed effects.

[Insert Table 3 here]

Table 3 reports the regression results. Column (2), which includes industry and year fixed effects, provides the cleanest cross-country identification and directly tests Hypothesis 1. Consistent with this hypothesis, the coefficient on *Transparency Index* is positive and highly significant, suggesting that when disclosure systems are more credible and informative, human analysts gain a clear advantage over purely data-driven algorithms. The estimated coefficient of 0.005 implies that a one-point increase in the *Transparency Index* (ranging from 0 to 100) is associated with a 0.005 increase in relative forecast performance, equivalent to 0.5 percentage points, as shown in Panel A. Consistently, the probability that analysts outperform machines (*Beat* ratio) rises by 0.2%, as reported in Panel B. Both dimensions thus point to the same conclusion: stronger and more transparent disclosure regimes systematically favor human analysts over machine forecasts.

Column (1), which includes industry fixed effects and a linear time trend, allows for an assessment of how the human–machine gap evolved during the early decades of institutional reform. The estimated coefficients on the time-trend term are 0.003 in Panel A and 0.002 in Panel B, both significant at the 10% level. The positive and significant time-trend coefficients indicate that human forecasts improved relative to machine forecasts over this period, reflecting analysts' growing ability to process expanding disclosure information and to integrate qualitative context into their predictions. This finding supports Hypothesis 2, suggesting that as transparency reforms advanced, analysts increasingly learned to interpret and synthesize newly available institutional data. Notably, the direction of this time trend contrasts with that in Tables 2 and 4, a pattern further discussed in Section 4.6.

From an economic perspective, these results suggest that institutional transparency reforms enhanced the informational foundation upon which human analysts build their qualitative judgments. Analysts benefit more from credible disclosure frameworks because they can contextualize quantitative signals and detect soft information embedded in corporate communication and governance structures. In contrast, algorithms remain limited to codified numerical inputs and cannot fully capture subtle institutional or narrative cues that accompany reform-driven disclosure improvements.

Among the significant accounting control variables, the signs and significance levels are broadly consistent with those reported in Table 2. Overall, these patterns align with the mechanisms discussed in subsection 4.3, indicating that transparency reforms strengthened analysts' ability to exploit soft information and that credible institutional disclosure enhances the value of human interpretation relative to machine prediction. For brevity, the detailed discussion is not repeated here.

4.5. Access to Information Index: Broadening Information Channels (2013–2021)

The *Access to Information Index* covers the intermediate period from 2013 to 2021, representing an era of rapid expansion in digital communication channels, open government data, and online disclosure platforms. This measure captures how easily various stakeholders, including investors, analysts, and firms, can obtain both public and private information in a timely and equitable manner. It therefore reflects the extent to which information accessibility, rather than institutional credibility, shapes national information environments.

To examine how the broadening of information channels influenced the relative forecasting performance of human analysts and machine learning models, I estimate the following panel regression at the firm i, analyst j, date t, and forecast horizon h levels:

$$Y_{i,j,t,h} = \alpha + \beta \, \text{AccessInfoIndex}_{c,t} + \gamma X_{i,t} + \mu_f + \epsilon_{i,j,t,h}, \tag{10}$$

where $Y_{i,j,t,h}$ is the outcome variable. In **Panel A**, $Y_{i,j,t,h} = \text{RelPerf}_{i,j,t,h}$, defined as the difference between the machine's and the analyst's forecast errors, normalized by the larger of their absolute values, such that positive values indicate greater human accuracy. In **Panel B**, $Y_{i,j,t,h} = \text{Beat}_{i,j,t,h}$, an indicator equal to one if the analyst's forecast error is smaller than the machine's and zero otherwise. The key explanatory variable, $\text{AccessInfoIndex}_{c,t}$, measures the degree of accessibility and openness of financial and institutional information for country c in year t. The control vector $X_{i,t}$ includes firm and macro-level variables as described in Section 2, and μ_f represents industry and year fixed effects.

[Insert Table 4 here]

Table 4 reports the regression results. Column (2), which includes both industry and year fixed effects, provides the most stringent cross-country specification and directly tests Hypothesis 1. Consistent with this hypothesis, the coefficient on *Access to Information Index* is positive and highly significant at the 1% level in both panels, with $\beta = 0.029$ in Panel A and $\beta = 0.014$ in Panel B. These results indicate that broader and more equitable access to information systematically favors human analysts, as they are better able to interpret contextual nuances and integrate diverse data sources into coherent forecasts.

Column (1), which includes industry fixed effects and a linear time trend, allows for an assessment of the human–machine performance gap over time during the digital expansion period. The estimated coefficient on the time-trend term is -0.004 in Panel A (significant at the 1% level) and insignificant in Panel B, indicating that human analysts' relative advantage declined modestly over time as digital data access became increasingly standardized and algorithm friendly. The direction of this time trend is consistent with that reported in Tables 2, suggesting a continuation of the recent phase in which technological progress and real-time data availability have gradually narrowed the human–machine performance gap.

From an economic perspective, these findings imply that when information becomes broadly accessible yet remains partially unstructured, human analysts continue to retain a comparative advantage through contextual judgment and the integration of qualitative signals. However, as digital access further improves and data availability becomes more automated, the marginal benefit of human interpretation diminishes relative to machine-driven forecasts.

Among the accounting control variables, the signs and significance levels are broadly consistent with those reported in Table 2. Overall, these patterns align with the mechanisms discussed in subsection 4.3, and the detailed discussion is omitted here for brevity.

4.6. Interpreting Opposite Time Trends

The regression results reveal opposite time-trend coefficients across information environment measures, reflecting distinct historical phases of human–machine forecasting dynamics.

The **Transparency Index (1980–2010)** captures the early period of institutional transparency reforms, during which improvements in disclosure quality and governance openness enhanced analysts' relative advantage through richer qualitative and contextual information.

In contrast, the **Data Service Score (2016–2022)** represents the recent digital era, characterized by the expansion of real-time data infrastructure, standardized APIs, and the rapid diffusion of machine-learning tools, which substantially reduced the informational edge of human analysts.

Situated between these two stages, the **Access to Information Index (2013–2021)** marks a transitional phase in which both institutional openness and digital access expanded, but with persistent cross-country heterogeneity in data usability and coverage.

Taken together, these patterns suggest that the sign reversal of time trends mirrors **two distinct stages of human–machine evolution**: the institutional transparency stage that favored human expertise, and the digital data stage that increasingly leveled the playing field for algorithmic forecasts.

5. Causal and Dynamic Extensions

To deepen the causal interpretation of the preceding results, this chapter examines how major regulatory reforms can serve as quasi-natural experiments that reveal the mechanisms behind the human–machine forecasting asymmetry. The analysis focuses on two landmark information reforms that both reduced analysts' relative advantage, but through distinct channels of information transmission.

The first is the *IFRS adoption*, which strengthens the *hard-information* channel by standardizing accounting data and improving cross-country comparability. This reform enhances the structure and consistency of available data, making it easier for algorithms to process and learn from quantitative information.

The second is the 2018 implementation of *MiFID II*, which weakens the *soft-information* channel by restricting analyst–firm interactions and reducing the flow of qualitative insights. This reform limits analysts' access to contextual and narrative signals that previously underpinned their comparative advantage.

Together, these two reforms disadvantage human analysts through opposite mechanisms: one by enhancing machines through greater data standardization, and the other by constraining humans through reduced qualitative access. Examining these shocks in tandem provides a clean causal lens to understand when and why humans outperform or fall behind machines in complex information environments.

5.1. Information Shock: Evidence from MiFID II (2018 Reform)

The 2018 implementation of the *Markets in Financial Instruments Directive II* (*MiFID II*) provides a quasi-natural experiment to test the causal mechanism underlying the human–machine forecasting asymmetry. By unbundling research payments from trading commissions, the reform sharply reduced sell-side research coverage and the dissemination of qualitative insights across European markets. This setting thus constitutes a negative shock to the supply of *soft information*, allowing for a direct test of whether human analysts' advantage depends on the availability of contextual and narrative information.

Hypothesis. The MiFID II reform weakened analysts' access to soft information, thereby narrowing the human–machine performance gap in affected European markets.

To assess the causal effect of this reform on the relative forecasting performance of humans and machines, I estimate the following difference-in-differences specification at the firm–year level:

$$Y_{i,c,t} = \alpha + \beta \, did\_mifid_{c,t} + \gamma X_{i,t} + \mu_{\text{industry}\times t} + \epsilon_{i,c,t}, \tag{11}$$

where $Y_{i,c,t}$ denotes the *machine win rate*, defined as the proportion of forecasts in which the machine's absolute error is smaller than the analyst's. The treatment indicator $did\_mifid_{c,t}$ equals one for treated European countries (EU, EEA, the United Kingdom, and Switzerland) in post-2018 years and zero otherwise. The control vector $X_{i,t}$ includes the same firm- and macro-level covariates used in the baseline cross-country regressions, and $\mu_{\text{industry}\times t}$ denotes industry–year fixed effects. Standard errors are clustered at the country level or two-way clustered by country and firm.

[Insert Table 5 here]

Table 5 reports the results. In the baseline specification (Panel A), the coefficient on did_mifid is positive and statistically significant ($\beta = 0.030$), indicating that following the MiFID II reform, machines more frequently outperformed human analysts in treated markets. This result is consistent with the interpretation that restricting qualitative research channels weakened analysts' informational advantage. In the alternative specification (Panel B), which excludes low coverage firm years and expands the event window to ± 5 years, the estimated treatment effect becomes even larger ($\beta = 0.040$), reinforcing the robustness of the causal pattern.

From an economic standpoint, these results underscore the pivotal role of soft information in sustaining human analysts' advantage. When regulatory reforms reduce access to narrative and contextual insights, the precision of human forecasts declines, whereas machine learning models, which rely primarily on structured numerical data, remain largely unaffected or even relatively advantaged. This pattern supports the central mechanism of this paper: **human analysts excel when soft information is abundant but underperform when qualitative signals are constrained.**

Among the firm-level control variables, the signs and magnitudes are consistent with those reported in earlier cross-sectional analyses (Section 4.3). For instance, higher *current assets* (*ACT*) and stronger *depreciation* (*DP*) correlate with greater machine wins, while greater *cash holdings* (*CHE*) and higher *stock returns* (*RET*) are associated with stronger analyst performance. These patterns reinforce the interpretation that MiFID II primarily affected the qualitative component of financial information rather than the quantitative fundamentals captured by standard accounting variables.

Taken together, the MiFID II evidence provides causal validation for the cross-country findings in in Sections 4.3, 4.4, and 4.5. The reduction in soft-information availability shifted the balance of forecasting power toward machines, demonstrating that the human advantage hinges critically on access to contextual and interpretive signals.

5.2. Information Reform: Evidence from IFRS Adoption (Hard-information Shock)

To complement the preceding analysis of the MiFID II reform as a soft-information shock, this subsection examines the adoption of the *International Financial Reporting Standards* (*IFRS*) as a hard-information shock.

Together, these two quasi-natural experiments offer a unified perspective on how changes in the information environment shape the relative forecasting performance of humans and machines—one reform limiting qualitative insights, the other improving the structure and comparability of quantitative information. Between 2005 and the present, more than one hundred countries have adopted IFRS, standardizing accounting rules and enhancing the comparability of financial statements across markets. This reform improved the structure, transparency, and credibility of firm-level data, providing an exogenous enhancement of hard information available to both humans and algorithms.

Hypothesis. IFRS adoption improves the quality and consistency of accounting information, benefiting both humans and machines. However, if standardization also makes quantitative information easier to interpret and integrate with contextual judgment, human forecasts may improve even more. Thus, the relative advantage of humans is expected to strengthen following IFRS adoption.

To estimate the causal effect of IFRS adoption on the relative forecasting performance of humans and machines, I implement a staggered difference-in-differences design across countries with varying adoption years:

$$Y_{i,c,t} = \alpha + \beta\, did\_ifrs_{c,t} + \gamma X_{i,t} + \mu_{\text{country}} + \mu_t + \epsilon_{i,c,t}, \tag{12}$$

where $Y_{i,c,t}$ denotes the firm-year *machine win rate*, defined as the proportion of forecasts in which the machine's absolute error is smaller than the analyst's. The treatment indicator $did\_ifrs_{c,t}$ equals one for firm–year observations in countries that have adopted IFRS by year t and zero otherwise. The control vector $X_{i,t}$ includes the same firm- and macro-level covariates as in the baseline cross-country regressions, and standard errors are clustered at the country level. For robustness, I also estimate a two-way fixed-effects specification as a benchmark.

[Insert Table 6 here]

The staggered DiD estimates show a significant negative post-adoption effect on the machine win rate. The estimated average treatment effect is -0.191 (s.e. = 0.073, $p = 0.009$). This indicates that after IFRS adoption, the machine win rate decreased by about 19 percent, a statistically significant effect at the 1 percent level. In other words, humans improved their relative forecasting performance following the reform.

From an economic perspective, these findings suggest that while IFRS adoption improved the structure and comparability of quantitative information, it also enhanced the interpretability and usefulness of such data for human analysts. By increasing transparency and credibility, IFRS strengthened the informational foundation upon which analysts apply contextual judgment, thereby amplifying the human advantage in processing standardized yet economically nuanced information.

Taken together, IFRS adoption provides a complementary causal test to the MiFID II analysis. While MiFID II reduced the supply of soft information and disadvantaged human analysts, IFRS adoption expanded the supply of hard information yet unexpectedly strengthened human performance. Both reforms thus affect the human–machine relative performance through distinct channels: one by reducing qualitative inputs, the other by enhancing the structure of quantitative data in ways that improve human interpretability.

5.3. Discussion: Interpreting the Human Advantage

The empirical evidence shows that humans outperform machines in richer and more transparent information environments. This section discusses a possible mechanism that rationalizes this pattern. Conceptually, humans' advantage can be viewed as the ability to flexibly interpret signals under different mental models of how the world works, a form of soft information advantage that algorithms with fixed data-generating assumptions cannot easily replicate.

