

Online Appendix to “*Forecasting GDP Growth Rates Using Accounting Earnings: A Large Panel Micro Data Approach*”

Yumeng Cui* Yongmiao Hong[†] Naijing Huang[‡] Yicheng Wang[§]

February 6, 2026

This Online Appendix provides supplementary materials for the paper entitled “Forecasting GDP Growth Rates Using Accounting Earnings: A Large Panel Micro Data Approach” by Yumeng Cui, Yongmiao Hong, Naijing Huang and Yicheng Wang. The supplement is organized as follows: Section [A](#) describes the data processing procedures, Section [B](#) presents details of the models used in the analysis, and Section [C](#) reports additional tables and figures that support our findings.

A Details on Data Processing Procedures

A.1 Data Processing Procedures for the Main Results

In this section, we provide details about data processing of individual accounting earnings. The accounting earnings of an individual firm i at time t , denoted by $X_{i,t}$, are measured by scaling quarterly net income by sales ([Konchitchki and Patatoukas, 2014](#)). The procedure is as follows:

- **Raw data source:** The raw net income and sales data is sourced from the *Compustat Preliminary History Quarterly* dataset.

*School of Economics, Central University of Finance and Economics.

[†]National Key Laboratory of Mathematical Sciences and Center for Forecasting Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences; School of Economics and Management and MOE Social Science Laboratory of Digital Economic Forecasts and Policy Simulation, University of Chinese Academy of Sciences.

[‡]School of Economics, Central University of Finance and Economics. Corresponding author, e-mail: huang.naijing@cufe.edu.cn.

[§]HSBC Business School, Peking University.

- **Fiscal-year-ends:** To align fiscal quarters with calendar quarters, only firms with December fiscal year-ends are included.
- **Reporting weeks:** Both accounting earnings and GDP releases may experience significant delays. This timing is likely to introduce future information into our pseudo out-of-sample predictions. Therefore we only keep the samples released within 16 weeks after the quarter end. We also test robustness using alternative cutoff windows of 8, 12, and 20 weeks in Section 5.3.
- **New and exiting firms:** In constructing the database, we include both active and inactive firms, including those that later exit the sample. Within any given rolling window, a firm is retained if it is either currently active or was active at any point during the window and has accounting earnings available. For firms that exit the sample (e.g., due to mergers or delistings), available accounting earnings from their active periods are retained, while other observations remain missing. As the rolling window advances, once no records remain for an exited firm, it is naturally excluded from the predictor set by data availability, rather than by any manual selection.
- **The geographic segments of the firms:** The sample is drawn from all publicly listed firms in the U.S. However, GDP is measured on a territorial basis, meaning that only the value added generated within U.S. borders is included. Consequently, some income components, such as foreign firms' revenues generated outside the U.S. or U.S. firms' oversea revenues, may introduce noise to the forecasting. To address this concern, we provide a more detailed discussion and robustness check in Section 5.3.
- Finally, the sample retains 19,053 firms in total.

Then, we calculate accumulated accounting earnings growth as $\Delta x_{i,t} = X_{i,t} - X_{i,t-h}$. Missing values in the predictive panel of accounting earnings growth are set to zero.

Given the vast number of listed firms, we preselect 1,500 firms following [Zou and Hastie \(2005\)](#) to reduce computational costs. The procedure is as follows:

- For each firm i , we perform a linear regression of the accumulated GDP growth rate g_{t+h} on the accounting earning of firm i in quarter t , denoted by $\Delta x_{i,t}$, and its one-

period lag $\Delta x_{i,t-1}$, as well as the GDP growth rate in quarter t , denoted by g_t . The regression is defined as follows:

$$g_{t+h} = \mu_h^i + \alpha_h^i g_t + \beta_{1,h}^i \Delta x_{i,t} + \beta_{2,h}^i \Delta x_{i,t-1} + \varepsilon_{t+h}^i$$

$$i = 1, 2, \dots, n,$$
(A.1)

where n is the number of listed firms in the U.S., and $g_t = 100 \times (GDP_t - GDP_{t-h}) / GDP_{t-h}$.

- The t -statistic score for each firm i is the mean value of the absolute t -statistics for $\{\beta_{1,h}^i\}_{i=1}^n$ and $\{\beta_{2,h}^i\}_{i=1}^n$.
- We select the top 1,500 firms with the highest t -statistic scores.

A.2 Data Processing Procedures for Aggregate Accounting Earnings

In Section 6.1, we compare the forecasting accuracy of using individual and aggregate accounting earnings in the same forecasting framework.

Following [Konchitchki and Patatoukas \(2014\)](#), we construct value-weighted aggregate accounting earnings, denoted by $X_t^{Aggregate}$ and the corresponding accumulated growth rates, denoted by $\Delta x_t^{Aggregate}$. In each quarter t , we retain firms with non-missing accounting earnings $X_{i,t}$ and market value of equity $MVE_{i,t}$. To mitigate the influence of outliers, we trim the sample by dropping observations in the bottom and top 1% of the cross-sectional distribution within each quarter. The aggregate accounting earnings $X_t^{Aggregate}$ is defined as follows:

$$X_t^{Aggregate} = \sum_{i=1}^n \left(\frac{MVE_{i,t}}{\sum_{i=1}^n MVE_{i,t}} \times X_{i,t} \right)$$
(A.2)

where $MVE_{i,t}$ represents the market value of equity of firm i at the beginning of quarter t , n is the total number of firms. The accumulated growth $\Delta x_t^{Aggregate}$ is defined similar to the firm-level ones.

B Details on Forecasting Models

B.1 An Illustration for Rolling Window Procedure

The rolling window procedure is shown in Figure B.1.

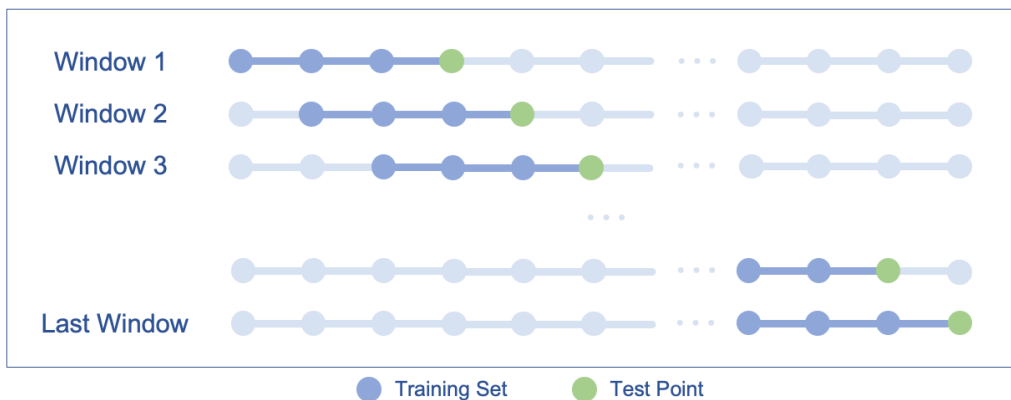


Figure B.1: Rolling Window Procedure.

Note: This figure is for illustration purposes only and the number of data points involved is different from those of real data.

B.2 Penalized Linear Regression Models

This section provides details on penalized linear regression models that are not covered in Section 4.

Penalized regression, also referred to as the regularization or shrinkage method, involves incorporating a penalty in the loss function to obtain a parsimonious linear regression model that discourages the inclusion of too many predictors. The outcome of this penalty is the reduction or shrinkage of coefficient values towards zero. As a result, less influential variables can have a coefficient close to or equal to zero, thus minimizing their contribution to the model. Specifically, the loss function of these models is defined as follows:

$$L(\beta_h) = (T - h)^{-1} \sum_{t=1}^{T-h} (g_{t+h} - \beta^{h'} Z_t)^2 + \sum_{i=1}^N p(\beta_i^h, \lambda, w_i) \quad (\text{B.1})$$

where $p(\beta_i^h, \lambda, w_i)$ is a penalty, $\beta^h = (\beta_1^h, \dots, \beta_N^h)'$ is an $N \times 1$ parameter vector, w_i is a weight for β_i^h , and λ is a tuning parameter. There are different forms of penalty to control the intensity of penalization on regression coefficients, such as shrinking towards

zero (e.g., *Ridge*), exactly to zero (e.g., *LASSO* and *adaLASSO*), or a combination of both (e.g., *Elastic Net*).

LASSO algorithm, proposed by Tibshirani (1996), imposes a strong shrinkage to the variables with an ℓ_1 -norm penalty. This leads to sparse solutions where the coefficients of the least important variables are set to zero. The penalty function is given as

$$\sum_{i=1}^N p(\beta_i^h, \lambda, \omega_i) = \lambda \sum_{i=1}^N |\beta_i^h|. \quad (\text{B.2})$$

The **adaLASSO**, proposed by Zou (2006), aims to reduce estimation bias and to improve variable selection accuracy. Given an initial estimate $\tilde{\beta}^h = (\tilde{\beta}_1^h, \dots, \tilde{\beta}_N^h)'$, usually the standard *LASSO* estimate, *adaLASSO* will assign a relatively large penalty for small coefficients and a small penalty for large coefficients. The penalty function is given as

$$\sum_{i=1}^N p(\beta_i^h, \lambda, \omega_i) = \lambda \sum_{i=1}^N \omega_i |\beta_i^h|, \quad (\text{B.3})$$

where the weight $\omega_i = |\tilde{\beta}_i^h|^{-1}$. We set $\omega_i = \frac{1}{|\tilde{\beta}_i^h| + \frac{1}{\sqrt{T}}}$ to avoid a zero denominator.

Ridge regression shrinks regression coefficients, bringing those of variables with minor contributions to the outcome close to zero. The shrinkage of coefficients is achieved by penalizing the regression model with an ℓ_2 -norm penalty, which is given as

$$\sum_{i=1}^N p(\beta_i^h, \lambda, \omega_i) = \lambda \sum_{i=1}^N \beta_i^{h2}. \quad (\text{B.4})$$

Elastic Net produces a regression model that is penalized with both the ℓ_1 - and ℓ_2 -norms. This effectively shrinks coefficients (as in *Ridge regression*) and sets some coefficients to zero (as in *LASSO*). The penalty function is given as

$$\sum_{i=1}^N p(\beta_i^h, \lambda, \omega_i) = \alpha \lambda \sum_{i=1}^N \beta_i^{h2} + (1 - \alpha) \lambda \sum_{i=1}^N |\beta_i^h|, \quad (\text{B.5})$$

where $\alpha \in [0, 1]$. *Elastic Net* aims to merge the benefits of *Ridge* and *LASSO*, which involve shrinkage and sparsity respectively. By incorporating the *Ridge* regularization component, *Elastic Net* can effectively deal with high correlations among predictors, which

is a limitation of *LASSO*. On the other hand, the ℓ_1 -norm regularization of *Elastic Net* enables it to exploit sparsity.

B.3 Random Forests and Gradient Boosting Regression Trees

The *RF* algorithm is widely used in *ML* to tackle classification and regression problems. It builds upon a regression tree model and can capture predictors' interaction effects. Instead of relying on a single tree, *RF* takes predictions from multiple trees and averages them to produce the final forecast. This approach is a variant of the bootstrap aggregation technique called “bagging” (Breiman, 2001). Each decision tree in the forest is constructed using a different subset of the training data and a random selection of features, reducing the risk of overfitting and increasing model generalization. The final prediction is typically determined by a majority vote (for classification tasks) or averaging (for regression tasks) of the individual tree predictions, making *RF* robust to noise and outliers.

The steps to implement *RF* to predict GDP growth rates is as follows:

- Step 1. Draw B different bootstrap samples of the original data. For each bootstrap sample, denote it as b , $b = 1, 2, 3, \dots, B$, and grow a tree K_b as follows:
 - (1) for sample b , randomly select a subset of M predictors from whole set \mathbf{X}_t .
 - (2) based on one of the selected predictors in (1), the tree splits the space of predictors into two rectangles.
 - (3) a new “branch” sorts the data leftover from the preceding “split” based on one of the predictor variables, which slices the predictor space into three rectangular parts according to *split rule*.
 - (4) repeat (3) until no further split is allowed by the stopping rules, such as a maximum tree depth (L). Denote the number of terminal nodes as K , and each terminal node by N_k , where $k = 1, 2, \dots, K$. Given the dependent variable g_{t+h} , a set of predictors \mathbf{X}_t , θ_k is solved by minimizing the sum of squared errors of the following regression:

$$g_{t+h} = \sum_{k=1}^K \theta_k I_{\{\mathbf{X}_t \in N_k\}} + \varepsilon_{t+h}, \quad (\text{B.6})$$

which yields $\hat{\theta}_k = \text{mean}(g_{t+h}^k)$, the sample average of g_{t+h}^k , where g_{t+h}^k is the corresponding outcome value for \mathbf{X}_t in terminal node N_k .

- Step 2. The final forecast is the average of the forecasts of each tree K_b applied to the original data:

$$\hat{g}_{t+h} = \frac{1}{B} \sum_{b=1}^B \left[\sum_{k=1}^{K_b} \hat{\theta}_k I_{\{\mathbf{X}_t \in N_k\}} \right]. \quad (\text{B.7})$$

The tuning parameters for our *RF* model are the number of bootstrap samples (B), the number of predictors in subset (M), the depth of the trees (L), and the *split rule* $\in \{\text{extratree}, \text{variance}\}$. *extratree* relies on random split selection to promote diversity and reduce inter-tree correlation, whereas *variance* follows an optimization-based rule that selects splits to minimize variance.

The idea of gradient boosting is first introduced by Breiman (2001), while Friedman (2001) later develops explicit regression gradient boosting algorithms. *GBRT* is a form of regularization that involves combining predictions from multiple “shallow trees”, which are also known as “weak learners”. These trees are recursively combined to create a powerful combination that is difficult to be beaten by many other algorithms. *GBRT* offers high predictive accuracy, robustness to outliers and noise, feature importance assessment, and the ability to handle nonlinear relationships and missing data.

