# Forecasting Earnings from Early Announcers: A Latent Factor Approach

Zhenping Wang

Emory University

Nov, 2017

**Abstract**

I propose a new method to predict non-announcing firms' earnings using the cross section of all available early announcers' earnings, the number of which can be as large as thousands. The method assumes common latent factors driving the earnings of non-announcing firms and early announcers and thus efficiently reduces the dimension of announced earnings. Empirical tests show that the extracted measure strongly predicts earnings surprise and earnings announcement return with both statistical and economic significance. A long-short trading strategy based on the extracted information realizes a 15% alpha annually, indicating a delayed reaction of investors. Large firms incorporate the news from early announcers faster than small firms. Controlling a series of documented information channels has little impact on the predictive power of this extracted measure.

Thousands of firms release their earnings quarterly for the same operating periods but on different dates. How can we update earnings expectations for late announcers by learning from numerous announced earnings when the number of early announcers is often relatively larger than the length of the earnings time series? The literature usually simplifies all announced earnings to an aggregated measure such as the averaged value. Instead, I apply a new method to more effectively extract relevant information for the earnings of each non-announcing firm from the cross section of early announcers' earnings. This method takes into consideration the possibility that each predictor can be associated in a different way with the forecasting target, while such associations may possess commonality. Empirical tests show that the extracted measure forecasts earnings surprise and announcement return with both statistical and economic significance.

Using all relevant information in this context presents challenges when the number of predictors is relatively larger than the length of a given time series. A simple solution is to calculate the average of all predictors. However, such an aggregated measure is potentially very noisy. It also ignores the relationships among all the predictors and overlooks how they are related to the forecasting target. To increase similarity among predictors and to strengthen their connection with the forecasting target, previous studies usually enforce a grouping criterion motivated by economic connections. For example, a series of studies, starting with that of Foster (1981), attempted to forecast non-announcers' earnings using the averaged earnings news of early announcers belonging to the same industry. In this setting, the aggregation measure is intended to capture systematic industry news. The effectiveness of this measure highly depends on the relationship between the forecasting target and predictors. For example, the launch of the iPhone induced a smart phone revolution in the mobile telephone industry. The launch also boosted revenues for companies, such as Apple, the creator of iPhone, and Samsung, a close follower of iPhone that produces similar smart phones. It also signaled an end to the dominant market status of traditional mobile phone companies such as Nokia and BlackBerry. When companies such as Apple and Nokia are early announcers, their earnings may offset each other. In this situation, an average measure could not capture the related technology shock.

In this study, I address the above issue using a method developed by Kelly and Pruitt

(2015) called the three-pass regression filter (3PRF), which is a generalized version of partial least squares (PLS). The central assumption of this method is that there is a common latent factor structure between the forecasting target and the predictors. This factor assumption enables the estimation of common information using a wealth of independent variables. In this setting, earnings of Apple and Nokia have opposite loadings on the factor that captures the industry technology shock of smart phones. Instead of being diminished in the averaged measure, the technology shock is estimated from the cross section of early announcers earnings given that earnings will load differently on factors. As presented in the empirical part of this study, most forecasting targets, or the earnings of late announcers, are assembled by two latent factors that are common with the earnings of early announcers. A simple average of all announced earnings would not completely consider the dynamics between the two factors.

Furthermore, firms could be connected in multiple ways, either transparently or implicitly. More recent studies have gone beyond the industry, such as investigating firm-level supplier-customer linkage (Cohen and Frazzini, 2008) and industry-level supplier-customer relationships (Menzly and Ozbas, 2010). In addition to these direct economic connections, the earnings of all firms are potentially linked through their sensitivity to marketwide cash flow news, as noted by Da and Warachka (2009). However, data availability hinders potential studies that examine all types of relationships between firms. Cohen and Frazzini (2008) manually identify the supplier-customer relationship because mapping firm names to unique identifications is difficult.

Instead of associating firms via specific economic links, the factor structure assumed by 3PRF captures the statistical covariance among individual firms earnings. The extract measure from 3PRF sums up all types of potential economic links. Specifically, I find that the predicted earnings from 3PRF forecast earnings surprises with both statistical and economic significance, with a series of stock characteristics controlled. For standardized unexpected earnings (SUE), a measure of earnings surprise, one standard deviation change in the 3PRF measure predicts a 20% standard deviation change in SUE. For analyst forecast error (FE), another measure of earnings surprise, one standard deviation change in the 3PRF measure predicts a 8% standard deviation change in FE. Supporting results demonstrate the power and validity of 3PRF.

How does the market react to the information captured by 3PRF? In an efficient market, we should not observe any return predictability if investors promptly and properly react to early announcers' news; this occurs by incorporating related information into the non-announcers' prices. Notably, when we can estimate the information using 3PRF with sufficient early announcers, investors may already incorporate such news for related stocks if they possess more sophisticated skills in the collection and analysis of earnings news. However, in reality, investors always face many limitations. Theoretically and empirically (Peng and Xiong, 2006; Hirshleifer, Lim, and Teoh, 2009), investors have been documented to suffer from limited attention; they prioritize salient information, such as market news, but display delayed reactions to other types of news, such as firm-specific news. The cost of extensive information collection and model testing is possibly too expensive for investors (Hong, Stein, and Yu, 2007). They must also address market frictions, such as short-sale constraints, which impede arbitrage activity involving negative information (Miller, 1977). Professional money managers may not fully react to situations in which their investment faces arbitrage risk (Shleifer and Vishny, 1997). As a result, the earnings news captured by the 3PRF method may not be absorbed into prices at the time of estimation.

Consistent with a delayed reaction of investors, empirical tests demonstrate the strong power of the 3PRF measure in predicting earnings announcement return. One standard deviation change of the extracted earnings news predicts a change of 10 basis points in the same direction for the earnings announcement return in excess of market return, which is approximately 5% annualized. Furthermore, an event study reveals that investors react to the news contained by the 3PRF measure substantially earlier for large firms than for small firms, a finding that conforms to the literature documenting that large firms incorporate information faster than small firms (Lo and MacKinlay, 1990). A long-short trading strategy that exploits the market reaction from the estimation of 3PRF to earnings announcement earns a 15% annual alpha calculated using the capital asset pricing model, the Fama and French (1993) three-factor model, and the Carhart (1997) four-factor model. This strategy can be traded during one third of the sample period from January 1979 to December 2016. Furthermore, a strategy that only trades around the earnings announcement day earns a 20% alpha annually, and it can be actively traded for one fifth of the sample period. These findings indicate that the 3PRF measure captures valuable information before the market fully incorporates such news into stock prices, thus providing evidence against market efficiency.

A vital question is the degree of the additional contribution by the 3PRF measure that extracts information from early announcers' earnings compared with other information channels recognized in the literature. To answer this question, I examine the predictability of the 3PRF measure in the face of additional controls from other informed stock groups, including stocks of the same industry (Foster, 1981) or of related customer and supplier industries (Menzly and Ozbas, 2010), large firms (Lo and MacKinlay, 1990), as well as firms with a higher institutional ownership (Badrinath, Kale, and Noe, 1995), with analyst coverage (Brennan, Jegadeesh, and Swaminathan, 1993), or with a higher turnover (Chordia and Swaminathan, 2000). Two types of controls are constructed. The first type takes the average of announced earnings surprises of stocks belonging to a certain informed group. The second type averages lagged five-day cumulative returns of all stocks belonging to a specific informed group. This construction attempts to capture information from stock prices even if some of those firms have not yet announced their earnings. The contribution of the 3PRF measure in predicting SUE and FE is not affected by the addition of the controls listed above[1]. Furthermore, none of the additional controls predict earnings announcement return with 3PRF measure added to the regression, which indicates that investors fully react to earnings news captured by the average measure but fail to completely incorporate information extracted by 3PRF before the earnings announcement.