Machine-learning models typically assume a single, stable data-generating process (DGP):

$$y_t = f(X_t) + \varepsilon_t,$$

and estimate the mapping $f(\cdot)$ globally across samples. Human forecasters, by contrast, may entertain multiple plausible models of the world,

$$y_t = f_i(X_t) + \varepsilon_t, \quad i \in \{\text{cyclical, structural, policy-driven, ...}\},$$

and use contextual cues to infer which data-generating process is currently relevant. This model-switching interpretation aligns with the rational-beliefs framework of Kurz (1994b,a), in which agents hold internally consistent yet heterogeneous beliefs about the structure of the economy. Even when public information is shared, small uncertainty about the signal distribution can generate persistent heterogeneity in interpretations (Acemoglu, Chernozhukov, and Yildiz 2016).

In richer information environments, analysts can more accurately identify the correct model, reducing mis-specification and over-reaction. In opaque environments, they are more likely to misclassify shocks, leading to bias and noise. Therefore, improvements in data quality and transparency enhance humans' forecasting accuracy more than machines', consistent with the empirical findings.

One interpretation of the results is that human forecasters possess a comparative advantage in flexibly adapting their mental models of the data-generating process (DGP). As emphasized by Kurz (1994b,a) and Acemoglu, Chernozhukov, and Yildiz (2016), even small uncertainty about the signal distribution can generate persistent heterogeneity in beliefs. Analysts may therefore rationally differ in how they interpret the same public signals, depending on which DGP they consider most plausible at a given point in time. In richer and more transparent information environments, analysts can more effectively identify which model of the world is relevant, for example by distinguishing temporary shocks from structural changes and updating their forecasts accordingly. In contrast, machine-learning algorithms, which rely on a fixed functional mapping from inputs to outcomes, implicitly assume a stable DGP and therefore adapt more slowly to shifts in economic regimes or institutional structures. This perspective helps explain why better information environments favor humans, since credible and interpretable data enable them to update or switch across DGPs in ways that leverage contextual understanding rather than noise (see also Kandel and Pearson 1995).

5.4. Policy Implications

This paper offers several policy implications for financial regulation, institutional development, and AI governance.

Open Data Does Not Imply Human Redundancy

The evidence shows that better information environments strengthen, rather than replace, human expertise. When countries improve disclosure quality, data openness, and transparency, human analysts benefit more than algorithms. Open data initiatives, IFRS adoption, and disclosure reforms therefore enhance the effectiveness of human judgment rather than render it obsolete. Transparency should be viewed as a complement to human analysis, as it amplifies the value of interpretation, contextual understanding, and qualitative reasoning.

Lessons from MiFID II

The 2018 MiFID II reform, which unbundled research payments, led to reduced sell-side coverage and a decline in qualitative insights. This episode shows that when soft-information channels are constrained, human forecast accuracy deteriorates relative to machines. Regulators should therefore balance cost transparency with incentives for analysts to produce contextual and interpretative research. Sustaining a healthy ecosystem for human expertise is essential for maintaining informational efficiency in capital markets.

Institutional Transparency and Market Development

In highly transparent institutional environments, analysts are more effective at interpreting macroeconomic and policy signals, while in opaque or noisy environments, machine forecasts tend to be relatively more reliable due to their consistency and data-processing strength. In the short run, emerging markets may rely on algorithmic forecasts to compensate for weak data infrastructure. In the long run, investments in institutional transparency and information systems strengthen human capital and enhance the comparative advantage of professional judgment. Humans and machines are complementary, not substitutes, across stages of development.

AI Governance and Market Design

The findings indicate that differences in AI performance arise primarily from the structure of information environments rather than from algorithmic sophistication itself. Governing AI in finance requires governing the information environment by ensuring both data availability and standardization, while preserving the soft dimensions of information such as narrative and context. The optimal direction is to integrate human and algorithmic strengths: letting machines process what is measurable and humans interpret what is meaningful. Encouraging human-in-the-loop systems and training analysts to collaborate effectively with AI tools can maximize social and informational welfare.

6. Conclusion

This paper documents a fundamental asymmetry in how humans and machines respond to richer information environments. Across countries, I find that in settings with stronger data infrastructure, higher institutional transparency, and broader access to information, **human analysts consistently outperform machine forecasters**. These environments do not merely increase the quantity of available data; they enhance its quality and interpretability. When credible and contextual information becomes abundant, humans can better exploit their comparative advantage in judgment, synthesis, and narrative understanding. In contrast, machines, which are designed to detect structured numerical patterns, benefit less from such qualitative improvements.

Over time, both humans and machines improve as technology and data accessibility advance, but **machines catch up faster**. This temporal convergence does not contradict the cross-sectional evidence; rather, it highlights two dimensions of the same process. Technological progress strengthens machine learning models' ability to absorb structured signals, while institutional reforms strengthen humans' ability to interpret soft and contextual information. Together, they reveal an evolving pattern of **human–machine complementarity**: better institutions favor humans cross-sectionally, whereas technological progress favors machines over time.

Evidence from quasi-natural experiments, such as the 2018 MiFID II reform, reinforces this mechanism. When soft information channels contract, the human advantage weakens and machines gain relative ground, consistent with the view that analysts' strength lies in contextual insight. Similarly, the adoption of IFRS in treated countries, which expanded and standardized public disclosure, helped machines narrow their gap with human analysts by improving the quality and comparability of structured financial information. By contrast, when broader transparency reforms enhance both data credibility and interpretability, human forecasters regain their edge.

Taken together, these findings refine our understanding of how technology interacts with information institutions. Algorithms perform well in data-poor yet stable environments by exploiting historical regularities, while humans excel in data-rich and transparent settings by interpreting meaning and context. Improvements in information quality therefore shift, rather than erase, the frontier of human–machine complementarity. From a policy perspective, disclosure reforms, open data initiatives, and transparency improvements enhance market efficiency while preserving the value of human expertise. As financial information systems evolve, the most productive path forward is not substitution but **integration**: allowing machines to process what is measurable and humans to interpret what is meaningful.

Looking ahead, a natural next step is to examine how new generations of predictive models can further bridge the gap between human and machine forecasts. Future research will compare human forecasts, traditional statistical learning models, and models augmented with language-based sentiment indicators, while carefully addressing look-ahead bias. Although machines may never fully capture the soft information available to human analysts, advances in language modeling now make it possible to incorporate part of the contextual and qualitative cues that humans use. This extension will make it possible to assess whether richer contextual learning improves machine forecasts and whether the incremental benefit varies across information environments.

Another promising direction is to combine human and machine forecasts. Preliminary experiments suggest that using analysts' forecasts as an additional input feature enables the model to capture complementarities between qualitative judgment and quantitative learning. This line of analysis can reveal how human–machine integration improves forecast accuracy and under what conditions such collaboration is most effective.

Beyond these extensions, future work can also investigate how the relative performance of humans and machines changes across economic cycles, especially during periods of crisis and recovery. Comparing these regimes will help clarify whether adaptive human reasoning or algorithmic pattern recognition performs better in turbulent environments. In addition, examining heterogeneity across analysts, such as differences in *experience* and *coverage breadth*, can shed light on who adapts most effectively to automation and which human skills remain most complementary to computational forecasting.

References

Acemoglu, Daron. 2021. *Redesigning Ai*.: MIT Press.

Acemoglu, Daron, Victor Chernozhukov, and Muhamet Yildiz. 2016. "Fragility of asymptotic agreement under Bayesian learning." *Theoretical Economics* 11 (1): 187–225.

Acemoglu, Daron, and Pascual Restrepo. 2018. "Artificial intelligence, automation, and work." In *The economics of artificial intelligence: An agenda*, 197–236: University of Chicago Press.

Agrawal, Ajay, Joshua Gans, and Avi Goldfarb. 2018. *Prediction machines: the simple economics of artificial intelligence*.: Harvard Business Press.

Agrawal, Ajay, Joshua Gans, and Avi Goldfarb. 2019. *The economics of artificial intelligence: An agenda*.: University of Chicago Press.

Agrawal, Ajay, Joshua Gans, and Avi Goldfarb. 2022. *Prediction machines, updated and expanded: The simple economics of artificial intelligence*.: Harvard Business Press.

Autor, David H. 2015. "Why Are There Still So Many Jobs? The History and Future of Workplace Automation." *Journal of Economic Perspectives* 29 (3): 3–30.

Bae, Kee-Hong, René M Stulz, and Hongping Tan. 2008. "Do local analysts know more? A cross-country study of the performance of local analysts and foreign analysts." *Journal of financial economics* 88 (3): 581–606.

Ball, Ryan T, and Eric Ghysels. 2018. "Automated earnings forecasts: Beat analysts or combine and conquer?" *Management Science* 64 (10): 4936–4952.

Barrot, Jean-Noël, and Julien Sauvagnat. 2016. "Input Specificity and the Propagation of Idiosyncratic Shocks in Production Networks." *The Quarterly Journal of Economics* 131 (3): 1543–1592.

Black, Fischer, and Robert Litterman. 1990. "Asset allocation: combining investor views with market equilibrium." *Goldman Sachs Fixed Income Research* 115 (1): 7–18.

Black, Fischer, and Robert Litterman. 1992. "Global portfolio optimization." *Financial analysts journal* 48 (5): 28–43.

Bonelli, Maxime. 2022. "The Adoption of Artificial Intelligence by Venture Capitalists."

Born, Benjamin, and Johannes Pfeifer. 2009. "Policy Risk and the Business Cycle." *Journal of Monetary Economics* 56 (4): 623–633.

Campbell, John Y., and Robert J. Shiller. 1997. "Valuation Ratios and the Long-Run Stock Market Outlook." *The Journal of Portfolio Management* 24 (2): 11–26.

Cao, Jian, and Mark Kohlbeck. 2011. "Analyst quality, optimistic bias, and reactions to major news." *Journal of Accounting, Auditing & Finance* 26 (3): 502–526.

Cao, Sean, Wei Jiang, Junbo L Wang, and Baozhong Yang. 2021. "From man vs. machine to man+ machine: The art and ai of stock analyses."Technical report, National Bureau of Economic Research.

Cao, Sean, Wei Jiang, Junbo Wang, and Baozhong Yang. 2024. "From man vs. machine to man+ machine: The art and AI of stock analyses." *Journal of Financial Economics* 160: 103910.

Chen, Nai-Fu, Richard Roll, and Stephen A Ross. 1986. "Economic forces and the stock market." *Journal of business*: 383–403.

Cowen, Amanda, Boris Groysberg, and Paul Healy. 2006. "Which types of analyst firms are more optimistic?" *Journal of Accounting and Economics* 41 (1-2): 119–146.

Dahlquist, Magnus, Lee Pinkowitz, René M Stulz, and Rohan Williamson. 2003. "Corporate governance and the home bias." *Journal of financial and quantitative analysis* 38 (1): 87–110.

De Silva, Tim, and David Thesmar. 2021. "Noise in expectations: Evidence from analyst forecasts."Technical report, National Bureau of Economic Research.

De Silva, Tim, and David Thesmar. 2024. "Noise in expectations: Evidence from analyst forecasts." *The Review of Financial Studies* 37 (5): 1494–1537.

Dessaint, Olivier, Thierry Foucault, and Laurent Frésard. 2021. "Does alternative data improve financial forecasting? the horizon effect."

Djankov, Simeon, Rafael La Porta, Florencio Lopez-de Silanes, and Andrei Shleifer. 2007. “The Law and Economics of Self-Dealing.” *Journal of Financial Economics* 88 (3): 430–465.

Fama, Eugene F, and Kenneth R French. 1989. “Business conditions and expected returns on stocks and bonds.” *Journal of financial economics* 25 (1): 23–49.

Fedyk, Anastassia, Ali Kakhbod, Peiyao Li, and Ulrike Malmendier. 2024. “Chatgpt and perception biases in investments: An experimental study.” *Available at SSRN* 4787249.

Forbes, Kristin J., and Francis E. Warnock. 2014. “Capital Flow Waves: Surges, Stops, Flight, and Retrenchment.” *Journal of International Economics* 92 (2): 235–251.

Fouliard, Jeremy, Michael Howell, and Hélène Rey. 2021. “Answering the queen: Machine learning and financial crises.”Technical report, National Bureau of Economic Research.

Fuster, Andreas, Matthew Plosser, Philipp Schnabl, and James Vickery. 2019. “The Role of Technology in Mortgage Lending.” *The Review of Financial Studies* 32 (5): 1854–1899.

Gabaix, Xavier, and Ralph S. J. Koijen. 2020. “In Search of the Origins of Financial Fluctuations: The Inelastic Markets Hypothesis.” *NBER Working Paper* (27367).

Grennan, Jillian, and Roni Michaely. 2021. “Fintechs and the market for financial analysis.” *Journal of Financial and Quantitative Analysis* 56 (6): 1877–1907.

Groysberg, Boris, Lex Lee, and Ashish Nanda. 2011. “Does Individual Performance Affect Entrepreneurial Mobility? Evidence from the Financial Analysis Market.” *Journal of Financial Economics* 100 (3): 581–595.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu. 2020. “Empirical Asset Pricing via Machine Learning.” *The Review of Financial Studies* 33 (5): 2223–2273.

Hale, Thomas, Noam Angrist, Rafael Goldszmidt, Beatriz Kira, Anna Petherick, Toby Phillips, Samuel Webster, Emily Cameron-Blake, Laura Hallas, Saptarshi Majumdar, and Helen Tatlow. 2021. “A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker).” *Nature Human Behaviour* 5: 529–538.

Hansen, Jorge W, and Christoffer Thimsen. 2021. “Forecasting corporate earnings with machine learning.” *It takes considerable knowledge just to realize the extent of your own ignorance.*: 57.

Hong, Harrison, and Jeffrey D Kubik. 2003a. “Analyzing the analysts: Career concerns and biased earnings forecasts.” *The Journal of Finance* 58 (1): 313–351.

Hong, Harrison, and Jeffrey D. Kubik. 2003b. “Analyzing the Analysts: Career Concerns and Biased Earnings Forecasts.” *The Journal of Finance* 58 (1): 313–351.

Kandel, Eugene, and Neil D Pearson. 1995. “Differential interpretation of public signals and trade in speculative markets.” *Journal of Political Economy* 103 (4): 831–872.

Kasparov, Garry. 2020. “Deep Thinking: Where Machine Intelligence Ends and Human Creativity Begins..” *Revista Empresa y Humanismo* 23 (2): 139–143.

Katona, Zsolt. 2018. *On the capital market consequences of alternative data: Evidence from outer space*.: eScholarship, University of California.

Kothari, SP. 2001. “Capital markets research in accounting.” *Journal of accounting and economics* 31 (1-3): 105–231.

Kurz, Mordecai. 1994a. “On rational belief equilibria.” *Economic Theory* 4 (6): 859–876.

