The European Journal of Finance

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rejf20

From text to treasure: the predictive superiority of a FinTech index in stock market returns

Yangli Guo, Feng Ma, Yizhi Wang & Juandan Zhong

To cite this article: Yangli Guo, Feng Ma, Yizhi Wang & Juandan Zhong (15 Sep 2024): From text to treasure: the predictive superiority of a FinTech index in stock market returns, The European Journal of Finance, DOI: 10.1080/1351847X.2024.2399773

To link to this article: https://doi.org/10.1080/1351847X.2024.2399773

Published online: 15 Sep 2024.

THE EUROPEAN JOURNAL OF FINANCE
https://doi.org/10.1080/1351847X.2024.2399773

From text to treasure: the predictive superiority of a FinTech index in stock market returns

Yangli Guo[a], Feng Ma[a], Yizhi Wang[b] and Juandan Zhong[a]

[a]School of Economics and Management, Southwest Jiaotong University, Chengdu, People's Republic of China; [b]Cardiff Business School, Cardiff University, UK

ABSTRACT

This study employs a text analysis methodology to construct a Financial Technology (FinTech) Index, utilizing textual data from The New York Times. The primary aim is to investigate the correlation between financial technology and stock market performance. Our findings provide compelling evidence that the FinTech Index possesses substantial predictive capability for excess returns in the US stock market, a feature that becomes particularly pronounced during economic downturns. Notably, when compared with traditional macroeconomic indicators, the FinTech Index offers valuable incremental insights. Moreover, this study expands to include sector-level and international market analyses, demonstrating the broad applicability and robust performance of the FinTech Index. Importantly, through the use of out-of-sample testing, we substantiate that the FinTech Index demonstrates superior predictive accuracy, presenting opportunities for investors to achieve higher economic returns.

ARTICLE HISTORY
Received 22 April 2024
Accepted 26 August 2024

KEYWORDS
FinTech index; text analysis; predictive power; stock market returns; economic downturns

JEL CLASSIFICATIONS
G17; G12; C53; G14; O33

1. Introduction

The development of internet technology has made financial services and products more accessible, gradually evolving into the fintech industry. The outbreak of COVID-19 in 2020 accelerated the application of digital technology in financial services. By 2023, 11% of the world's unicorn companies were fintech firms.[1] According to KPMG's report, a staggering $164.1 billion flooded into global FinTech investments in 2022, distributed across 6,006 deals. Utilizing state-of-the-art instruments like blockchain, cloud computing, artificial intelligence, and big data analysis, FinTech is reshaping the traditional contours of the financial sector (Goldstein, Jiang, and Karolyi 2019; Hendershott et al. 2021; Wang, Liu, and Luo 2021), offering fresh avenues for investment, trading, and wealth management. However, it has also introduced potential risks such as market manipulation and cybersecurity threats. As investors seek to capitalize on these advancements, comprehending the intricate interplay between the equity market and technology has become paramount. Simultaneously, their dynamic relationship has emerged as a focal point for scholarly inquiry and practical application (Garlappi and Song 2020; Gârleanu, Kogan, and Panageas 2012a, Garleanu, Panageas, and Yu 2012b; Hirshleifer, Hsu, and Li 2013; Hirshleifer, Hsu, and Li 2018; Hsu and Huang 2010; Kaltenbrunner and Lochstoer 2010; Kogan and Papanikolaou 2014; Lin 2012; Sharma and Narayan 2022). Hsu and Huang (2010) argue that technological advancements can elucidate stock returns. Garleanu, Panageas, and Yu (2012b) suggest that asset prices swiftly respond to the emergence of new technologies. Kaltenbrunner and Lochstoer (2010), Garlappi and Song (2020), and Sharma and Narayan (2022) delve into modeling the impact of technology shocks on market returns and premiums. Inspired by the aforementioned literature, the purpose of this study is to look into the connection between stock market performance and financial technologies.

CONTACT Yizhi Wang wangy510@cardiff.ac.uk

We construct a FinTech Index based on abstract texts from The New York Times spanning from January 2000 to June 2022. Shiller (2015) highlights the significance of news media, as they often emphasize stories that resonate with investors when reporting investment-related narratives. Numerous studies utilise information from news media to analyze stock markets (An et al. 2020; Garcia 2013; Tetlock 2007). However, there remains debate over the most appropriate methods for studying and analyzing language (Pennebaker, Mehl, and Niederhoffer 2003). One approach suggests that language is inherently context-dependent, thus necessitating comprehensive contextual analysis. Another approach advocates for the use of strategies such as word frequency statistics (e.g. DICTION), as human judges are often influenced by the content when reading texts and may not fully control the selection of words (Hart 2001). Young and Soroka (2012) demonstrate that basic word frequency statistics can provide robust and reliable analysis of text themes and emotional composition if there is a well-defined, comprehensive dictionary. In this study, we utilise the dictionary method, conducting simple word counts of key words from a predefined dictionary within the text, to construct the FinTech Index.

We initiate our study by employing traditional predictive estimation to analyze the association between the lagged FinTech Index and the excess returns of the US stock market. The FinTech Index demonstrates a significant in-sample R^2 of 1.965%, indicating its substantial explanatory power. With each unit increase in the FinTech Index, the subsequent month's excess returns in the US market increase by 0.654%. Furthermore, the FinTech Index continues to exhibit predictive power for quarterly US excess returns. However, for for longer intervals such as half a year or a year, the regression coefficients of the FinTech Index lose significance. We also look into how predictable returns are at various points in the business cycle. The FinTech Index shows a significant correlation with US excess returns during both recessionary and expansionary periods, although its impact is more pronounced during recessions. This finding suggests a time-varying influence of the FinTech Index. Therefore, inspired by the finding of Sharma and Narayan (2022), we further explore the dynamic predictability of the FinTech Index for the US stock market.

We also assess the predictability of returns using the FinTech Index in comparison to various macroeconomic forecasting indicators (Jiang et al. 2019; Rapach, Ringgenberg, and Zhou 2016). Specifically, we employ 14 commonly used macroeconomic variables as outlined by Welch and Goyal (2008). These include Log dividend-price ratio, Log dividend yield, Log earnings-price ratio, Log dividend-payout ratio, Excess stock return volatility, Book-to-market ratio, Net equity expansion, Treasury bill rate, Long-term yield, Long-term return, Term spread, Default yield spread, Default return spread, and Inflation. Our analysis reveals that the FinTech Index's capacity in forecasting outperforms that of most macroeconomic forecasting indicators. Furthermore, even when considering bivariate tests controlling for macroeconomic indicators, the FinTech Index exhibits significant predictive ability.

To gain a holistic perspective on the FinTech Index's standing, we explore its association with stock returns at the industry level. We scrutinize the ten sector indices within the US S&P 500, which include consumer discretionary, consumer staples, health care, industrials, information technology, materials, telecommunications services, utilities, financials, and energy. Our investigation reveals compelling insights: the constructed FinTech Index demonstrates substantial predictive prowess across seven sectors, with the exceptions of consumer staples, telecommunications services, and energy. Particularly noteworthy is its robust predictive performance within the information technology sector, boasting an impressive in-sample R^2 of 5.410%. Furthermore, we extend our analysis to ascertain the predictive efficacy of the FinTech Index in forecasting excess returns across the stock markets of ten developed countries: Australia, Canada, France, Germany, Italy, Japan, the Netherlands, Sweden, Switzerland, and the United Kingdom. The findings underscore the broad applicability of the US FinTech Index, demonstrating a commendable fit with the majority of these countries' stock markets, excluding Canada, Japan, and Switzerland.

Welch and Goyal (2008) note that prominent forecasting factors do not predict stock risk premiums based on out-of-sample testing, despite a wealth of data supporting their in-sample predictive power. Therefore, we additionally investigate the out-of-sample prediction power of the FinTech Index (Campbell and Thompson 2008). We find that compared to 14 well-known macroeconomic indicators, our constructed FinTech Index exhibits the highest out-of-sample R^2. In order to assess the FinTech Index's informational value in comparison to the 14 macroeconomic indicators, we conduct a forecast encompassing test. The results indicate that, compared to most macroeconomic indicators, the FinTech Index contains incremental information.

We also investigate the economic value of FinTech Index-based stock market forecasts. Following Campbell and Thompson (2008), Rapach, Strauss, and Zhou (2010), and Ferreira and Santa-Clara (2011), we use out-of-sample forecasts to calculate the Certainty Equivalent Returns (CER) and Sharpe ratio for mean-variance investors who optimally allocate their wealth between stocks and risk-free assets. We discover that the FinTech Index can yield higher economic returns for investors than the majority of macroeconomic indices. The Sharpe ratio is also higher than that of most macroeconomic indicators.

Our research adds significantly to the expanding body of knowledge on financial technology (fintech) and its impact on stock return predictability. Previous research has often used proxies such as fintech stock prices, patent counts, or R&D investments to assess the impact of technology on financial markets. For instance, Lin (2012) employed a dynamic equilibrium model to examine cross-sectional variations in stock returns related to technological advancements. In this model, technological progress is driven by R&D investments and includes both product innovation and new tangible capital. Hirshleifer, Hsu, and Li (2013) discovered that innovative efficiency (IE), quantified by the ratio of patents or citations to R&D expenditures, is a robust positive indicator of future stock market returns, even when accounting for firm characteristics and risk factors. Sharma and Narayan (2022) utilized patent data dating back to 1870 to calculate local and global technology shocks. Tiwari et al. (2023) examined the interconnections between fintech stocks, green financial assets, and energy markets, exploring how fluctuations in fintech stock prices during boom and bust periods impact the values of environmentally sustainable assets and energy markets. In this study, we construct a novel fintech index using text data extracted from New York Times summaries to reflect the level of attention given to fintech. We rigorously examine the predictive accuracy of this index within the US stock market, extending our analysis to encompass sector-specific returns across the ten sectors in the S&P 500. Additionally, we assess the index's efficacy in forecasting returns in international markets. Our findings reveal that the FinTech Index not only enhances our understanding of return forecasting ability in the US but also offers solid predictive insights globally, demonstrating its strong overall performance and broad applicability.