This study relates to the literature that applies innovative methodologies to reduce the dimension of a large panel of forecasters (relatively, compared to the sample size) for a better prediction. One closely related method is principal component analysis (PCA), which has widespread applications, such as macroeconomic time series forecasting (Stock and Watson, 2002). The PCA approach selects orthogonal statistical factors that optimally explain the covariance of predictors that does not necessarily relate to the forecasting target. In contrast, the 3PRF method of Kelly and Pruitt (2015) searches for the optimal factors that expand the covariance between predictors and the target, thus establishing a more connected predictive

[1] I also examine whether announced earnings or stock returns of firm-level customers, as in Cohen and Frazzini (2008), lead earnings of non-announcing suppliers. Using data provided by the authors, these additional controls have no significant impact on the predictive power of 3PRF measures. However, these controls reduce the sample size considerably. Thus, the test results are not tabulated.

relationship. This method has been applied in several empirical scenarios. Kelly and Pruitt (2013) improve the predictability, especially for out-of-sample tests, of market return and cash flow growth using a factor extracted from the cross section of the portfolio level or even the stock-level book-to-market ratios. Light, Maslov, and Rytchkov (2017) aggregate the common latent factor relevant to the individual stock expected return from a large number of stock characteristics. These studies demonstrate that 3PRF is quite effective in abstracting the relevant portion of a rich information set to a few latent factors.

Future profitability is a vital input to the valuation analysis. This study contributes to the literature that seeks to predict individual firms' accounting earnings. This stream of literature generally uses two types of statistical methods[2]. The first type applies time series models to the firm-specific history of earnings. Time series models, such as the autoregressive moving average model, require sufficient data history. Because firms announce earnings quarterly, a large proportion of the sample cannot be included in the time series study. This issue also raises a concern regarding survivorship bias, meaning only firms that survive for a long period can be examined. The second type estimates pooled cross section regression to capture the average relationship between predictors and earnings. For example, Fama and French (2000) examine predicted earnings using a series of stock characteristics. Pooled cross section regression waives the requirement of sufficient earnings history by sacrificing a firm-specific relationship but utilizing all firms in the cross section. In addition, adding a relatively large number of stock characteristics is easy due to the large sample size. As a result, I use the pooled cross section regression method in this study.

The topic of market efficiency continues to produce heated debate in the field of financial research. Scholars continue to discover more anomalies while studies advance to explain them[3]. The information spillover literature focuses on information events and examines whether investors react to them in a timely and properly consistent manner with the market efficiency assumption. If the market is inefficient, investors may have a delayed reaction to public signals or misreact to such information[4]. This paper is closely related to studies

[2]See Kothari (2001)

[3]For example, Hou, Xue, and Zhang (2015) review 286 anomalies, and the q-factor model rejects 85% of these in terms of statistical significance.

[4]Cohen and Frazzini (2008) provide a general review of theories and empirical findings on how investors react to different types of information events.

that examine assets indirectly affected by news shocks of other assets. Cohen and Frazzini (2008) evaluate the price impact of news transmitted from firms customers. Menzly and Ozbas (2010) examine how firms incorporate shocks from related supplier and customer industries. A series of accounting studies, starting with that of Foster (1981), explores the market reaction of non-announcing firms to the earnings news of announcers within the same industry. In this study, I also investigate information diffusion among stocks. Rather than search for various economic links between firms, I take advantage of their statistical links of earnings, which sum up all connections despite the manner in which firms are fundamentally related. With the assumption of a common latent factor structure, the 3PRF method can extract valuable information relevant to each individual non-announcer if there are sufficient announcers. In this manner, precise and timely information is captured when investors may not have yet fully reacted, which is consistent with the empirical findings.

This paper is organized as follows. Section 1 describes the methodological aspects. Section 2 provides a summary of data. Section 3 presents the empirical findings. Section 4 concludes.

# 1 Methodology

This section discusses assumptions, such as the latent factor structure of earnings, for using three-pass regression filter (3PRF) methodology. The estimation procedures of 3PRF developed by Kelly and Pruitt (2015) are also described.

To take seasonality into account (Livnat and Mendenhall, 2006), I denote unexpected earnings for firm $i$ and calendar quarter $q$ as the difference between quarter $q$'s earnings per share and that of four quarters ago:

$$UE_{i,q} = E_{i,q} - E_{i,q-4} \tag{1}$$

$E_{i,q}$ denotes earnings per share for firm $i$, calendar quarter $q$. This definition assumes that earnings follow a seasonal random walk. $EAD_{i,q}$ is the announcement date for firm $i$'s earnings of the fiscal quarter that overlaps with calendar quarter $q$. This study attempts to forecast quarterly earnings of late announcers using earnings news of all early announcers. To ensure that a set of common latent factors drives both targets and predictors, I require that the earnings news of all firms involved, either as forecasting targets or predictors, cover the same calendar quarter. As a result, to predict $UE_{i,q}$, the corresponding predictors include all $UE_{j,q}$ for $j \neq i$ and $EAD_{j,q} < EAD_{i,q}$.

To implement the 3PRF method by Kelly and Pruitt (2015), a latent factor structure is imposed for unexpected earnings. I assume that

$$UE_{i,q} = \mu_i + \boldsymbol{\beta}_i^T \mathbf{F}_q + \epsilon_{i,q} \tag{2}$$

$\mathbf{F}_q$ denotes the vector of latent factors constituting firm $i$'s unexpected earnings at quarter $q$, the dimension of which is $K_F$. For any firm $j$ that announces quarter $q$'s earnings earlier than firm $i$, I suppose that

$$UE_{j,q} = \mu_j + \boldsymbol{\beta}_j^T \mathbf{F}_q + \boldsymbol{\gamma}_j^T \mathbf{G}_q + \epsilon_{j,q} \tag{3}$$

Thus, predictors' or early announcers earnings are expanded by the same latent factors of firm $i$'s earnings, as well as some additional factors $\mathbf{G}_q$ with dimension $K_G$. Essentially, predictors may be driven by factors that are not related to the forecasting target. This scenario also highlights the power of 3PRF compared with PCA. If $\mathbf{G}_q$ drives a larger variation of predictors than $\mathbf{F}_q$, PCA will attract more attention to $\mathbf{G}_q$ instead of $\mathbf{F}_q$. In contrast, 3PRF will extract $\mathbf{F}_q$ from predictors and ignore $\mathbf{G}_q$, which has nothing to do with the forecast target $UE_{i,q}$[5].

A linear factor model is commonly used in the asset pricing literature because of its parsimony and analytical convenience. Possibly, the assumptions I make cannot fully capture the dynamics of earnings. If such is the case, the power of 3PRF is diminished. The empirical section shows that information extracted using 3PRF has strong predictive power of late announcers' earnings. Thus, the concern for a linear factor structure is insignificant.

Factors are assumed to be latent, meaning that we cannot identify them explicitly. Furthermore, they are extracted using statistical procedures; thus, assigning certain economic meanings is not realistic. However, we can still make a reasonable guess regarding the fun-

[5]The factor structures in equation 2 and 3 are specific to the scenario when $UE_{i,q}$ is the forecasting target. If another $UE_{k,q}$ is the forecast target, factors will be different depending on which common factors expand $UE_{k,q}$ and earnings news of its corresponding early announcers. A more rigorous expression will have subscripts indicating different factors and loadings for each quarterly non-announcing earnings. I omit such subscripts for abbreviation.

damentals that expand the same information set with statistical factors. Da and Warachka (2009) demonstrate the pricing implications of loadings on systematic cash flow innovations. A series of accounting studies starting with that of Foster (1981) attempts to identify the predictability of early announcers' earnings to late announcers' earnings when they belong to the same industry. Menzly and Ozbas (2010) show the cross-predictability of returns for firms connected via an inter-industry supplier-customer relationship. Counting all potential fundamentals that connect individual firms is not possible. Including all fundamental channels in one test is also unrealistic because of data limitations. Instead, I seek an alternative solution, 3PRF, which determines statistical connections as a sum of all potential economic relationships.

Kelly and Pruitt (2015) explain that 3PRF is a generalized version of the PLS method. To discipline the dimension reduction of predictors, 3PRF can either use the forecast target, which is similar to PLS, or a group of new variables motivated by economics theory. The authors develop two methods to estimate 3PRF. The first is a closed-form solution derived from an equivalent constrained least squares problem under the assumption that only relevant factors influence the forecasting target. The second method is a series of three ordinary least squares (OLS) regressions. In this study, I apply the estimation procedures following the second approach, which is more intuitive and is capable of handling unbalanced panels.