Kurz, Mordecai. 1994b. “On the structure and diversity of rational beliefs.” *Economic theory* 4 (6): 877–900.

La Porta, Rafael, Florencio Lopez-de Silanes, and Andrei Shleifer. 2006. “What Works in Securities Laws?” *The Journal of Finance* 61 (1): 1–32.

La Porta, Rafael, Florencio Lopez-de Silanes, Andrei Shleifer, and Robert W. Vishny. 1998. “Law and Finance.” *Journal of Political Economy* 106 (6): 1113–1155.

Landier, Augustin, Yueran Ma, and David Thesmar. 2017. “New experimental evidence on expectations formation.”

Lang, Mark, Karl V Lins, and Mark Maffett. 2012. “Transparency, liquidity, and valuation: International evidence on when transparency matters most.” *Journal of Accounting Research* 50 (3): 729–774.

Liberti, José María, and Mitchell A Petersen. 2019. “Information: Hard and soft.” *Review of Corporate Finance Studies* 8 (1): 1–41.

Lim, Terence. 2001. “Rationality and analysts’ forecast bias.” *The journal of Finance* 56 (1): 369–385.

Lucas, Robert E. 1976. “Econometric policy evaluation: A critique.” *Carnegie-Rochester Conference Series on Public Policy* 1: 19–46.

Mola, Simona, and Massimo Guidolin. 2009. “Affiliated mutual funds and analyst optimism.” *Journal of Financial Economics* 93 (1): 108–137.

Obizhaeva, Anna A, and Yajun Wang. 2019. “Trading in crowded markets.” *Available at SSRN 31527*.

Pinkowitz, Lee, René M Stulz, and Rohan Williamson. 2001. “Corporate governance and the home bias.”

Porta, Rafael La, Florencio Lopez-de Silanes, Andrei Shleifer, and Robert W Vishny. 1998. “Law and finance.” *Journal of political economy* 106 (6): 1113–1155.

Price, S. McKay, James S. Doran, David R. Peterson, and Brett A. Bliss. 2012. “Earnings Conference Calls and Stock Returns: The Incremental Informativeness of Textual Tone.” *Journal of Banking & Finance* 36 (4): 992–1011.

Reinhart, Carmen M., and Kenneth S. Rogoff. 2009. *This Time Is Different: Eight Centuries of Financial Folly.*: Princeton University Press.

Sims, Christopher A. 2003. “Implications of rational inattention.” *Journal of monetary Economics* 50 (3): 665–690.

Stein, Jeremy C. 2002. “Information production and capital allocation: Decentralized versus hierarchical firms.” *The journal of finance* 57 (5): 1891–1921.

Susan Athey. 2023. “The Brookings Institution Webinar: ChatGPT and the Future of Work.” https://www.brookings.edu/wp-content/uploads/2023/03/es_20230315_chatgpt_work_transcript.pdf. Accessed on 15 March 2023.

Teorell, Jan, Aksel Sundström, Sören Holmberg, Bo Rothstein, Natalia Alvarado Pachon, Cem Mert Dalli, Rafael Lopez Valverde, Victor Saidi Phiri, and Lauren Gerber. 2025. “The Quality of Government Standard Dataset, version Jan25.” Available at https://www.gu.se/en/quality-government.

Tetlock, Paul C. 2007. “Giving Content to Investor Sentiment: The Role of Media in the Stock Market.” *The Journal of Finance* 62 (3): 1139–1168.

Van Binsbergen, Jules H, Xiao Han, and Alejandro Lopez-Lira. 2020. “Man vs. machine learning: The term structure of earnings expectations and conditional biases.”Technical report, National Bureau of Economic Research.

Van Binsbergen, Jules H, Xiao Han, and Alejandro Lopez-Lira. 2023. “Man versus machine learning: The term structure of earnings expectations and conditional biases.” *The Review of financial studies* 36 (6): 2361–2396.

Woodford, Michael. 2003. “nImperfect Common Knowledge and the Effects of Monetary Policy. oIn Philippe Aghion, Roman Frydman, Joseph E. Stiglitz and Michael Wood $ford, eds., Knowledge, Information, and Expectations in Modern Macroeconomics : In Honor of Edmund S. Phelps.''$

Figure 1. Visualized Construction of the Machine Forecast Matched to Each Human Forecast

This figure provides a visual illustration of how machine forecasts are constructed and paired with each individual human forecast in a globally consistent yet locally trained framework. For every analyst prediction of next-year earnings per share (EPS), a corresponding machine forecast is generated on the exact same *issue forecast day*, ensuring strict temporal alignment between humans and algorithms. The machine analyst is trained separately for each country, using only that country's firm-level and macroeconomic data. Within each country, the model is re-estimated every year in a rolling-window manner, allowing it to incorporate newly available information while avoiding any look-ahead bias. The machine model's training set includes only information observable prior to the forecast day. The *forecast horizon* is precisely measured in calendar days as the interval between the issue forecast day and the subsequent earnings announcement date. Both human and machine forecast errors are computed relative to the realized EPS, and all variables are scaled by the latest stock price available before the forecast day to ensure comparability across firms, years, and countries. This design guarantees that the forecast comparisons capture genuine differences in predictive ability rather than artifacts of timing, information leakage, or heterogeneous data availability.

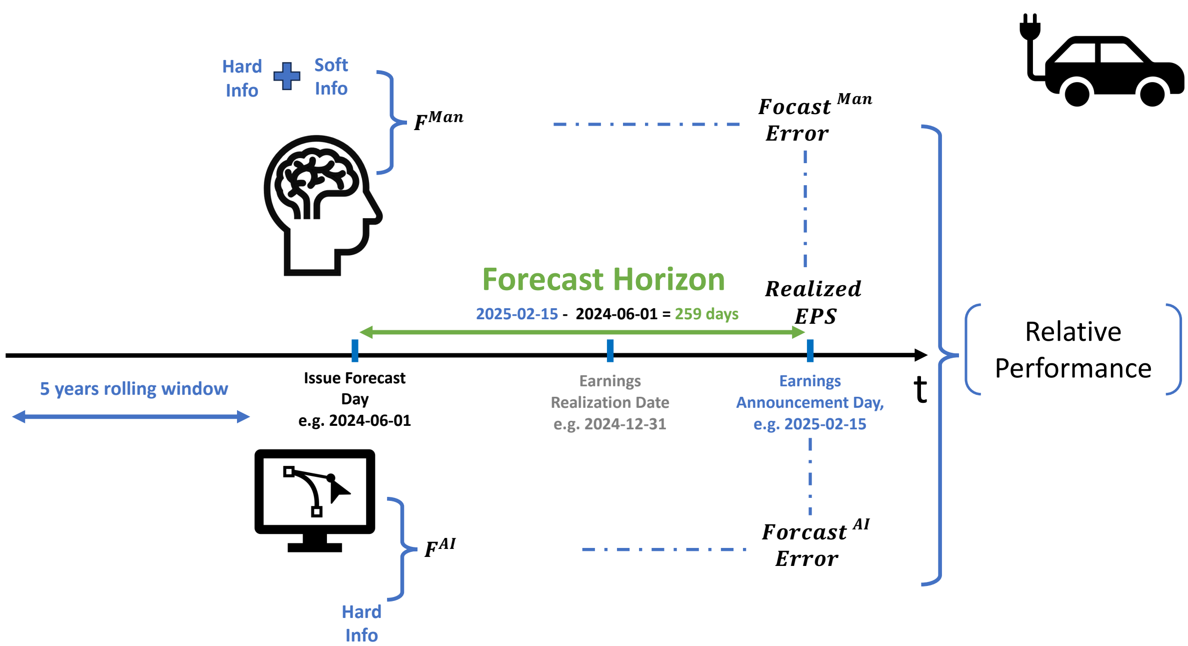


Figure 2. Average Beat Rate and Data Services Quality (All Countries)

This figure illustrates how the average beat rate varies across countries with different levels of data services quality. The beat rate captures the proportion of firm-level forecasts where the human analyst's absolute forecast error is smaller than that of the machine, reflecting the frequency with which humans outperform algorithms. The underlying *beat* variable takes the value of one when the analyst's forecast is more accurate than the machine's and zero otherwise. For each country, the beat ratio is calculated as the mean of all firm–analyst–machine forecast pairs with horizons of 365 days or less, focusing on short-horizon earnings-per-share forecasts, and for firms with fiscal year-ends on December 31. The horizontal axis reports the *Data Services Quality* score from the World Bank's *Data Services Pillar* (2016–2022), which aggregates four aspects of national statistical capacity: the timeliness and quality of official data, the openness of online access, the strength of advisory and analytical support, and the responsible provision of microdata.

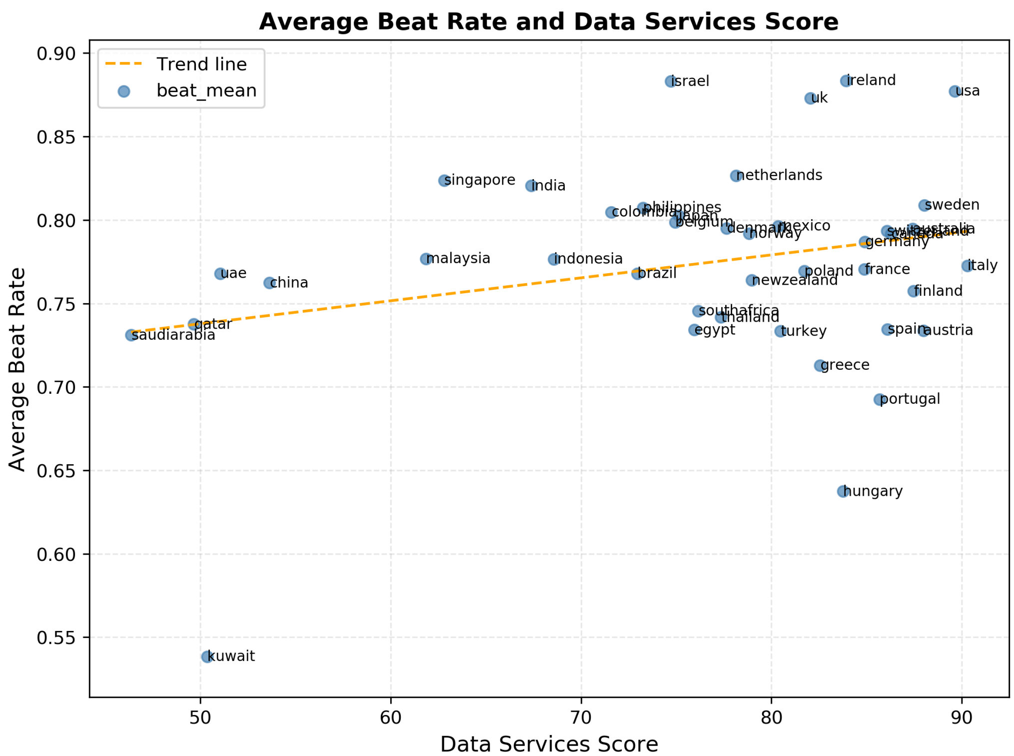


Figure 3. Mean Relative Performance and Data Services Quality (All Countries)

This figure presents the association between the mean relative performance of human analysts and the quality of data services. Relative performance is defined as the forecast-error difference between the machine and the analyst, divided by the larger of the two absolute forecast errors. A positive value indicates that analysts are, on average, more accurate than machines. The metric is averaged across all short-horizon (365 days) earnings-per-share forecasts for each country, and for firms with fiscal year-ends on December 31. The beat indicator underlying the comparison equals one if the human forecast is more precise than the machine forecast and zero otherwise. The *Data Services Quality* index, drawn from the World Bank's *Data Services Pillar* for 2016–2022, summarizes information availability, access, and analytic capacity in national data systems.

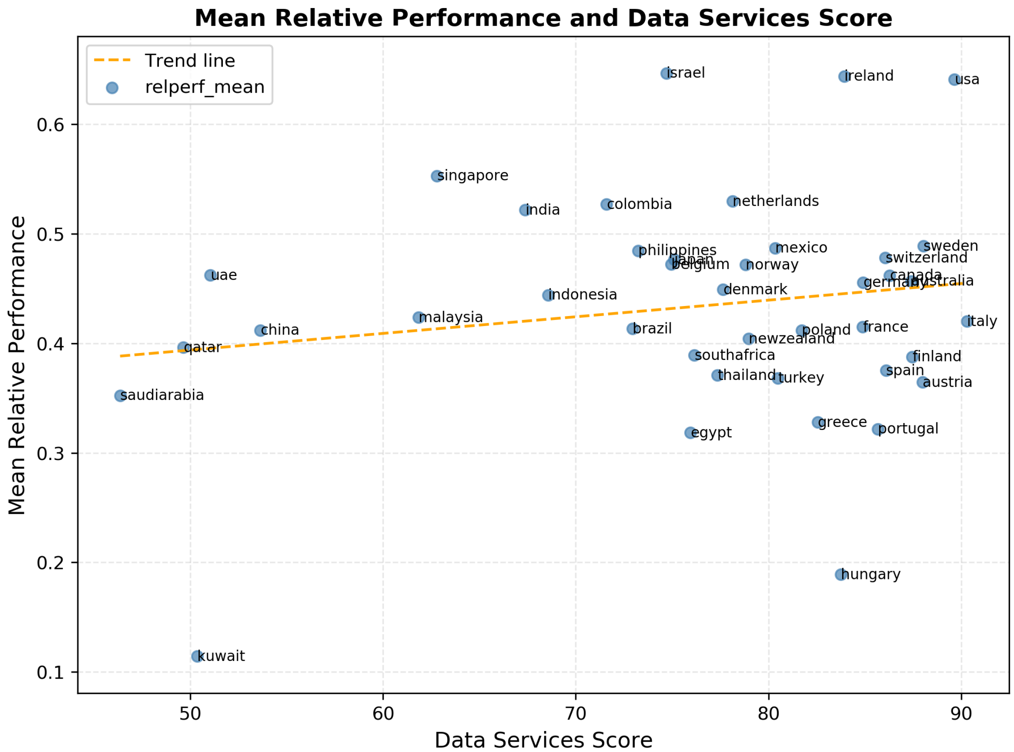


FIGURE 4. Median Relative Performance and Data Services Quality (All Countries)

This figure replicates the analysis in Figure 3 using median instead of mean aggregation. The median relative performance summarizes the central tendency of the human–machine forecast accuracy gap while reducing the influence of outliers. Each observation corresponds to a country-level median of all firm–analyst–machine pairs with forecast horizons no longer than 365 days, and for firms with fiscal year-ends on December 31. The beat variable equals one when the analyst's forecast outperforms the machine's and zero otherwise. The measure is plotted against the same *Data Services Quality* index described in Figure 2.