This paper is organized as follows in the subsequent sections: We describe our data in Section 2 and elucidate the process involved in constructing the financial technology index. Section 3 provides empirical analysis, presenting both in-sample and out-of-sample outcomes of predictive regression for the financial technology index alongside 14 frequently employed predictor variables, accompanied by robustness checks. Section 4 delves into the findings of our asset allocation analysis. Ultimately, Section 5 is the conclusion.

2. Data

2.1. Construction of the FinTech index

We construct the financial technology index using abstracts from The New York Times, spanning from January 2000 to June 2022. Following Engle et al. (2020), we initially curate a comprehensive dictionary of terms related to the financial technology. This dictionary encompassed a wide range of concepts, including FinTech, Artificial Intelligence, Blockchain, Cloud computing, Big data, Cryptocurrency, DeFi, CBDC, Non-fungible tokens, Stablecoin, Metaverse, P2P lending, and Crowdfunding, constituting 13 fundamental aspects, each associated with a specific list of keywords. Follow Tvinnereim and Fløttum (2015) and Ma et al.(2023), we then search for these keywords within The New York Times articles and tally the number of occurrences for each keyword daily. The monthly FinTech Index is determined by averaging the total relevant keywords for each month, expressed by the following formula:

$$\text{FinTech}_m = \frac{\sum_{t=1}^{N_m} f_{t,m}}{N_m}, \tag{1}$$

where FinTech_m signifies the FinTech Index for month m; $f_{t,m}$ denotes the number of keywords for day t in month m; whereas N_m signifies the total number of days in month m.

Additionally, to capture fintech trends without underlying fluctuations, we detrend the FinTech Index by subtracting a three-month moving average, following Campbell, Grossman, and Wang (1993) and Baltussen, van Bekkum, and Da (2019). This approach ensures our index reflects genuine fintech trends, making it a robust tool for analyzing the impact of fintech on financial markets.

Table 1. Descriptive statistics.

Variable	Mean	Median	Std.dev.	Minimum	Maximum	Skewness	Kurtosis	J-B
USA_r	0.004	0.009	0.044	−0.170	0.127	−0.521	3.940	21.995***
FinTech	−0.003	−0.008	0.938	−3.252	7.189	3.303	30.245	8776.154***
DP	−4.004	−3.974	0.193	−4.524	−3.281	0.028	4.735	33.646***
DY	−4.000	−3.968	0.194	−4.531	−3.295	−0.117	4.532	26.825***
EP	−3.158	−3.100	0.376	−4.836	−2.566	−2.098	9.049	605.323***
DE	−0.846	−0.949	0.440	−1.244	1.380	3.027	13.900	1735.928***
ROVL	0.147	0.138	0.057	0.055	0.317	0.587	2.883	15.552***
BM	0.280	0.286	0.068	0.121	0.446	−0.368	2.758	6.716***
NTIS	−0.004	−0.002	0.018	−0.056	0.029	−0.551	3.042	13.575***
TBL	1.453	0.920	1.704	0.010	6.170	1.204	3.354	66.187***
LTY	3.640	3.641	1.423	0.620	6.400	−0.129	2.018	11.505***
LTR	0.505	0.540	3.172	−11.240	14.430	0.107	4.807	36.992***
TMS	2.187	2.240	1.372	−0.410	4.530	−0.090	1.815	16.032***
DFY	1.031	0.920	0.413	0.550	3.380	3.074	15.159	2072.989***
DFR	0.070	0.050	1.971	−9.760	7.370	−0.778	8.673	386.481***
INFL	0.206	0.201	0.389	−1.915	1.335	−0.652	6.175	131.606***

Notes: This table presents descriptive statistics for the data utilized in this study, including excess returns on the US stock market, the FinTech Index, and 14 macroeconomic variables proposed by Welch and Goyal (2008). The 14 macroeconomic variables listed in the first column of the table include: Log dividend-price ratio (DP), Log dividend yield (DY), Log earnings-price ratio (EP), Log dividend-payout ratio (DE), Excess stock return volatility (ROVL), Book-to-market ratio (BM), Net equity expansion (NTIS), Treasury bill rate (TBL), Long-term yield (LTY), Long-term return (LTR), Term spread (TMS), Default yield spread (DFY), Default return spread (DFR), and Inflation (INFL). St. dev. is an abbreviation for standard deviation; J-B represents the statistic of the Jarque-Bera test, which is used to test whether the sequence conforms to the normal distribution; Asterisk '***' indicates significance at the 1% level.

2.2. Other data

Our empirical analysis is largely based on monthly excess returns within the US stock market. Following Rapach, Strauss, and Zhou (2013), excess returns are defined as the difference between the monthly returns of the S&P 500 Index and the previous month's risk-free rate. The risk-free rate used in this study is the three-month Treasury bill rate. The data utilised in this study is sourced from Global Financial Data.

For comparison, and following the methodologies of Rapach, Ringgenberg, and Zhou (2016) and Jiang et al. (2019), we also consider 14 macroeconomic variables from Welch and Goyal (2008). These variables include Log dividend-price ratio (DP), Log dividend yield (DY), Log earnings-price ratio (EP), Log dividend-payout ratio (DE), Excess stock return volatility (ROVL), Book-to-market ratio (BM), Net equity expansion (NTIS), Treasury bill rate (TBL), Long-term yield (LTY), Long-term return (LTR), Term spread (TMS), Default yield spread (DFY), Default return spread (DFR), and Inflation (INFL).[2] These variables are commonly employed as controlled parameters to evaluate the predictability of the objective variable (Chen et al. 2022; Ma et al. 2022; Wang et al. 2019). All the data utilized in this study covers the period from March 2000 to June 2022.

Table 1 provides the summary statistics regarding the excess returns observed in the U.S. stock market, the FinTech Index we constructed, and 14 macroeconomic variables. From left to right, the table displays the Mean, Median, Standard Deviation (Std. Dev.), Minimum, Maximum, Skewness, Kurtosis, and results of the Jarque-Bera Test (J-B) for the data. Notably, LTR exhibits the highest standard deviation, indicating heightened volatility compared to other variables. Furthermore, the variation in skewness and kurtosis values across the variables indicates diversity among them. Specifically, USA_r, DY, EP, BM, NTIS, LTY, MS, DFR, and INFL exhibit left-skewness and leptokurtosis, while other variables, including the FinTech Index, demonstrate right-skewness and leptokurtosis. Additionally, all time series significantly deviate from the assumption of a normal distribution assumptions at the 1% level, as evidenced by the Jarque-Bera test, confirming that none of the variables conform to a normal distribution pattern.

3. Empirical results

3.1. Predictive regression analysis

This section focuses on examining the in-sample forecasting capability of the financial technology index for excess returns in the US stock market, and this can be conducted simply by a standard univariate predictive regression as follow:

$$r_{t+h} = \alpha + \beta F_t + \varepsilon_{t+h}, \tag{2}$$

where r_{t+h} is the excess return of the US S&P 500 index; $r_{t+h} = (1/h)(r_{t+1} + \ldots + r_{t+h})$, $h = 1, 3, 6, 9, 12$; F_t represents the FinTech Index; α represents the intercept term; β represents the estimated coefficient for the FinTech Index. ε_{t+h} represents the random disturbance term. We identify the predictive power of FinTech Index by examining the statistical significance of β through full-sample estimation. When the null hypothesis $\beta = 0$ is rejected, it implies that The FinTech Index possesses predictive insights regarding stock returns., otherwise, FinTech Index fails to predict stock returns. Specifically, following Jiang et al. (2019), we apply the heteroskedasticity- and autocorrelation-robust Newey-West t-statistic.

Table 2 presents the regression findings regarding the FinTech Index. It is noted that the FinTech Index significantly predicts market excess returns over both monthly and quarterly periods. Specifically, in forecasting the upcoming month, our regression analysis reveals a significant coefficient of 0.654% for the FinTech Index, supported by a robust t-statistic of 3.587, signifying its remarkable significance at the 1% level. Similarly, when extending our analysis to a three-month horizon, the significance of the FinTech Index persists, evidenced by a coefficient of 0.310% and a corresponding t-statistic of 2.736, thus further affirming its influential role at the 1% level of significance. However, the predictive capability of the FinTech Index diminishes over longer horizons of 6, 9, and 12 months, as indicated by the weak t-statistics of β estimates. Garleanu, Panageas, and Yu (2012b) argue that asset prices respond swiftly to the emergence of new technologies. In the early stages of technological shocks, there is an increase in the risk premium. However, over time, the risk premium on their stocks decreases. Therefore, FinTech Index is difficult to accurately predict stock excess returns over longer time horizons. Furthermore, the in-sample R^2 evaluates the predictive significance of the FinTech Index from an economic standpoint. At the monthly and quarterly horizons, the in-sample R^2 values are 1.965% and 1.261%, respectively. Essentially, this implies that the FinTech Index contributes 1.965% and 1.261% to the variation of stock market excess return over time, respectively. However, when we extend our analysis to longer horizons, all in-sample R^2 values fall below 1%. This consistency aligns with the β estimate results, further reinforcing the FinTech Index's predictive potency in the short term while indicating its diminishing impact over longer periods.