Figure 1 illustrates 3PRF procedures for one latent factor model. To forecast $UE_{i,q}$ using all $UE_{j,q}$ for $j \neq i$ and $EAD_{j,q} < EAD_{i,q}$, the estimation of 3PRF with one relevant factor can be calculated following three-step OLS regressions. I denote the latent factor as $F_{t,1}$ for $t \leq q$. In the first step, a time series regression is run for each pair of $UE_{i,q}$ and $UE_{j,q}$. Specifically, for each $j$, I estimate the regression

$$UE_{j,t} = \psi_{j,0} + \psi_j UE_{i,t} + \nu_{j,t} \tag{4}$$

with all available common history for $UE_{i,t}$ and $UE_{j,t}$ before quarter $q$ or $t \leq q-1$. I require a minimum length of 15 quarters' history in this regression. $UE_{i,t}$ is used as a proxy for the common latent factor. The estimated loading $\hat{\psi}_j$ captures the degree of dependence of $UE_{j,t}$ on the common factor with the forecast target.

In the second step, for each quarterly cross section of predictors $UE_{j,t}$ for $t \leq q$, I estimate the regression

$$UE_{j,t} = \phi_{t,0} + \phi_{t,1}\hat{\psi}_j + \omega_{j,t} \tag{5}$$

The resulting time series of the slope estimates $\hat{\phi}_{1,1}, \hat{\phi}_{2,1}, ..., \hat{\phi}_{q,1}$ captures the time variation of the common latent factor $F_{t,1}$ for $t \leq q$. Essentially, the second step backs out the time-varying factor by utilizing the cross section of predictors. A minimum number of 30 predictors are enforced for each cross section estimation.

In the last step, the time series regression

$$UE_{i,t} = \alpha_i + \beta_i\hat{\phi}_{t,1} + \upsilon_{i,t} \tag{6}$$

is estimated for $t \leq q-1$, which captures the relationship between our forecast target and the factor. The prediction for $UE_{i,q}$ is $\hat{\alpha}_i + \hat{\beta}_i\hat{\phi}_{q,1}$.

If the factor $F_{t,1}$ is known, we can simply proceed to the third step. However, the factor is unobserved and latent. Kelly and Pruitt (2015) name the forecast using true factors "infeasible best forecast". They prove that the forecast obtained using 3PRF is consistent with the infeasible best forecast under assumptions of linear factor structures and regular technical conditions. First, both the target and predictors follow a linear factor structure in which the target can be composed by a subset of factors spanning the predictors. Technical conditions require finite second moments and the probability convergence of factors, loadings, and residuals. Both cross section correlation and serial correlation are allowed for the residuals of predictors and the forecasting target.

An automatic proxy-selection algorithm can be performed to consider more factors. To generate the second factor, I use residuals $\hat{\upsilon}_{i,t} = UE_{i,t} - \hat{UE}_{i,t}$ for $t \leq q-1$ from the third step of one-factor estimation above. In the first step, for each pair of firm $j$ and $i$, I run time series regressions of $UE_{j,t}$ on both $UE_{i,t}$ and $\hat{\upsilon}_{i,t}$ for $t \leq q-1$. The estimated coefficients are denoted as $\hat{\psi}_{j,1}$ and $\hat{\psi}_{j,2}$. $UE_{i,t}$ and $\hat{\upsilon}_{i,t}$ serve as proxies for the two latent factors. In the second step, for each quarter $t \leq q$, a cross section regression of $UE_{j,t}$ on $\hat{\psi}_{j,1}$ and $\hat{\psi}_{j,2}$ are estimated to obtain factor proxies $\hat{\phi}_{t,1}$ and $\hat{\phi}_{t,2}$. The third step estimates the time series regression $UE_{i,t} = \alpha_i + \beta_{i,1}\hat{\phi}_{t,1} + \beta_{i,2}\hat{\phi}_{t,2} + \upsilon_{i,t}$ using $t \leq q-1$. The prediction with two factors is $\hat{UE}_{i,q} = \hat{\alpha}_i + \hat{\beta}_{i,1}\hat{\phi}_{q,1} + \hat{\beta}_{i,2}\hat{\phi}_{q,2}$. Similarly, an N factor model can be estimated by adding the residuals from the third step of the N-1 factor estimation to another round of three-step OLS regressions.

In this study, to avoid look-ahead bias, I perform an out-of-sample implementation of 3PRF in that all estimations are calculated on the samples before the forecast target. In addition, not all individual firms have correlated earnings; I thus apply a coarse filter in selecting predictors. As proven by Kelly and Pruitt (2015), 3PRF requires at least a subset of predictors with non-zero loadings on relevant factors. This filter requires that the estimated loading of regressing each predictor on the target, as in the first step of one latent factor model, is statistically significant at 10%, the standard error of which is calculated following Newey and West (1987). Considering that the number of relevant factors may vary between forecast targets, I perform 3PRF estimations recursively in search of three potential factors maximally according to the automatic proxy selection algorithm. Only three factors are examined because a number of firms have a short time series of quarterly earnings. The procedures will end with three forecasting models with one to three factors. The selected model will satisfy two conditions. First, all factors are statistically significant (10% level, calculated by Newey and West (1987) standard errors) in the third step. Second, the number of factors of the selected model is the largest among all models that satisfy the first condition.

# 2 Data and summary statistics

## 2.1 Data and variables

Data on stock prices, returns, and volumes are obtained from the Center for Research in Securities Prices (CRSP). Quarterly earnings and other accounting variables are gathered from Compustat. The main sample, an intersection of CRSP and Compustat, consists of stocks with a share code of 10 or 11, spanning the time period from January 1979 to December 2016[6]. Analyst forecasts are from Institutional Brokers' Estimate System (IBES) provided by Thomson and Reuters. Because of data availability, tests involving analyst forecasts are

[6]The quarterly accounting variables of Compustat can be traced back to the 1960s. In the estimation of 3PRF, a minimum length of 15-quarter history is required for the first step, and a minimum number of 30 qualified early announcers are required for the second step. Given such conditions, the estimation can be performed for a large proportion of the cross section from the late 1970s. In this study, I use the sample from January 1979; including earlier samples has little impact on the empirical results.

performed on a sample from January 1985 to December 2016.

Earnings from Compustat and IBES are usually not exactly the same. IBES earnings are considered street earnings in the sense that they are usually what investors see in analyst reports or the media. Bradshaw and Sloan (2002) finds that the special items explain the differences between earnings reported by Compustat and IBES. I use street earnings throughout the paper, which subtracts special items multiplied by 0.65 from the Compustat net income before extraordinary items[7]. I calculate two measures to capture earnings surprises. The first is the standardized unexpected earnings or SUE (Jegadeesh, Kim, Krische, and Lee, 2004), defined as

$$SUE_{i,q} = \frac{UE_{i,q}}{Std_{i,q}(UE)} \tag{7}$$

where $Std_{i,q}(UE)$ is the standard deviation of unexpected earnings for firm $i$ using the eight preceding quarters from $q-7$ to $q$. The second is the analyst forecast error or FE, defined as

$$FE_{i,q,d} = \frac{E_{i,q} - Med_{i,q,d}}{P_{i,q}} \tag{8}$$

where $d$ denotes the date on which we collect analyst forecasts in the calculation of $Med_{i,q,d}$, which is the median of all available analyst earnings forecasts up to date $d$ for firm $i$ at quarter $q$. Only the most recent forecast from each analyst is included. Analyst forecasts made more than 90 days earlier than the announcement day are considered stale forecasts and deleted from the sample. $P_{i,q}$ is the stock price of firm $i$ at the end of quarter $q$. To measure the market reaction to the earnings announcement, I define the earnings announcement return (AR) as the cumulative return, in excess of market returns, from two trading days before the announcement ($EAD_{i,q}$) to two trading days after the announcement:

$$AR_{i,q} = \prod_{t=EAD_{i,q}-2}^{EAD_{i,q}+2} (1 + r_{i,t} - r_{mkt,t}) - 1 \tag{9}$$