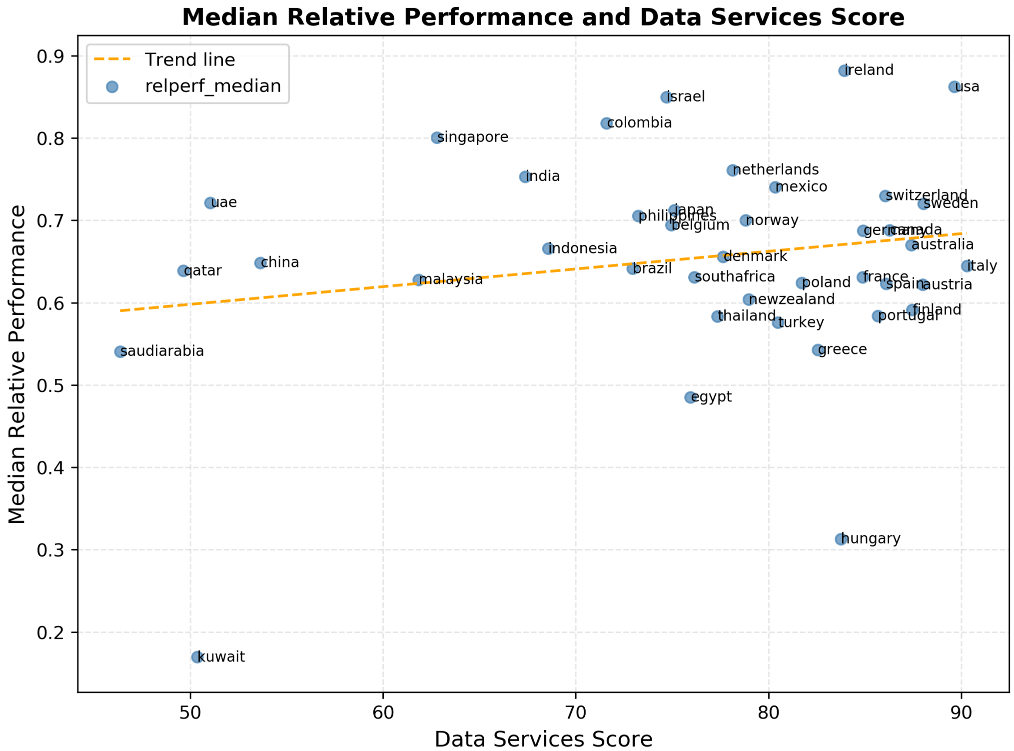


FIGURE 5. Average Beat Rate and Transparency Index (All Countries)

This figure displays the cross-country relationship between the average beat rate and national transparency levels. The beat rate represents the proportion of firm-level forecasts in which the human analyst's absolute forecast error is lower than that of the machine. The *beat* variable equals one if the analyst outperforms the algorithm and zero otherwise. For each country, the beat ratio is computed as the mean of all short-horizon (365 days) earnings-per-share forecasts comparing human and machine predictions, and for firms with fiscal year-ends on December 31. The horizontal axis reports the *Transparency Index*, a composite measure of *Information Transparency* and *Accountability Transparency* for the period 1980–2010, capturing both the openness of information dissemination and the degree of institutional accountability.

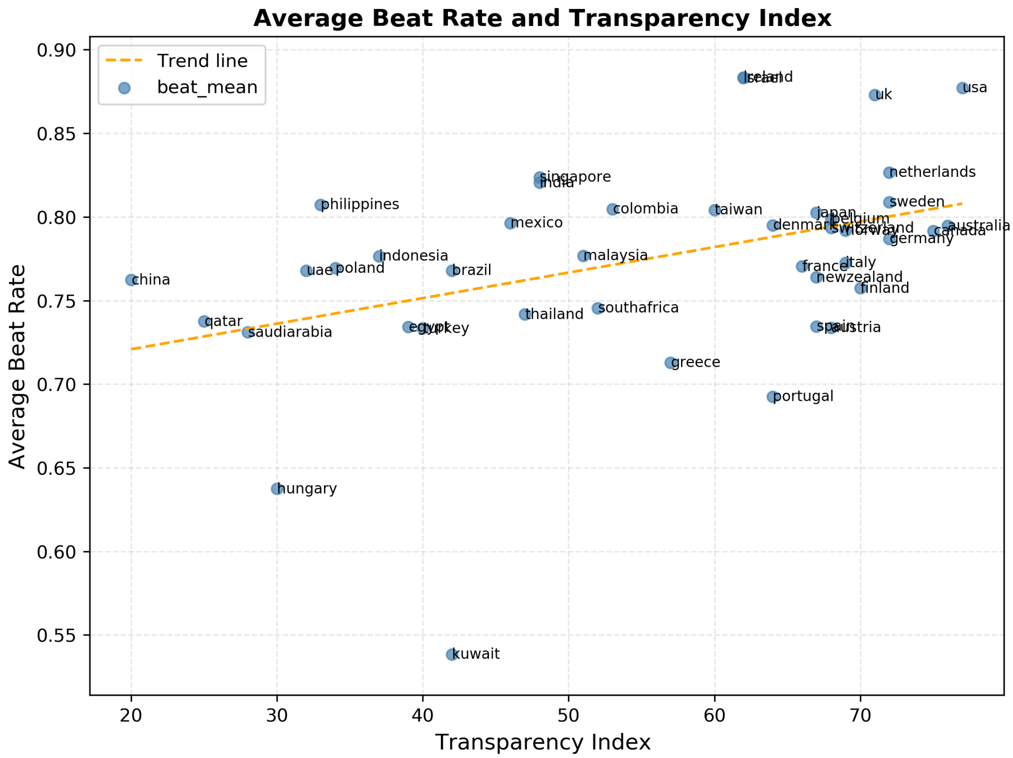


Figure 6. Mean Relative Performance and Transparency Index (All Countries)

This figure shows how the mean relative performance of human analysts compared with machines changes with institutional transparency. Relative performance is measured as the difference between the machine's and the analyst's forecast errors, normalized by the larger of their absolute values. Positive values indicate greater human accuracy. Each country's score averages all firm-level short-horizon (365 days) earnings-per-share forecasts, and for firms with fiscal year-ends on December 31. The underlying beat variable equals one when the analyst provides a more precise forecast than the machine and zero otherwise. The *Transparency Index*, combining information and accountability dimensions (1980–2010), is defined as in Figure 5.

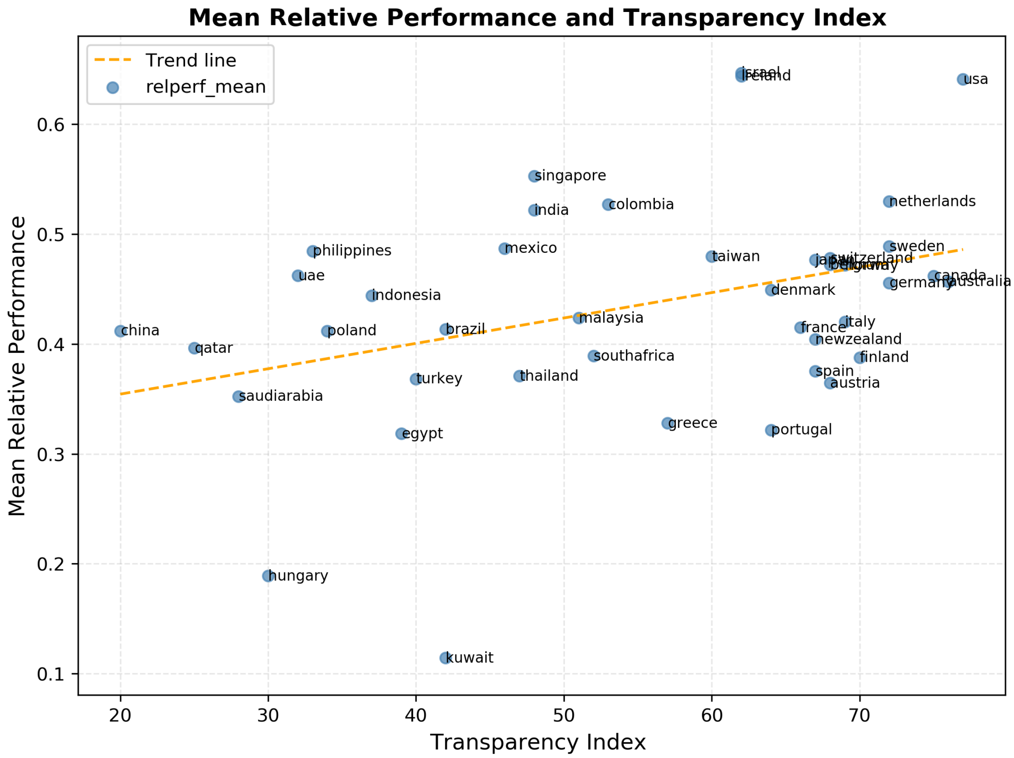


FIGURE 7. Median Relative Performance and Transparency Index (All Countries)

This figure reproduces the analysis in Figure 6 using the median rather than the mean. Median relative performance captures the typical magnitude of the human–machine forecast accuracy gap while mitigating the impact of extreme errors. Country-level medians are computed from firm–analyst–machine forecast pairs with horizons of up to 365 days, and for firms with fiscal year-ends on December 31. The beat variable equals one when the human forecast is more accurate and zero otherwise. The *Transparency Index* combines *Information Transparency* and *Accountability Transparency* as explained in Figure 5.

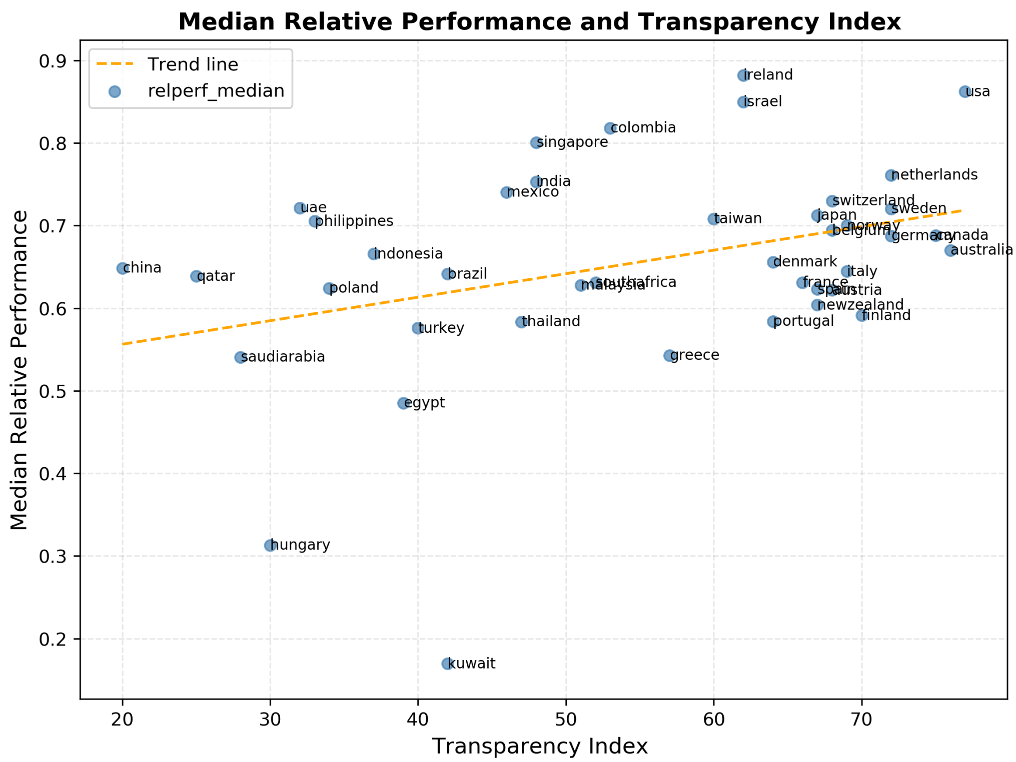


Figure 8. Average Beat Rate and Access to Information Index (All Countries)

This figure presents the cross-country relationship between the average beat rate and the degree of access to information. The beat rate measures how often human analysts outperform machines in absolute forecast accuracy. The underlying *beat* variable equals one if the analyst's forecast error is smaller than the machine's and zero otherwise. Each country's beat ratio is the average of all firm–analyst–machine forecast pairs with horizons up to 365 days, focusing on short-horizon earnings-per-share forecasts, and for firms with fiscal year-ends on December 31. The horizontal axis reports the *Access to Information Index*, corresponding to the subpillar of the Robust Democracy category for 2013–2021. This index reflects media independence, pluralism, and public access to government information, encompassing measures such as media freedom, diversity, and openness of official communication.