Numerous studies have explored how economic variables perform differently in predicting outcomes across economic expansion and contraction periods (Rapach, Strauss, and Zhou 2010; Rapach and Zhou 2013; Zhu and Zhu 2013). Interestingly, it has been observed that factors used for predicting returns often demonstrate more robust performance during recessions, which conforms to the concept of countercyclical risk premiums (Fama and French 1989). Similarly, financial technology also has a close relationship with economic cycle (Beck et al. 2016; Fuerst 1995; Wang and Tan 2021). Governments and firms often adopt new technologies to enhance productivity and stimulate economic growth. However, during recessions, the pace at which these new technologies are integrated into production processes tends to slow down (Anzoategui et al. 2019). Under these circumstances, we explore the predictability of returns using the FinTech Index across different business cycles. Following Sharma and Narayan (2022), we conduct a regression model as follow:

$$r_{t+1} = \alpha + \beta^{rec} * F_t * D_t^{rec} + \beta^{exp} * F_t * (1 - D_t^{rec}) + \varepsilon_{t+1}, \tag{3}$$

where r_{t+1} is the excess returns, β^{rec} and β^{exp} signify the regression coefficients for recessionary and expansionary periods, respectively, D_t^{rec} is a dummy variable representing the economic cycle (A value of 1 indicates the recession phase, while 0 indicates expansion phase). According to Panel B of Table 2, it is observed that the FinTech Index predicts stock index returns with a significant t-statistic during both recessions and economic expansions. When incorporating the economic cycle into the regression equation, the in-sample R^2 is higher compared to models that do not differentiate between economic cycles. Additionally, the impact of the FinTech Index is slightly greater during recessions than during boom periods, consistent with findings from existing studies. This also indicates the time-varying nature of the influence of the FinTech Index on stock market excess returns. Therefore, inspired by the work of Sharma and Narayan (2022), we further explore the dynamic predictability of the FinTech Index for the US stock market. We select observations from the first ten years as the

Table 2. In-sample predictive regression estimation results.

Panel A: Multi-horizon predictability

Horizon	α(%)	t-stat	β(%)	t-stat	R^2(%)
h = 1	0.326	1.127	0.654	3.587	1.965
h = 3	0.360	1.370	0.310	2.736	1.261
h = 6	0.385	1.647	0.133	1.092	0.422
h = 9	0.409	1.963	0.082	0.768	0.229
h = 12	0.425	2.246	0.025	0.221	0.027

Panel B: Business cycle

α(%)	t-stat	$\beta^{\exp}$(%)	t-stat	β^{rec}(%)	t-stat	R^2(%)
0.315	1.074	0.554	2.100	0.875	6.114	2.066

Panel C: Time-varying predictability results.

	1% CV	5% CV	10% CV
Total	29	45	47
%	19.333	30.000	31.333
Total windows	150	150	150

Notes: This table reports the in-sample predictive ability of the FinTech Index for excess returns. Panel A displays the results of ordinary least squares estimation for the predictive regression model: $r_{t+1\to t+h} = \alpha + \beta F_t + \varepsilon_{t+1\to t+h}$, where $r_{t+1\to t+h} = (1/h)(r_{t+1} + \ldots + r_{t+h})$, $h = 1, 3, 6, 9, 12$, r_{t+1} is the excess return of the US S&P 500 index, F_t represents the FinTech Index, α denotes the intercept term, β is the estimated coefficient of the FinTech Index, and t-stat represents Newey–West heteroskedasticity and autocorrelation-robust t-statistics,. Panel B illustrates the FinTech Index's predictive prowess for excess returns across diverse business cycles. The table delineates the outcomes of ordinary least squares regression using the formula $r_{t+1} = \alpha + \beta^{rec} * F_t * D_t^{rec} + \beta^{exp} * F_t * (1 - D_t^{rec}) + \varepsilon_{t+1}$. Here β^{rec} and β^{exp} signify the regression coefficients for recessionary and expansionary periods, respectively, the variable D_t^{Rec} acts as a binary indicator, assuming the value of 1 during recessions and 0 during expansions. The expansionary and recessionary periods are delineated based on the NBER-dated business-cycle. Panel C reports the total number of statistically significant windows at the 10% significance level or higher, based on estimating time-varying predictability regression models, along with their percentage relative to the total number of windows. Specifically, we utilize the fixed rolling window approach with observations from the preceding 10 years and roll the fixed window until considering the entire sample. The 1% CV, 5% CV, and 10% CV represent critical values at the 1%, 5%, and 10% levels, respectively.

initial rolling window. Subsequently, we progressively shift this window throughout the dataset and use a univariate regression equation to examine the evolution of the FinTech Index's predictive performance over time until the entire sample is considered. This approach allows us to evaluate the predictability of the model across different time frames.

Panel C of Table 2 displays the total number of statistically significant windows at or above the 10% significance level, along with their percentage relative to the overall number of windows. According to the results, it can be observed that 31.333% of the windows are significant at the 10% level or higher. Figure 1 presents a visual representation of our predictive regression analysis, employing a rolling window method, showcasing the time-series t-statistics and R^2 values. The dashed lines in the graph represent critical thresholds at the 10% significance level. As shown in Figure 1, we identified significant estimates of the FinTech Index's impact on stock excess returns during two time intervals: from January 2010 to July 2012 and from April 2021 to June 2022. Furthermore, the trend in the in-sample R^2 closely aligns with the t-statistics.

3.2. Comparison with economic predictors

This section compares the FinTech Index's predictive power to the economic indicators outlined by Welch and Goyal (2008). To kick things off, we embark on an exploration of univariate regression models tailored to economic variables. Here's the formula we used:

$$r_{t+1} = \alpha + \psi E_t^k + \varepsilon_{t+1}, \tag{4}$$

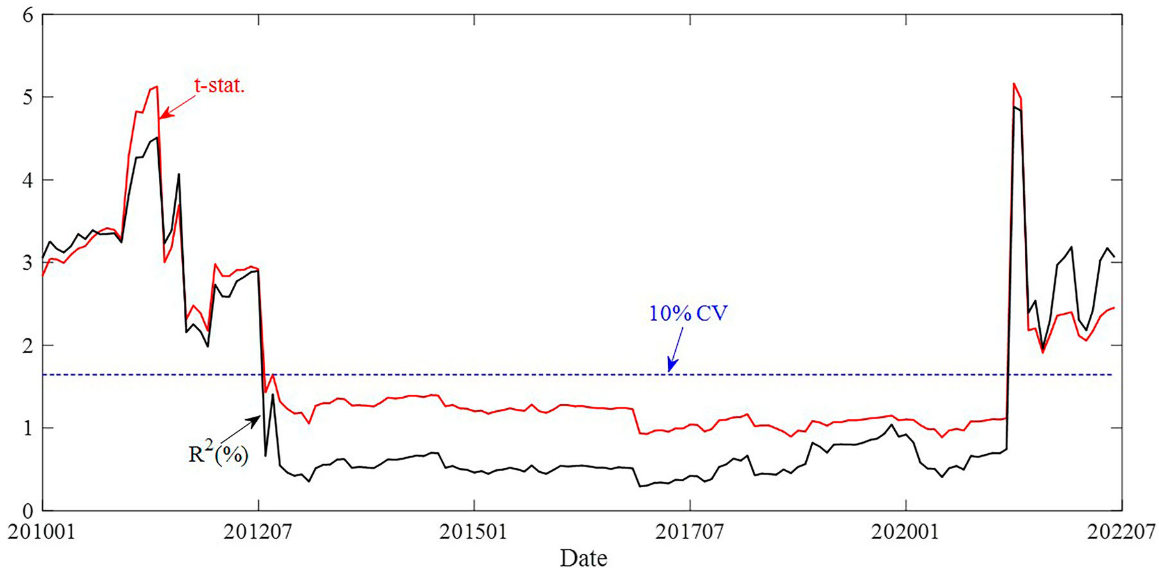

Figure 1. Time-varying predictability results.

Notes: This figure illustrates the time-series *t*-statistics and *R*-squared values derived from our analysis of the predictability regression using a rolling window method. Specifically, we estimate the following equation: $r_{t+1} = \alpha + \beta F_t + \varepsilon_{t+1}$, where r_{t+1} represents excess returns in the U.S. stock market and F_t represents the FinTech Index. To conduct our analysis, we adopt a fixed rolling window technique, initially utilising data from the preceding 10 years. Subsequently, we incrementally move this window through the dataset until the entire sample is considered, allowing us to evaluate the predictability of the model across different time frames.

Here,r_{t+1} represents the excess return of the US stock market in month t+1; E_t^kstands for the k-th economic variable, $k = 1, \ldots, 14$; and ψ represents the estimated coefficient associated with the economic indicator.

Panel A in Table 3 presents the estimates from the regression of Equation (4). Within the array of 14 economic variables scrutinized, it emerges that only the Log dividend-price ratio (DP), Log dividend yield (DY), Book-to-market ratio (BM), Treasury bill rate (TBL), and Long-term yield (LTY) demonstrate noteworthy accuracy in forecasting market performance, attaining significance levels of 10% or higher.