We define the *GBRT* model for predicting GDP growth rates as follows:

- Step 1. Start by formalizing a “shallow” tree K_1^{GBRT} with depth L , which is usually a small number. The forecast for g_{t+h} made by K_1^{GBRT} is denoted as $f_1^L(\mathbf{X}_t)$.
- Step 2. The second “shallow” tree K_2^{GBRT} grows by fitting the prediction residuals from $f_1^L(\mathbf{X}_t)$. The forecasts made by K_2^{GBRT} is denoted as $f_2^L(\mathbf{X}_t)$. We ensemble $f_1^L(\mathbf{X}_t)$ and $f_2^L(\mathbf{X}_t)$ as follows:

$$f_1^L(\mathbf{X}_t) + \gamma f_2^L(\mathbf{X}_t), \quad (\text{B.8})$$

where $\gamma \in (0, 1)$ is a shrinkage parameter preventing the model from overfitting.

- Step 3 to Step B . Repeat Step 2 until the B -th tree K_B^{GBRT} has grown up. At each repeating step, a “shallow” tree is fitted with the residuals from the previous one,

and its residual forecast is added to the total at a shrinkage rate γ . Therefore, the final forecast for g_{t+h} is defined as an additive model of B “shallow” trees:

$$\hat{g}_{t+h} = f_1^L(\mathbf{X}_t) + \gamma f_2^L(\mathbf{X}_t) + \dots + \gamma f_B^L(\mathbf{X}_t). \quad (\text{B.9})$$

The hyperparameters L , B , γ are determined by the validation procedure, in which we consider a range of candidate values for each of these parameters. The in-sample dataset is partitioned into a training set and a validation set, with the latter being used to tune the hyperparameters. By fitting the model to the training set and making predictions for the validation set, we can calculate the forecast errors and search for the best combination of hyperparameters that optimize the validation objective.

B.4 Tuning Parameter Grids for Machine Learning Models

Table B.1 presents the set of hyperparameters and their corresponding value ranges used for tuning each *ML* model.

Table B.1: Hyperparameters for All Models

Model	Tuning Grid
<i>adaLASSO</i>	$\lambda \in [0, 2]$
<i>LASSO</i>	$\lambda \in [0, 2]$
<i>Elastic Net</i>	$\alpha = 0.5, \lambda \in [0, 1]$
<i>Ridge</i>	$\lambda \in [0.1, 2000]$
<i>RF</i>	<i>regressor</i> = { <i>variance</i> , <i>extratrees</i> }
	<i>max_features</i> = {402, 702, 1002, 1302}
	<i>n_estimators</i> = {200, 500, 1000, 2000}
	<i>max_depth</i> = {3, 5, 8, 10}
<i>GBRT</i>	<i>depth</i> = {3, 4, 5, 6}
	<i>n.trees</i> = {50, 60, 70, 80, 90, 100}
	<i>shrinkage</i> = {0.1, 0.01}
	<i>minobsinnode</i> = {4, 5, 6}

Note: This table reports the ranges of hyperparameters over which we tune each *ML* method.

B.5 Standard Principal Component Analysis

In this paper, our standard Principal Component Analysis (standard *PCA*) follows a two-step process.

- Step 1. For a panel $X_{T \times N}$ with N predictors and T observations, demean every column of the predictor matrix to get the centered matrix \tilde{X} . Calculate the covariance matrix $\Sigma_{N \times N} = (T-1)^{-1} \tilde{X}^T \tilde{X}$. Then extract the eigenvalues λ and eigenvectors v via eigenvalue decomposition. Rearrange the eigenvectors according to the descending order of their eigenvalues, the top k eigenvectors are used to construct $W_{N \times k}$.¹
- Step 2. Calculate the principal component as $F = \tilde{X}W$. These principal components are then estimated using the Ordinary Least Squares (*OLS*) method to forecast GDP growth:

$$g_{t+h} = \beta_0 + \sum_{i=1}^k \beta_i f_{i,t} + \varepsilon_{t+h}. \quad (\text{B.10})$$

¹In our analysis, we set $k = 4$ for our standard *PCA* model.

C Additional Results

This section provides additional results that are not included in the main manuscript due to space constraint.

C.1 Additional Results for Micro-Forecasting Models vs. Benchmark Model

Table C.1 reports the MSE values for the U.S. GDP growth rate forecasts.

Table C.1: Forecasting Performance of Micro-Forecasting Models: U.S. Real GDP Growth (MSE)

Micro-Forecasting Model	Forecast Horizon				Avg.
	h=1	h=2	h=4	h=8	
Panel (a): Benchmark					
<i>AR</i>	4.061	4.802	7.054	14.125	7.510
Panel (b): Micro-Forecasting Models					
<i>adaLASSO</i>	2.046	4.288	6.245	12.639	6.304
<i>LASSO</i>	2.044	4.336	6.165	12.488	6.258
<i>Elastic Net</i>	2.041	4.258	6.594	12.683	6.394
<i>Ridge</i>	2.240	3.977	6.337	12.396	6.238
<i>RF</i>	2.059	3.518	6.169	10.853	5.650
<i>GBRT</i>	2.736	3.948	5.948	11.419	6.013

Note: This table reports the MSE of each model for each forecast horizon. The forecast target is the U.S. real GDP growth rate, while the predictors are individual firm accounting earnings. Panels (a) and (b) show the results of the benchmark *AR* model and micro-forecasting models, respectively. The out-of-sample period is 2000Q1-2023Q4. The ‘‘Avg.’’ column shows the average MSE across horizons. The MSE for the best-performing model at each horizon is highlighted in bold.

In the main result, missing values in the predictive panel of accounting earnings growth are set to zero. We follow common practice in the recent machine learning finance literature (Kelly et al., 2019; Feng et al., 2020; Gu et al., 2020; Cong et al., 2025), and impute missing values using the cross-sectional median within each forecasting rolling window. The results are in Table C.2.

Table C.2: Forecasting Performance of Micro-Forecasting Models: Changing Missing Value Imputation (MSE Ratio)

Micro-Forecasting Model	Forecast Horizon				Avg.
	h=1	h=2	h=4	h=8	
<i>adaLASSO</i>	0.503	0.909	0.893	0.922	0.807
<i>LASSO</i>	0.503	0.900	0.882	0.903	0.797
<i>Elastic Net</i>	0.504	0.882	0.937	0.928	0.813
<i>Ridge</i>	0.551	0.827	0.896	0.899	0.793
<i>RF</i>	0.499	0.729	0.876	0.770	0.718
<i>GBRT</i>	0.689	0.861	0.849	0.814	0.803

Note: This table reports the MSE ratio of each micro-forecasting model relative to the *AR* benchmark, for each forecast horizon. The forecast target is the U.S. real GDP growth rate, while the predictors are individual firm accounting earnings. The missing values are filled with cross-sectional median within each rolling window. The out-of-sample period is 2000Q1-2023Q4. The “Avg.” column shows the average MSE ratio across horizons. The MSE ratio for the best-performing model at each horizon is highlighted in bold.

We further do [Clark and West’s \(2007\)](#) test to compare the forecasting performance of the micro-forecasting models against the benchmark *AR* model. The CW test adjusts for the upward bias in the larger model’s mean squared prediction error that arises from estimating redundant parameters under the null, thereby providing valid inference in nested model settings. This framework is particularly well suited to our application, since our micro-forecasting models extend the *AR* specification by incorporating firm-level accounting earnings. The CW test has been widely adopted in the forecasting literature (e.g., [Huang et al., 2015](#); [Anesti et al., 2022](#); [Rossi, 2013](#)), and its use here provides an additional check on the statistical significance of our findings.

Table [C.3](#) reports the *p*-values from the CW test for the main results. All results are significant at the 10% level or better, indicating that relative to the parsimonious *AR* model, the additional micro-level accounting earnings in the micro-forecasting models contain useful information for forecasting.

Moreover, forecasting literature (e.g., [Timmermann, 2006](#)) suggests model averaging across various *ML* models may improve forecast performance. We consider six commonly used model averaging methods ([Timmermann, 2006](#)):

Table C.3: Forecasting Performance Comparison: Micro-Forecasting vs. *AR* Model (CW Test p -value)

Micro-Forecasting Model	Forecast Horizon			
	h=1	h=2	h=4	h=8
<i>adaLASSO</i>	0.026	0.050	0.080	0.008
<i>LASSO</i>	0.026	0.049	0.077	0.009
<i>Elastic Net</i>	0.027	0.051	0.093	0.009
<i>Ridge</i>	0.026	0.042	0.088	0.079
<i>RF</i>	0.019	0.027	0.062	0.036
<i>GBRT</i>	0.021	0.030	0.047	0.054

Note: This table reports p -values of Clark and West’s (2007) test comparing each micro-forecasting model against the *AR* benchmark, for each forecast horizon. The null hypothesis is that the two models have equal predictive accuracy, while the alternative hypothesis is that the micro-forecasting model delivers superior forecasts. The p -value for the best-performing model at each horizon is highlighted in bold.

- **Simple averaging** (*Simple*): In each rolling window, the predictions of the six models are averaged using equal weights.
- **OLS averaging** (*OLS*): In each rolling window, the model averaging equation is defined as

$$\hat{g}_{t+h}^{avg} = \alpha + \sum_{m=1}^6 \beta^m \hat{g}_{t+h}^m, \quad (\text{C.1})$$

where \hat{g}_{t+h}^{avg} denotes the forecast series resulting from model averaging across individual models, $m = 1, \dots, 6$ denotes the six *ML* models, and the parameters α and β^m are estimated using the in-sample data within the window. Specifically, the forecast target g_{t+h}^{in} in the in-sample dataset is regressed on the in-sample fitted values $\hat{g}_{t+h}^{in,m}$ from the six *ML* models:

$$g_{t+h}^{in} = \alpha + \sum_{m=1}^6 \beta^m \hat{g}_{t+h}^{in,m} + \varepsilon_{t+h}, \quad (\text{C.2})$$

and the estimated coefficients α and β^m serve as the weights for model averaging.

- **Robust averaging** (*Robust*): It’s similar to OLS averaging, but the parameters in the regression Eq.C.2 are estimated using median regression instead of ordinary least squares to reduce the sensitivity to outliers.

- **Constrained least squares (CLS):** It’s similar to OLS, but the regression equation excludes the constant term and imposes constraints on the coefficients β^m

$$g_{t+h}^{train} = \sum_{m=1}^6 \beta^m g_{t+h}^m + \varepsilon_{t+h} \quad s.t. \quad \sum_{m=1}^6 \beta^m = 1. \quad (C.3)$$

- **Variance-based weighting (Variance-based):** For each rolling window, the weight β^m of model m in Eq.C.1 is determined based on its in-sample MSE (MSE^m):

$$\beta^m = \frac{1/MSE^m}{\sum_{i=1}^6 1/MSE^i} \quad (C.4)$$

- **Best-performing model (Best):** For each rolling window, the model with the lowest in-sample MSE is selected, and its out-of-sample prediction is used for that window.

We apply six model-averaging methods and compute the average MSE ratio relative to the *AR* benchmark across four forecast horizons (as reported in Table C.4).

Table C.4: Forecasting Performance of Model Averaging Methods: U.S. Real GDP Growth (MSE Ratio)

Model	Forecast Horizon				Avg.
	h=1	h=2	h=4	h=8	
Panel (a): Individual Model					
<i>RF</i>	0.507	0.733	0.875	0.768	0.721
Panel (b): Model Averaging					
<i>Simple</i>	0.502	0.802	0.871	0.816	0.748
<i>OLS</i>	0.617	0.744	0.894	0.792	0.762
<i>Robust</i>	0.653	0.741	0.892	0.793	0.770
<i>Variance-based</i>	0.514	0.749	0.858	0.794	0.729
<i>CLS</i>	0.535	0.754	0.862	0.761	0.728
<i>Best</i>	0.533	0.764	0.859	0.763	0.730

Note: This table reports the MSE ratio of each micro-forecasting model relative to the *AR* benchmark, for each forecast horizon. The forecast target is the U.S. real GDP growth rate, while the predictors are individual firm accounting earnings. Panel (a) reports the best-performing individual model *RF*, while Panel (b) presents results for six model-averaging methods. The out-of-sample period is 2000Q1-2023Q4. The “Avg.” column shows the average MSE ratio across horizons. The MSE ratio for the best-performing model at each horizon is highlighted in bold.

C.2 Additional Results for Time-Varying Predictability

C.2.1 Additional Figures and Regression Results for Time-Varying Predictability

Figure C.1 shows the time-varying predictability of $h = 2$ and 4.