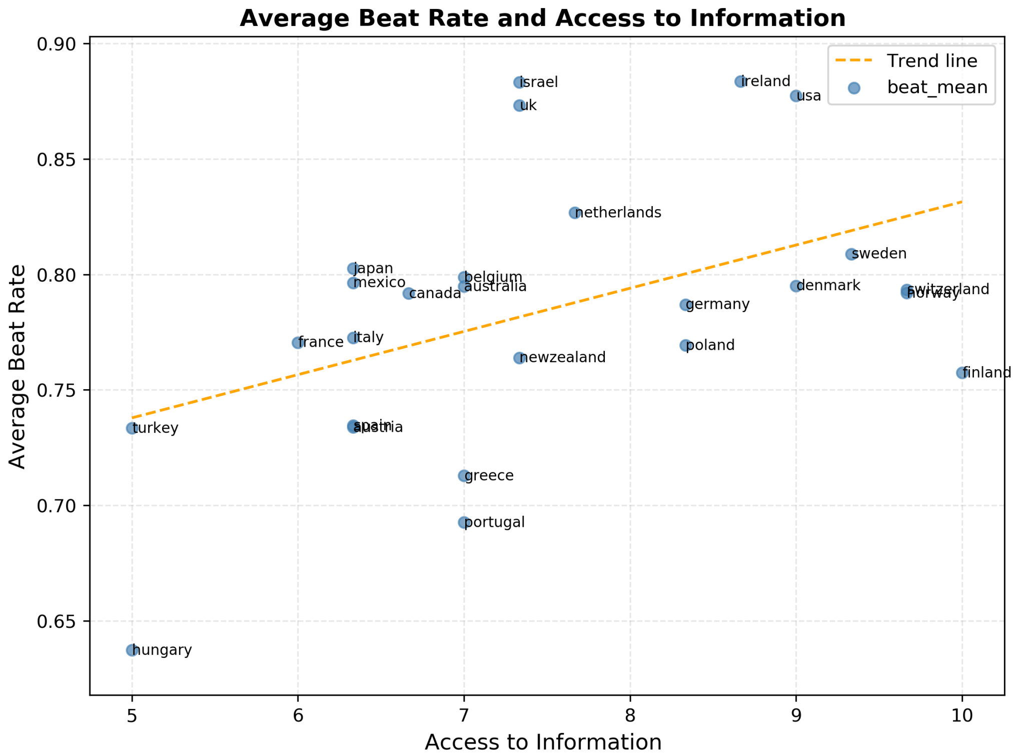


FIGURE 9. Mean Relative Performance and Access to Information Index (All Countries)

This figure examines how the mean relative performance of human analysts compared with machines varies with the level of access to information. Relative performance is calculated as the machine–analyst forecast-error difference divided by the maximum of their absolute errors. Positive values indicate higher accuracy of human forecasts. Each country's value represents the average across all firm-level short-horizon (365 days) earnings-per-share forecasts, and for firms with fiscal year-ends on December 31. The beat indicator equals one if the analyst's forecast is more accurate than the machine's and zero otherwise. The *Access to Information Index*, described in Figure 8, spans 2013–2021 and captures media freedom, pluralism, and accessibility of official information.

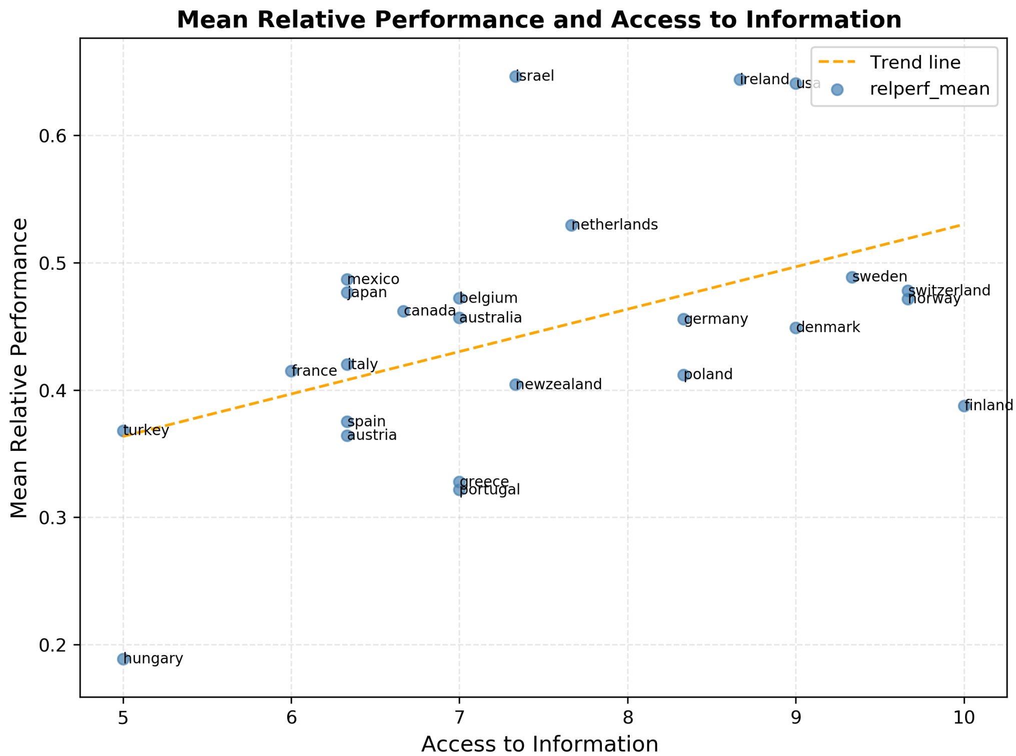


FIGURE 10. Median Relative Performance and Access to Information Index (All Countries)

This figure repeats the analysis from Figure 9 using the median instead of the mean. The median relative performance captures the central human–machine accuracy gap while limiting the effect of outliers. For each country, the median is computed from firm–analyst–machine forecast pairs with horizons not exceeding 365 days, and for firms with fiscal year-ends on December 31. The *beat* variable takes the value of one when the analyst's forecast outperforms the machine's and zero otherwise. The *Access to Information Index* represents the 2013–2021 subpillar of the Robust Democracy category reflecting media independence, diversity, and the openness of government data.

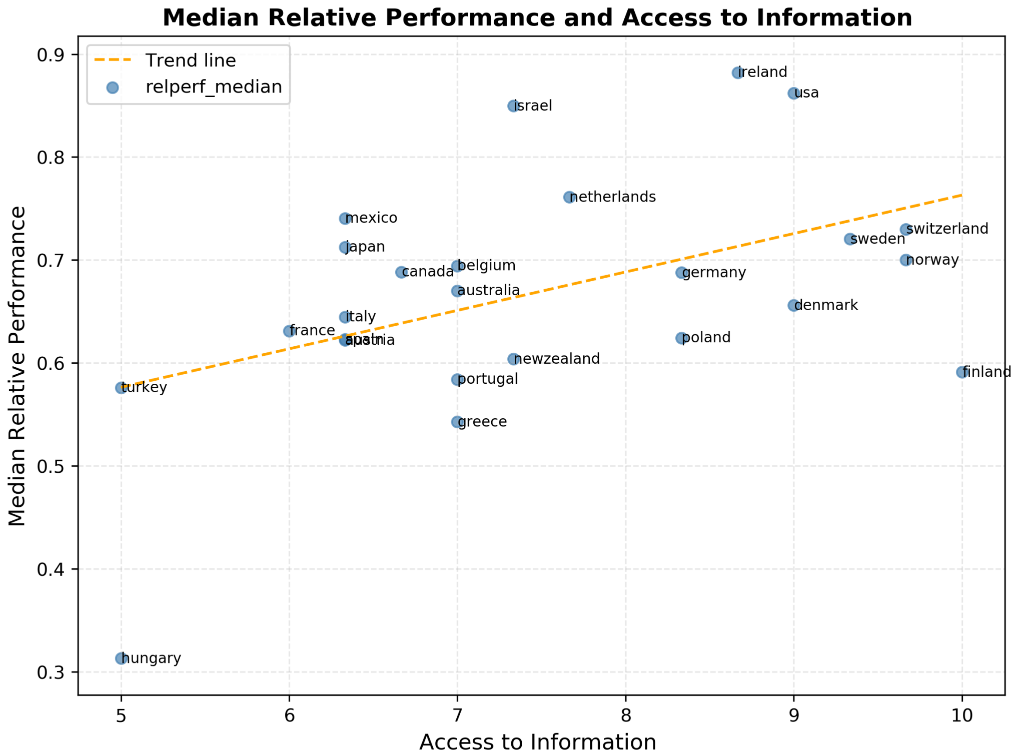


Figure 11. Beat Ratio Over Time Across Countries

This figure plots the evolution of the average beat ratio over time for four representative countries (USA, France, Indonesia, and Thailand). These countries are selected to illustrate economies at different stages of development, given space limitations. The beat ratio measures the fraction of firm-level forecasts in which the human analyst's absolute forecast error is smaller than that of the machine, indicating the share of cases where humans outperform algorithms. The underlying *beat* indicator is a binary variable that equals one if the human forecast is more accurate than the machine forecast and zero otherwise. For each country and year, the beat ratio is computed as the mean of all firm–analyst–machine forecast pairs with forecast horizons of no more than 365 days (short-horizon earnings-per-share forecasts), and for firms with fiscal year-ends on December 31. . The solid lines display yearly averages, while the dashed lines show fitted linear trends over the sample period.

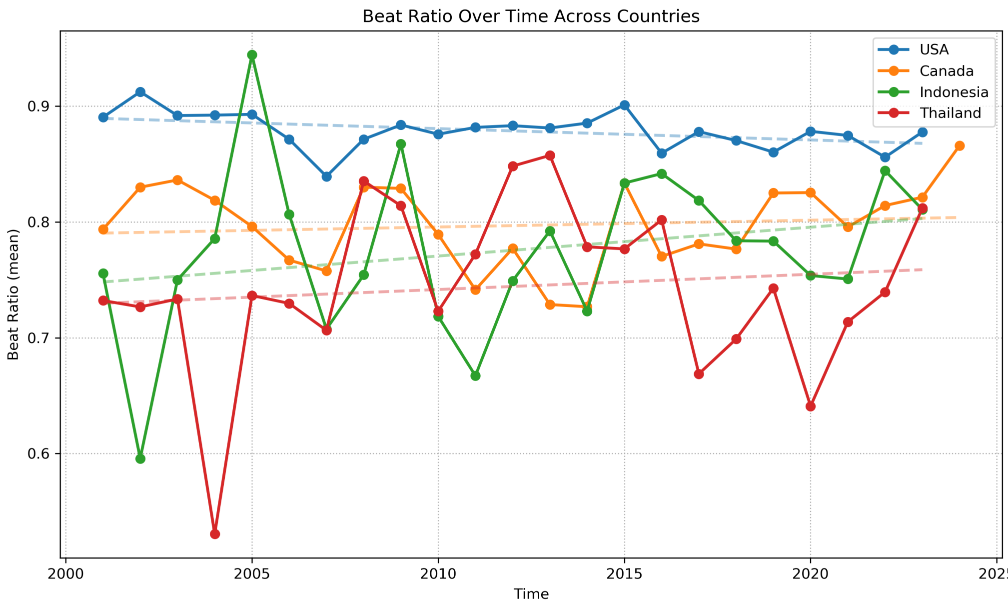
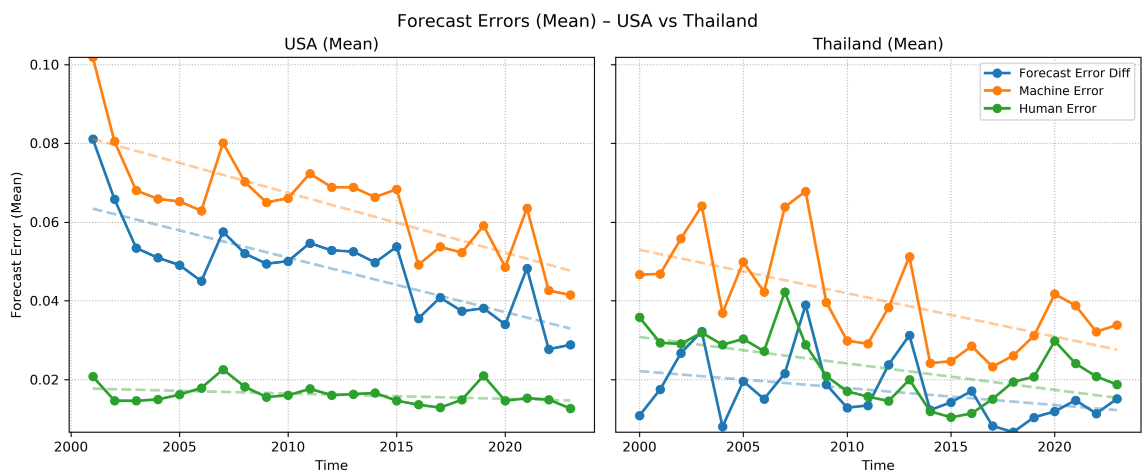


FIGURE 12. Forecast Errors over Time: USA vs Thailand

This figure compares the evolution of forecast errors between human analysts and machine models for the United States and Thailand over time. The upper panels display results based on mean values across all firm–analyst–machine forecast pairs within each year, and the lower panels present the corresponding medians. Each point in the plots represents the average or median forecast error across all short-horizon (up to 365 days) earnings-per-share forecasts for that country and year, and for firms with fiscal year-ends on December 31. Forecast errors are defined as the absolute value of the difference between the forecasted EPS and the realized EPS, where the difference is taken first and then its absolute value is computed. All forecasted and realized EPS values are scaled by the latest available stock price prior to the forecast day to ensure comparability across firms and time. The orange and green lines show the absolute forecast errors of the machine and the human, respectively. Each captures the magnitude of deviation between predicted and realized earnings, reflecting the absolute accuracy achieved by algorithms versus human analysts. The blue line represents the forecast error difference (*machine error minus human error*), which measures the relative performance gap between machines and humans. A positive blue value indicates that machine forecasts are less accurate on average, while a negative value implies that machines outperform human analysts in predicting firm-level earnings. The dashed lines illustrate fitted linear trends over time. Together, these panels highlight how the accuracy of human and machine forecasts evolves across two economies at different stages of development, shedding light on the changing balance between human judgment and algorithmic prediction in financial markets.