Following this, we delve deeper into assessing the predictive prowess of the FinTech Index while taking economic variables into account. To achieve this, we embark on bivariate regressions that combine the FinTech Index with one of the commonly utilized macroeconomic variables, as outlined below:

$$r_{t+1} = \alpha + \beta F_t + \psi E_t^k + \varepsilon_{t+1}, \tag{5}$$

where F_t represents the FinTech Index at month t, and β represents the estimated coefficient associated with the FinTech Index. The significance of the estimated coefficient (β) for the FinTech Index validates whether the FinTech Index remains predictive of stock returns even when economic variables are controlled.

Panel B in Table 3 showcases the estimation outcomes derived from Equation (5). Throughout these estimations, the FinTech Index consistently exhibits a positive estimated coefficient (β), hovering around 0.6%. This consistency mirrors the results obtained from Equation (2) when $h = 1$. Remarkably, even after accounting for economic indicators, the significance of β persists at the 1% level. Moreover, the range of R^2 values spans from 1.965% to 5.630%, surpassing the R^2 obtained from the univariate regression solely reliant on economic variables as depicted in Panel A. These findings compellingly suggest that financial technology significantly enriches the informational content of economic variables, thereby markedly enhancing the accuracy in predicting excess returns within the US market. By integrating FinTech metrics, investors and analysts can more effectively decipher complex market dynamics and anticipate future movements with greater precision. This integration not only bolsters the robustness of predictive models but also implies a transformative shift in how economic data is utilized to forecast financial outcomes. Consequently, these insights advocate for the increased adoption of

Table 3. Comparison with economic variables.

	Panel A: Univariate regressions			Panel B: Bivariate regressions				
	$r_{t+1} = \alpha + \psi E_t^k + \varepsilon_{t+1}$			$r_{t+1} = \alpha + \beta F_t + \psi E_t^k + \varepsilon_{t+1}$				
Variable	ψ(%)	t-stat	R^2(%)	β(%)	t-stat	ψ(%)	t-stat	R^2(%)
DP	3.788	1.670	2.800	0.638	3.291	3.722	1.629	4.667
DY	4.055	2.155	3.219	0.667	3.314	4.105	2.144	5.263
EP	0.271	0.235	0.054	0.653	3.534	0.258	0.225	2.014
DE	0.532	0.582	0.286	0.653	3.663	0.529	0.576	2.248
ROVL	6.212	1.364	0.646	0.644	3.542	5.925	1.279	2.552
BM	12.021	3.181	3.524	0.645	2.970	11.930	3.075	5.435
NTIS	12.805	0.543	0.266	0.652	3.655	12.564	0.529	2.220
TBL	−0.408	−3.134	2.531	0.653	3.414	−0.408	−2.996	4.491
LTY	−0.601	−3.690	3.822	0.628	3.053	−0.588	−3.565	5.630
LTR	0.087	1.119	0.401	0.625	3.254	0.061	0.742	2.159
TMS	−0.016	−0.077	0.003	0.654	3.607	−0.003	−0.016	1.965
DFY	−0.176	−0.144	0.028	0.652	3.561	−0.115	−0.093	1.977
DFR	0.098	0.440	0.195	0.684	3.492	0.133	0.589	2.320
INFL	0.593	0.587	0.273	0.637	3.395	0.451	0.432	2.121

Notes: This table contrasts the predictive efficacy of the FinTech Index against various macroeconomic variables, showcasing regression estimation coefficients, New-West *t*-statistics, and R^2. Panel A reports the in-sample estimation results for the univariate predictive regression of lagged macroeconomic variable E_t^k on excess returns in the U.S. stock market. The regression equation is $r_{t+1} = \alpha + \psi E_t^k + \varepsilon_{t+1}$, where E_t^k represents the *k*-th macroeconomic variable, $k = 1, \ldots, 14$. Panel B reports the in-sample estimation results for the bivariate predictive regressions on both the lagged FinTech Index F_t and macroeconomic variable E_t^k. The regression equation is $r_{t+1} = \alpha + \beta F_t + \psi E_t^k + \varepsilon_{t+1}$, where F_t represents the lagged FinTech Index.

FinTech tools in strategic investment decisions, potentially leading to more informed and effective portfolio management strategies across diverse economic cycles.

3.3. Sector-level return predictability testing

The existing literature has explored a range of indicators to forecast sector returns, striving for a comprehensive understanding of stock market dynamics (Bao, Hou, and Zhang 2023; Phan, Sharma, and Tran 2018; Salisu, Ogbonna, and Adediran 2021). Therefore, this section explores the correlation between the FinTech Index and the excess returns across the ten sectors of the S&P 500 index. These sectors encompass consumer discretionary (CD), consumer staples (CS), health care (HEAL), industrials (IND), information technology (IT), materials (MAT), telecommunication services (TEL), utilities (UTI), financials (FIN), and energy (ENER). This relationship can be represented as follows:

$$r_{t+1}^s = \alpha + \beta F_t + \varepsilon_{t+1}, \tag{6}$$

where r_{t+1}^s denotes the excess return of one of the ten sectors.

Table 4 presents the regression estimates for the relationship between FinTech at the sector level and stock returns. Among the ten sectors, the FinTech Index significantly predicts excess returns in seven industries, excluding consumer staples, telecommunication services, and energy, at a statistical significance threshold of 10% or higher. Notably, the estimated coefficients of the FinTech Index for the excess returns of the utilities and energy sectors are negative, while they are positive for the remaining eight industries. The R^2 values exhibit a range from 0.057% to 5.410%. Particularly noteworthy is the information technology sector, which demonstrates the highest R^2, demonstrating that the FinTech Index is particularly effective at predicting variations in excess returns within this industry.

The differences in predictive ability across industries can be attributed to several factors. Firstly, industries vary in their sensitivity to technological innovations. For example, sectors like information technology and industrials, which directly benefit from FinTech advancements, exhibit higher predictive power. Secondly, the adoption rates of FinTech solutions differ, with industries such as utilities and energy adopting FinTech more slowly due to regulatory barriers and longer investment cycles, resulting in negative coefficients. Thirdly, the regulatory environment and industry structure influence predictive ability; the financial sector, deeply integrated

Table 4. Excess return predictability for ten sector indexes.

Sector	α(%)	t-stat	β(%)	t-stat	R^2(%)
CD	0.525	1.517	0.910	3.939	2.487
CS	0.490	2.499	0.085	0.475	0.057
HEAL	0.532	2.304	0.499	3.749	1.389
IND	0.403	1.188	0.613	2.995	1.181
IT	0.464	1.080	1.715	3.255	5.410
MAT	0.507	1.364	0.510	1.847	0.659
TEL	−0.152	−0.445	1.036	1.236	2.879
UTI	0.302	1.079	−0.603	−3.644	1.532
FIN	0.254	0.590	0.653	2.871	0.985
ENER	0.473	1.137	−0.202	−0.818	0.073

Notes: This table reports the in-sample estimation results of lagged FinTech Index on excess returns across ten sector indices within the US S&P 500. The model equation is expressed as $r^s_{t+1} = \alpha + \beta F_t + \varepsilon_{t+1}$, where r^s_{t+1} denotes the excess return of one of the ten sectors. The first column of the table lists the ten sectors included, with CD representing consumer discretionary, CS representing consumer staples, HEAL representing health care, IND representing industrials, IT representing information technology, MAT representing materials, TEL representing telecommunication services, UTI representing utilities, FIN representing financials, and ENER representing energy.

with FinTech innovations, shows strong predictive power. Fourthly, market dynamics and cyclicality play a role, with sectors like consumer discretionary benefiting from improved financial services and demonstrating significant predictive ability. Lastly, risk management and efficiency gains are crucial, as seen in the healthcare sector, where FinTech innovations improve operational efficiency, leading to stronger predictive power. These factors collectively explain the differences in the FinTech Index's predictive ability across industries.

3.4. Country-level return predictability testing

Technological innovations originating in one country can reverberate across the international economy through mechanisms such as cross-border technology diffusion, interconnected industrial chains, and shared shocks (Aysun 2024). To test whether the FinTech Index constructed based on The New York Times can influence stock markets in other countries, following the methodology of Rapach, Strauss, and Zhou (2013), regression estimations were performed using the FinTech Index to forecast stock market excess returns of 10 major developed countries, including Australia (AUS), Canada (CAN), France (FRA), Germany (DEU), Italy (ITA), Japan (JPN), the Netherlands (NLD), Sweden (SWE), Switzerland (CHE), and the United Kingdom (GBR). The regression equation employed in our analysis is as follows:

$$r^c_{t+1} = \alpha + \beta F_t + \varepsilon_{t+1}, \tag{7}$$

where r^c_{t+1} represents the excess return of one of the ten countries.

Table 5 displays the regression results of the FinTech Index on excess returns for various countries Table 5 presents the regression results of the FinTech Index on excess returns across various countries. Through examination of the t-statistics associated with estimated slopes, it is evident that the FinTech Index yields promising in-sample estimation outcomes for Australia, France, Germany, Italy, New Zealand, Sweden, and the United Kingdom, with significantly positive coefficients. Furthermore, the R^2 values vary from 0.001% to 3.683%. Notably, Sweden boasts the highest R^2 (3.683%), closely trailed by Germany (3.399%), while Japan presents the lowest. In summary, the FinTech Index, derived from textual data sourced from *The New York Times*, exhibits a robust predictive capability across the majority of developed countries. This demonstrates not only the universality of the index's applicability but also its potential to provide valuable insights into financial market behaviors on a global scale. The index's effectiveness in forecasting financial outcomes across diverse economic environments underscores its utility as a powerful analytical tool for international investors and policymakers. By integrating this index into financial models, stakeholders can enhance their understanding of market dynamics and improve their strategic decision-making processes. Furthermore, the success of the FinTech Index in diverse

Table 5. Excess return predictability for international stock markets.