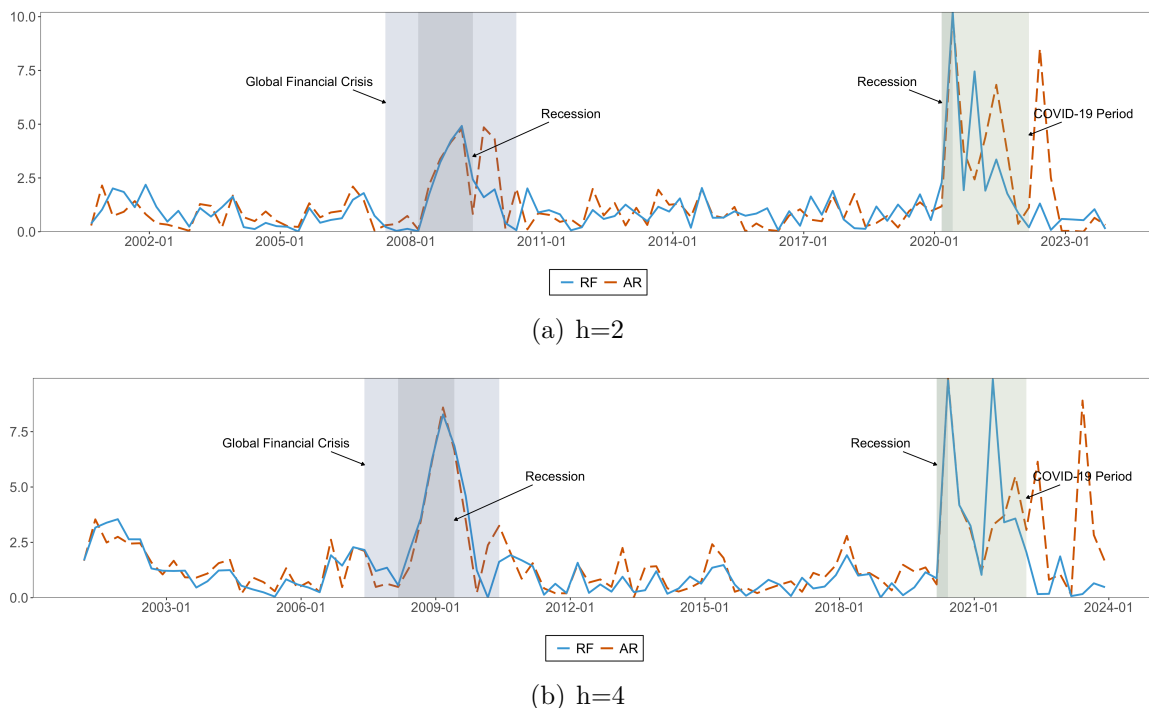


Figure C.1: Absolute Forecast Errors of the Micro-Forecasting vs. AR Model: U.S. GDP Growth.

Note: These figures present the absolute forecast errors of the micro-forecasting and AR models for $h = 2$ and 4 in forecasting the U.S. real GDP growth rate. We select the best-performing RF to represent the micro-forecasting model. The periods of NBER recession periods, the 2008 global financial crisis, and COVID-19 are highlighted in gray, purple and green shading, respectively.

We further follow [Medeiros et al. \(2021\)](#) and run the following regression to confirm whether our micro-forecasting approach maintains its superior performance under volatile economic conditions:

$$\Delta(\hat{e}_{t+h}^{AR,2}, \hat{e}_{t+h}^{Micro,2}) = \beta_0 I_{t+h} + \beta_1 (1 - I_{t+h}) + v_{t+h}, \quad (C.5)$$

where $\hat{e}_{t+h}^{m,2} = (g_{t+h} - \hat{g}_{t+h,m})^2$ represents the squared forecasting error of model m (AR for the AR model and $Micro$ for the micro-forecasting model), and $\hat{g}_{t+h,m}$ is the predicted value of GDP growth rate g_{t+h} generated by model m . The difference between the squared fore-

casting errors, $\hat{e}_{t+h}^{AR,2}$ and $\hat{e}_{t+h}^{Micro,2}$, denoted as $\Delta(\hat{e}_{t+h}^{AR,2}, \hat{e}_{t+h}^{Micro,2})$, quantifies the improvement achieved by the micro-forecasting model over the AR model. The regressor I_{t+h} indicates the presence of a volatile period, which is equal to 1 during periods of economic volatility (as defined in Section 5.2 in the main manuscript) and 0 otherwise.

Table C.5 presents the estimated coefficients $\hat{\beta}_0$ and $\hat{\beta}_1$ from Eq.C.5, along with their corresponding p -values in parentheses. The results show that $\hat{\beta}_0$ is positive and statistically significant at the 10% level in all cases, with some results significant at the 5% and 1% levels, where the micro-forecasting framework demonstrates stronger improvement over AR . This indicates that the superiority of the micro-forecasting model becomes more prominent during periods of high economic volatility.

Table C.5: Predictive Ability during Volatile Periods: Micro-Forecasting Model vs. AR Model

	Forecast Horizon			
	h=1	h=2	h=4	h=8
$\hat{\beta}_0$	3.095*** (0.002)	3.323*** (0.001)	2.351** (0.010)	2.437* (0.084)
$\hat{\beta}_1$	-0.005 (0.994)	-0.164 (0.807)	0.702 (0.281)	4.179*** (0.000)

Note: This table reports the estimated coefficients $\hat{\beta}_0$ and $\hat{\beta}_1$ in Eq.C.5, with the corresponding p -values in parentheses, for each forecast horizon. “***”, “**”, and “*” denote statistical significance at the 1%, 5%, and 10% levels, respectively. $\hat{\beta}_0 > 0$ indicates predictive ability of micro-forecasting approach during volatile periods, with $\hat{\beta}_1 > 0$ for non-volatile periods. The volatile periods include NBER recession periods, the global financial crisis, and the COVID-19 pandemics.

C.2.2 Time-Varying Analysis on How Forecasting Performance Correlates to Investor Sentiment

Furthermore, we examine whether our forecasting performance is related to investor sentiment, as a large literature examines earnings management, with several studies highlighting its connection to sentiment, a key factor in finance.

We follow Huang et al. (2015) and employ their investment sentiment index to investigate the relationship between the model’s forecast errors and investment sentiment. The scatter plots, as shown in Figure C.2, indicate a positive correlation: when investment

sentiment is higher, the forecast errors for GDP are relatively larger, especially at medium- to long-term horizons ($h = 4$ and 8).

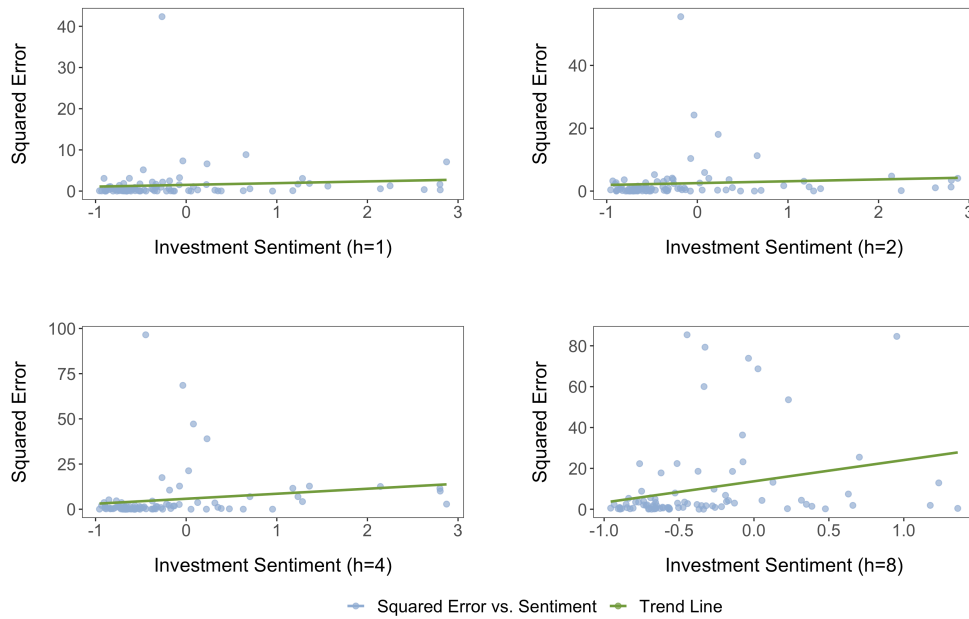


Figure C.2: Relationship Between Forecast Errors and Investment Sentiment.

Note: These figures present scatter plots and fitted trend lines between forecast squared error and investment sentiment at each forecast horizon.

We further conduct a regression analysis. First, we define a dummy variable H_t indicating high investment sentiment periods:

$$H_t = \begin{cases} 1 & \text{if } is_t > \text{median}(is) \\ 0 & \text{otherwise} \end{cases} \quad (\text{C.6})$$

where is_t denotes investment sentiment at time t , and $\text{median}(is)$ denotes the median level of investment sentiment over the sample period. The following regression is then estimated to examine the predictive power during high-sentiment periods:

$$\hat{e}_{t+h}^2 = \beta_0 H_t + \beta_1 (1 - H_t) + \varepsilon_{t+h} \quad (\text{C.7})$$

where \hat{e}_t denotes the squared forecast error.

The results, as shown in Table C.6, are broadly consistent with the scatter plot evidence. It is worth emphasizing that although our model's forecast errors are relatively larger in

high-sentiment periods, the forecasting performance still remains significantly better than that of the benchmark model.

Table C.6: Predictive Ability during High Sentiment Periods

	Forecast Horizon			
	h=1	h=2	h=4	h=8
$\hat{\beta}_0$	2.360*** (0.001)	4.147*** (0.000)	10.067*** (0.000)	18.913*** (0.000)
$\hat{\beta}_1$	0.601 (0.361)	0.843 (0.372)	0.991 (0.606)	2.997 (0.266)

Note: This table reports the estimated coefficients $\hat{\beta}_0$ and $\hat{\beta}_1$ in Eq.C.7, with the corresponding p -values in parentheses, for each forecast horizon. “***”, “**”, and “*” denote statistical significance at the 1%, 5%, and 10% levels, respectively. $\hat{\beta}_0 > 0$ indicates predictive ability of micro-forecasting approach during high sentiment periods, with $\hat{\beta}_1 > 0$ for low sentiment periods. The high and low sentiment periods are defined following the investor sentiment index proposed by [Huang et al. \(2015\)](#).

C.2.3 Figures for Time-Varying Comparisons

As a further complement to the analysis in Section 5.2, we also compare the time-varying predictive performance of different *ML* models, as well as the time-varying predictability of micro-level versus aggregate predictors in Section 6.1. The corresponding results are presented in Figure C.3 and Figure C.4, respectively.

C.3 Additional Results for Robustness Check

C.3.1 Forecasting Results for Different Sample Selection Criteria

First, in the main analysis, firms are retained if they report accounting earnings within 16 weeks after the fiscal quarter end. In this section, we examine the distribution of reporting weeks across all firms and vary this cutoff window as a robustness check.

For each firm, the average reporting weeks are computed as the number of weeks between the fiscal quarter end and the actual reporting date across the sample period, and firms are then grouped into five categories: within 8 weeks, 8-12 weeks, 12-16 weeks, 16-20 weeks, and beyond 20 weeks. Figure C.5 presents the frequency histogram and cumulative distribution of these average reporting weeks, while Figure C.6 shows the patterns over time.

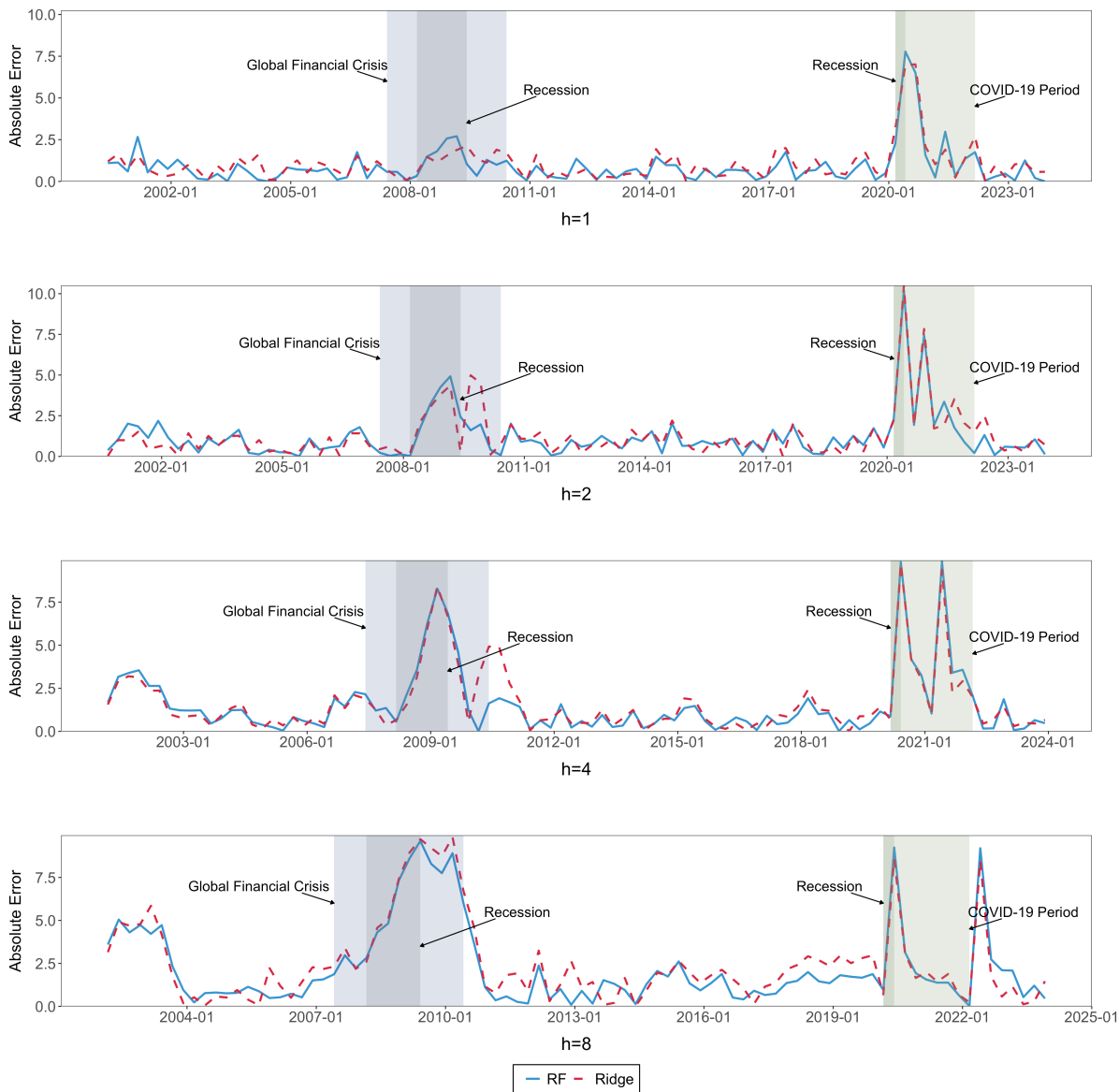


Figure C.3: Absolute Forecast Errors of Linear vs. Non-Linear Models: U.S. Real GDP Growth.