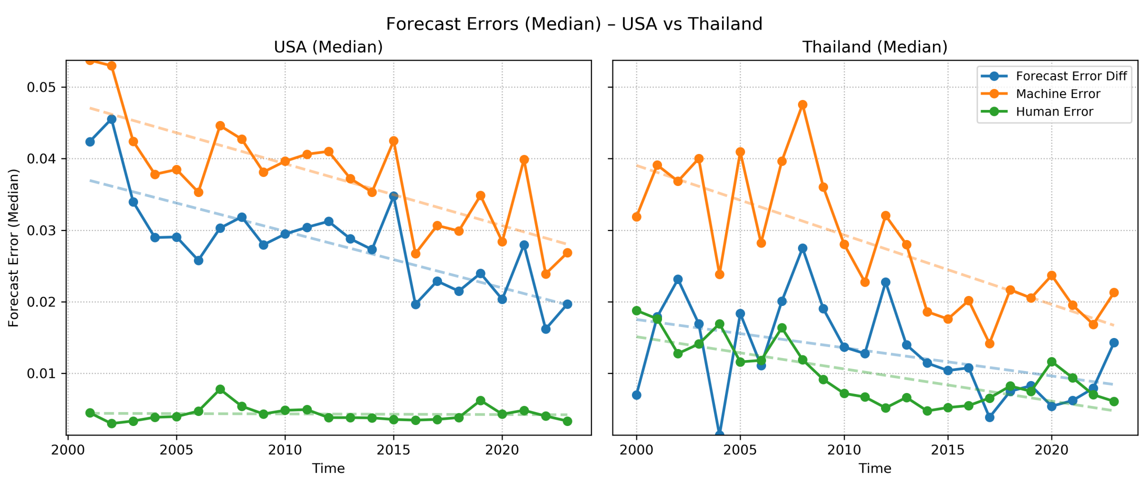


Table 1. Summary of Analyst–Machine Forecast Performance by Country

This table reports the summary statistics of human–machine forecast comparisons across 47 countries in the global sample. Two countries, Chile and Peru, are excluded due to insufficient merged observations. Each observation corresponds to an analyst forecast of the next annual earnings per share (EPS) to be announced. The sample includes only forecasts with horizons shorter than 365 days and for firms with fiscal year-ends on December 31. Firm-year observations are retained only if they can be successfully merged with accounting data from Compustat and market data from Datastream (or CRSP for the United States). All forecasted and realized earnings values are scaled by the latest available stock price prior to the forecast day to ensure comparability across firms and over time. Forecast errors are winsorized at the 1st and 99th percentiles to mitigate the influence of outliers. The last four columns report the mean and median absolute forecast errors for both machines and analysts. The forecast-error difference is defined as the machine's absolute forecast error minus the analyst's absolute forecast error; thus, a positive value indicates that machine forecasts are less accurate. The relative performance measure scales this forecast-error difference by the maximum of the two absolute errors, providing a normalized measure of relative accuracy. Finally, the beat ratio represents the proportion of firm-level forecasts in which the analyst's absolute error is smaller than the machine's. A beat ratio above 0.5 indicates that analysts outperform the machine on average.

Country	Obs.	Beat ratio	Rel. perf (mean)	Rel. perf (median)	FE diff (mean)	FE diff (median)	$\lvert FE \rvert$ (Machine, mean)	$\lvert FE \rvert$ (Machine, median)	$\lvert FE \rvert$ (Human, mean)	$\lvert FE \rvert$ (Human, median)
australia	10791	0.795	0.457	0.670	0.041	0.022	0.072	0.039	0.030	0.012
austria	2971	0.734	0.364	0.622	0.028	0.016	0.049	0.030	0.021	0.011
belgium	4764	0.799	0.472	0.694	0.051	0.020	0.074	0.032	0.023	0.008
brazil	21604	0.768	0.413	0.641	0.022	0.017	0.047	0.031	0.025	0.010
canada	387908	0.792	0.462	0.688	0.024	0.017	0.045	0.029	0.021	0.008
china	37996	0.762	0.412	0.648	0.011	0.008	0.021	0.015	0.010	0.005
colombia	256	0.805	0.527	0.818	0.019	0.017	0.032	0.021	0.012	0.004
czech	358	0.869	0.601	0.819	0.064	0.053	0.084	0.067	0.020	0.012
denmark	7162	0.795	0.449	0.656	0.014	0.009	0.026	0.016	0.012	0.004
egypt	474	0.734	0.318	0.485	0.025	0.014	0.054	0.035	0.029	0.015
finland	12174	0.757	0.388	0.591	0.021	0.013	0.041	0.026	0.020	0.009
france	48754	0.770	0.415	0.631	0.016	0.011	0.033	0.021	0.017	0.007

Continued on next page

Table 1. Summary of Analyst–Machine Forecast Performance by Country (continued)

See the notes below Table 1 for definitions of variables.

Country	Obs.	Beat ratio	Rel. perf (mean)	Rel. perf (median)	FE diff (mean)	FE diff (median)	$\lvert FE \rvert$ (Machine, mean)	$\lvert FE \rvert$ (Machine, median)	$\lvert FE \rvert$ (Human, mean)	$\lvert FE \rvert$ (Human, median)
germany	42289	0.787	0.456	0.688	0.023	0.016	0.043	0.027	0.020	0.007
greece	1790	0.713	0.328	0.543	0.031	0.019	0.070	0.041	0.040	0.018
hk	17644	0.789	0.463	0.712	0.030	0.017	0.048	0.031	0.018	0.008
hungary	364	0.637	0.189	0.313	0.021	0.007	0.051	0.025	0.030	0.016
india	4268	0.821	0.522	0.753	0.007	0.005	0.012	0.008	0.004	0.002
indonesia	7870	0.776	0.444	0.666	0.016	0.011	0.028	0.020	0.012	0.006
ireland	747	0.884	0.644	0.882	0.019	0.012	0.024	0.017	0.005	0.002
israel	2746	0.883	0.647	0.850	0.017	0.010	0.024	0.015	0.007	0.002
italy	9683	0.773	0.420	0.645	0.019	0.012	0.034	0.021	0.016	0.006
japan	10289	0.803	0.477	0.713	0.014	0.013	0.027	0.021	0.012	0.006
kuwait	78	0.538	0.114	0.170	0.005	0.002	0.020	0.014	0.014	0.009
malaysia	5986	0.777	0.424	0.628	0.026	0.014	0.048	0.026	0.022	0.009
mexico	11143	0.796	0.487	0.741	0.019	0.014	0.033	0.022	0.013	0.005
netherlands	8100	0.827	0.530	0.761	0.023	0.016	0.040	0.025	0.017	0.005
newzealand	538	0.764	0.404	0.604	0.025	0.014	0.045	0.024	0.020	0.008
norway	16277	0.792	0.472	0.700	0.034	0.019	0.058	0.034	0.024	0.009
philippines	3649	0.807	0.485	0.705	0.020	0.014	0.034	0.021	0.014	0.005
poland	3728	0.769	0.412	0.624	0.031	0.022	0.060	0.043	0.029	0.013
portugal	2287	0.693	0.322	0.584	0.014	0.008	0.027	0.020	0.013	0.007
qatar	320	0.738	0.396	0.639	0.019	0.011	0.031	0.020	0.012	0.009

Continued on next page

Table 1. Summary of Analyst–Machine Forecast Performance by Country (continued)

See the notes below Table 1 for definitions of variables.

Country	Obs.	Beat ratio	Rel. perf (mean)	Rel. perf (median)	FE diff (mean)	FE diff (median)	$\|FE\|$ (Machine, mean)	$\|FE\|$ (Machine, median)	$\|FE\|$ (Human, mean)	$\|FE\|$ (Human, median)
saudiarabia	2975	0.731	0.352	0.541	0.011	0.008	0.029	0.020	0.018	0.007
singapore	3449	0.824	0.553	0.801	0.077	0.032	0.099	0.046	0.022	0.009
southafrica	3245	0.745	0.389	0.631	0.025	0.020	0.055	0.038	0.030	0.013
southkorea	48990	0.777	0.401	0.595	0.026	0.022	0.058	0.040	0.032	0.015
spain	9988	0.735	0.375	0.623	0.022	0.012	0.041	0.025	0.019	0.008
sweden	18359	0.809	0.489	0.720	0.026	0.016	0.041	0.025	0.015	0.006
switzerland	13929	0.793	0.478	0.730	0.014	0.010	0.022	0.017	0.009	0.004
taiwan	41993	0.804	0.480	0.708	0.016	0.014	0.029	0.023	0.012	0.005
thailand	9999	0.742	0.371	0.584	0.017	0.011	0.038	0.024	0.022	0.008
turkey	4276	0.733	0.368	0.576	0.024	0.019	0.054	0.040	0.030	0.014
uae	461	0.768	0.462	0.721	0.023	0.013	0.035	0.023	0.012	0.004
uk	35866	0.873	0.649	0.881	2.109	1.325	2.697	1.760	0.588	0.181
usa	1058816	0.877	0.641	0.862	0.048	0.027	0.064	0.036	0.016	0.004

TABLE 2. Relative Advantages of Human vs. Machine Forecasts: Data Service Score

This table examines how national information environments, measured by the **Data Service Score** available during 2016–2022, relate to the relative forecasting performance of humans and machines. Panel A reports results using a continuous measure of relative forecast performance (machine minus analyst forecast error), while Panel B uses a binary indicator equal to one when the machine forecast outperforms the analyst (*Beat* = 1). The *Data Service Score* captures the quality and accessibility of national data services, and the time trend represents a linear term in calendar year. All regressions include a full set of firm- and macro-level control variables[3] as well as industry fixed effects, though, due to space constraints, not all accounting variables are reported. Column (2) additionally includes year fixed effects to absorb global time trends. Country, firm, and analyst fixed effects are excluded because they would absorb the cross-country variation in information environments that identifies Hypothesis 1 (see Appendix E for detailed discussions of fixed-effect choices and specifications). Standard errors are clustered at the firm level. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A. Relative Forecast Performance and National Data Service Score

	(1) Industry FE + Trend			(2) Industry + Year FE		
Variables	*b*	*se*	*p*	*b*	*se*	*p*
Information Environment Variable						
Data Service Score	0.005***	(0.001)	[0.000]	0.005***	(0.001)	[0.000]
Time trend	−0.014***	(0.003)	[0.000]			
Accounting Variables						
ACT (Current Assets)	−0.079***	(0.016)	[0.000]	−0.080***	(0.017)	[0.000]
AQC (Acquisitions)	0.005	(0.006)	[0.430]	0.005	(0.006)	[0.449]
AT (Total Assets)	−0.007	(0.065)	[0.910]	−0.001	(0.064)	[0.883]
CAPX (Capital Expenditures)	−0.006	(0.019)	[0.738]	−0.005	(0.018)	[0.807]
CEQ (Common Equity)	0.017	(0.024)	[0.473]	0.018	(0.024)	[0.442]
CHE (Cash and Short-Term Investments)	0.080***	(0.020)	[0.001]	0.076***	(0.021)	[0.001]
IB (Income Before Extraordinary Items)	0.051*	(0.029)	[0.089]	0.046	(0.030)	[0.121]
LEV (Leverage Ratio)	−0.016	(0.013)	[0.207]	−0.018	(0.014)	[0.183]
PRSTKC (Purchase of Common and Preferred Stock)	0.032***	(0.008)	[0.000]	0.033***	(0.008)	[0.000]
SALE (Net Sales)	0.008	(0.005)	[0.110]	0.007	(0.005)	[0.145]
SPI (Special Items)	−0.011**	(0.005)	[0.018]	−0.011**	(0.005)	[0.018]
RET (Return on Stock Price)	0.123***	(0.043)	[0.006]	0.130***	(0.041)	[0.003]
Constant	0.398***	(0.064)	[0.000]	0.120**	(0.054)	[0.031]
Observations	465,451			465,451		
R^2	0.058			0.059		
Adj. R^2	0.057			0.059		

Panel B. Beat Indicator and National Data Service Score

	(1) Industry FE + Trend			(2) Industry + Year FE		
Variables	*b*	*se*	*p*	*b*	*se*	*p*
Information Environment Variable						
Data Service Score	0.003***	(0.001)	[0.000]	0.003***	(0.001)	[0.000]
Time trend	−0.005**	(0.002)	[0.010]			
Accounting Variables						
ACT (Current Assets)	−0.042***	(0.009)	[0.000]	−0.043***	(0.009)	[0.000]
AQC (Acquisitions)	0.004	(0.005)	[0.401]	0.004	(0.005)	[0.405]
AT (Total Assets)	−0.009	(0.046)	[0.847]	−0.007	(0.045)	[0.873]
CAPX (Capital Expenditures)	−0.003	(0.011)	[0.783]	−0.002	(0.010)	[0.846]
CEQ (Common Equity)	0.012	(0.015)	[0.413]	0.013	(0.016)	[0.396]
CHE (Cash and Short-Term Investments)	0.038***	(0.010)	[0.001]	0.037***	(0.011)	[0.001]
IB (Income Before Extraordinary Items)	0.019	(0.017)	[0.272]	0.018	(0.018)	[0.305]
LEV (Leverage Ratio)	−0.010	(0.008)	[0.224]	−0.011	(0.008)	[0.193]
PRSTKC (Purchase of Common and Preferred Stock)	0.013***	(0.004)	[0.001]	0.014***	(0.004)	[0.001]
SALE (Net Sales)	0.005	(0.003)	[0.089]	0.004	(0.003)	[0.137]
SPI (Special Items)	−0.006**	(0.003)	[0.023]	−0.006**	(0.003)	[0.030]
RET (Return on Stock Price)	0.072***	(0.026)	[0.008]	0.073***	(0.025)	[0.005]
Constant	0.703***	(0.040)	[0.000]	0.604***	(0.033)	[0.000]
Observations		465,451			465,451	
R^2		0.026			0.026	
Adj. R^2		0.026			0.026	

Table 3. Relative Advantages of Human vs. Machine Forecasts: Transparency Index

This table examines how national information environments, measured by the **Transparency Index** relate to the relative forecasting performance of humans and machines. Panel A uses a continuous measure of relative forecast performance (machine minus analyst forecast error), while Panel B uses a binary indicator equal to one if the machine forecast outperforms the analyst (*Beat* = 1). The *Transparency Index* is a composite measure combining two sub-indices: the *Information Transparency Index*, which captures media freedom, fiscal transparency, and political constraints; and the *Accountability Transparency Index*, which measures accountability mechanisms, checks on political power, and transparency in governance processes. The dataset covers the period 1980–2010. All regressions include full sets of firm- and macro-level controls as well as industry fixed effects. Column (2) further adds year fixed effects to remove global time trends. Country, firm, and analyst fixed effects are excluded because they would absorb the cross-country variation in national information environments that identifies Hypothesis 1. (See Appendix E for detailed explanations of fixed effect choices and specifications.) Standard errors are clustered at the firm level. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels.