Section 1 describes the estimation procedures to acquire $\hat{UE}_{i,q}$. As shown by figure 2, two predicted unexpected earnings are estimated using different training samples. For firm $i$, quarter $q$, $\hat{UE}_{i,q,f}$ is estimated using the sample of early announcers up to the earliest

[7] I also conduct all tests using Compustat earnings. The results are very similar.

date when more than 30 qualified firms have announced. $f$, short for $first$, denotes the date on which this training sample ends. This measure can capture earnings news from early announcers as soon as sufficient early announcers are available. On the other hand, $\hat{UE}_{i,q,l}$ is estimated using the firms' earnings news available at the end of four trading days before the announcement day or at the end of two trading days before the time window of announcement returns. This additional one trading day gap is enforced because some firms announce earnings after trading hours. $l$, short for $last$, denotes the date on which this training sample ends. $\hat{UE}_{i,q,l}$ is potentially more informative than $\hat{UE}_{i,q,f}$. The last predicted UE is estimated based on a larger sample and thus may capture more information but less noise. To differentiate the first and last predicted UE, only last predictions generated at least one week after the first predictions are considered. I also require that the last predicted UE is estimated using at least 30 additional early announcers compared to the first prediction.

I derive two measures for predicted earnings surprise from 3PRF predictions. The first measure is the predicted SUE (PSUE). For quarter $q$ and firm $i$, PSUE is defined as

$$PSUE_{i,q} = \frac{\hat{UE}_{i,q}}{Std_{i,q-1}(UE)} \tag{10}$$

Compared to SUE, PSUE replaces UE with predicted UE from 3PRF. To use only data before the announcement of $E_{i,q}$, PSUE is scaled by the standard deviation of UE using observations from $q-8$ to $q-1$. Two types of PSUE are constructed. One takes the first predicted UE, which is denoted as $\mathrm{PSUE_f}$. The other takes the last predicted UE, which is denoted as $\mathrm{PSUE_l}$.

The other measure, the predicted analyst forecast error (PFE), is used to measure the predicted earnings surprise against analyst forecasts. For firm $i$, quarter $q$, I define PFE using analyst forecasts available at the end of date $d$ as

$$PFE_{i,q,d} = \frac{\hat{UE}_{i,q} + E_{i,q-4} - Med_{i,q,d}}{P_{i,q}} \tag{11}$$

where $E_{i,q}$ is replaced by $\hat{UE}_{i,q} + E_{i,q-4}$ in the definition of $FE_{i,q,d}$. Many firms announce earnings after trading hours. For PFE using the first predicted UE ($\mathrm{PFE_f}$), I use analyst forecasts available at the end of one trading day after the first estimation. Because some firms announce earnings after trading hours, I allow analysts to update their forecast with a one-day lag compared to the estimation date of 3PRF. Similarly, for PFE using the last predicted UE ($\text{PFE}_\text{l}$), I use analyst forecasts available at the end of one trading day before the time window of announcement return or at the end of three trading days before the announcement day.

## 2.2 Summary statistics

Panel A of table 1 presents summary statistics for the sample with 3PRF prediction. On average, the sample has 947 observations of SUE and $\text{PSUE}_\text{f}$ at each quarter from January 1979 to December 2016, while it has 551 observations of $\text{PSUE}_\text{l}$ quarterly. I report two types of analyst forecast errors. Corresponding to $\text{PFE}_\text{f}$, $\text{FE}_\text{f}$ is calculated using analyst forecasts available at the end of one trading day after the first estimation of UE by 3PRF. Similarly, $\text{FE}_\text{l}$ is calculated using analyst forecasts available at the end of three trading days before the announcement day, in line with $\text{PFE}_\text{l}$. On average, there are 645 quarterly observations of $\text{FE}_\text{f}$ and $\text{PFE}_\text{f}$ and 351 observations of $\text{FE}_\text{l}$ and $\text{PFE}_\text{l}$ from January 1985 to December 2016. Panel B of table 1 reports summary statistics for the sample with no predictions from 3PRF.

An observation from panels A and B indicates that the distributions of earnings surprises and announcement returns are very similar for samples with or without predictions generated by 3PRF. Figure 3 displays the histograms of the day gap between the fiscal quarter end and the announcement day for the two samples. Samples with no 3PRF predictions are primarily early announcers with insufficient data to estimate the 3PRF measure. Furthermore, a small portion of this sample consists of late announcers with an earnings history that is shorter than fifteen quarters or with earnings that are not highly correlated with most early announcers. In summary, the choice of announcement timing is not correlated with the performance of firms. My results in the empirical tests are not dependent on a selected sample with specific types of firm performance.

Table 2 summarizes the training sample size and the number of factors used in 3PRF measures. For the early sample from 1979 to 2000, the first predicted SUE is estimated by training samples of 28 quarterly time series and 37 early announcers on average. For the later sample from 2001 to 2016, the first PSUE is estimated using slightly larger training samples.

For both sub samples, most predictions are composed of two factors, which demonstrates the necessity of a multi-factor assumption. The last predicted SUE includes a substantially larger number of early announcers on average, which is 200 for the early sample and 361 for the later sample. A similar pattern is observed for the first and last predicted FE.

# 3 Empirical Tests

In this section, empirical tests are conducted to support the argument that measures from 3PRF have strong power in predicting earnings surprise. In addition, because of the delayed reaction of investors, news captured by 3PRF is not fully reflected in late announcers' stock prices. The tests in Section 3.2 demonstrate that 3PRF measures forecast earnings announcement return with both statistical and economic significance. Furthermore, large stocks incorporate 3PRF information faster than small stocks. Trading strategies exploiting this return predictability gain an annual alpha of greater than 15%. Section 3.3 conducts a series of tests to ensure that 3PRF measures have over-and-above contributions compared with other information channels documented in the literature.

## 3.1 Prediction of earnings surprise

To examine the predictive power of 3PRF measures, I conduct panel regressions of earnings surprise, SUE or analyst FE on the corresponding predicted measures while controlling for a number of stock characteristics. Controls include lag of the dependent variable (SUE or FE), firm size, book-to-market ratio, idiosyncratic volatility, total accruals scaled by total assets, and Amihud (2002) illiquidity. To control previous market reactions to the information, I also use two variables of historical abnormal returns following Tetlock, Saar-Tsechansky, and Macskassy (2008). One is the cumulative alpha of the Carhart (1997) four-factor model, for which loadings of factors are estimated using one-year daily returns before the most recent fiscal quarter end, as measured from the quarter end to one trading day after the estimation of the predicted UE. The other variable is the historical alpha, which is estimated using daily returns one year preceding the quarter end, of the Fama and French (1993) three-factor model[8].

Table 3 presents the estimation results of regressing SUE on the first PSUE. Model 1 shows the regression on $\mathrm{PSUE_f}$ only, while models 2-5 represent results with additional controls. In model 3, standard errors are clustered by firm and quarter following Petersen (2009), while in model 4 I also add firm and quarter fixed effects. Model 5 exhibits results using the Fama-MacBeth method. Specifically, for each quarterly cross section observation, I estimate the regression of SUE on $\mathrm{PSUE_f}$ and additional controls. The time series average of the regression coefficients is presented as the estimates in model 5. From model 2 to model 5, the estimated coefficients of $\mathrm{PSUE_f}$ are all statistically significant and have very similar values. One standard deviation change in the first PSUE predicts approximately 10% of one standard deviation change in SUE in the same direction.