Country	α(%)	t-stat	β(%)	t-stat	R^2(%)
AUS	0.077	0.294	0.309	2.059	0.540
CAN	0.411	1.444	−0.018	−0.074	0.002
FRA	0.080	0.236	0.824	2.874	2.349
DEU	0.307	0.801	1.124	3.625	3.399
ITA	0.125	0.345	0.695	2.903	1.373
JPN	0.245	0.649	−0.019	−0.050	0.001
NLD	0.276	0.773	0.675	1.829	1.581
SWE	0.291	0.787	1.091	3.687	3.683
CHE	0.200	0.708	0.248	1.186	0.370
GBR	0.280	1.085	0.556	3.085	1.721

Notes: This table presents the in-sample estimation findings regarding the lagged FinTech Index's impact on excess returns across ten developed nations. The model equation is represented as $r^c_{t+1} = \alpha + \beta F_t + \varepsilon_{t+1}$, where r^c_{t+1} signifies the excess return of one of the ten countries. The first column enumerates the included nations: namely Australia (AUS), Canada (CAN), France (FRA), Germany (DEU), Italy (ITA), Japan (JPN), the Netherlands (NLD), Sweden (SWE), Switzerland (CHE), the United Kingdom (GBR).

markets invites further exploration into its potential applications in emerging markets, potentially offering a new paradigm for assessing financial health and investment opportunities worldwide.

3.5. Out-of-sample tests

3.5.1. Out-of-sample R^2

The in-sample results provide parameter estimates for relevant variables, offering initial insights into the predictive capability of the FinTech Index. However, investors are primarily concerned with its out-of-sample performance, as this is more closely associated with real-time return predictability (Jones and Mo 2021; Li, Tsiakas, and Wang 2015; Li et al. 2023; Rapach, Strauss, and Zhou 2010). Therefore, this subsection analyzes the out-of-sample forecasting efficacy of the FinTech Index for excess returns. We divide the complete sample into periods for in-sample and out-of-sample analysis, selecting March 2000 to December 2009 as the initial period. We compute the first estimate of β based on regression (2) and obtain forecasts sequentially from January 2010 to June 2022.

The out-of-sample R^2 (R^2_{oos}), as proposed by Campbell and Thompson (2008), serves as a widely adopted metric to evaluate how well models predict performance beyond the sample period (Liang et al. 2023; Tang et al. 2021). This metric quantifies the reduction in mean square prediction error (MSPE) relative to a baseline model. To compute this statistic, follow this equation:

$$R^2_{oos} = 1 - \frac{MSPE_{model}}{MSPE_{bench}}, \tag{8}$$

where $MSPE_{model} = \frac{1}{q}\sum_{t=1}^{q}(r_t - \hat{r}_t)^2$ and $MSPE_{bench} = \frac{1}{q}\sum_{t=1}^{q}(r_t - \bar{r}_t)^2$; $\hat{r}_t$ represents the forecasting value of the predictive model with FinTech Index; $\bar{r}_t$ represents the average historical stock returns, denoting the forecasts of the benchmark model. When the R^2_{oos} statistic is greater than 0, it shows superior performance of the predictive model over the benchmark. We also employ the MSFE-adjusted measure proposed by Clark and West (2007) to assess whether the mean square prediction error (MSPE) of the predictive regression model is greater than that of the benchmark model.

Table 6 presents the out-of-sample results. The second column presents the R^2_{oos} statistics for the univariate models integrating the FinTech Index alongside each of the 14 economic variables. Notably, the model incorporating the FinTech Index exhibits the highest out-of-sample R^2_{oos} value of 1.278%, and the MAFE-adjusted statistic illustrates it is statistically significant at the 10% level. This finding suggests that the FinTech Index contributes to predicting stock market performance even in out-of-sample scenarios. However, among the 14 economic variables, only the R^2_{oos} statistics for DFY and LYT are both positive and significant, while for the other 12 economic variables, the R^2_{oos} statistics are either negative or insignificant.

Table 6. Out-of-sample results.

Variable	R^2_{oos}	MSFE-adj	p
Fintech	1.278	1.599	0.055
DP	−2.313	0.822	0.205
DY	−1.218	0.668	0.252
EP	−2.302	−1.337	0.909
DE	0.462	0.650	0.258
ROVL	−0.112	0.681	0.248
BM	−1.773	1.127	0.130
NTIS	−2.163	−0.576	0.718
TBL	−2.266	0.250	0.401
LYT	0.870	1.764	0.039
LTR	0.048	0.449	0.327
TMS	−1.327	−1.326	0.908
DFY	1.229	1.885	0.030
DFR	−1.571	−0.331	0.630
INFL	−2.510	−1.052	0.853

Notes: The table showcases the out-of-sample performance in predicting the excess market return of the US stock market, leveraging the FinTech Index and 14 macroeconomic variables. The out-of-sample forecasting interval spans from January 2010 to June 2022. R^2_{oos} is the out-of-sample R^2 statistics proposed by Campbell and Thompson (2008), which assesses the percentage reduction in mean squared forecast error (MSFE) achieved by the competing model compared to the benchmark model (historical average). A positiveR^2_{oos} statistic signifies the superior performance of the competing model over the benchmark. MSFE-adj is the MSFE adjustment statistic by Clark and West (2007), utilised to test for significant differences between the competing model and the benchmark model.

3.5.2. Forecast encompassing tests

In this subsection, we employ forecast encompassing tests to conduct a direct comparison of the predictive information contained in FinTech-based forecasts with that derived from individual prediction regressions utilizing 14 widely recognized economic variables. The encompassing test, as proposed by Harvey, Leybourne, and Newbold (1998), stands as a widely employed tool for assessment (Jiang et al. 2019; Lin, Wu, and Zhou 2018; You and Liu 2020). Specifically, through regressing the actual values against predictions from different models, the ensuing equation is derived as follows:

$$r_t = \lambda \hat{r}_{f,t} + (1-\lambda)\hat{r}_{eco,t}, 0 \leq \lambda \leq 1, \tag{9}$$

where $\hat{r}_{f,t}$ and $\hat{r}_{eco,t}$ denote the forecasted values from the FinTech Index's univariate model and one of the macroeconomic variables' univariate model, with r_t representing the actual stock return. When $\lambda(1-\lambda)$ equals 0, it signifies an inclusion relationship between the two models. Specifically, when λ equals 0, it indicates that the predictive capacity of the single-variable for the FinTech Index surpasses that of the economic index's individual model. In contrast, if $(1-\lambda)$ equals 0, it indicates that the economic index's univariate model fully includes the FinTech Index's univariate model. To evaluate this, we employ a statistical test proposed by Harvey, Leybourne, and Newbold (1998). This test examines whether the model under study does not offer any additional information beyond what is present in the compared models, contrasting with the alternative hypothesis that the model of interest indeed furnishes additional information compared to the comparison model.

Table 7 provides the significance levels corresponding to forecast encompassing examinations. A significant p-value in the λ column suggests that the single-variable model for the FinTech Index holds additional incremental knowledge compared to the economic index's univariate model. Conversely, a significant p-value in $(1-\lambda)$ indicates that the economic index's univariate model incorporates more predictive information. Examining the table findings, it becomes clear that, with the exception of the three economic variables DE, LTR, and DFY, the FinTech Index contains additional extra information related to the other ten economic indicators. Noteworthy, although the FinTech Index cannot fully encompass the predictive information of DE, LTR, and DFY economic variables, similarly, these economic variables cannot encapsulate the predictive information of the FinTech Index. These findings underscore the superior performance of the FinTech Index in providing valuable

Table 7. Forecast encompassing test results.

	Fintech	
Variable	λ	$1-\lambda$
DP	0.040	0.223
DY	0.057	0.288
EP	0.008	0.800
DE	0.114	0.409
ROVL	0.042	0.385
BM	0.014	0.163
NTIS	0.006	0.655
TBL	0.008	0.454
LYT	0.044	0.063
LTR	0.104	0.422
TMS	0.012	0.728
DFY	0.207	0.246
DFR	0.041	0.602
INFL	0.004	0.823

Notes: The table displays *p*-values for Harvey et al.'s (1998) statistic regarding the FinTech Index and 14 macroeconomic indices. In the column labeled 'λ', these *p*-values gauge the null hypothesis that whether the FinTech Index forecasts effectively encompass those of the 14 macroeconomic indices. contrasting with the alternative hypothesis that the FinTech forecasts fail to encompass those of the macroeconomic indices. Similarly, the *p*-values in the '1-λ' column evaluate the null hypothesis that the forecasts of the 14 macroeconomic indices encompass the FinTech forecasts, against the alternative hypothesis that the forecasts of the 14 macroeconomic indices do not encompass the FinTech forecasts.

information, surpassing that of traditional macroeconomic variables. This advantage is particularly evident in the alignment of our results with those derived from rigorous out-of-sample testing, reinforcing the robustness and reliability of the FinTech Index as a predictive tool. The index's ability to capture subtle market signals offers a refined understanding of financial dynamics, which macroeconomic variables alone may fail to reveal. This enhancement in predictive accuracy not only bolsters confidence in financial forecasting but also opens new avenues for developing advanced analytical frameworks. Furthermore, the consistency of these findings with out-of-sample studies suggests that the FinTech Index could serve as a cornerstone for future research into financial market predictability. This could lead to more sophisticated investment strategies and improved risk management practices, ultimately contributing to more stable and efficient financial markets.