Panel A. Relative Forecast Performance and Transparency Index

	(1) Industry FE + Trend			(2) Industry + Year FE		
Variables	*b*	*se*	*p*	*b*	*se*	*p*
Information Environment Variable						
Transparency Index	0.005***	(0.001)	[0.000]	0.005***	(0.001)	[0.000]
Time trend	0.003*	(0.002)	[0.070]			
Accounting Variables						
ACT (Current Assets)	−0.015	(0.017)	[0.380]	−0.019	(0.016)	[0.230]
AQC (Acquisitions)	−0.003	(0.006)	[0.640]	−0.000	(0.005)	[0.970]
AT (Total Assets)	−0.038**	(0.017)	[0.030]	−0.038**	(0.019)	[0.050]
CAPX (Capital Expenditures)	0.020*	(0.015)	[0.090]	0.020*	(0.010)	[0.060]
CEQ (Common Equity)	0.020*	(0.010)	[0.050]	0.020*	(0.009)	[0.060]
CHE (Cash and Short-Term Investments)	0.022***	(0.007)	[0.002]	0.019***	(0.007)	[0.005]
IB (Income Before Extraordinary Items)	0.044**	(0.016)	[0.010]	0.039***	(0.010)	[0.001]
LEV (Leverage Ratio)	−0.016	(0.013)	[0.207]	−0.018	(0.014)	[0.183]
PRSTKC (Purchase of Common and Preferred Stock)	0.012*	(0.006)	[0.060]	0.015**	(0.006)	[0.020]
SALE (Net Sales)	−0.923	(2.043)	[0.650]	−0.299	(2.100)	[0.890]
SPI (Special Items)	0.004	(0.004)	[0.330]	0.007	(0.005)	[0.150]
RET (Return on Stock Price)	0.201***	(0.036)	[0.000]	0.193***	(0.035)	[0.000]
Constant	0.175***	(0.039)	[0.000]	0.193***	(0.040)	[0.000]
Observations	525,067			525,067		
R^2	0.065			0.072		
Adj. R^2	0.065			0.072		

Panel B. Beat Indicator and Transparency Index

	(1) Industry FE + Trend			(2) Industry + Year FE		
Variables	*b*	*se*	*p*	*b*	*se*	*p*
Information Environment Variable						
Transparency Index	0.002***	(0.000)	[0.000]	0.002***	(0.000)	[0.000]
Time trend	0.002**	(0.001)	[0.030]			
Accounting Variables						
ACT (Current Assets)	−0.012	(0.010)	[0.230]	−0.012	(0.009)	[0.210]
AQC (Acquisitions)	−0.002	(0.003)	[0.520]	−0.001	(0.003)	[0.640]
AT (Total Assets)	−0.015*	(0.008)	[0.070]	−0.015*	(0.008)	[0.070]
CAPX (Capital Expenditures)	0.002	(0.003)	[0.510]	0.005*	(0.003)	[0.090]
CEQ (Common Equity)	−0.005	(0.005)	[0.290]	−0.005	(0.005)	[0.290]
CHE (Cash and Short-Term Investments)	0.012**	(0.005)	[0.020]	0.011**	(0.005)	[0.030]
IB (Income Before Extraordinary Items)	0.018**	(0.009)	[0.045]	0.016**	(0.008)	[0.050]
LEV (Leverage Ratio)	−0.023***	(0.009)	[0.010]	−0.023***	(0.009)	[0.010]
PRSTKC (Purchase of Common and Preferred Stock)	0.004	(0.004)	[0.310]	0.005	(0.004)	[0.240]
SALE (Net Sales)	−1.614	(1.432)	[0.260]	−1.301	(1.439)	[0.350]
SPI (Special Items)	0.001	(0.003)	[0.690]	0.002	(0.003)	[0.580]
RET (Return on Stock Price)	0.097***	(0.023)	[0.000]	0.093***	(0.022)	[0.000]
Constant	0.635***	(0.024)	[0.000]	0.651***	(0.025)	[0.000]
Observations		525,067			525,067	
R^2		0.029			0.032	
Adj. R^2		0.028			0.032	

TABLE 4. Relative Advantages of Human vs. Machine Forecasts: Access to Information

This table examines how national information environments, measured by the **Access to Information** index from the Robust Democracy Dataset (Varieties of Democracy Project) available during 2013–2021, relate to the relative forecasting performance of humans and machines. Panel A uses a continuous measure of relative forecast performance (machine minus analyst forecast error), while Panel B uses a binary indicator equal to one if the machine forecast outperforms the analyst (*Beat* = 1). *Access to Information* evaluates whether media are independent and express a diversity of opinions, and whether government information is accessible to the public. It incorporates key components such as media freedom, media pluralism, and citizens' access to government information. All regressions include full sets of firm- and macro-level controls as well as industry fixed effects. Column (2) further adds year fixed effects to remove global time trends. Country, firm, and analyst fixed effects are excluded because they would absorb the cross-country variation in national information environments that identifies Hypothesis 1. (See Appendix E for detailed explanations of fixed effect choices and specifications.) Standard errors are clustered at the firm level. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels.

Panel A. Relative Forecast Performance and Access to Information

	(1) Industry FE + Trend			(2) Industry + Year FE		
Variables	*b*	*se*	*p*	*b*	*se*	*p*
Information Environment Variable						
Access to Information	0.029***	(0.005)	[0.000]	0.031***	(0.007)	[0.000]
Time trend	−0.004***	(0.002)	[0.010]			
Accounting Variables						
ACT (Current Assets)	−0.057***	(0.020)	[0.005]	−0.058***	(0.020)	[0.004]
AQC (Acquisitions)	0.004	(0.007)	[0.580]	0.004	(0.007)	[0.590]
AT (Total Assets)	0.008	(0.038)	[0.830]	0.006	(0.038)	[0.870]
CAPX (Capital Expenditures)	−0.005	(0.017)	[0.770]	−0.007	(0.016)	[0.660]
CEQ (Common Equity)	−0.006	(0.016)	[0.710]	−0.007	(0.016)	[0.670]
CHE (Cash and Short-Term Investments)	0.081***	(0.024)	[0.001]	0.081***	(0.024)	[0.001]
IB (Income Before Extraordinary Items)	0.043**	(0.021)	[0.040]	0.040**	(0.021)	[0.050]
PRSTKC (Purchase of Common and Preferred Stock)	−0.038***	(0.007)	[0.000]	−0.037***	(0.008)	[0.000]
SPI (Special Items)	−0.007*	(0.004)	[0.080]	−0.008*	(0.004)	[0.070]
RET (Return on Stock Price)	0.123***	(0.031)	[0.000]	0.126***	(0.031)	[0.000]
Constant	0.415***	(0.059)	[0.000]	0.337***	(0.051)	[0.000]
Observations	534,684			534,684		
R^2	0.066			0.068		
Adj. R^2	0.066			0.068		

Panel B. Beat Indicator and Access to Information

	(1) Industry FE + Trend			(2) Industry + Year FE		
Variables	*b*	*se*	*p*	*b*	*se*	*p*
Information Environment Variable						
Access to Information	0.014***	(0.003)	[0.000]	0.015***	(0.004)	[0.000]
Time trend	−0.001	(0.001)	[0.340]			
Accounting Variables						
ACT (Current Assets)	−0.026**	(0.010)	[0.010]	−0.027**	(0.010)	[0.008]
AQC (Acquisitions)	0.003	(0.004)	[0.450]	0.003	(0.004)	[0.460]
AT (Total Assets)	0.008	(0.018)	[0.650]	0.007	(0.017)	[0.670]
CAPX (Capital Expenditures)	0.001	(0.009)	[0.880]	0.001	(0.009)	[0.890]
CEQ (Common Equity)	0.001	(0.009)	[0.920]	0.001	(0.009)	[0.910]
CHE (Cash and Short-Term Investments)	0.036***	(0.011)	[0.001]	0.037***	(0.011)	[0.001]
IB (Income Before Extraordinary Items)	0.023**	(0.009)	[0.015]	0.023**	(0.009)	[0.016]
PRSTKC (Purchase of Common and Preferred Stock)	0.016***	(0.004)	[0.000]	0.015***	(0.004)	[0.000]
SPI (Special Items)	−0.005**	(0.002)	[0.020]	−0.005**	(0.002)	[0.020]
TXDB (Deferred Taxes and Investment Tax Credit)	0.011***	(0.004)	[0.006]	0.011***	(0.004)	[0.007]
RET (Return on Stock Price)	0.057***	(0.019)	[0.003]	0.056***	(0.019)	[0.003]
Constant	0.750***	(0.039)	[0.000]	0.727***	(0.033)	[0.000]
Observations		534,684			534,684	
R^2		0.029			0.030	
Adj. R^2		0.029			0.030	

TABLE 5. Causal Evidence from MiFID II: Soft-Information Shock

This table presents difference-in-differences estimates evaluating the causal impact of the 2018 *Markets in Financial Instruments Directive II* (*MiFID II*) reform on the relative forecasting performance of machines versus human analysts. The dependent variable is the firm-year machine win rate, defined as the proportion of forecasts in which the machine's absolute error is smaller than the analyst's. The regression uses winsorized data and includes both the main difference-in-differences specification and a small-window event-study over ±3 years around the reform. Industry and year fixed effects are included, and treatment covers EU countries plus the United Kingdom and Switzerland. Standard errors are clustered at the country level (Column 1) and two-way clustered by country and firm (Column 2). A positive coefficient on *did_mifid* indicates that machine forecasts outperform human analysts more strongly after the MiFID II reform in treated markets. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A. MiFID II Difference-in-Differences (Main Specification)

	(1) Clustered by Country			(2) Two-Way Cluster (Country × Firm)		
Variables	*b*	*se*	*p*	*b*	*se*	*p*
Treatment Effect						
did_mifid	0.030*	(0.014)	[0.030]	0.030*	(0.014)	[0.030]
Selected Accounting Controls						
ACT (Current Assets)	0.027***	(0.007)	[0.000]	0.027***	(0.007)	[0.000]
CHE (Cash Holdings)	−0.036***	(0.008)	[0.001]	−0.036***	(0.008)	[0.001]
DP (Depreciation)	0.015***	(0.004)	[0.001]	0.015***	(0.004)	[0.001]
IB (Income Before Extraordinary Items)	−0.022***	(0.004)	[0.000]	−0.022***	(0.004)	[0.000]
PRSTKC (Share Repurchase)	−0.016***	(0.003)	[0.000]	−0.016***	(0.003)	[0.000]
SPI (Special Items)	0.005**	(0.001)	[0.031]	0.005**	(0.001)	[0.031]
RET (Stock Return)	−0.060***	(0.003)	[0.000]	−0.060***	(0.003)	[0.000]
Constant	0.146***	(0.009)	[0.000]	0.146***	(0.009)	[0.000]
Observations	509,254			509,254		
R^2	0.096			0.096		
Adj. R^2	0.096			0.096		

This panel provides a complementary specification using non-winsorized data and excluding firm-year observations with fewer than ten forecasts. The analysis estimates both the main difference-in-differences regression and an event-study over ±5 years to verify dynamic effects around the reform. Treatment includes all EU and EEA markets (Switzerland, the United Kingdom, Norway, Iceland, and Liechtenstein). All regressions include a comprehensive set of firm- and macro-level control variables (approximately 150 in total, not all reported due to space constraints). Standard errors are clustered at the country level.

Panel B. MiFID II Difference-in-Differences (Alternative Specification)

	Industry × Year FE		
Variables	*b*	*se*	*p*
Treatment Effect			
did_mifid	0.040**	(0.015)	[0.012]
Selected Accounting Controls			
ACT (Current Assets)	0.030***	(0.006)	[0.000]
CAPX (Capital Expenditures)	0.012***	(0.004)	[0.005]
CHE (Cash Holdings)	−0.036***	(0.009)	[0.001]
DP (Depreciation)	0.015***	(0.004)	[0.001]
IB (Income Before Extraordinary Items)	−0.031***	(0.004)	[0.000]
PRSTKC (Share Repurchase)	−0.019***	(0.004)	[0.000]
SPI (Special Items)	0.004**	(0.002)	[0.031]
RET (Stock Return)	−0.058***	(0.004)	[0.000]
Constant	0.146***	(0.009)	[0.000]
Observations	508,999		
R^2	0.149		
Adj. R^2	0.148		

Table 6. Causal Evidence from IFRS Adoption: Hard Information Shock

This table reports the staggered difference-in-differences estimates evaluating the impact of the mandatory adoption of *International Financial Reporting Standards* (IFRS) on the relative forecasting performance of human analysts versus machine-learning models. The dependent variable is the firm-year *machine win rate*, defined as the proportion of forecasts in which the machine's Column (1) reports the Callaway–Sant'Anna (2021) staggered DiD (CSDID) estimates, which control for the full set of firm-level accounting variables (coefficients not reported). Column (2) reports the traditional two-way fixed-effects (TWFE) specification with country and year fixed effects, where coefficients for selected accounting controls are shown. All regressions include the full set of firm-level accounting controls; standard errors are clustered by country. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1) CSDID: Firm and Year FE			(2) TWFE: Country and Year FE		
Variables	*b*	*se*	*p*	*b*	*se*	*p*
Treatment Effect						
Average ATT (CSDID) / Post × Treated (TWFE)	-0.192***	(0.073)	[0.009]	-0.029	(0.020)	[0.162]
Selected Accounting Controls						
DD1 (Depreciation and Depletion)				0.007*	(0.004)	[0.086]
DLC (Debt in Current Liabilities)				0.004	(0.004)	[0.329]
DLTT (Long-Term Debt)				0.013**	(0.006)	[0.032]
DP (Depreciation)				0.008	(0.008)	[0.364]
DVC (Cash Dividends Declared)				-0.012***	(0.003)	[0.001]
Fixed Effects	Firm & Year			Country & Year		
Observations	18,206			18,206		
Firms	5,257			5,257		
Countries	44			44		
Treated Countries	40			40		
Mean Adoption Year	2007.8			2007.8		

Appendix A. Machine Learning Techniques

This section provides a more detailed description of the supervised learning techniques I explore for forecasting firm earnings. I begin with the class of penalized linear estimators, followed by tree-based methods.

A.1. Quasi-Linear Models

Quasi-linear models are a class of supervised learning algorithms that combine linear regression with regularization techniques. These models aim to balance the trade-off between model complexity and prediction accuracy. The following three quasi-linear models are commonly used in financial analysis:

Lasso

Lasso (Least Absolute Shrinkage and Selection Operator) is a penalized linear estimator that adds an L1 penalty term to the mean squared error (MSE) loss function. The objective function for Lasso is defined as:

$$L(\beta, \alpha_1, \alpha_2) = \sum \left[(EPS - X'\beta)^2 \right] + \alpha_1 \|\beta\|_1 + \alpha_2 \|\beta\|_2^2,$$

where β represents the coefficient vector, α_1 and α_2 are the penalty parameters controlling the amount of regularization.