Note: These figures present the absolute forecast errors of the linear model (*Ridge*) and non-linear model (*RF*) using individual firm accounting earnings for each forecast horizon. The periods of NBER recession periods, the global financial crisis, and COVID-19 are highlighted in gray, purple and green, respectively.

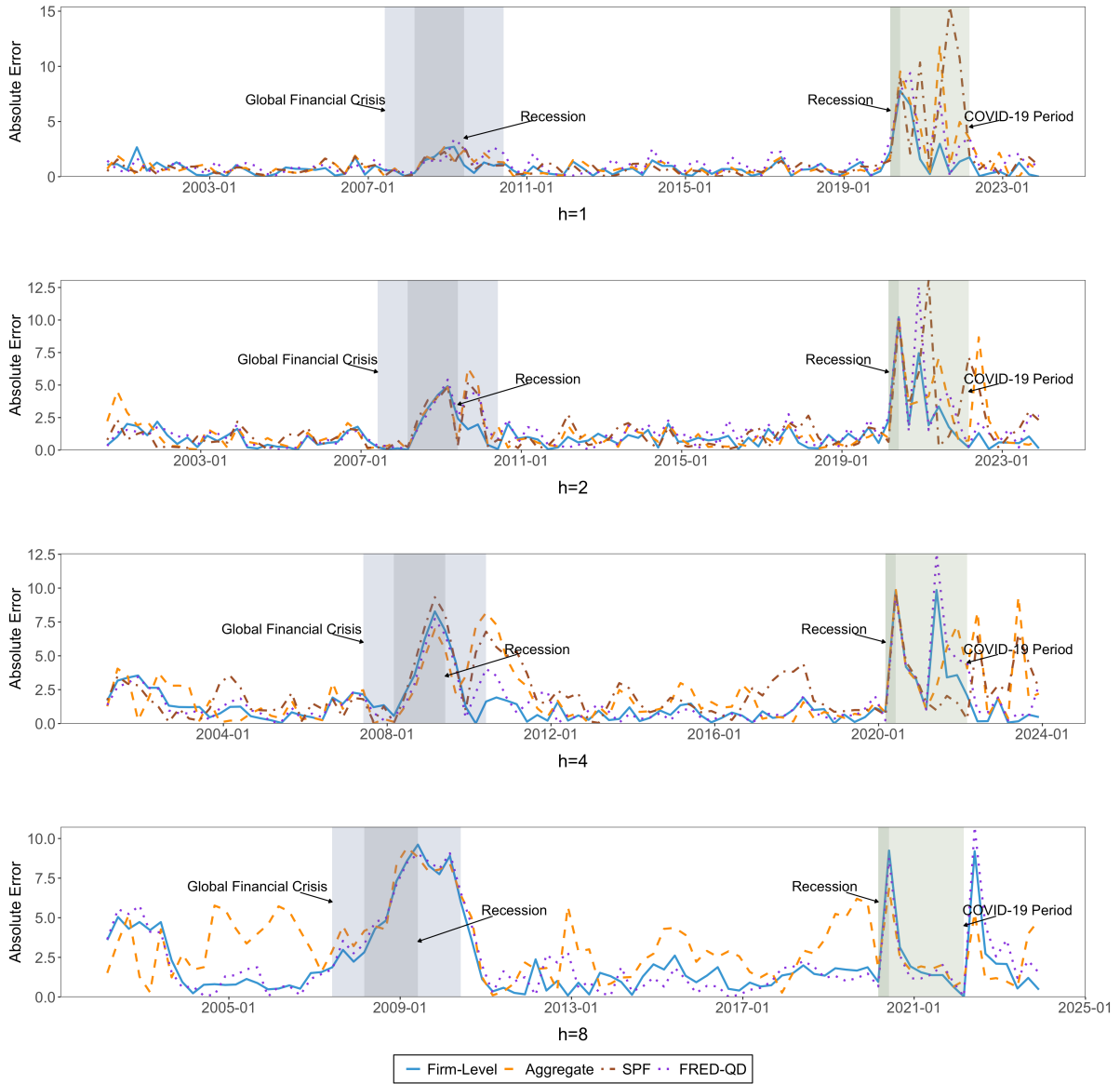


Figure C.4: Absolute Forecast Errors of Micro-Level vs. Aggregate Predictors: U.S. Real GDP Growth.

Note: These figures present the absolute forecast errors of the models using aggregate accounting earnings (“Aggregate”), FRED-QD database (“FRED-QD”), individual accounting earnings (“Firm-Level”), and the Survey of Professional Forecasters (“SPF”) in U.S. real GDP growth rate forecasting. The periods of NBER recession, the global financial crisis, and COVID-19 are highlighted in gray, purple and green, respectively.

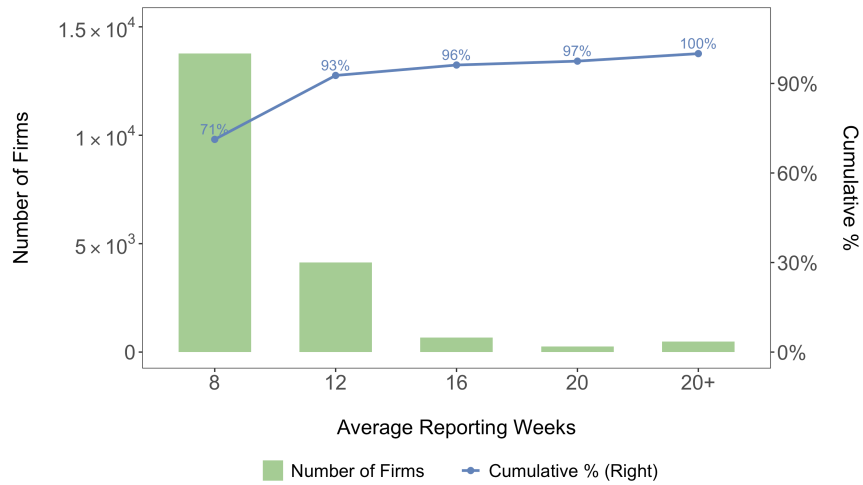


Figure C.5: Distribution of Firms' Average Reporting Weeks.

Note: The figure shows the frequency histogram (bars, left axis) and cumulative distribution (line, right axis) of firms' average reporting weeks, measured as the number of weeks between the fiscal quarter-end and the earnings announcement date. Firms are grouped into five categories: within 8 weeks (*8*), 8-12 weeks (*12*), 12-16 weeks (*16*), 16-20 weeks (*20*), and beyond 20 weeks (*20+*).

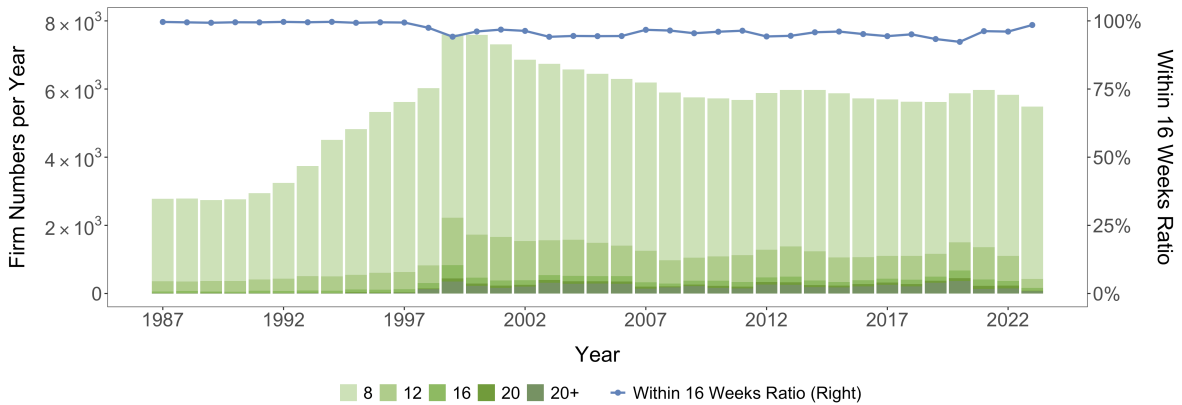


Figure C.6: Annual Distribution of Firms' Reporting Week Groups.

Note: The figure shows the annual distribution of firms' reporting weeks. In each year, firms are grouped into five categories: within 8 weeks (*8*), 8-12 weeks (*12*), 12-16 weeks (*16*), 16-20 weeks (*20*), and beyond 20 weeks (*20+*). The line (right axis) illustrates the ratio of firms that release their earnings within 16 weeks after the fiscal quarter end for each year.

Based on this grouping, we re-estimate the results using subsamples restricted to firms reporting within 8, 12, and 20 weeks, respectively. As shown in Table C.8 Panel (a), the findings remain robust across all selections.

Second, the main sample includes all firms publicly listed in the U.S., including foreign firms. Since U.S. GDP is measured on a territorial basis, only the value added generated within U.S. borders contributes to U.S. GDP, while overseas operations of foreign firms do not. Including such firms may therefore introduce noise and weaken the empirical link to U.S. GDP.

To address this concern, we construct a revised sample that excludes foreign firms. The country of incorporation for each firm is identified using the *Compustat-Capital IQ/Compustat/North America/Fundamentals Quarterly* database, where the variable *fic* provides the ISO country code. Out of 19,053 firms in the main result sample, 18,675 are successfully matched: 16,016 incorporated in the U.S. (*fic = USA*) and 2,659 identified as foreign firms. The remaining 378 unmatched firms are also excluded. We thus retain 16,016 U.S.-incorporated firms as the revised sample.

We then re-estimate the pre-selection and prediction procedures using this U.S.-only sample. As shown in Table C.8 Panel (b), the predictive performance confirms that the main findings are robust to the exclusion of foreign firms.

Third, while the main sample includes all income of all firms, some U.S.-incorporated firms derive a substantial share of their earnings from abroad, which does not directly contribute to U.S. GDP and may also introduce additional noise into the forecasts.

To account for this, we use information from the *Compustat-Capital IQ/Compustat/Historical Segments/Historical Segments* database to approximate the extent of foreign operations. The key variables used include (see Table C.7 for example records from **Apple Inc.** (U.S.) and **Bayer** (Germany)):

- *gvkey*: the unique company identifier in *Compustat*
- *datadate*: the source date of the segment-level annual fundamentals
- *conm*: the company name
- *soptp1*: segment type; we keep only *soptp1 = GEO* indicating the segment information is presented by geographic operations

- *snms*: segment name; for geographic records, this variable indicates the location where the financial item occurs
- *geotp*: geographical type; the type of geographical segment, with 2 representing domestic and 3 representing non-domestic segments
- *sales*: gross sales of a segment, representing actual billings to customers for regular sales completed during the period, net of cash discounts, trade discounts, and returned sales or allowances for which credit is given to customers
- *nis*: net income of a segment, representing the income or loss after all expenses and losses have been subtracted from revenues and gains for the fiscal period, including extraordinary items and discontinued operations

Table C.7: A Sample of Historical Segments Database

gvkey	datadate	conm	soptp1	snms	geotp	sales	nis
001690	2000-09-30	APPLE INC	GEO	Japan	3	1345.000	NA
001690	2000-09-30	APPLE INC	GEO	Americas	2	4298.000	NA
001690	2000-09-30	APPLE INC	GEO	Europe	3	1817.000	NA
001690	2000-09-30	APPLE INC	GEO	Other	3	523.000	NA
001690	2014-09-30	APPLE INC	GEO	United States	2	68909	NA
001690	2014-09-30	APPLE INC	GEO	China	3	30638	NA
001690	2014-09-30	APPLE INC	GEO	Other Countries	3	83248	NA
100080	1998-12-31	BAYER AG	GEO	Europe	2	18052.190	NA
100080	1998-12-31	BAYER AG	GEO	North America	3	9226.155	NA
100080	1998-12-31	BAYER AG	GEO	Latin America	3	2231.553	NA
100080	1998-12-31	BAYER AG	GEO	Asia, Africa, Australia	3	3413.917	NA

We retain only records with $soptp1 = GEO$, indicating that segment information is reported by geographic operations. Ideally, segment-level net income would allow us to distinguish domestic from foreign earnings, but these data are largely missing. We therefore approximate the split using sales. For each firm-year pair, the share of overseas activity is measured as the ratio of non-domestic sales ($geotp = 3$) to total sales, and a firm-level measure is obtained by averaging across time. For example, **Apple Inc.** in 2000 reports overseas sales of 3,685 (Japan, Europe, Others) out of total sales of 7,983, yielding an overseas revenue share of 46.16%.

For foreign firms, a similar procedure could in principle be applied to extract U.S.-sourced sales. However, due to missing values and heterogeneous reporting of geographic

segments (e.g., “North America” without further breakdown), reliable measurement is not feasible. We therefore exclude all foreign firms from this exercise.

Among the 16,016 U.S. firms, 3,290 report a positive overseas share (20.54% of the sample), while 1,556 have an average share above 50% (9.72%). Based on this distribution, we conduct two robustness checks: (1) excluding all firms with any overseas sales, and (2) excluding only those with a mean overseas share greater than 50%.

As shown in Table C.8 Panel (c), the results remain robust overall. This indicates that isolating foreign-sourced earnings does not materially affect the main findings.