Models 6 and 7 of table 3 control for analyst forecasts. I construct a variable $\mathrm{PSUE_a}$ defined as

$$PSUEA_{i,q,d} = \frac{U\hat{E}A_{i,q}}{Std_{i,q-1}(UE)} \tag{12}$$

where

$$U\hat{E}A_{i,q,d} = Med_{i,q,d} - E_{i,q-4} \tag{13}$$

$E_{i,q}$ denotes earnings per share for firm $i$, quarter $q$. $Std_{i,q-1}(UE)$ is the standard deviation of unexpected earnings for firm $i$ using eight preceding quarters from $q-8$ to $q-1$. $Med_{i,q,d}$ is the median of all available analyst earnings forecasts up to date $d$ for firm $i$ at quarter $q$. Essentially, $\mathrm{PSUE_a}$ is the predicted SUE using analyst forecasts instead of a 3PRF extracted measure. For models 5 and 6 of table 3, $d$ is one trading day after the estimation of $\mathrm{PSUE_f}$, allowing one additional day for analysts to update their forecasts when firms announce earnings after trading hours. For stocks covered by analysts, $\mathrm{PSUE_f}$ still have independent information in forecasting SUE.

Table 4 and 5 show analyst reactions to information captured by 3PRF before and after the estimation of $\mathrm{PFE_f}$. Analyst revision is defined as the difference between the number of positive and negative forecast revisions divided by the total number of forecast changes during

[8]These two measures of the cumulative alpha are intended to control for the effect of short return reversal and momentum. I also use alternative controls, which are the one-month lag return and cumulative returns of the previous eleven months. Regression results are quantitatively similar.

a given time period. Table 4 presents regressions of analyst revisions from the corresponding fiscal quarter end to one trading day after the estimation of $\mathrm{PFE_f}$. The results indicate that analyst revisions prior to $\mathrm{PFE_f}$ are negatively correlated with $\mathrm{PFE_f}$. Table 5 shows the regression results of analyst revisions from two trading days after the estimation of $\mathrm{PFE_f}$ to three trading days before earnings announcement days. We do not observe any updates from analysts responding to the information captured by 3PRF. Table 6 shows the tests of regressing analyst FE on the first PFE. The strong forecasting power of the first PFE confirms that analysts fail to utilize all information from early announcers.

## 3.2 Prediction of market reaction

Table 7 demonstrates that after a series of stock characteristics is controlled, a change in the first PSUE of one standard deviation predicts a change in the earnings announcement return by 10 basis points in the same direction, which is approximately 5% annualized return in excess of market returns. Models 6 and 7 of table 7 examine the predictive magnitude of the first PSUE among stocks with analyst coverage. The estimated coefficient is smaller but still significant in model 6, which is a pooled cross section regression with both firm and quarter fixed effects. The standard errors are clustered by both firm and quarter. The estimated coefficient of $\mathrm{PSUE_f}$ is not statistically significant for stocks with analyst coverage using the Fama-MacBeth method. In general, the regression results prove that the first PSUE has both statistical and economic significance in terms of predicting earnings announcement returns.

To examine whether investors respond to the news of early announcers before the earnings announcement day, I perform an event study of stock returns. Specifically, stocks are classified into five groups in each quarter according to their first PSUE value. Table 9 presents the cumulative returns of the five groups for the following three time windows: Q to [First+1], [First+2] to [EA-3], and [EA-2] to [EA+2]. Q denotes the quarter end of the corresponding SUE. [X+n] or [X-n] stands for n trading days after or before timing X. If X is first, it is the calculation day of the first PSUE. If X is EA, it is the earnings announcement day. I examine stock returns, raw or abnormal, from the quarter end to the first SUE calculation, from the post of the first PSUE to three trading days in advance of the announcement day, and upon the time window of earnings announcement.

As shown in Table 8, the stock characteristics are very similar across all five groups. Surprisingly, they are primarily larger stocks with higher liquidity and lower volatility. This empirical fact indicates that more stocks that are small, with earnings that are more idiosyncratic, have no 3PRF predictions. Characteristic similarity across stock groups formed by PSUE value excludes the possibility that any event return pattern is caused by certain stock characteristics instead of news from early announcers.

For each group, I obtain a quarterly time series of the averaged cumulative return or the averaged cumulative alpha of the four-factor model. Specifically, factor loadings are estimated by the daily returns for one year prior to the quarter end. Table 9 presents the value-weighted and equal-weighted average of the time series. The results show that the stock prices across all five groups do not significantly respond to early announcers before the determination of the first PSUE. After a sufficient number of early announcers, we observe a significant value-weighted return difference in terms of cumulative raw return and Carhart (1997) four-factor alpha in the period after the estimation of the first PSUE until the earnings announcement. In contrast, equal-weighted returns do not exhibit a significant difference until earnings announcement, indicating that the early reaction of stock prices is driven by large stocks, while small stocks incorporate the information only when the actual earnings are announced.

To determine whether investors behave differently to stocks of different sizes, I perform the event analysis for two sub samples split by firm size: one with firms of the market capitalization above the 50th percentile among NYSE stocks and one sample with other smaller firms. Table 10 shows that for large firms, the market responds to the news of early announcers immediately after the determination of the first PSUE, resulting in an insignificant return difference around the earnings announcement window. However, no price change is observed for small firms before earnings announcement.

Due to the asymmetric response of investors to large and small firms, I develop two trading strategies, one of which captures information spillover immediately after the determination of the first PSUE until the announcement and the other utilizes the predictability of the earnings announcement return only. The first trading strategy is presented in Table 11. Specifically, the long (short) side buys (shorts) stocks with the highest (lowest) first PSUE value among the five groups formed at the end of [first+1]. The position is liquidated at the end of [EA+1]. PSUE groups are classified using the distribution of the first PSUE four quarters ago to avoid look-ahead bias. A minimum number of 10 stocks are required to maintain the portfolios. Only stocks with a price larger than $5 on the formation day are considered. For a sample of approximately 38 years, this trading strategy is active for 3,399 trading days, which is above one third of the sample period. The average holding period is 11 trading days. The long minus short strategy gains a 15% value-weighted alpha and a 13% equal-weighted alpha annually. The last four columns of panels B and C present the factor loadings of long minus short returns.

The other trading strategy forms portfolios at the end of [EA-3] according to the last PSUE, which liquidates at the end of [EA+1]. The holding period is only a maximum of four trading days, so the active trading days number 2,607, which is above one fourth of the sample period. Finding a stronger equal-weighted abnormal return than the value-weighted return, which is 27% and 20% annually, is expected, given the empirical fact that small stocks react to the information captured by the 3PRF measure only when earnings are announced. Table 13 considers trading costs with a maximum of 10 basis points assumed for a round-trip trading. Both trading strategies survive, with the exception of the value-weighted one for the second strategy.

## 3.3 Robustness tests

### 3.3.1 Compare the first and the last PSUE

As shown in table 2, the last predicted SUE and FE are estimated using much more early announcers than their first peers. As a result, the last predictions from 3PRF potentially capture more earnings news. Table 14 presents the regression results of SUE, FE, and AR using the first or the last prediction. As expected, the estimated coefficients of the last predicted SUE(FE) more than double those of the first predicted SUE(FE) in forecasting SUE(FE). The last predicted SUE also has stronger power in forecasting AR compared to the first predicted SUE.

### 3.3.2 Subsample analysis

I perform predictive regressions using two subsamples: one early sample before the year 2000 and another starting from the year 2001. Table 15 presents the estimation results. For the forecasting of SUE, the magnitude of the PSUE coefficient is very similar for the two subsamples. For the analyst forecast errors, PFE has more predictive information during the later sample starting from 2001. The magnitude of the PSUE in forecasting the earnings announcement return of the later sample doubles that of the early sample, indicating that investors do not react more quickly during the later sample period.

### 3.3.3 Alternative information channels

A number of lead-lag relationships between stock groups have been identified in the literature. Return predictability arises under the condition in which a group of stocks incorporates certain news faster than other stocks. In this section, I examine whether news captured by 3PRF survives the controls of other information channels.

The first type of information channel includes stocks connected by fundamentals. Several accounting studies, starting with that of Foster (1981), find that late announcers' earnings can be predicted using the average earnings of early announcers within the same industry. However, due to the noisy measure of industry information transfer, none of the studies have developed a profitable trading strategy as in this paper. In the tests, I classify 48 industries as defined by Fama and French (1997). Menzly and Ozbas (2010) study information spillover from industries as customers or suppliers to the forecast target. I follow their identification using the BEA Input-Output Surveys to identify industry relationships.