3.6. Robustness

Different estimation intervals yield distinct out-of-sample estimates because they incorporate various breakpoints, leading to variations in the effectiveness of out-of-sample predictions, as noted in seminal works by Pesaran and Timmermann (2007) and Pesaran and Pick (2011). To explore this variability, we adjusted the initial sample estimate period to March 2000 through December 2004 and subsequently tested the predictive power of relevant indicators across differing forecasting window sizes. Table 8 displays the results obtained from out-of-sample testing spanning January 2005 to June 2022. Our findings reveal that even after altering the rolling window size, the FinTech Index retains significant capability to predict stock returns, evidenced by a notable R^2_{oos} statistic of 1.103%. In stark contrast, the R^2_{oos} statistics for all 14 macroeconomic variables are negative, highlighting their ineffectiveness in predicting returns. This stark difference further emphasizes the robustness and superior predictive capability of the FinTech Index across various prediction window sizes, suggesting its potential as a more reliable tool in financial forecasting in contrast to conventional macroeconomic indicators.

4. Asset allocation

In this part, we look at the financial implications of stock return predictability using the FinTech Index in terms of asset allocation strategies. Following Campbell and Thompson (2008), Rapach, Ringgenberg, and Zhou (2016),

Table 8. Out-of-sample results for different window sizes.

Variable	R^2_{oos}	MSFE-adj	p
Fintech	1.103	1.674	0.047
DP	−7.370	0.839	0.201
DY	−7.485	0.409	0.341
EP	−14.400	0.539	0.295
DE	−9.903	0.795	0.213
ROVL	−8.846	0.213	0.416
BM	−1.686	1.374	0.085
NTIS	−0.937	1.077	0.141
TBL	−5.864	−0.611	0.730
LYT	−2.238	1.539	0.062
LTR	−2.346	−0.506	0.694
TMS	−5.018	−1.660	0.952
DFY	−6.733	0.749	0.227
DFR	−7.664	−0.725	0.766
INFL	−3.171	−0.742	0.771

Notes: This table revises the window size based on Table 6 and re-evaluates the out-of-sample performance of forecasting excess market returns of the US stock market using the FinTech Index and 14 macroeconomic variables. The out-of-sample forecasting interval spans from January 2005 to June 2022. R^2_{oos} is the out-of-sample R^2 statistics proposed by Campbell and Thompson (2008), which assesses the percentage reduction in mean squared forecast error (MSFE) achieved by the competing model compared to the benchmark model (historical average). A positive R^2_{oos} statistic signifies the superior performance of the competing model over the benchmark. MSFE-adj is the MSFE adjustment statistic by Clark and West (2007), utilised to test for significant differences between the competing model and the benchmark model.

Jiang et al. (2019), and Liang, Wang, and Duong (2024), we compute the certainty equivalent return (CER) gain as well as the Sharpe ratio. We examine a mean-variance investor that uses predictive regression of stock excess returns to determine asset allocation between stocks and risk-free assets. The investor allocates the following ratios to the stocks in the ideal portfolio at the end of the t-th month:

$$w_t = \frac{1}{\gamma}\frac{\hat{r}_{t+1}}{\hat{\sigma}^2_{t+1}}, \tag{10}$$

where γ represents the risk aversion coefficient of the investor, $\hat{r}_{t+1}$ represents predicted excess returns, and $\hat{\sigma}^2_{t+1}$ denotes predicted excess return variance. Subsequently, the investor's portfolio return $R_{p,t+1}$ is determined by:

$$R_{p,t+1} = w_t r_{t+1} + (1 - w_t)R_{f,t+1}, \tag{11}$$

where $R_{f,t+1}$ represents the risk-free return. Employing a moving five-year window of historical monthly returns, in line with the approach of Campbell and Thompson (2008) and Neely et al. (2014), we limit w to an interval of 0 to 1.5, preventing short sales and limiting leverage to a maximum of 50%. The portfolio's certainty equivalent return (CER) is computed as follows:

$$CER = \hat{\mu}_p - 0.5\gamma\,\hat{\sigma}_p, \tag{12}$$

where $\hat{\mu}_p$ represents the average and $\hat{\sigma}_p$ denotes the variability of the portfolio returns throughout the forecast evaluation period. For the purpose of determining the mean-variance investor's ideal allocation between equities and risk-free assets, we examine the Certainty Equivalent Return (CER) and Sharpe ratio. This analysis utilizes out-of-sample forecast regressions to compute these metrics. The difference between a given investment portfolio's CER and a benchmark portfolio's is what yields the CER gain. Following this, the CER gain is amplified by 1200 to ascertain the annual percentage investors are prepared to invest for a particular return projection, rather than settling for a mere average forecast. Furthermore, we compute the monthly Sharpe ratio, which measures the average return on the portfolio in relation to the risk-free rate divided by the excess return standard deviation.

Table 9. Asset allocation results.

	$\gamma = 2$		$\gamma = 3$		$\gamma = 4$		$\gamma = 5$	
Variable	CER gain (%)	SR	CER gain (%)	SR	CER gain (%)	SR	CER gain (%)	SR
Fintech	4.355	0.223	3.540	0.219	2.183	0.183	1.141	0.153
DP	3.334	0.226	2.961	0.213	1.922	0.163	1.514	0.161
DY	3.684	0.248	3.298	0.245	2.563	0.223	2.019	0.212
EP	0.908	0.039	0.868	0.054	0.536	0.051	0.132	0.037
DE	4.059	0.220	3.355	0.217	2.518	0.204	1.577	0.174
ROVL	0.397	0.000	−0.253	−0.012	−1.232	−0.028	−1.661	−0.013
BM	4.084	0.234	3.753	0.265	2.673	0.238	2.161	0.239
NTIS	0.865	0.021	1.157	0.079	0.102	0.053	−0.792	0.023
TBL	1.336	0.048	0.691	0.050	−0.050	0.043	−0.510	0.023
LYT	7.563	0.335	6.768	0.373	4.955	0.356	3.621	0.357
LTR	2.726	0.127	3.238	0.203	2.347	0.193	1.284	0.164
TMS	3.412	0.166	1.992	0.127	0.543	0.073	−0.486	0.025
DFY	3.263	0.156	1.941	0.124	1.228	0.125	1.290	0.169
DFR	2.068	0.098	1.858	0.120	1.305	0.116	0.919	0.115
INFL	1.507	0.062	0.424	0.031	0.107	0.035	−0.132	0.034

Notes: This table provides an insight into portfolio performance metrics tailored for a mean-variance investor with varying degrees of risk aversion coefficients (2, 3, 4, and 5). These investors allocate their investments monthly between equities and risk-free bills, utilizing out-of-sample predictive regression forecasts derived from the FinTech Index and 14 macroeconomic variables to inform their decisions. The CER gain represents the annualized certainty equivalent return gain. Additionally, the monthly Sharpe ratio calculated as the mean portfolio return exceeding the risk-free rate divided by the standard deviation of the excess portfolio return. The out-of-sample forecasting interval spans from January 2010 to June 2022.

Table 9 presents the portfolio performance of various models. Notably, irrespective of varying degrees of risk aversion, the FinTech Index demonstrates a noteworthy positive annualized Compound Excess Return (CER) gain, surpassing the majority of macroeconomic indices. Nonetheless, the CER gain of the FinTech Index increases as the risk aversion coefficient decreases. Additionally, the FinTech Index's monthly Sharpe ratio ranges from 0.153 to 0.223 for various risk aversion thresholds, consistently outpacing the majority of macroeconomic indices. Overall, Table 9 vividly illustrates the FinTech Index's capacity to generate substantial economic value for investors, an advantage that becomes particularly pronounced under conditions of low risk aversion. This finding highlights the index's strategic utility in optimizing investment outcomes during favorable market conditions, where investors are more inclined to assume higher risks in pursuit of greater returns. The enhanced performance of the FinTech Index in these scenarios suggests it effectively captures and leverages market dynamics that are typically overlooked by traditional financial indicators. Consequently, this ability can guide investors towards more informed decision-making processes, maximizing their return potential while managing risk exposure. Additionally, the significant value generated by the FinTech Index under varied risk appetites underscores its adaptability and relevance in diverse investment strategies, affirming its role as a pivotal tool in contemporary financial management and planning.

5. Conclusions

In this study, we leveraged text data from The New York Times to construct a FinTech Index and explored its relationship with stock market returns. Our rigorous analysis has unearthed several pivotal findings that underscore the substantial predictive capabilities of the FinTech Index within both the US stock market and its sectors, as well as its applicability to international markets.

Firstly, our findings reveal that the FinTech Index has a strong capacity to forecast US stock market returns both in-sample and out-of-sample, particularly during recessionary periods, where its influence markedly exceeds that during expansionary phases. This enhanced effectiveness in volatile economic times suggests that the FinTech Index could serve as a crucial tool for risk management, aiding investors in navigating uncertain markets. Secondly, our sector-specific analysis indicates that the FinTech Index excels in predicting stock performance across seven of the ten sectors within the US S&P 500, with the most pronounced impact within the information technology sector. This insight is particularly valuable for sector-specific funds and investors focusing on technological advancements. Thirdly, the index's relevance extends beyond the US, demonstrating significant predictive accuracy in several developed countries. This international applicability makes the FinTech Index a useful instrument for global investment strategies, offering a consistent metric for cross-border investment decisions. Moreover, our evaluation through Certainty Equivalent Returns (CER) and Sharpe ratio calculations highlights the FinTech Index's capacity to substantially enhance economic returns and improve risk-adjusted performance compared to traditional macroeconomic indicators. These aspects prove the FinTech Index's utility in optimizing portfolio strategies, reinforcing its value to both individual investors and institutional portfolio managers.