Table C.8: Robustness Check: Using Different Sample Selection Criteria (MSE Ratio)

Predictor	Forecast Horizon			
	h=1	h=2	h=4	h=8
Panel (a): Selection by Reporting Window Cutoff				
8 Week	0.578	0.851	0.869	0.739
12 Week	0.540	0.812	0.892	0.761
16 Week	0.507	0.733	0.875	0.768
20 Week	0.558	0.817	0.883	0.770
Panel (b): Selection by Incorporating Countries				
All Firms	0.507	0.733	0.875	0.768
U.S. Firms Only	0.551	0.828	0.894	0.764
Panel (c): Selection by Income Abroad				
All Firms	0.507	0.733	0.875	0.768
Oversea Share=0	0.570	0.793	0.880	0.759
Oversea Share<50%	0.550	0.829	0.898	0.759

Note: This table reports the MSE ratio of each model relative to the *AR* benchmark, for each forecast horizon. The forecast target is the U.S. real GDP growth rate, while each panel presents results based on samples selected using different criteria. Panel (a) presents the results of firms selected by different window cutoffs for earnings reporting (8, 12, 16, and 20 weeks), where the 16-week cutoff corresponds to the main results. Panel (b) shows the results of firms after excluding foreign firms. Row “All Firms” corresponds to the full sample (same as the main results), while Row “U.S. Firms Only” restricts the sample to U.S. firms only. Panel (c) reports the results of firms after excluding firms by the share of foreign income: Row “All Firms” corresponds to the full sample (same as the main results), Row “Overseas Share=0” includes only firms without foreign income, and Row “Overseas Share<50%” includes firms with less than half of their income from abroad. All results are based on the best-performing *RF* model.

C.3.2 Forecasting Results for Other Output Measures

Moreover, we use various output measures for robustness checks, namely, U.S. nominal GDP growth rate, and U.S. industrial production (IP) growth rate. Tables C.9 and C.10 show the MSE ratios relative to *AR*, while Tables C.11 and C.12 report the MSE values.

Table C.9: Robustness Check: Forecasting U.S. Nominal GDP Growth (MSE Ratio)

Micro-Forecasting Model	Forecast Horizon				Avg.
	h=1	h=2	h=4	h=8	
<i>adaLASSO</i>	0.456	0.545	0.439	0.596	0.509
<i>LASSO</i>	0.423	0.527	0.416	0.592	0.489
<i>Elastic Net</i>	0.424	0.524	0.422	0.597	0.492
<i>Ridge</i>	0.384	0.575	0.425	0.734	0.530
<i>RF</i>	0.440	0.501	0.358	0.683	0.495
<i>GBRT</i>	0.606	0.599	0.356	0.656	0.554

Note: This table reports the MSE ratio of each micro-forecasting model relative to the *AR* benchmark, for each forecast horizon. The forecast target is the U.S. nominal GDP growth rate, while the predictors are individual firm accounting earnings. The out-of-sample period is 2000Q1-2023Q4. The “Avg.” column shows the average MSE ratio across horizons. The MSE ratio for the best-performing model at each horizon is highlighted in bold.

Table C.10: Robustness Check: Forecasting U.S. IP Growth (MSE Ratio)

Micro-Forecasting Model	Forecast Horizon				Avg.
	h=1	h=2	h=4	h=8	
<i>adaLASSO</i>	0.405	0.597	1.039	1.087	0.782
<i>LASSO</i>	0.405	0.582	0.955	1.078	0.755
<i>Elastic Net</i>	0.405	0.647	1.161	1.087	0.825
<i>Ridge</i>	0.400	0.583	0.941	1.100	0.756
<i>RF</i>	0.411	0.586	0.883	1.135	0.754
<i>GBRT</i>	0.446	0.603	0.916	1.186	0.788

Note: This table reports the MSE ratio of each micro-forecasting model relative to the *AR* benchmark, for each forecast horizon. The forecast target is the U.S. IP growth rate, while the predictors are individual firm accounting earnings. The out-of-sample period is 2000Q1-2023Q4. The “Avg.” column shows the average MSE ratio across horizons. The MSE ratio for the best-performing model at each horizon is highlighted in bold.

Table C.11: Robustness Check: Forecasting U.S. Nominal GDP Growth (MSE)

Micro-Forecasting Model	Forecast Horizon				Avg.
	h=1	h=2	h=4	h=8	
Panel (a): Benchmark					
<i>AR</i>	5.696	9.616	30.769	41.686	21.942
Panel (b): Micro-Forecasting Models					
<i>adaLASSO</i>	2.599	5.238	13.498	24.863	11.549
<i>LASSO</i>	2.409	5.070	12.786	24.661	11.232
<i>Elastic Net</i>	2.413	5.036	12.976	24.900	11.331
<i>Ridge</i>	2.190	5.528	13.066	30.611	12.849
<i>RF</i>	2.504	4.813	11.026	28.486	11.707
<i>GBRT</i>	3.451	5.762	10.965	27.330	11.877

Note: This table reports the MSE of each model for each forecast horizon. The forecast target is the U.S. nominal GDP growth rate, while the predictors are individual firm accounting earnings. Panels (a) and (b) show the results of the benchmark *AR* model and micro-forecasting models, respectively. The out-of-sample period is 2000Q1-2023Q4. The “Avg.” column shows the average MSE across horizons. The MSE for the best-performing model at each horizon is highlighted in bold.

Table C.12: Robustness Check: Forecasting U.S. IP Growth (MSE)

Micro-Forecasting Model	Forecast Horizon				Avg.
	h=1	h=2	h=4	h=8	
Panel (a): Benchmark					
<i>AR</i>	12.861	18.612	29.452	53.749	28.669
Panel (b): Micro-Forecasting Models					
<i>adaLASSO</i>	5.209	11.106	30.601	58.419	26.334
<i>LASSO</i>	5.207	10.835	28.117	57.935	25.524
<i>Elastic Net</i>	5.209	12.048	34.200	58.440	27.474
<i>Ridge</i>	5.144	10.858	27.714	59.150	25.716
<i>RF</i>	5.283	10.909	26.017	60.998	25.802
<i>GBRT</i>	5.736	11.227	26.977	63.765	26.926

Note: This table reports the MSE of each model for each forecast horizon. The forecast target is the U.S. IP growth rate, while the predictors are individual firm accounting earnings. Panels (a) and (b) show the results of the benchmark *AR* model and micro-forecasting models, respectively. The out-of-sample period is 2000Q1-2023Q4. The “Avg.” column shows the average MSE across horizons. The MSE for the best-performing model at each horizon is highlighted in bold.

C.3.3 Forecasting GDP Components

In this section, we complement the analysis on forecasting GDP component growth rate using individual accounting earnings. According to the expenditure approach, GDP consists of four main components: personal consumption expenditures (*consumption*), gross private domestic investment (*investment*), government consumption expenditures and gross investment (*government spending*), and net exports of goods and services (*net exports*). The data is sourced from the Bureau of Economic Analysis (BEA). Understanding the predictive value of individual accounting earnings for each of these components could provide valuable insights into their dynamic relationship with future GDP growth.

Like the GDP growth rate, we calculate the accumulated growth rates of these components as:

$$g_{t+h}^c = 100 \times (GDP_{t+h}^c - GDP_t^c) / GDP_t^c, \quad (\text{C.8})$$

where GDP_t^c represents the value of different GDP components at time t , and c stands for *consumption*, *investment*, *net exports* and *government spending*, respectively. We use the important firms selected from the GDP growth forecast exercise in Section 5.1 to predict the growth rate of each GDP component g_{t+h}^c .

For each GDP component, Table C.13 reports the MSE ratios of the micro-forecasting approach relative to the *AR* model. Several important points stand out. First, Table C.13 shows that micro-forecasting approach using individual accounting earnings significantly improves the forecasting of all components at all horizons (except for *government spending*). This further highlights the advantage of the micro-forecasting approach.

Moreover, by comparing the forecasting performance across different GDP components, it is evident that the micro-forecasting approach performs even better when the GDP component is more volatile and the forecast horizon is longer. For example, it is well known that *net exports* and *investment* are much more volatile than other components, with standard deviations of their annual growth rates being 0.516 and 0.082, respectively. In Table C.13, we find that for these two components, the MSE ratios are smaller than 1. Additionally, for *net exports* forecasting, the MSE ratio decreases further as the forecast horizon extends, from 0.701 and 0.712 for $h = 1$ and $h = 2$, respectively, to 0.668 and 0.575 for $h = 4$ and 8, respectively.

These findings are consistent with our earlier results, which show that our micro-forecasting approach forecasts GDP growth more accurately during volatile periods across business cycles. Again, this highlights the advantage of *ML* in exploiting individual-level information in more complex environments. On the other hand, we also observe from Table C.13 that for *government spending* – a relatively smooth component, with a standard deviation 0.024 of annual growth rates – our methodology cannot improve upon the *AR* model. In this case, the *AR* model based on historical information already performs well in predicting a relatively smooth GDP component.

Lastly, we note that for *consumption*, the dominant component of GDP, the MSE ratios can be even lower, with a value of 0.334 for $h = 1$. This is possibly because a firm’s accounting earnings, driven by profitability, are closely correlated with their employees’ wages and labor income. Moreover, as discussed in Section 2.4 in the main manuscript, the liquidity and wealth channel can be important for investors and consumers. As a result, aggregate household consumption can be forecasted more accurately by individual accounting earnings.

Table C.13: Forecasting Performance of Micro-Forecasting Approach: GDP Components (MSE Ratio)

Component	Std.	Forecast Horizon			
		h=1	h=2	h=4	h=8
Net Exports	0.516	0.701	0.712	0.668	0.575
Investment	0.082	0.710	0.755	0.829	0.864
Gov. Spending	0.024	1.071	0.898	1.028	1.147
Consumption	0.030	0.334	0.323	0.242	0.801

Note: This table reports the MSE ratio of each micro-forecasting model relative to the *AR* benchmark, for each forecast horizon. The forecast target is the U.S. real GDP component growth rate, while the predictors are individual firm accounting earnings. The out-of-sample period is 2000Q1-2023Q4. The “Std.” column shows the standard deviation of the GDP component growth rates. The “Avg.” column shows the average MSE ratio across horizons. The MSE ratio for the best-performing model at each horizon is highlighted in bold.

Moreover, a sector-level variable importance analysis (Table C.14) shows that earnings from *manufacturing*, *finance*, and *transportation* are consistently most influential, though their weights vary. *Manufacturing* earnings broadly affect net exports (29.96%), consump-

tion, and government spending ($\approx 27\%$ each); *finance* earnings mainly drive investment (20.96%) versus other components; *retail* earnings impact government spending (5.58%) and consumption (4.41%) but little else.

These patterns reflect transmission channels: *manufacturing* signals production capacity and competitiveness, *finance* underpins credit and capital allocation, and *retail* captures household demand and fiscal spending. Overall, sector-level earnings align with the structural roles of these sectors in shaping aggregate output.

Table C.14: Sectoral Importance Analysis for GDP Component Forecasts

Net Export		Investment		Gov. Spending		Consumption	
Sector	Importance	Sector	Importance	Sector	Importance	Sector	Importance
Manufacturing	29.96%	Manufacturing	28.93%	Manufacturing	26.47%	Transportation	27.66%
Transportation	22.37%	Finance	20.96%	Transportation	24.52%	Manufacturing	27.22%
Finance	19.52%	Transportation	20.96%	Finance	17.71%	Finance	15.22%
Services	12.76%	Services	13.15%	Services	12.92%	Services	11.70%
Mining	4.19%	Mining	4.67%	Retail	5.58%	Retail	4.41%

Note: This table reports the results of the sectoral importance analysis using the micro-forecasting approach to forecast GDP components. The results are based on the best-performing *RF* model. Specifically, all firms are grouped into sectors according to their SIC codes, and the share of variable importance of each sector is calculated. The four GDP components considered are “Net Exports”, “Investment”, Government Spending (“Gov. Spending”), and “Consumption”. For each component, the first column lists the top five sectors, and the second column reports their corresponding shares of sectoral variable importance.

C.3.4 Forecasting Results Using Another Type of Accounting Variable

We use another accounting variables – micro-level incomes before extraordinary items (IB) – to forecast U.S. GDP growth rate. The MSE ratios and values are reported in Tables C.15 and C.16, respectively.

Table C.15: Robustness Check: Using Firm-Level IB (MSE Ratio)

Micro-Forecasting Model	Forecast Horizon				Avg.
	h=1	h=2	h=4	h=8	
<i>adaLASSO</i>	0.490	0.932	0.843	0.817	0.770
<i>LASSO</i>	0.475	0.896	0.846	0.821	0.760
<i>Elastic Net</i>	0.501	0.899	0.930	0.822	0.788
<i>Ridge</i>	0.480	0.801	0.898	0.865	0.761
<i>RF</i>	0.505	0.721	0.863	0.781	0.717
<i>GBRT</i>	0.628	0.796	0.832	0.808	0.766

Note: This table reports the MSE ratio of each micro-forecasting model relative to the *AR* benchmark, for each forecast horizon. The forecast target is the U.S. real GDP growth rate, while the predictors are individual firm incomes before extraordinary items (IB). The out-of-sample period is 2000Q1-2023Q4. The “Avg.” column shows the average MSE ratio across horizons. The MSE ratio for the best-performing model at each horizon is highlighted in bold.

Table C.16: Robustness Check: Using Firm-Level IB (MSE)

Micro-Forecasting Model	Forecast Horizon				Avg.
	h=1	h=2	h=4	h=8	
Panel (a): Benchmark					
<i>AR</i>	4.061	4.802	7.054	14.125	7.510
Panel (b): Micro-Forecasting Models					
<i>adaLASSO</i>	1.991	4.474	5.944	11.544	5.988
<i>LASSO</i>	1.929	4.300	5.970	11.601	5.950
<i>Elastic Net</i>	2.035	4.318	6.559	11.606	6.129
<i>Ridge</i>	1.950	3.845	6.334	12.224	6.088
<i>RF</i>	2.049	3.461	6.087	11.036	5.659
<i>GBRT</i>	2.552	3.824	5.867	11.413	5.914

Note: This table reports the MSE of each model for each forecast horizon. The forecast target is the U.S. real GDP growth rate, while the predictors are individual firm incomes before extraordinary items (IB). Panels (a) and (b) show the results of the benchmark *AR* model and micro-forecasting models, respectively. The out-of-sample period is 2000Q1-2023Q4. The “Avg.” column shows the average MSE across horizons. The MSE for the best-performing model at each horizon is highlighted in bold.