The second type of information advantage exists among stocks with certain stock characteristics. Lo and MacKinlay (1990) show that large stocks lead small stocks. Badrinath et al. (1995) claim that the information leadership of large stocks is a result of higher institutional ownership. Brennan et al. (1993) assign the source of early news incorporation to analyst coverage, whereas Chordia and Swaminathan (2000) document the important role of higher turnover.

Two types of controls are constructed to handle such information channels. The first type is calculated as the average of earnings surprises or announcement returns of stock groups with an information advantage. For example, in the regression of SUE, the average SUE of the early announcers of the same industry with the forecast target is used to control for the industry channel. This measure restricts information from the intersection of the informed group and early announcers. The second type is the average of the cumulative returns of five trading days preceding [first+1]. This measure considers the possibility that some stocks reflect information faster even when they have not yet announced their earnings.

Tables 16 to 18 present the results of forecasting SUE, FE, and announcement return after controlling for these two types of measures as well as a series of stock characteristics. For the forecast of SUE as shown in table 16, the estimated coefficients of $PSUE_f$ have similar value from model 1 to model 8. Specifically, each regression from model 1 to model 7, in either panel A or panel B, adds a control of one informed group as listed in the first column, while model 8 adds all controls to the regression[9]. According to the results of model 8 as shown in panel A of table 16, the averaged SUE of early announcers within the same industry (IND_SUE) and that of early announcers belonging to the customer industries (CUSIND_SUE) have independent information in predicting the SUE of non-announcing firms. 3PRF extracts news that is common among predictors. For earnings news that only a few early announcers contain, 3PRF may fail to capture such news since most of the predictors do not have loadings on the news. Thus, although the average measure is noisy, it may still offer complement information to the 3PRF measure since such measures can be calculated with a few early announcers of certain informed group. Table 17 shows the regression results of forecasting analyst forecast errors using the predicted FE and additional controls. As presented in model 8 of panel A, only the averaged FE of early announcers belonging to the customer industries offer independent information among all additional controls. Table 18 presents the results of regressing announcement returns on the predicted SUE and additional controls. None of the additional controls contribute to the prediction of announcement returns with $PSUE_f$ added. It seems that investors have already incorporated information captured by the average measure before the earnings announcement, while they fail to fully react to the information captured by the 3PRF measure. In conclusion, the robustness of information

[9]For controls of some informed groups, the value is highly correlated with stocks of market capitalization above 66th percentile among NYSE listed stocks (the informed group denoted by BIG). Thus, they are omitted in model 8 from table 16 to 18

derived from 3PRF predictions proves the advantage of this method when facing a large number of predictors.

# 4 Conclusion

This study deals with a difficult situation in empirical studies, in which the number of predictors is much larger than the length of the time series. Traditionally, we can simply average all the predictors. However, as we enter an information explosion era, a better method is required. In this study, I apply the generalized PLS method designed by Kelly and Pruitt (2015) in the scenario of numerous firms earnings announcement. What is the best way to collect information from hundreds of early announcers, which is relevant to firms that have not announced yet? The method developed by Kelly and Pruitt (2015) is a perfect tool for handling this problem. With the assumption of latent common factors that expand firm earnings, information can be extracted from early announcers considering the dynamics among both early announcers and late announcers. The success of predicting earnings surprise and returns demonstrates the effectiveness of the adopted methodology.

## Table 1 Summary Statistics

Panel A (Panel B) reports quarterly average of summary statistics for variables listed in the first column. Panel A covers the sample with predicted UEs by 3PRF. SUE is the standardized unexpected earnings. $\text{PSUE}_\text{f}$ ($\text{PSUE}_\text{l}$) is predicted SUE using first (last) predicted UEs by 3PRF. $\text{FE}_\text{f}$ ($\text{FE}_\text{l}$) is analyst forecast error using analyst forecasts up to one trading day after the first (last) estimation day of 3PRF. $\text{PFE}_\text{f}$ ($\text{PFE}_\text{l}$) is predicted analyst forecast error using first (last) predicted UEs by 3PRF. AR is earnings announcement return adjusted by market return. Panel B covers the sample without predicted UEs by 3PRF. FE is the analyst forecast error using data up to three trading days before the announcement day. All variables except AR are winzorized by 1%. The sample for SUE and predicted SUE is from Jan 1979 to Dec 2016, while the sample for FE and predicted FE is from Jan 1985 to Dec 2016.

Panel A: summary statistics for sample with 3PRF prediction

| Variable | N | Mean | STD | Median | Skew | Kurt |
|---|---|---|---|---|---|---|
| SUE | 947 | 0.30 | 1.37 | 0.22 | 0.46 | 0.72 |
| $\text{PSUE}_\text{f}$ | 947 | 0.30 | 1.37 | 0.18 | 0.80 | 4.30 |
| $\text{PSUE}_\text{l}$ | 551 | 0.24 | 1.00 | 0.15 | 0.77 | 2.23 |
| $\text{FE}_\text{f}$ | 645 | -0.19% | 1.26% | 0.00% | -3.39 | 21.78 |
| $\text{PFE}_\text{f}$ | 645 | -0.14% | 2.30% | -0.03% | -1.37 | 13.08 |
| $\text{FE}_\text{l}$ | 351 | -0.28% | 1.65% | -0.01% | -3.66 | 23.92 |
| $\text{PFE}_\text{l}$ | 351 | -0.19% | 2.08% | -0.03% | -1.81 | 13.16 |
| AR | 999 | 0.23% | 7.07% | 0.08% | 0.40 | 6.77 |

Panel B: summary statistics for sample with no 3PRF prediction

| Variable | N | Mean | STD | Median | Skew | Kurt |
|---|---|---|---|---|---|---|
| SUE | 728 | 0.46 | 1.42 | 0.34 | 0.48 | 0.64 |
| FE | 850 | -0.17% | 1.22% | 0.01% | -3.13 | 18.33 |
| AR | 1146 | 0.27% | 8.76% | 0.07% | 0.58 | 13.63 |

Table 2 Summary of 3PRF Predictions

Panel A (Panel B) reports summarized information about $\mathrm{PSUE_f}$ ($\mathrm{PSUE_l}$), with sub samples from Jan, 1979 to Dec, 2000 and from Jan, 2001 to Dec, 2016. Panel C (Panel D) reports summarized information about $\mathrm{PFE_f}$ ($\mathrm{PFE_l}$), with sub samples from Jan, 1985 to Dec, 2000 and from Jan, 2001 to Dec, 2016. The column "Sample" indicates the sample period reported. The second column with header "T" presents average length of quarterly time series for step 1 regression of 3PRF. The column "N" is the averge number of cross section observations for step 2 of 3PRF. The last three columns summarize the quarterly average of numbers of 3PRF predictions with one, two and three latent factors.

Panel A: summary of $\mathrm{PSUE_f}$

| | Predictors | | Factors | | |
|---|---|---|---|---|---|
| Sample | T | N | One | Two | Three |
| 1979 to 2000 | 28 | 37 | 18 | 544 | 36 |
| 2001 to 2016 | 35 | 40 | 40 | 1304 | 91 |

Panel B: summary of $\mathrm{PSUE_l}$

| | Predictors | | Factors | | |
|---|---|---|---|---|---|
| Sample | T | N | One | Two | Three |
| 1979 to 2000 | 27 | 200 | 24 | 297 | 14 |
| 2001 to 2016 | 32 | 361 | 87 | 728 | 38 |

Panel C: summary of $\mathrm{PFE_f}$

| | Predictors | | Factors | | |
|---|---|---|---|---|---|
| Sample | T | N | One | Two | Three |
| 1985 to 2000 | 31 | 37 | 9 | 368 | 29 |
| 2001 to 2016 | 35 | 40 | 32 | 1028 | 73 |

Panel D: summary of $\mathrm{PFE_l}$

| | Predictors | | Factors | | |
|---|---|---|---|---|---|
| Sample | T | N | One | Two | Three |
| 1986 to 2000 | 30 | 182 | 13 | 166 | 11 |
| 2001 to 2016 | 33 | 348 | 61 | 557 | 30 |