In conclusion, our research underscores the transformative impact of FinTech on enhancing predictability within the stock market and refining the outcomes of investments. By developing a pioneering FinTech Index based on textual data, our study offers a groundbreaking perspective on the dynamic interplay between technological innovations and financial systems. This approach not only contributes substantial theoretical advancements but also offers actionable perspectives for investors and decision-makers in finance and governance. The practical implications of our findings are particularly significant in the context of today's rapidly digitizing financial landscape. As financial markets become increasingly driven by digital technologies, the ability to integrate and leverage technological insights becomes crucial. Our FinTech Index serves as a vital tool in this integration, enabling more informed and effective market strategies and economic policies. By facilitating a deeper understanding of the correlations between FinTech developments and market behaviors, our research helps stakeholders capitalize on opportunities and navigate challenges in the digital era, ultimately leading to more robust and agile financial environments.

Notes

1. https://www.svb.com/trends-insights/reports/fintech-industry-report/
2. For a detailed introduction and explanation of 14 macroeconomic variables, please refer to Welch and Goyal (2008). For the latest data, please visit the website athttps://sites.google.com/view/agoyal145/?redirpath = /

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by National Natural Science Foundation of China [grant number 72071162].

Data availability statement

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Notes on contributors

Yang Guo is a Ph.D. candidate at Southwest Jiaotong University. Her research interests focus on financial forecasting.

Feng Ma, Associate Professor and Doctoral Supervisor. He obtained his Ph.D. in Management from the School of Economics and Management at Southwest Jiaotong University in June 2016 and began working at the same university in September of the same year. In 2017, he was selected as a postgraduate supervisor, and in December 2018, he was exceptionally promoted to Associate Professor. In 2019, he was selected as a doctoral supervisor. In November 2019, he was appointed Director of the Graduate Education Center of the School. During his Ph.D. studies, he received Southwest Jiaotong University's highest personal honor, the "Qishi Yanghua Medal," as well as the title of "Outstanding Graduate of Sichuan Province for 2016," the "Outstanding Doctoral Dissertation Award of 2016" at the university level, and the "National Scholarship for Doctoral Students." In December 2019, he was selected for the Sichuan Province Talent Program, and in June 2020, he was selected for the university's "Chuying Program." In March 2021, he was chosen as a reserve candidate for the 13th batch of academic and technical leaders in Sichuan Province. In May 2022, he was awarded the "Siguo Award" by Southwest Jiaotong University. Currently, he is leading one project each from the National Natural Science Foundation of China (General and Youth Programs) and the Ministry of Education's Humanities and Social Sciences Program, and he has participated in several national-level projects. He has published over 100 academic journal articles, with his main research work appearing in journals such as the *Risk Analysis, Journal of Banking & Finance, Journal of Empirical Finance, International Journal of Forecasting, Journal of Forecasting, Quantitative Finance, Energy Economics, International Review of Financial Analysis*, Six of his papers are highly cited in the Economics & Business (ESI) category. He has been listed as an "Elsevier Highly Cited Chinese Researcher" for two consecutive years, being selected in the field of Applied Economics in 2021 (among 50 scholars nationwide) and in the field of Theoretical Economics in 2022 (among 25 scholars nationwide).

Yizhi Wang, FHEA, is a Lecturer (assistant professor) in Finance at Cardiff University, Cardiff Business School, UK. Yizhi Wang has been immersed in the field of Finance for many years. He has accumulated excellent research experience. He widely published in leading academic journals and independently designed an R-based tutorial book related to financial econometrics (500+ pages). Moreover, he has taken part in numerous academic activities. His research interests include Financialization and Econometrics. As a single author, first author or corresponding author, Yizhi's research is widely published in ABS-ranked journals, such as, amongst others, *Risk Analysis, the Journal of Economic Behavior & Organization, Energy, Energy Economics, Journal of Commodity Markets, Technological Forecasting and Social Change, Journal of International Financial Markets, Institutions and Money, International Review of Financial Analysis*. He was awarded the Shanghai Institute of International Finance and Economics Best Paper Award. Furthermore, Yizhi Wang is the Editor-in-Chief of *Elsevier's Finance Research Letters* (ABS2*), and an Associate Editor for the *International Review of Financial Analysis* (ABS3*), and the *Research in International Business and Finance* (ABS2*). Yizhi Wang also serves as a peer reviewer for more than ten leading academic outlets in Finance.

Juandan Zhong is a Ph.D. candidate at Southwest Jiaotong University. Her research interests lie in the study of tail risk within financial markets.

References

An, Z., C. Chen, V. Naiker, and J. Wang. 2020. "Does Media Coverage Deter Firms from Withholding bad News? Evidence from Stock Price Crash Risk." *Journal of Corporate Finance* 64:101664. https://doi.org/10.1016/j.jcorpfin.2020.101664

Anzoategui, D., D. Comin, M. Gertler, and J. Martinez. 2019. "Endogenous Technology Adoption and R&D as Sources of Business Cycle Persistence." *American Economic Journal: Macroeconomics* 11 (3): 67–110. https://doi.org/10.1257/mac.20170269

Aysun, U. 2024. "Technology Diffusion and International Business Cycles." *Journal of International Money and Finance* 140:102974. https://doi.org/10.1016/j.jimonfin.2023.102974

Baltussen, G., S. van Bekkum, and Z. Da. 2019. "Indexing and Stock Market Serial Dependence Around the World." *Journal of Financial Economics* 132 (1): 26–48. https://doi.org/10.1016/j.jfineco.2018.07.016

Bao, J., K. Hou, and S. Zhang. 2023. "Systematic Default and Return Predictability in the Stock and Bond Markets." *Journal of Financial Economics* 149 (3): 349–377. https://doi.org/10.1016/j.jfineco.2023.05.006

Beck, T., T. Chen, C. Lin, and F. M. Song. 2016. "Financial Innovation: The Bright and the Dark Sides." *Journal of Banking & Finance* 72:28–51. https://doi.org/10.1016/j.jbankfin.2016.06.012

Campbell, J. Y., S. J. Grossman, and J. Wang. 1993. "Trading Volume and Serial Correlation in Stock Returns." *The Quarterly Journal of Economics* 108 (4): 905–939. https://doi.org/10.2307/2118454

Campbell, J. Y., and S. B. Thompson. 2008. "Predicting Excess Stock Returns out of Sample: Can Anything Beat the Historical Average?" *Review of Financial Studies* 21 (4): 1509–1531. https://doi.org/10.1093/rfs/hhm055

Chen, J., G. Tang, J. Yao, and G. Zhou. 2022. "Investor Attention and Stock Returns." *Journal of Financial and Quantitative Analysis* 57 (2): 455–484. https://doi.org/10.1017/S0022109021000090

Clark, T. E., and K. D. West. 2007. "Approximately Normal Tests for Equal Predictive Accuracy in Nested Models." *Journal of Econometrics* 138 (1): 291–311. https://doi.org/10.1016/j.jeconom.2006.05.023

Engle, R. F., S. Giglio, B. Kelly, H. Lee, and J. Stroebel. 2020. "Hedging Climate Change News." *The Review of Financial Studies* 33 (3): 1184–1216. https://doi.org/10.1093/rfs/hhz072

Fama, E. F., and K. R. French. 1989. "Business Conditions and Expected Returns on Stocks and Bonds." *Journal of Financial Economics* 25 (1): 23–49. https://doi.org/10.1016/0304-405X(89)90095-0

Ferreira, M. A., and P. Santa-Clara. 2011. "Forecasting Stock Market Returns: The sum of the Parts is More Than the Whole." *Journal of Financial Economics* 100 (3): 514–537. https://doi.org/10.1016/j.jfineco.2011.02.003

Fuerst, T. S. 1995. "Monetary and Financial Interactions in the Business Cycle." *Journal of Money, Credit and Banking* 27 (4): 1321–1338. https://doi.org/10.2307/2078053

Garcia, D. 2013. "Sentiment During Recessions." *The Journal of Finance* 68 (3): 1267–1300. https://doi.org/10.1111/jofi.12027

Garlappi, L., and Z. Song. 2020. "Investment Shocks and Asset Prices: An Investment-Based Approach." *Journal of Financial and Quantitative Analysis* 55 (8): 2665–2699. https://doi.org/10.1017/S0022109019000796

Gârleanu, N., L. Kogan, and S. Panageas. 2012a. "Displacement Risk and Asset Returns." *Journal of Financial Economics* 105 (3): 491–510. https://doi.org/10.1016/j.jfineco.2012.04.002

Garleanu, N., S. Panageas, and J. Yu. 2012b. "Technological Growth and Asset Pricing." *The Journal of Finance* 67 (4): 1265–1292. https://doi.org/10.1111/j.1540-6261.2012.01747.x

Goldstein, I., W. Jiang, and G. A. Karolyi. 2019. "To FinTech and Beyond." *The Review of Financial Studies* 32 (5): 1647–1661. https://doi.org/10.1093/rfs/hhz025

Hart, R. P. 2001. "Redeveloping DICTION: Theoretical Considerations." In *Progress in Communication Sciences*, 43–60.