C.3.5 Forecasting Results for Other Countries

We examine the effectiveness of our micro-forecasting approach on two other OECD countries, U.K. and Japan. The accounting data of these countries are sourced from the *S&P Capital IQ* database. The real GDP growth rates are calculated as the accumulated change rate of real GDP obtained from the *FRED* database. Due to data limitations, the full sample periods for U.K. and Japan are 1998Q1-2023Q4 and 1997Q1-2023Q4, respectively. The out-of-sample starting points of both countries are 2006Q1. The MSE ratios of these countries are presented in Tables C.17 and C.18, while the MSE values are reported in Tables C.19 and C.20.

Table C.17: Robustness Check: Forecasting U.K. Real GDP Growth (MSE Ratio)

Micro-Forecasting Model	Forecast Horizon				Avg.
	h=1	h=2	h=4	h=8	
<i>adaLASSO</i>	0.499	0.419	1.128	0.899	0.736
<i>LASSO</i>	0.498	0.420	1.115	0.893	0.732
<i>Elastic Net</i>	0.465	0.400	1.293	0.891	0.762
<i>Ridge</i>	0.478	0.407	0.975	0.875	0.684
<i>RF</i>	0.455	0.381	0.934	0.875	0.661
<i>GBRT</i>	0.493	0.409	0.855	0.889	0.662

Note: This table reports the MSE ratio of each micro-forecasting model relative to the *AR* benchmark, for each forecast horizon. The forecast target is the U.K. real GDP growth rate, while the predictors are individual firm accounting earnings. The out-of-sample period is 2006Q1-2023Q4. The “Avg.” column shows the average MSE ratio across horizons. The MSE ratio for the best-performing model at each horizon is highlighted in bold.

Table C.18: Robustness Check: Forecasting Japan Real GDP Growth (MSE Ratio)

Micro-Forecasting Model	Forecast Horizon				Avg.
	h=1	h=2	h=4	h=8	
<i>adaLASSO</i>	0.658	0.997	0.579	0.743	0.744
<i>LASSO</i>	0.610	1.033	0.581	0.742	0.741
<i>Elastic Net</i>	0.704	0.851	0.736	0.825	0.779
<i>Ridge</i>	0.623	0.905	0.601	0.674	0.701
<i>RF</i>	0.598	0.673	0.559	0.713	0.636
<i>GBRT</i>	0.704	0.903	0.559	0.720	0.722

Note: This table reports the MSE ratio of each micro-forecasting model relative to the *AR* benchmark, for each forecast horizon. The forecast target is the Japan real GDP growth rate, while the predictors are individual firm accounting earnings. The out-of-sample period is 2006Q1-2023Q4. The “Avg.” column shows the average MSE ratio across horizons. The MSE ratio for the best-performing model at each horizon is highlighted in bold.

Table C.19: Robustness Check: Forecasting U.K. Real GDP Growth (MSE)

Micro-Forecasting Model	Forecast Horizon				Avg.
	h=1	h=2	h=4	h=8	
Panel (a): Benchmark					
<i>AR</i>	30.636	58.483	38.822	55.522	45.866
Panel (b): Micro-Forecasting Models					
<i>adaLASSO</i>	15.284	24.497	43.774	49.930	33.371
<i>LASSO</i>	15.248	24.569	43.292	49.604	33.178
<i>Elastic Net</i>	14.244	23.379	50.215	49.464	34.326
<i>Ridge</i>	14.657	23.794	37.839	48.571	31.215
<i>RF</i>	13.949	22.270	36.257	48.574	30.263
<i>GBRT</i>	15.106	23.940	33.185	49.362	30.398

Note: This table reports the MSE of each model for each forecast horizon. The forecast target is the U.K. real GDP growth rate, while the predictors are individual firm accounting earnings. Panels (a) and (b) show the results of the benchmark *AR* model and micro-forecasting models, respectively. The out-of-sample period is 2006Q1-2023Q4. The “Avg.” column shows the average MSE across horizons. The MSE for the best-performing model at each horizon is highlighted in bold.

Table C.20: Robustness Check: Forecasting Japan Real GDP Growth (MSE)

Micro-Forecasting Model	Forecast Horizon				Avg.
	h=1	h=2	h=4	h=8	
Panel (a): Benchmark					
<i>AR</i>	6.098	9.529	17.786	25.385	14.699
Panel (b): Micro-Forecasting Models					
<i>adaLASSO</i>	4.014	9.498	10.297	18.864	10.668
<i>LASSO</i>	3.719	9.839	10.327	18.839	10.681
<i>Elastic Net</i>	4.293	8.106	13.084	20.943	11.606
<i>Ridge</i>	3.801	8.627	10.697	17.112	10.059
<i>RF</i>	3.645	6.412	9.946	18.088	9.523
<i>GBRT</i>	4.296	8.609	9.942	18.288	10.284

Note: This table reports the MSE of each model for each forecast horizon. The forecast target is the Japan real GDP growth rate, while the predictors are individual firm accounting earnings. Panels (a) and (b) show the results of the benchmark *AR* model and micro-forecasting models, respectively. The out-of-sample period is 2006Q1-2023Q4. The ‘‘Avg.’’ column shows the average MSE across horizons. The MSE for the best-performing model at each horizon is highlighted in bold.

C.3.6 Results for Alternative Evaluation Metrics

In this section, we provide supplementary metrics for evaluating the performance of the micro-forecasting approach.

First, as a complement to the MSE ratios relative to the *AR* model, we compute the out-of-sample R^2 relative to the *AR* model (Campbell and Thompson 2008):

$$R_{oos}^2 = 1 - \frac{\sum_{t=1}^{T-h} (g_{t+h} - \hat{g}_{t+h})^2}{\sum_{t=1}^{T-h} (g_{t+h} - \hat{g}_{t+h}^{AR})^2} \quad (\text{C.9})$$

where T denotes the length of the out-of-sample period and $T-h$ is the number of available forecasts at horizon h , \hat{g}_{t+h} denotes the predicted GDP growth rate by micro-forecasting approach, \hat{g}_{t+h}^{AR} denotes the the predicted GDP growth rate by *AR* model. **A positive $oosR^2$ indicates that our micro-forecasting model outperforms the benchmark, and higher values reflect better predictive performance.** The results are shown in Table C.21.

Second, we examine whether our micro-forecasting approach can detect the sign of GDP

growth using directional accuracy (DA):

$$DA = \frac{1}{T_{oos} - h} \sum_{t=1}^{T_{oos}-h} \mathbf{1}_{\{\hat{g}_{t+h}, g_{t+h} > 0\}} \quad (\text{C.10})$$

A higher DA indicates a greater accuracy in predicting the direction of growth.

The results are shown in Table C.22.

Third, we use the area under the Receiver Operating Characteristic (ROC) curve (AUC) to evaluate the ability of the micro-forecasting approach to classify the sign of GDP growth. The ROC curve plots the True Positive Rate (TPR) against the False Positive Rate (FPR) for all possible classification thresholds (Bradley 1997):

$$TPR(\tau) = \frac{TP(\tau)}{TP(\tau) + FN(\tau)}, FPR(\tau) = \frac{FP(\tau)}{FP(\tau) + TN(\tau)} \quad (\text{C.11})$$

where TP , FP , TN , FN denote the numbers of true positives, false positives, true negatives, and false negatives at threshold τ . Since presenting all ROC curves would be excessive, we summarize the results by AUC. An AUC of 0.5 indicates no discriminative ability, equivalent to random guessing, whereas an AUC between 0.5 and 1 reflects increasing predictive performance—the closer the AUC is to 1, the better the model is at distinguishing between positive and negative classes. The results are reported in Table C.23.

Last, we implement the Hansen and Nason’s (2011) Model Confidence Set (MCS) test to evaluate the relative predictive accuracy of competing models. Let $L_{i,t}$ denote the loss of model i at time t (e.g., squared forecast error). The MCS procedure tests the null hypothesis of equal predictive ability across models and sequentially eliminates inferior models until the remaining set cannot be statistically distinguished at a given confidence level. The results of MCS is shown in Table C.24.

Table C.21: Robustness Check: U.S. Real GDP Growth ($oosR^2$)

Micro-Forecasting Model	Forecast Horizon				Avg.
	h=1	h=2	h=4	h=8	
<i>adaLASSO</i>	0.496	0.107	0.115	0.105	0.206
<i>LASSO</i>	0.497	0.097	0.126	0.116	0.209
<i>Elastic Net</i>	0.497	0.113	0.065	0.102	0.194
<i>Ridge</i>	0.448	0.172	0.102	0.122	0.211
<i>RF</i>	0.493	0.267	0.125	0.232	0.279
<i>GBRT</i>	0.326	0.178	0.157	0.192	0.213

Note: This table reports the out-of-sample R-squared ($oosR^2$) of each micro-forecasting model relative to the *AR* benchmark, for each forecast horizon. The forecast target is the U.S. real GDP growth rate, while the predictors are individual firm accounting earnings. The out-of-sample period is 2000Q1-2023Q4. The “Avg.” column shows the average $oosR^2$ across horizons. The $oosR^2$ for the best-performing model at each horizon is highlighted in bold.

Table C.22: Robustness Check: U.S. Real GDP Growth (Directional Accuracy)

Micro-Forecasting Model	Forecast Horizon				Avg.
	h=1	h=2	h=4	h=8	
<i>AR</i>	0.853	0.777	0.848	0.909	0.847
<i>adaLASSO</i>	0.895	0.691	0.880	0.909	0.844
<i>LASSO</i>	0.895	0.713	0.880	0.909	0.849
<i>Elastic Net</i>	0.895	0.713	0.880	0.909	0.849
<i>Ridge</i>	0.884	0.777	0.870	0.909	0.860
<i>RF</i>	0.937	0.809	0.913	0.909	0.892
<i>GBRT</i>	0.905	0.766	0.913	0.909	0.873

Note: This table reports the directional accuracy (DA) of each micro-forecasting model relative to the *AR* benchmark, for each forecast horizon. The forecast target is the U.S. real GDP growth rate, while the predictors are individual firm accounting earnings. The out-of-sample period is 2000Q1-2023Q4. The “Avg.” column shows the average DA across horizons. The DA for the best-performing model at each horizon is highlighted in bold.

Table C.23: Robustness Check: U.S. Real GDP Growth (Area under ROC Curve)

Micro-Forecasting Model	Forecast Horizon				Avg.
	h=1	h=2	h=4	h=8	
<i>AR</i>	0.897	0.850	0.637	0.525	0.727
<i>adaLASSO</i>	0.926	0.884	0.618	0.600	0.757
<i>LASSO</i>	0.926	0.888	0.629	0.616	0.765
<i>Elastic Net</i>	0.924	0.888	0.545	0.592	0.737
<i>Ridge</i>	0.941	0.882	0.628	0.653	0.776
<i>RF</i>	0.916	0.888	0.662	0.644	0.778
<i>GBRT</i>	0.918	0.891	0.656	0.620	0.771

Note: This table reports the area under the ROC curve (AUC) of each micro-forecasting model relative to the *AR* benchmark, for each forecast horizon. The forecast target is the U.S. real GDP growth rate, while the predictors are individual firm accounting earnings. The out-of-sample period is 2000Q1-2023Q4. The “Avg.” column shows the average AUC across horizons. The AUC for the best-performing model at each horizon is highlighted in bold.

Table C.24: Robustness Check: U.S. Real GDP Growth (Model Confidence Set)

Model	Forecast Horizon			
	h=1	h=2	h=4	h=8
<i>AR</i>				
<i>adaLASSO</i>				
<i>LASSO</i>	*		*	
<i>Elastic Net</i>	*			
<i>Ridge</i>		*		*
<i>RF</i>	*	*	*	*
<i>GBRT</i>		*	*	*

Note: This table reports the forecasting performance of micro-forecasting model ranking based on the model confidence set (MCS) procedure, among the *AR* benchmark and micro-forecasting models for U.S. GDP growth rate forecast. The top three best-performing models at each forecast horizon is marked by a star. The best-performing model is highlighted in bold.

C.3.7 Changing the Number of Pre-Selected Firms

We increase the number of pre-selected firms to 2,000 and also consider the full sample for forecasting. The Diebold-Mariano test is conducted to compare the predictive performance of the 1,500-firm model with the 2,000-firm and all-firm models. The results are reported in Table C.25.

Table C.25: Robustness Check: Changing the Number of Pre-Selected Firms (DM Test Statistics)

Micro-Forecasting Model	Forecast Horizon			
	h=1	h=2	h=4	h=8
2,000 Firms	-0.442	-0.469	-1.129	-1.696
All Firms	-2.342	-2.964	-0.727	-1.387

Note: This table reports the [Diebold and Mariano \(2002\)](#) test statistics for the forecasting performance difference, for each forecast horizon. All results are based on best-performing *RF* model. Row “2,000 Firms” shows the results between 2,000-firm model and 1,500-firm model. The null hypothesis is H_0 : The two models have the same prediction accuracy and the alternative hypothesis is H_1 : The 2,000 – firm model predicts more accurately than 1,500 – firm model. Row “All Firms” shows the results between all-firm model and 1,500-firm model. The null hypothesis is H_0 : The two models have the same prediction accuracy and the alternative hypothesis is H_1 : The all – firm model predicts more accurately than 1,500 – firm model. All p -values are greater than 10%.