Table 3 Forecast of Quarterly Earnings Surprise

This table presents the regression results in forecasting standardized unexpected earnings (SUE). Column 1 shows regressors. $PSUE_f$ is the first predicted SUE by 3PRF. $PSUE_a$ is predicted SUE using analyst forecasts available at the end of one trading day after the calculate day of $PSUE_f$. Size is the market capitalization of the stock. BM is the ratio of book equity and market capitalization. Cum_alpha is the cumulative alpha, calculated using four factor model, from most recent fiscal quarter end to one trading day after calculate day of $PSUE_f$. Hist_alpha is the historical alpha of daily returns for one year before the most recent fiscal quarter end, calculated using three-factor model. TA is the ratio of total accrual scaled by total asset. IVOL is the idiosyncratic volatility using three-factor model. Amihud is the market adjusted illiquidity as in Amihud (2002). Intercept is not reported for abbreviation. Standard errors are reported in parentheses. All variables are winzorized by 1% and standardized to mean 0 and standard deviation 1. The notations ***, ** and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| $PSUE_f$ | 0.370*** | 0.110*** | 0.110*** | 0.094*** | 0.125*** | 0.032*** | 0.072*** |
| | (0.002) | (0.003) | (0.010) | (0.009) | (0.006) | (0.004) | (0.011) |
| $PSUE_a$ | | | | | | 0.723*** | 0.507*** |
| | | | | | | (0.014) | (0.060) |
| Lag(SUE) | | 0.447*** | 0.447*** | 0.368*** | 0.429*** | 0.160*** | 0.229*** |
| | | (0.003) | (0.010) | (0.009) | (0.007) | (0.007) | (0.029) |
| Log(Size) | | 0.039*** | 0.039*** | -0.023 | 0.043*** | -0.114*** | 0.040** |
| | | (0.003) | (0.006) | (0.014) | (0.005) | (0.019) | (0.020) |
| Log(BM) | | -0.061*** | -0.061*** | -0.089*** | -0.073*** | -0.074*** | -0.044 |
| | | (0.003) | (0.005) | (0.008) | (0.006) | (0.008) | (0.029) |
| Cum_alpha | | 0.042*** | 0.042*** | 0.040*** | 0.053*** | 0.031*** | 0.036*** |
| | | (0.002) | (0.004) | (0.004) | (0.003) | (0.004) | (0.009) |
| Hist_alpha | | 0.132*** | 0.132*** | 0.152*** | 0.145*** | 0.082*** | 0.109*** |
| | | (0.002) | (0.005) | (0.005) | (0.004) | (0.004) | (0.039) |
| TA | | -0.002 | -0.002 | -0.020*** | -0.008* | -0.005* | 0.102 |
| | | (0.002) | (0.004) | (0.004) | (0.004) | (0.003) | (0.093) |
| Log(Amihud) | | -0.041*** | -0.041*** | -0.063*** | -0.032*** | -0.026*** | -0.062** |
| | | (0.003) | (0.007) | (0.008) | (0.005) | (0.009) | (0.029) |
| IVOL | | -0.054*** | -0.054*** | -0.023*** | -0.036*** | -0.019*** | -0.015 |
| | | (0.003) | (0.008) | (0.004) | (0.004) | (0.004) | (0.057) |
| N | 143065 | 131332 | 131332 | 131058 | 131332 | 83217 | 83501 |
| Adj. $R^2$ | 0.133 | 0.346 | 0.346 | 0.372 | 0.345 | 0.616 | 0.511 |
| Start | 1979 | 1979 | 1979 | 1979 | 1979 | 1985 | 1985 |
| End | 2016 | 2016 | 2016 | 2016 | 2016 | 2016 | 2016 |
| Fixed effect | No | No | No | F & Q | FM | F & Q | FM |
| Clustered Std Error | No | No | F & Q | F & Q | No | F & Q | No |

Table 4 Reaction of Analysts to Information Captured by 3PRF

This table reports contemporaneous regressions of analyst revisions on the first predicted forecast errors ($PFE_f$ ) by 3PRF. The dependent variable is the analyst revision from fiscal quarter end to one trading day after the calculation date of $PFE_f$, requiring that the end date is at least one week after the beginning date of the revision period. Analyst revision is the difference between the number of positive forecast revisions and the number of negative forecast revisions scaled by total forecast revisions given the time period. FE is the analyst forecast error. Size is the market capitalization of the stock. BM is the ratio of book equity and market capitalization. Cum_alpha is the cumulative alpha, calculated using four factor model, from most recent fiscal quarter end to one trading day after calculate day of $PSUE_f$. Hist_alpha is the historical alpha of daily returns for one year before the most recent fiscal quarter end, calculated using three-factor model. TA is the ratio of total accrual scaled by total asset. IVOL is the idiosyncratic volatility using three-factor model. Amihud is the market adjusted illiquidity as in Amihud (2002). Intercept is not reported for abbreviation. Standard errors are reported in parentheses. All variables except the dependent variable are winzorized by 1% and standardized to mean 0 and standard deviation 1. The notations ***, ** and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. The sample is from Jan 1986 to Dec 2016.

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $PFE_f$ | -0.013*** | -0.014*** | -0.014*** | -0.014*** | -0.014*** |
| | (0.001) | (0.001) | (0.002) | (0.002) | (0.002) |
| Lag(FE) | | 0.005*** | 0.005*** | 0.004*** | 0.007*** |
| | | (0.001) | (0.001) | (0.001) | (0.001) |
| Log(Size) | | -0.004*** | -0.004* | -0.019*** | -0.002 |
| | | (0.001) | (0.002) | (0.005) | (0.002) |
| Log(BM) | | -0.004*** | -0.004*** | -0.004* | -0.003** |
| | | (0.001) | (0.002) | (0.002) | (0.001) |
| Cum_alpha | | 0.036*** | 0.036*** | 0.036*** | 0.034*** |
| | | (0.001) | (0.002) | (0.002) | (0.002) |
| Hist_alpha | | 0.024*** | 0.024*** | 0.025*** | 0.024*** |
| | | (0.001) | (0.002) | (0.002) | (0.002) |
| TA | | -0.001 | -0.001 | -0.001 | 0.000 |
| | | (0.001) | (0.001) | (0.001) | (0.001) |
| Log(Amihud) | | -0.001 | -0.001 | -0.009*** | -0.005*** |
| | | (0.001) | (0.002) | (0.003) | (0.002) |
| IVOL | | -0.013*** | -0.013*** | -0.003* | -0.005*** |
| | | (0.001) | (0.003) | (0.002) | (0.002) |
| N | 89646 | 75500 | 75500 | 75200 | 75500 |
| Adj. $R^2$ | 0.003 | 0.039 | 0.039 | 0.067 | 0.039 |
| Start | 1985 | 1985 | 1985 | 1985 | 1985 |
| End | 2016 | 2016 | 2016 | 2016 | 2016 |
| Fixed effect | No | No | No | F & Q | FM |
| Clustered Std Error | No | No | F & Q | F & Q | No |

Table 5 Reaction of Analysts to Information Captured by 3PRF

This table reports regressions of future analyst revisions on the first predicted forecast errors ($PFE_f$) by 3PRF. The dependent variable is the analyst revision from one trading day after the calculation date of $PFE_f$ to the end of [EA-3], requiring that the end date is at least one week after the beginning date of the revision period. EA is earnings announcement day. EA-n represents n trading days before EA. Analyst revision is the difference between the number of positive forecast revisions and the number of negative forecast revisions scaled by total forecast revisions given the time period. FE is the analyst forecast error. Size is the market capitalization of the stock. BM is the ratio of book equity and market capitalization. Cum_alpha is the cumulative alpha, calculated using four factor model, from most recent fiscal quarter end to one trading day after calculate day of $PSUE_f$. Hist_alpha is the historical alpha of daily returns for one year before the most recent fiscal quarter end, calculated using three-factor model. TA is the ratio of total accrual scaled by total asset. IVOL is the idiosyncratic volatility using three-factor model. Amihud is the market adjusted illiquidity as in Amihud (2002). Intercept is not reported for abbreviation. Standard errors are reported in parentheses. All variables except the dependent variable are winzorized by 1% and standardized to mean 0 and standard deviation 1. The notations ***, ** and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. The sample is from Jan 1985 to Dec 2016.