Ridge

Ridge is another penalized linear estimator that introduces an L2 penalty term to the MSE loss function. The objective function for Ridge is given by:

$$L(\beta, \alpha_1, \alpha_2) = \sum \left[(EPS - X'\beta)^2 \right] + \alpha_1 \|\beta\|_1 + \alpha_2 \|\beta\|_2^2,$$

where the terms have the same meaning as in Lasso.

Elastic Net

Elastic Net combines the L1 and L2 penalties of Lasso and Ridge, respectively, in order to leverage the benefits of both regularization techniques. The objective function for Elastic Net is defined as:

$$L(\beta, \alpha_1, \alpha_2) = \sum \left[(EPS - X'\beta)^2 \right] + \alpha_1 \|\beta\|_1 + \alpha_2 \|\beta\|_2^2,$$

where α_1 and α_2 control the amount of regularization.
The hyperparameters α_1 and α_2 are chosen using cross-validation on the training set to avoid introducing any look-ahead bias.

A.2. Non-Linear Models

Non-linear models offer greater flexibility in capturing complex relationships between predictor variables and firm earnings. I consider two popular non-linear models:

Random Forest

Random Forest is an ensemble learning method that combines multiple regression trees. Each tree is built using a random subset of predictor variables and a random subset of observations. The final prediction is obtained by averaging the predictions of all the trees. Random Forest is regularized through the averaging of trees with different structures, reducing prediction variance and limiting overfitting. The hyperparameters, such as the number of trees and the maximum number of splits, can be chosen using cross-validation on the training set.

Gradient-Boosted Trees

Gradient-Boosted Trees (GBT) is another tree-based method that builds an ensemble of regression trees in a sequential manner. GBT starts by fitting a shallow tree of depth d to the data and calculates the residuals from this tree. Then, another shallow tree of depth d is fitted to the residuals, and this process is repeated for a specified number of iterations. The predicted values are formed by adding the predicted values from each tree, with a regularization factor λ applied to shrink the predicted values from subsequent trees. By growing trees sequentially on the residuals from previous trees, GBT reduces correlation among the trees and limits overfitting.

GBT has three hyperparameters: the number of iterations B, the depth of each tree d, and the regularization factor λ. These hyperparameters can be chosen using cross-validation on the training set.

Appendix B. Forecasts formation

B.1. Analyst Forecast Processing

We implement several procedures to ensure the robustness of analyst forecast data:

- **Outlier Control**: Earnings-per-share (EPS) forecasts are winsorized at the 5% level within each country on an annual basis to remove extreme values that may distort analysis.
- **Date Alignment**: Forecast dates and earnings release dates are standardized using UTC timestamps to ensure temporal consistency across countries and time zones.
- **Forecast Horizon Classification**: Forecasts are grouped into short-, mid-, and long-term horizons depending on the number of calendar days between the forecast date and the actual earnings announcement. The classification procedure follows Dessaint, Foucault, and Frésard (2021), and details are provided in Section 2.5.

B.2. Machine Forecast Formation

Following the common framework of current literature Van Binsbergen, Han, and Lopez-Lira (2020), De Silva and Thesmar (2021), Cao et al. (2021), I build the machine analyst using the following framework:

- **Build a Machine Analyst using algorithms**: I construct the machine analyst by leveraging supervised learning algorithms. Specifically, I utilize quasi-linear models such as Lasso, Ridge, Elastic Net, as well as non-linear models such as Random Forest and Gradient-Boosted Trees. These algorithms are chosen for their ability to capture complex patterns and relationships in the data.
- **Feed public information to the machine and generate forecasts**: The machine analyst is trained using a combination of publicly available information. This includes financial statements, market data, macro information and other relevant variables. The trained machine analyst then generates forecasts based on this input.
- **Compare Man vs. Machine in forecasting accuracy**: I assess the forecasting accuracy of the machine analyst by comparing its predictions against the historical forecasts made by human analysts. This

comparison allows us to evaluate the performance of the machine analyst in terms of accuracy and reliability.

- **Feed (analyst forecasts + public information) to the machine to build a (Man+Machine) analyst**: To further enhance the forecasting process, I combine the forecasts made by the human analysts with the information provided to the machine analyst. This fusion of inputs creates a hybrid forecasting approach, referred to as the (Man+Machine) analyst.

- **Compare Man vs. Machine vs (Man+Machine)**: Finally, I compare the forecasting performance of the human analyst, machine analyst, and the hybrid (Man+Machine) analyst. This comparison enables us to evaluate the relative strengths and weaknesses of each approach and identify the most effective forecasting strategy.

Appendix C. International Datasets

C.1. List of Country/Region

The study considers the countries included in the MSCI ACWI Index, which represents stocks from 23 Developed Markets and 24 Emerging Markets, covering about 85% of the global investable equity market. See Table 7 for the full list of countries and regions.

TABLE 7. List of Countries and Regions in the MSCI ACWI Index

MSCI World Index (Developed Markets)	
Americas	Canada, United States
Europe & Middle East	Austria, Belgium, Denmark, Finland, France, Germany, Ireland, Israel, Italy, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom
Pacific	Australia, Hong Kong, Japan, New Zealand, Singapore
MSCI Emerging Markets Index	
Americas	Brazil, Chile, Colombia, Mexico, Peru
Europe, Middle East & Africa	Czech Republic, Egypt, Greece, Hungary, Kuwait, Poland, Qatar, Saudi Arabia, South Africa, Turkey, United Arab Emirates
Asia	Mainland China, India, Indonesia, Korea, Malaysia, Philippines, Taiwan, Thailand

Appendix D. Data Inputs

D.1. Fundamental Variables

Compustat

Owing to the variations in the available variables lists between Compustat North America (covering the United States and Canada) and Compustat Global (covering all other countries/regions), the fundamental variables gathered for the United States and Canada differ slightly from those in other countries within the international datasets. Efforts were made to collect comparable variables to minimize the differences in fundamental variables as much as feasible.

It's important to mention that for some variables, the naming conventions in the Compustat Global dataset differ from those in the Compustat North America dataset.

The comprehensive lists of Computstat variables for North America and other countries/regions in the international datasets are detailed in Table 8 and Table 9, respectively.

TABLE 8. Fundamental Variables from Compustat North America

Variable	**Required non-missing?**
Total assets	✓
Total liabilities	✓
Revenue	✓
SG&A expense	
R&D expense	
Cost of goods sold	✓
Current assets	
Current liabilities	
Cash	
Cash and short-term investments	
Income tax expense	
Total long-term debt	
Total long-term debt due within one-year	
Debt in current liabilities	
Depreciation expense	
EBIT	
EBITDA	✓
Interest expense	
Interest paid	
Capital expenditures	
Goodwill	
Income tax payable	

Continued on next page

Variable	**Required non-missing?**
Income tax expense	
Total income tax	
Net income	
Common dividends	
Purchase of common and preferred stock	
Sale of common and preferred stock	
Subordinated debt	
Gross profit	✓
Operating cash flow	✓
Common shares outstanding	
Stock price at fiscal year end	✓
Extraordinary items	
Common ESOP obligation	
Special items	
Acquisitions	
Capitalized leases (due within two-years)	
Capitalized leases (due within three-years)	
Capitalized leases (due within four years)	
Capitalized leases (due within five years)	
Interest and related income (total)	
Total intangible assets	
Marketable securities adjustment	
Net PPE	✓
Nonoperating income	

Continued on next page

Variable	Required non-missing?
Tax loss carryforward	
Pension and retirement expense	
Preferred stock value	

Table 9. Fundamental Variables from Compustat Global

Compustat Name	Variable	Required non-missing?
act	Total Current Assets	
aqc	Acquisitions	
at	Total assets	✓
capfl	Capital Element of Finance Lease Rental Payments	
capx	Capital Expenditures	
ceq	Total Common/Ordinary Equity	
ch	Cash	
che	Cash and Short-Term Investments	
cogs	Cost of Goods Sold	✓
dd1	ong-Term Debt Due in One Year	
dlc	Total Debt in Current Liabilities	
dltt	Total Long-Term Debt	
dp	Depreciation and Amortization	✓
dvc	Dividends Common/Ordinary	
ebit	Earnings Before Interest and Taxes	✓
gdwl	Goodwill	
ib	Income Before Extraordinary Items	
idit	Total Interest and Related Income	
intan	Total Intangible Assets	
intpn	Net Interest Paid	
lct	Total Current Liabilities	
lt	Total Liabilities	✓

Continued on next page

Compustat Name	Variable	Required non-missing?
nopi	Nonoperating Income (Expense)	
oancf	Operating Activities - Net Cash Flow	✓
opprft	Operating Profit	
pi	Pretax Income	
ppent	Property, Plant and Equipment - Total (Net)	✓
prstkc	Purchase of Common and Preferred Stock	
pstk	Preferred/Preference Stock (Capital) - Total	
pstkn	Preferred/Preference Stock - Nonredeemable	
pstkr	Preferred/Preference Stock - Redeemable	
revt	Total Revenue	✓
sale	Sales/Turnover (Net)	
spi	Special Items	
sstk	Sale of Common and Preferred Stock	
tx	Taxation	
txdb	Deferred Taxes (Balance Sheet)	
txp	Income Taxes Payable	
txpd	Income Taxes Paid	
txt	Income Taxes - Total	
unnp	Unappropriated Net Profit (Stockholders' Equity)	
xi	Extraordinary Items	
xint	Interest and Related Expense - Total	
xpr	Pension and Retirement Expense	
xrd	Research and Development Expense	
xsga	Selling, General and Administrative Expense	

Continued on next page

Compustat Name	Variable	Required non-missing?
ajexi	Adjustment Factor (International Issue)-Cumulative by Ex-Date	
cshoi	Com Shares Outstanding - Issue	
cshpria	Common Shares Used to Calculate Earnings Per Share (Basic) - As Reported	
epsexcon	Earnings Per Share (Basic) - Excluding Extraordinary Items - Consolidated	
epsexnc	Earnings Per Share (Basic) - Excluding Extraordinary Items - Nonconsolidated	
epsincon	Earnings Per Share (Basic) - Including Extraordinary Items - Consolidated	
epsinnc	Earnings Per Share (Basic) - Including Extraordinary Items - Nonconsolidated	

CRSP and Datastream

For gathering firm-specific data on stock prices and market capitalization, I utilized resources from CRSP and Datastream. Following the literature, CRSP is commonly employed for accessing stock price data for firms within North America, specifically the United States and Canada. However, given CRSP's inaccessibility outside of North America, this study relies on Datastream to obtain information on stock prices and market valuations, which are crucial inputs for the econometric forecasting algorithms.

The comprehensive lists of price-related variables used for constructing market valuation inputs are detailed in Table 10 for CRSP (North America) and Table 11 for Datastream (international markets).

TABLE 10. Fundamental Variables from CRSP

Variable	**Required non-missing?**
SIC 2-digit industry code dummie	✓
Return over prior month to fiscal year end t	✓
Return over year prior to fiscal year end t, excluding last month	✓
Market capitalization at the end of year t	✓

TABLE 11. Fundamental Variables from Datastream

Variable	**Required non-missing?**
Adjusted close price on market date t	✓
Return over the day prior to market date t	✓
Market capitalization on market date t	✓

D.2. Macroeconomic Variables

To gather information on the macroeconomic conditions at the country level, I sourced data from Trading Economics and FactSet, serving as crucial inputs for the econometric forecasting algorithms.

Table 12. Macro Variables

Variable	**Required non-missing?**
balance_of_trade	✓
consumer_confidence	✓
consumer_price_index_cpi	✓
crude_oil_rigs	✓
currency	✓
exports	✓
gdp_growth_rate	✓
government_bond_10y	✓
government_debt_to_gdp	✓
gross_fixed_capital_formation	✓
gross_national_product	✓
housing_index	✓
imports	✓
inflation_rate	✓
interbank_rate	✓
interest_rate	✓
labor_force_participation_rate	✓
labour_costs	✓

Continued on next page

Variable	**Required non-missing?**
manufacturing_pmi	✓
money_supply_m0	✓
money_supply_m1	✓
money_supply_m2	✓
money_supply_m3	✓
productivity	✓
retail_sales_mom	✓
services_pmi	✓
stock_market	✓
unemployment_rate	✓

Note: This is the standard list of macroeconomic variables used in our analysis. For most countries in my sample, these variables are fully available. However, due to limitations in public data disclosure, a few countries may have incomplete macroeconomic coverage.

Appendix E. Fixed-Effects Structure and Identification Strategy

This appendix explains why the main specifications do not include country, firm, or analyst fixed effects, and how identification is achieved.

Country Fixed Effects.. The main research question in Hypothesis 1 concerns how *cross-country differences* in national information environments affect the relative forecasting performance of humans and machines. Including country fixed effects would absorb all time-invariant cross-country variation, rendering the information-environment variables, which are defined at the country-year level, collinear with the fixed effects.[4] My identification therefore relies on within-year, across-country differences (Year FE in) that capture contemporaneous cross-sectional variation.

Analyst Fixed Effects.. Country-level regressors cannot be identified when analyst fixed effects are included because the vast majority of analysts cover firms within a single country. In such cases, the national information environment is effectively constant within analyst, leaving no within-analyst variation for estimation. When I include analyst fixed effects, the coefficient on the information-environment index becomes mechanically under-identified and converges toward zero. My preferred specification for Hypothesis 1 thus includes industry and year fixed effects, exploiting within-year, cross-country variation in information environments. In a subsample of cross-border analysts, that is, those issuing forecasts across multiple countries, the coefficient re-emerges with the expected positive sign and magnitude, confirming that the attenuation arises purely from lack of within-analyst variation rather than misspecification.

Firm Fixed Effects.. Each firm in the global sample is assigned to a single country based on its headquarters location. That country determines the institutional information environment relevant for the firm, which remains constant over time. Including firm fixed effects would therefore eliminate the cross-country variation that identifies the effect of information environments. Instead, I account for firm-level heterogeneity through a rich set of controls such as size, profitability, leverage, and market capitalization, and include industry and year fixed effects. This approach preserves the country-level

[4] Formally, the country fixed effects fully span the space of country-level indicators, leaving no remaining within-country variation to identify β_{info}.

variation necessary for identifying how institutional information quality shapes the relative advantage of human versus machine forecasts.

Overall, the empirical identification of Hypothesis 1 relies on differences in national information environments across countries within the same year, while controlling for time trends, firm characteristics, and industry heterogeneity. The fixed-effects structure is therefore designed to balance robustness and interpretability without mechanically removing the key source of variation.