C.4 Additional Results for Variable Importance Analysis

C.4.1 Details on Calculating Variable Importance

We employ the best-performing *RF* model. The calculation of variable importance at each out-of-sample point t proceeds as follows:

- For each predictor x_j ($j = 1, \dots, 3005$), the importance is computed as the total reduction in node impurity (variance reduction) that it contributes across all splits in the forest. Specifically, whenever x_j is used to split a node t , we record the decrease in impurity, defined as $\Delta I(t, x_j) = I(t) - (\frac{n_L}{n_t} I(t_L) + \frac{n_R}{n_t} I(t_R))$, where $I(\cdot)$ denotes the impurity measure (mean squared error), n_t is the number of samples in node t , and t_L, t_R are its left and right child nodes. The raw importance of x_j is then obtained by summing $\Delta I(t, x_j)$ over all nodes and all trees in the forest.

- The raw importance is normalized by dividing through their sum, such that $\sum_j VI_j = 1$ at each out-of-sample point.
- For each firm $i, i = 1, \dots, 1500$, the variable importance of accounting earnings is defined as the sum of the current and lagged terms.
- The importance of the AR terms is obtained by summing the variable importance of all AR terms.

In the sectoral analysis (Section 7.1 in the main manuscript), we categorize all firms into eleven groups according to the Standard Industrial Classification (SIC) code and further include the AR terms, resulting in twelve categories: (1) agriculture, forestry, and fishing (*agriculture*); (2) *mining*; (3) *construction*; (4) *manufacturing*; (5) transportation and public utilities (*transportation*); (6) wholesale trade (*wholesale*); (7) retail trade (*retail*); (8) finance, insurance, and real estate (*finance*); (9) *services*; (10) *public administration*; and (11) *AR terms*. Within each sector, we sum the variable importance across all firms.

C.4.2 Additional Results for Analysis of Supply-Chain Centrality and GDP Growth Rate Forecasting

We then complement the sectoral variable importance analysis in Section 7.1 in the main manuscript by presenting scatter plots of the relationship between sectoral variable importance and supply-chain centrality (Figure C.7).

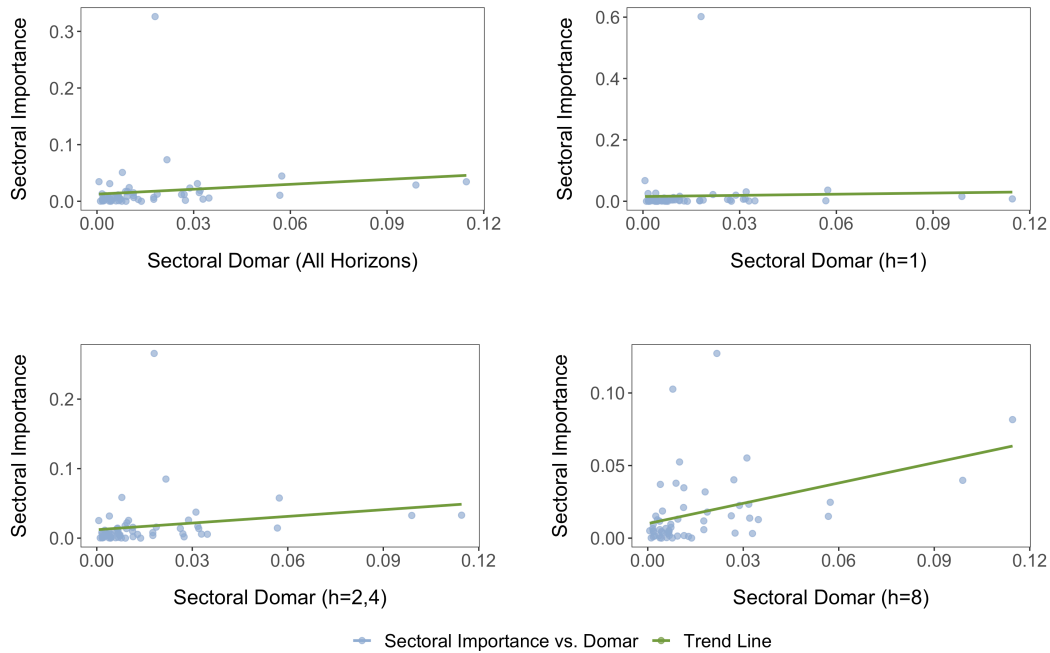


Figure C.7: Relationship Between Sectoral Variable Importance and Domar Weight.

Note: These figures present scatter plots and fitted trend lines between sectoral variable importance and Domar weight at each forecast horizon.

C.4.3 Additional Results for Relationship Between Firm-Level Variable Importance and Firm Characteristics

This section reports the additional results of relationship between firm-level variable importance and firm characteristics across forecast horizons that are not presented in Section 7.2 of the main manuscript. Tables C.26 to C.29 show detailed results focusing on the contribution of each single firm characteristic variable. In addition, we also examine the results using the logarithm of variable importance, which are presented in Table C.30.

Table C.26: Firm-Level Variable Importance and Firm Characteristics Inspections (All Forecast Horizons, $h = 1, 2, 4, 8$)

	(1)	(2)	(3)	(4)	(5)
Age	0.235*** (0.0136)	0.228*** (0.0137)	0.240*** (0.0139)	0.241*** (0.0139)	0.239*** (0.0151)
Age Squared	-0.00151*** (0.000105)	-0.00149*** (0.000105)	-0.00155*** (0.000107)	-0.00150*** (0.000106)	-0.00157*** (0.000116)
Assets (log)		0.178*** (0.0270)			0.0354 (0.0310)
Sales Growth			-0.370 (0.252)		
Investment Rate				0.980* (0.537)	
BM Ratio				-0.243*** (0.0633)	
Tobin's q				0.0904*** (0.0288)	
Debt/Assets					-1.29*** (0.218)
Dummy: Paying Dividend					1.94*** (0.156)
Tangible/Assets					3.82*** (0.399)
Observations	380,585	380,433	366,298	314,118	344,359
R-squared	0.024	0.025	0.025	0.019	0.026
Time FE	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes

Note: This table reports the results of regressing firm-level relative variable importance ($\times 10^4$) on firm characteristics over all forecast horizons. The sample includes those firms with positive forecasting importance (for any value of the forecast horizon h). Standard errors are presented in parentheses. “***”, “**” and “*” indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table C.27: Firm-Level Variable Importance and Firm Characteristics Inspections ($h = 1$)

	(1)	(2)	(3)	(4)	(5)
Age	0.447*** (0.0329)	0.424*** (0.0330)	0.453*** (0.0336)	0.430*** (0.0317)	0.427*** (0.0360)
Age Squared	-0.00269*** (0.000255)	-0.00259*** (0.000255)	-0.00274*** (0.000260)	-0.00252*** (0.000243)	-0.00258*** (0.000277)
Assets (log)		0.642*** (0.0683)			0.430*** (0.0785)
Sales Growth			-0.258 (0.629)		
Investment Rate				-1.64 (1.37)	
BM Ratio				-0.276* (0.151)	
Tobin's q				0.123 (0.0752)	
Debt/Assets					-4.10*** (0.592)
Dummy: Paying Dividend					2.66*** (0.377)
Tangible/Assets					8.86*** (0.991)
Observations	101,988	101,942	98,963	83,539	93,604
R-squared	0.061	0.062	0.061	0.049	0.063
Time FE	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes

Note: This table reports the results of regressing firm-level relative variable importance ($\times 10^4$) on firm characteristics at forecast horizon $h = 1$. The sample includes those firms with positive forecasting importance. Standard errors are presented in parentheses. “***”, “**” and “*” indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table C.28: Firm-Level Variable Importance and Firm Characteristics Inspections ($h = 2, 4$)

	(1)	(2)	(3)	(4)	(5)
Age	0.160*** (0.0160)	0.155*** (0.0161)	0.166*** (0.0164)	0.172*** (0.0171)	0.163*** (0.0183)
Age Squared	-0.00108*** (0.000122)	-0.00106*** (0.000122)	-0.00111*** (0.000125)	-0.00112*** (0.000129)	-0.00114*** (0.000140)
Assets (log)		0.137*** (0.0322)			-0.0282 (0.0374)
Sales Growth			-0.256 (0.298)		
Investment Rate				0.755 (0.663)	
BM Ratio				-0.0999 (0.0791)	
Tobin's q				0.0595* (0.0352)	
Debt/Assets					-0.982*** (0.259)
Dummy: Paying Dividend					2.54*** (0.192)
Tangible/Assets					2.55*** (0.487)
Observations	200,100	200,030	192,089	164,611	180,212
R-squared	0.021	0.021	0.021	0.018	0.022
Time FE	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes

Note: This table reports the results of regressing firm-level relative variable importance ($\times 10^4$) on firm characteristics at forecast horizon $h = 2$ and 4. The sample includes those firms with positive forecasting importance (for any value of the forecast horizon h). Standard errors are presented in parentheses. “***”, “**” and “*” indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table C.29: Firm-Level Variable Importance and Firm Characteristics Inspections ($h = 8$)

	(1)	(2)	(3)	(4)	(5)
Age	0.0748*** (0.0283)	0.0801*** (0.0284)	0.0829*** (0.0286)	0.0921*** (0.0298)	0.0981*** (0.0287)
Age Squared	-0.000584*** (0.000223)	-0.000601*** (0.000223)	-0.000636*** (0.000225)	-0.000679*** (0.000232)	-0.000722*** (0.000221)
Assets (log)		-0.214*** (0.0509)			-0.183*** (0.0543)
Sales Growth			-0.793 (0.485)		
Investment Rate				3.83*** (0.990)	
BM Ratio				-0.467*** (0.123)	
Tobin's q				0.0649 (0.0530)	
Debt/Assets					0.749** (0.363)
Dummy: Paying Dividend					0.0202 (0.274)
Tangible/Assets					-0.366 (0.692)
Observations	78,497	78,461	75,246	65,968	70,543
R-squared	0.002	0.002	0.002	0.003	0.003
Time FE	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes

Note: This table reports the results of regressing firm-level relative variable importance ($\times 10^4$) on firm characteristics at forecast horizon $h = 8$. The sample includes those firms with positive forecasting importance. Standard errors are presented in parentheses. “***”, “**” and “*” indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table C.30: Firm-Level Variable Importance (in Logs) and Firm Characteristics Inspections (Comparison over Horizons)

	(1) All h	(2) h=1	(3) h=2,4	(4) h=8
Age	0.224*** (0.0145)	0.378*** (0.0325)	0.158*** (0.0183)	0.102*** (0.0305)
Age Squared	-0.00143*** (0.000111)	-0.00221*** (0.000248)	-0.00107*** (0.000138)	-0.000741*** (0.000235)
Assets (log)	0.0627* (0.0329)	0.414*** (0.0773)	0.0386 (0.0416)	-0.242*** (0.0632)
Sales Growth	-0.333 (0.265)	-0.113 (0.634)	-0.186 (0.331)	-0.903* (0.509)
Investment Rate	2.60*** (0.586)	1.17 (1.47)	1.97*** (0.732)	4.67*** (1.06)
BM Ratio	-0.357*** (0.0690)	-0.493*** (0.161)	-0.157* (0.0874)	-0.541*** (0.132)
Tobin's q	0.217*** (0.0330)	0.469*** (0.0834)	0.159*** (0.0412)	-0.0360 (0.0597)
Debt/Assets	-1.93*** (0.259)	-5.61*** (0.636)	-1.21*** (0.325)	0.554 (0.476)
Dummy:Paying Dividend	2.69*** (0.161)	4.00*** (0.372)	3.03*** (0.207)	-0.0506 (0.305)
Tangible/Assets	2.61*** (0.409)	7.09*** (0.946)	0.991* (0.520)	0.222 (0.782)
Observations	291,458	78,524	151,990	60,944
R-squared	0.021	0.055	0.020	0.004
Time FE	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes

Note: This table reports the results of regressing firm-level relative variable importance ($\times 10^4$) on firm characteristics over all forecast horizons. Firm-level variable importance is transformed in logs (defined as $\log(1 + y)$). The sample includes those firms with positive forecasting importance (for any value of the forecast horizon h). Standard errors are presented in parentheses. “***”, “**” and “*” indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

C.4.4 Additional Results for Variable Importance Analysis Based on SHAP

In the main result, the importance of predictors is measured using the impurity-based approach. To validate these results, we assess predictor importance using an alternative method, SHAP (Rozemberczki et al. 2022). The results at the sectoral level are reported in Figure C.8 and Table C.31, while firm-level results are presented in Tables C.32-C.37.

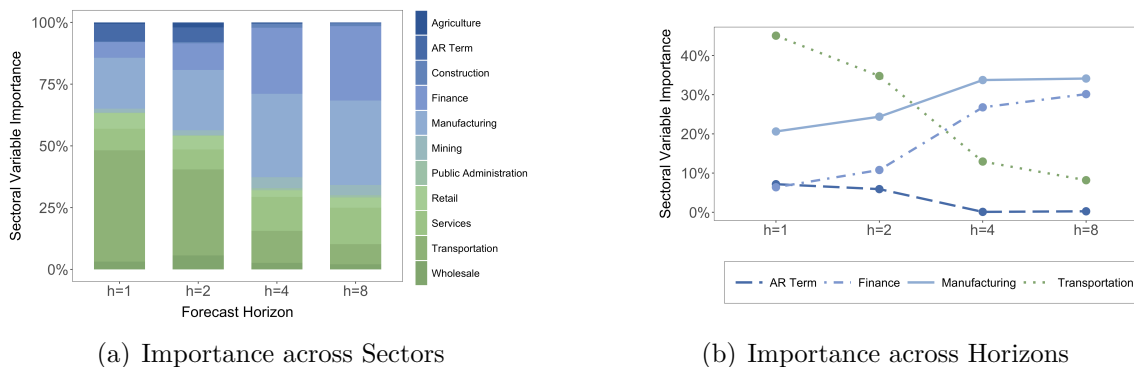


Figure C.8: Sectoral Importance for U.S. GDP Forecast over Sectors and Horizons: *RF*, SHAP-Based.