Harvey, D. I., S. J. Leybourne, and P. Newbold. 1998. "Tests for Forecast Encompassing." *Journal of Business & Economic Statistics* 16 (2): 254–259. https://doi.org/10.1080/07350015.1998.10524759

Hendershott, T., X. Zhang, J. L. Zhao, and Z. Zheng. 2021. "FinTech as a Game Changer: Overview of Research Frontiers." *Information Systems Research* 32 (1): 1–17. https://doi.org/10.1287/isre.2021.0997

Hirshleifer, D., P. H. Hsu, and D. Li. 2013. "Innovative Efficiency and Stock Returns." *Journal of Financial Economics* 107 (3): 632–654. https://doi.org/10.1016/j.jfineco.2012.09.011

Hirshleifer, D., P. H. Hsu, and D. Li. 2018. "Innovative Originality, Profitability, and Stock Returns." *The Review of Financial Studies* 31 (7): 2553–2605. https://doi.org/10.1093/rfs/hhx101

Hsu, P. H., and D. Huang. 2010. "Technology Prospects and the Cross-Section of Stock Returns." *Journal of Empirical Finance* 17 (1): 39–53. https://doi.org/10.1016/j.jempfin.2009.08.001

Jiang, F., J. Lee, X. Martin, and G. Zhou. 2019. "Manager Sentiment and Stock Returns." *Journal of Financial Economics* 132 (1): 126–149. https://doi.org/10.1016/j.jfineco.2018.10.001

Jones, C. S., and H. Mo. 2021. "Out-of-sample Performance of Mutual Fund Predictors." *The Review of Financial Studies* 34 (1): 149–193. https://doi.org/10.1093/rfs/hhaa026

Kaltenbrunner, G., and L. A. Lochstoer. 2010. "Long-run Risk Through Consumption Smoothing." *The Review of Financial Studies* 23 (8): 3190–3224. https://doi.org/10.1093/rfs/hhq033

Kogan, L., and D. Papanikolaou. 2014. "Growth Opportunities, Technology Shocks, and Asset Prices." *The Journal of Finance* 69 (2): 675–718. https://doi.org/10.1111/jofi.12136

Li, Y., L. Luo, C. Liang, and F. Ma. 2023. "The Role of Model Bias in Predicting Volatility: Evidence from the US Equity Markets." *China Finance Review International* 13 (1): 140–155. https://doi.org/10.1108/CFRI-04-2020-0037

Li, J., I. Tsiakas, and W. Wang. 2015. "Predicting Exchange Rates out of Sample: Can Economic Fundamentals Beat the Random Walk?" *Journal of Financial Econometrics* 13 (2): 293–341. https://doi.org/10.1093/jjfinec/nbu007

Liang, C., Q. Luo, Y. Li, and L. D. T. Huynh. 2023. "Global Financial Stress Index and Long-Term Volatility Forecast for International Stock Markets." *Journal of International Financial Markets, Institutions and Money* 88:101825. https://doi.org/10.1016/j.intfin.2023.101825

Liang, C., L. Wang, and D. Duong. 2024. "More Attention and Better Volatility Forecast Accuracy: How Does war Attention Affect Stock Volatility Predictability?" *Journal of Economic Behavior & Organization* 218:1–19. https://doi.org/10.1016/j.jebo.2023.12.009

Lin, X. 2012. "Endogenous Technological Progress and the Cross-Section of Stock Returns." *Journal of Financial Economics* 103 (2): 411–427. https://doi.org/10.1016/j.jfineco.2011.08.013

Lin, H., C. Wu, and G. Zhou. 2018. "Forecasting Corporate Bond Returns with a Large set of Predictors: An Iterated Combination Approach." *Management Science* 64 (9): 4218–4238. https://doi.org/10.1287/mnsc.2017.2734

Ma, F., J. Cao, Y. Wang, S. A. Vigne, and D. Dong. 2023. "Dissecting Climate Change Risk and Financial Market Instability: Implications for Ecological Risk Management." Risk Analysis. 4133.

Ma, F., X. Lu, J. Liu, and D. Huang. 2022. "Macroeconomic Attention and Stock Market Return Predictability." *Journal of International Financial Markets, Institutions and Money* 79:101603. https://doi.org/10.1016/j.intfin.2022.101603

Neely, C. J., D. E. Rapach, J. Tu, and G. Zhou. 2014. "Forecasting the Equity Risk Premium: The Role of Technical Indicators." *Management Science* 60 (7): 1772–1791. https://doi.org/10.1287/mnsc.2013.1838

Pennebaker, J. W., M. R. Mehl, and K. G. Niederhoffer. 2003. "Psychological Aspects of Natural Language use: Our Words, our Selves." *Annual Review of Psychology* 54 (1): 547–577. https://doi.org/10.1146/annurev.psych.54.101601.145041

Pesaran, M. H., and A. Pick. 2011. "Forecast Combination Across Estimation Windows." *Journal of Business & Economic Statistics* 29 (2): 307–318. https://doi.org/10.1198/jbes.2010.09018

Pesaran, M. H., and A. Timmermann. 2007. "Selection of Estimation Window in the Presence of Breaks." *Journal of Econometrics* 137 (1): 134–161. https://doi.org/10.1016/j.jeconom.2006.03.010

Phan, D. H. B., S. S. Sharma, and V. T. Tran. 2018. "Can Economic Policy Uncertainty Predict Stock Returns? Global Evidence." *Journal of International Financial Markets, Institutions and Money* 55:134–150. https://doi.org/10.1016/j.intfin.2018.04.004

Rapach, D. E., M. C. Ringgenberg, and G. Zhou. 2016. "Short Interest and Aggregate Stock Returns." *Journal of Financial Economics* 121 (1): 46–65. https://doi.org/10.1016/j.jfineco.2016.03.004

Rapach, D. E., J. K. Strauss, and G. Zhou. 2010. "Out-of-sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy." *The Review of Financial Studies* 23 (2): 821–862. https://doi.org/10.1093/rfs/hhp063

Rapach, D. E., J. K. Strauss, and G. Zhou. 2013. "International Stock Return Predictability: What is the Role of the United States?" *The Journal of Finance* 68 (4): 1633–1662. https://doi.org/10.1111/jofi.12041

Rapach, D., and G. Zhou. 2013. "Forecasting Stock Returns." In *Handbook of Economic Forecasting*. Vol. 2, 328–383. Elsevier.

Salisu, A. A., A. E. Ogbonna, and I. Adediran. 2021. "Stock-Induced Google Trends and the Predictability of Sectoral Stock Returns." *Journal of Forecasting* 40 (2): 327–345. https://doi.org/10.1002/for.2722

Sharma, S. S., and P. K. Narayan. 2022. "Technology Shocks and Stock Returns: A Long-Term Perspective." *Journal of Empirical Finance* 68:67–83. https://doi.org/10.1016/j.jempfin.2022.06.002

Shiller, R. J. 2015. Irrational Exuberance: Revised and Expanded Third Edition. In *Irrational Exuberance*. Princeton University Press.

Tang, Y., X. Xiao, M. I. M. Wahab, and F. Ma. 2021. "The Role of oil Futures Intraday Information on Predicting US Stock Market Volatility." *Journal of Management Science and Engineering* 6 (1): 64–74. https://doi.org/10.1016/j.jmse.2020.10.004

Tetlock, P. C. 2007. “Giving Content to Investor Sentiment: The Role of Media in the Stock Market.” *The Journal of Finance* 62 (3): 1139–1168. https://doi.org/10.1111/j.1540-6261.2007.01232.x

Tiwari, A. K., E. J. A. Abakah, X. Shao, T. L. Le, and M. N. Gyamfi. 2023. “Financial Technology Stocks, Green Financial Assets, and Energy Markets: A Quantile Causality and Dependence Analysis.” *Energy Economics* 118:106498. https://doi.org/10.1016/j.eneco.2022.106498

Tvinnereim, E., and K. Fløttum. 2015. “Explaining Topic Prevalence in Answers to Open-Ended Survey Questions About Climate Change.” *Nature Climate Change* 5 (8): 744–747. https://doi.org/10.1038/nclimate2663

Wang, R., J. Liu, and H. Luo. 2021. “Fintech Development and Bank Risk Taking in China.” *The European Journal of Finance* 27 (4-5): 397–418. https://doi.org/10.1080/1351847X.2020.1805782

Wang, Y., Z. Pan, L. Liu, and C. Wu. 2019. “Oil Price Increases and the Predictability of Equity Premium.” *Journal of Banking & Finance* 102:43–58. https://doi.org/10.1016/j.jbankfin.2019.03.009

Wang, R., and J. Tan. 2021. “Exploring the Coupling and Forecasting of Financial Development, Technological Innovation, and Economic Growth.” *Technological Forecasting and Social Change* 163:120466. https://doi.org/10.1016/j.techfore.2020.120466

Welch, I., and A. Goyal. 2008. “A Comprehensive Look at the Empirical Performance of Equity Premium Prediction.” *Review of Financial Studies* 21 (4): 1455–1508. https://doi.org/10.1093/rfs/hhm014

You, Y., and X. Liu. 2020. “Forecasting Short-run Exchange Rate Volatility with Monetary Fundamentals: A GARCH-MIDAS Approach.” *Journal of Banking & Finance* 116:105849. https://doi.org/10.1016/j.jbankfin.2020.105849

Young, L., and S. Soroka. 2012. “Affective News: The Automated Coding of Sentiment in Political Texts.” *Political Communication* 29 (2): 205–231. https://doi.org/10.1080/10584609.2012.671234

Zhu, X., and J. Zhu. 2013. “Predicting Stock Returns: A Regime-Switching Combination Approach and Economic Links.” *Journal of Banking & Finance* 37 (11): 4120–4133. https://doi.org/10.1016/j.jbankfin.2013.07.016