Note: The figures report the sectoral variable importance analysis results based on *RF* and SHAP method. Panel (a) shows the importance proportions of all sectors over all forecast horizons, while Panel (b) illustrates how the importance of the top three sectors and *AR Term* changes over the forecast horizons.

Table C.31: Sectoral Importance for U.S. GDP Growth Forecast: *RF*, SHAP-Based.

Sector	Forecast Horizon				Avg.
	h=1	h=2	h=4	h=8	
Manufacturing	20.62%	24.39%	33.74%	34.14%	28.74%
Transportation	45.07%	34.78%	12.94%	8.18%	26.89%
Finance	6.38%	10.78%	26.78%	30.15%	15.97%
Services	8.76%	8.15%	13.80%	14.70%	11.32%
Retail	6.32%	5.48%	2.70%	4.13%	4.51%
AR Term	3.11%	5.66%	2.60%	2.07%	3.80%
Wholesale	7.16%	5.90%	0.09%	0.24%	3.58%
Mining	1.57%	2.07%	4.43%	4.17%	3.01%
Construction	0.21%	0.49%	1.62%	1.10%	0.79%
Agriculture	0.59%	2.04%	0.42%	0.17%	0.73%
Public Administration	0.22%	0.25%	0.87%	0.96%	0.66%

Note: The table reports the detailed results of sectoral variable importance analysis based on best-performing *ML* model *RF* and SHAP-based method. The “Avg.” column represents the mean variable importance across forecast horizons. The three most important sectors are highlighted in bold. The intensity of the green color indicates the level of sectoral importance, with darker shades representing higher percentages.

Table C.32: Firm-Level Variable Importance and Firm Characteristics Inspections (Comparison across Horizons, SHAP-Based)

	(1) All h	(2) h=1	(3) h=2,4	(4) h=8
Age	0.238*** (0.0132)	0.420*** (0.0288)	0.168*** (0.0169)	0.0907*** (0.0284)
Age Squared	-0.00149*** (0.000101)	-0.00235*** (0.000219)	-0.00113*** (0.000128)	-0.000706*** (0.000219)
Assets (log)	0.0920*** (0.0300)	0.373*** (0.0684)	0.0423 (0.0384)	-0.0554 (0.0589)
Sales Growth	-0.202 (0.241)	0.777 (0.561)	-0.327 (0.306)	-1.03** (0.474)
Investment Rate	1.24** (0.533)	-0.204 (1.30)	0.870 (0.677)	2.80*** (0.992)
BM Ratio	-0.363*** (0.0627)	-0.539*** (0.142)	-0.222*** (0.0808)	-0.370*** (0.123)
Tobin's q	0.201*** (0.0300)	0.443*** (0.0738)	0.132*** (0.0381)	0.0131 (0.0556)
Debt/Assets	-2.17*** (0.235)	-5.80*** (0.563)	-1.28*** (0.301)	-0.300 (0.444)
Dummy: Paying Dividend	2.69*** (0.147)	3.56*** (0.329)	3.06*** (0.192)	0.514* (0.284)
Tangible/Assets	1.63*** (0.372)	5.41*** (0.837)	0.476 (0.481)	-0.747 (0.729)
Observations	291,452	78,522	151,986	60,944
R-squared	0.020	0.049	0.020	0.003
Time FE	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes

Note: This table reports the results of regressing firm-level relative variable importance ($\times 10^4$) on firm characteristics over forecast horizons. The variable importance is calculated based on SHAP. The sample includes those firms with positive forecasting importance (for any value of the forecast horizon h). Standard errors are presented in parentheses. “***”, “**” and “*” indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table C.33: Firm-Level Variable Importance and Firm Characteristics Inspections (All Forecast Horizons, SHAP-Based)

	(1)	(2)	(3)	(4)	(5)
Age	0.241*** (0.0119)	0.233*** (0.0120)	0.244*** (0.0122)	0.245*** (0.0123)	0.252*** (0.0132)
Age Squared	-0.00151*** (0.0000919)	-0.00148*** (0.0000920)	-0.00154*** (0.0000938)	-0.00150*** (0.0000940)	-0.00161*** (0.000101)
Assets (log)		0.215*** (0.0237)			0.0727*** (0.0271)
Sales Growth			-0.408* (0.221)		
Investment Rate				0.165 (0.477)	
BM Ratio				-0.245*** (0.0562)	
Tobin's q				0.0510** (0.0255)	
Debt/Assets					-1.46*** (0.190)
Dummy: Paying Dividend					1.98*** (0.136)
Tangible/Assets					2.92*** (0.348)
Observations	380,578	380,426	366,291	314,111	344,353
R-squared	0.023	0.023	0.023	0.017	0.025
Time FE	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes

Note: This table reports the results of regressing firm-level relative variable importance ($\times 10^4$) on firm characteristics over all forecast horizons. The variable importance is calculated based on SHAP. The sample includes those firms with positive forecasting importance (for any value of the forecast horizon h). Standard errors are presented in parentheses. “***”, “**” and “*” indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table C.34: Firm-Level Variable Importance and Firm Characteristics Inspections ($h = 1$, SHAP-Based)

	(1)	(2)	(3)	(4)	(5)
Age	0.448*** (0.0274)	0.428*** (0.0274)	0.451*** (0.0279)	0.444*** (0.0273)	0.449*** (0.0300)
Age Squared	-0.00258*** (0.000212)	-0.00249*** (0.000212)	-0.00261*** (0.000216)	-0.00249*** (0.000209)	-0.00259*** (0.000230)
Assets (log)		0.546*** (0.0568)			0.369*** (0.0653)
Sales Growth			-0.133 (0.523)		
Investment Rate				-1.87 (1.18)	
BM Ratio				-0.291** (0.130)	
Tobin's q				0.111* (0.0647)	
Debt/Assets					-4.06*** (0.492)
Dummy: Paying Dividend					2.35*** (0.313)
Tangible/Assets					6.74*** (0.824)
Observations	101,985	101,939	98,960	83,536	93,602
R-squared	0.056	0.057	0.056	0.044	0.058
Time FE	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes

Note: This table reports the results of regressing firm-level relative variable importance ($\times 10^4$) on firm characteristics at forecast horizon $h = 1$. The variable importance is calculated based on SHAP. The sample includes those firms with positive forecasting importance. Standard errors are presented in parentheses. “***”, “**” and “*” indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table C.35: Firm-Level Variable Importance and Firm Characteristics Inspections ($h = 2, 4$, SHAP-Based)

	(1)	(2)	(3)	(4)	(5)
Age	0.178*** (0.0146)	0.172*** (0.0146)	0.182*** (0.0149)	0.181*** (0.0155)	0.179*** (0.0166)
Age Squared	-0.00118*** (0.000111)	-0.00116*** (0.000111)	-0.00121*** (0.000114)	-0.00117*** (0.000117)	-0.00123*** (0.000126)
Assets (log)		0.159*** (0.0292)			-0.00799 (0.0339)
Sales Growth			-0.328 (0.271)		
Investment Rate				-0.100 (0.600)	
BM Ratio				-0.173** (0.0715)	
Tobin's q				0.0131 (0.0318)	
Debt/Assets					-1.07*** (0.234)
Dummy: Paying Dividend					2.54*** (0.173)
Tangible/Assets					2.19*** (0.441)
Observations	200,096	200,026	192,085	164,607	180,208
R-squared	0.021	0.022	0.022	0.018	0.023
Time FE	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes

Note: This table reports the results of regressing firm-level relative variable importance ($\times 10^4$) on firm characteristics at forecast horizon $h = 2$ and 4. The variable importance is calculated based on SHAP. The sample includes those firms with positive forecasting importance (for any value of the forecast horizon h). Standard errors are presented in parentheses. “***”, “**” and “*” indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table C.36: Firm-Level Variable Importance and Firm Characteristics Inspections ($h = 8$, SHAP-Based)

	(1)	(2)	(3)	(4)	(5)
Age	0.0542** (0.0261)	0.0534** (0.0262)	0.0559** (0.0265)	0.0728*** (0.0272)	0.0893*** (0.0263)
Age Squared	-0.000478** (0.000205)	-0.000475** (0.000205)	-0.000490** (0.000208)	-0.000575*** (0.000212)	-0.000712*** (0.000203)
Assets (log)		0.0237 (0.0469)			0.0157 (0.0499)
Sales Growth			-0.935** (0.450)		
Investment Rate				2.51*** (0.902)	
BM Ratio				-0.270** (0.112)	
Tobin's q				0.0222 (0.0483)	
Debt/Assets					0.0761 (0.334)
Dummy: Paying Dividend					0.465* (0.252)
Tangible/Assets					-0.712 (0.636)
Observations	78,497	78,461	75,246	65,968	70,543
R-squared	0.002	0.002	0.003	0.003	0.003
Time FE	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes

Note: This table reports the results of regressing firm-level relative variable importance ($\times 10^4$) on firm characteristics at forecast horizon $h = 8$. The variable importance is calculated based on SHAP. The sample includes those firms with positive forecasting importance. Standard errors are presented in parentheses. “***”, “**” and “*” indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table C.37: Firm-Level Variable Importance (in Logs) and Firm Characteristics Inspections (Comparison over Horizons, SHAP-Based)

	(1) All h	(2) h=1	(3) h=2,4	(4) h=8
Age	0.234*** (0.0129)	0.409*** (0.0281)	0.166*** (0.0166)	0.0908*** (0.0279)
Age Squared	-0.00146*** (0.0000985)	-0.00229*** (0.000214)	-0.00112*** (0.000125)	-0.000707*** (0.000215)
Assets (log)	0.0895*** (0.0293)	0.365*** (0.0667)	0.0416 (0.0377)	-0.0553 (0.0579)
Sales Growth	-0.198 (0.236)	0.761 (0.547)	-0.323 (0.300)	-1.01** (0.466)
Investment Rate	1.23** (0.522)	-0.175 (1.27)	0.840 (0.664)	2.80*** (0.975)
BM Ratio	-0.357*** (0.0614)	-0.527*** (0.139)	-0.218*** (0.0793)	-0.364*** (0.121)
Tobin's q	0.198*** (0.0294)	0.435*** (0.0719)	0.130*** (0.0374)	0.0123 (0.0547)
Debt/Assets	-2.13*** (0.231)	-5.69*** (0.549)	-1.27*** (0.295)	-0.289 (0.436)
Dummy: Paying Dividend	2.66*** (0.144)	3.52*** (0.321)	3.01*** (0.188)	0.503* (0.279)
Tangible/Assets	1.59*** (0.364)	5.31*** (0.816)	0.460 (0.472)	-0.746 (0.716)
Observations	291,452	78,522	151,986	60,944
R-squared	0.020	0.050	0.021	0.003
Time FE	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes

Note: This table reports the results of regressing firm-level relative variable importance ($\times 10^4$) on firm characteristics over all forecast horizons. The variable importance is calculated based on SHAP. Firm-level variable importance is transformed in logs (defined as $\log(1+y)$). The sample includes those firms with positive forecasting importance (for any value of the forecast horizon h). Standard errors are presented in parentheses. “***”, “**” and “*” indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

References

- Anesti, N., A. B. Galvão, and S. Miranda-Agrippino (2022). Uncertain Kingdom: Nowcasting Gross Domestic Product and Its Revisions. *Journal of Applied Econometrics* 37(1), 42–62.
- Breiman, L. (2001). Random Forests. *Machine Learning* 45(1), 5–32.
- 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.
- Cong, L. W., G. Feng, J. He, and X. He (2025). Growing the Efficient Frontier on Panel Trees. *Journal of Financial Economics* 167, 104024.
- Diebold, F. X. and R. S. Mariano (2002). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics* 20(1), 134–144.
- Feng, G., S. Giglio, and D. Xiu (2020). Taming the Factor Zoo: A Test of New Factors. *The Journal of Finance* 75(3), 1327–1370.
- Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics* 29(5), 1189–1232.
- Gu, S., B. Kelly, and D. Xiu (2020). Empirical Asset Pricing via Machine Learning. *Review of Financial Studies* 33(5), 2223–2273.
- Huang, D., F. Jiang, J. Tu, and G. Zhou (2015). Investor Sentiment Aligned: A Powerful Predictor of Stock Returns. *Review of Financial Studies* 28(3), 791–837.
- Kelly, B. T., S. Pruitt, and Y. Su (2019). Characteristics Are Covariances: A Unified Model of Risk and Return. *Journal of Financial Economics* 134(3), 501–524.
- Konchitchki, Y. and P. N. Patatoukas (2014). Accounting Earnings and Gross Domestic Product. *Journal of Accounting and Economics* 57(1), 76–88.
- Medeiros, M. C., G. F. Vasconcelos, Á. Veiga, and E. Zilberman (2021). Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods. *Journal of Business & Economic Statistics* 39(1), 98–119.
- Rossi, B. (2013). Exchange Rate Predictability. *Journal of Economic Literature* 51(4), 1063–1119.
- Tibshirani, R. (1996). Regression Shrinkage and Selection via the LASSO. *Journal of Royal Statistical Society: Series B (Methodological)* 58, 267–288.
- Timmermann, A. (2006). Forecast Combinations. *Handbook of Economic Forecasting* 1, 135–196.
- Zou, H. (2006). The Adaptive LASSO and Its Oracle Properties. *Journal of American Statistical Association* 101(476), 1418–1429.
- Zou, H. and T. Hastie (2005). Regularization and Variable Selection via the Elastic Net. *Journal of Royal Statistical Society Series B: Statistical Methodology* 67(2), 301–320.