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $PFE_f$ | 0.000 | 0.000 | 0.000 | -0.000 | 0.000 |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Lag(FE) | | 0.001** | 0.001* | 0.001 | 0.002 |
| | | (0.001) | (0.001) | (0.001) | (0.001) |
| Log(Size) | | -0.001 | -0.001 | -0.004 | 0.000 |
| | | (0.001) | (0.001) | (0.004) | (0.001) |
| Log(BM) | | -0.001 | -0.001 | 0.002 | -0.001 |
| | | (0.001) | (0.001) | (0.002) | (0.001) |
| Cum_alpha | | 0.005*** | 0.005*** | 0.004*** | 0.004*** |
| | | (0.001) | (0.001) | (0.001) | (0.001) |
| Hist_alpha | | 0.009*** | 0.009*** | 0.009*** | 0.008*** |
| | | (0.001) | (0.001) | (0.001) | (0.001) |
| TA | | -0.002*** | -0.002*** | -0.001 | -0.001 |
| | | (0.001) | (0.001) | (0.001) | (0.001) |
| Log(Amihud) | | -0.003*** | -0.003** | -0.008*** | -0.004*** |
| | | (0.001) | (0.001) | (0.002) | (0.002) |
| IVOL | | -0.005*** | -0.005*** | -0.002 | -0.002* |
| | | (0.001) | (0.001) | (0.001) | (0.001) |
| N | 49757 | 41459 | 41459 | 41015 | 41459 |
| Adj. $R^2$ | -0.000 | 0.007 | 0.007 | 0.018 | 0.007 |
| Start | 1985 | 1985 | 1985 | 1985 | 1985 |
| End | 2016 | 2016 | 2016 | 2016 | 2016 |
| Fixed effect | No | No | No | F & Q | FM |
| Clustered Std Error | No | No | F & Q | F & Q | No |

Table 6 Forecast of analyst forecast errors

This table reports regressions of analyst forecast errors on the first predicted forecast errors ($PFE_f$) by 3PRF. The dependent variable is the analyst forecast error, defined as the difference between realized quarterly earnings and analyst forecast median scaled by the corresponding fiscal quarter end price. Analyst forecast median is obtained at the end of one trading day after ($PFE_f$ calculation. Size is the market capitalization of the stock. BM is the ratio of book equity and market capitalization. Cum_alpha is the cumulative alpha, calculated using four factor model, from most recent fiscal quarter end to one trading day after calculate day of $PSUE_f$. Hist_alpha is the historical alpha of daily returns for one year before the most recent fiscal quarter end, calculated using three-factor model. TA is the ratio of total accrual scaled by total asset. IVOL is the idiosyncratic volatility using three-factor model. Amihud is the market adjusted illiquidity as in Amihud (2002). Intercept is not reported for abbreviation. Standard errors are reported in parentheses. All variables are winzorized by 1% and standardized to mean 0 and standard deviation 1. The notations ***, ** and * indicate statistical significance at the 1%, 5%, and 10% level, respectively. The sample is from Jan 1985 to Dec 2016.

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $PFE_f$ | 0.116*** | 0.071*** | 0.071*** | 0.049*** | 0.046*** |
| | (0.003) | (0.004) | (0.010) | (0.010) | (0.012) |
| Lag(FE) | | 0.181*** | 0.181*** | 0.093*** | 0.162*** |
| | | (0.004) | (0.014) | (0.013) | (0.011) |
| Log(Size) | | -0.009** | -0.009 | -0.024 | 0.014* |
| | | (0.004) | (0.011) | (0.030) | (0.008) |
| Log(BM) | | -0.073*** | -0.073*** | -0.157*** | -0.062*** |
| | | (0.004) | (0.011) | (0.018) | (0.009) |
| Cum_alpha | | 0.048*** | 0.048*** | 0.045*** | 0.056*** |
| | | (0.003) | (0.007) | (0.007) | (0.006) |
| Hist_alpha | | 0.121*** | 0.121*** | 0.123*** | 0.129*** |
| | | (0.003) | (0.008) | (0.008) | (0.009) |
| TA | | 0.010*** | 0.010** | 0.001 | 0.015*** |
| | | (0.003) | (0.004) | (0.005) | (0.005) |
| Log(Amihud) | | -0.071*** | -0.071*** | -0.110*** | -0.085*** |
| | | (0.005) | (0.011) | (0.017) | (0.009) |
| IVOL | | -0.124*** | -0.124*** | -0.097*** | -0.102*** |
| | | (0.004) | (0.017) | (0.011) | (0.009) |
| N | 97341 | 81003 | 81003 | 80734 | 81003 |
| Adj. $R^2$ | 0.013 | 0.096 | 0.096 | 0.151 | 0.09 |
| Start | 1985 | 1985 | 1985 | 1985 | 1985 |
| End | 2016 | 2016 | 2016 | 2016 | 2016 |
| Fixed effect | No | No | No | F & Q | FM |
| Clustered Std Error | No | No | F & Q | F & Q | No |

Table 7 Forecast of Earnings Announcement Return

This table presents regression results in forecasting earnings announcement returns (in percentage), which is the cumulative return in excess of market return from [EA-2] to [EA+2]. EA is earnings announcement day. EA-n (EA+n) represents n trading days before (after) EA. $PSUE_f$ is the first predicted SUE by 3PRF. $PSUE_a$ is predicted SUE using analyst forecasts available at the end of one trading day after the calculate day of $PSUE_f$. Size is the market capitalization of the stock. BM is the ratio of book equity and market capitalization. Cum_alpha is the cumulative alpha, calculated using four factor model, from most recent fiscal quarter end to one trading day after calculate day of $PSUE_f$. Hist_alpha is the historical alpha of daily returns for one year before the most recent fiscal quarter end, calculated using three-factor model. TA is the ratio of total accrual scaled by total asset. IVOL is the idiosyncratic volatility using three-factor model. Amihud is the market adjusted illiquidity as in Amihud (2002). Intercept is not reported for abbreviation. Standard errors are reported in parentheses. All variables except the depend variable are winzorized by 1% and standardized to mean 0 and standard deviation 1. The notations ***, ** and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| $PSUE_f$ | 0.119*** | 0.111*** | 0.111*** | 0.103*** | 0.085*** | 0.070** | 0.036 |
| | (0.020) | (0.025) | (0.028) | (0.029) | (0.032) | (0.034) | (0.045) |
| $PSUE_a$ | | | | | | 0.097** | 0.060 |
| | | | | | | (0.045) | (0.065) |
| Lag(SUE) | | 0.032 | 0.032 | -0.021 | 0.109*** | -0.155*** | -0.078 |
| | | (0.026) | (0.035) | (0.036) | (0.035) | (0.045) | (0.056) |
| Log(Size) | | 0.040 | 0.040 | -1.670*** | -0.024 | -1.977*** | -0.026 |
| | | (0.028) | (0.044) | (0.146) | (0.040) | (0.216) | (0.105) |
| Log(BM) | | 0.015 | 0.015 | 0.025 | 0.063 | -0.038 | -0.053 |
| | | (0.024) | (0.039) | (0.062) | (0.040) | (0.072) | (0.124) |
| Cum_alpha | | -0.119*** | -0.119*** | -0.154*** | -0.121*** | -0.131** | -0.209*** |
| | | (0.022) | (0.037) | (0.038) | (0.031) | (0.053) | (0.073) |
| Hist_alpha | | 0.049** | 0.049 | -0.123** | 0.065* | -0.197*** | 0.014 |
| | | (0.022) | (0.045) | (0.047) | (0.037) | (0.066) | (0.085) |
| TA | | -0.063*** | -0.063** | -0.077** | -0.073** | -0.024 | -0.121 |
| | | (0.022) | (0.028) | (0.031) | (0.033) | (0.037) | (0.124) |
| Log(Amihud) | | -0.109*** | -0.109* | -0.316*** | -0.151*** | -0.422*** | -0.081 |
| | | (0.026) | (0.060) | (0.074) | (0.053) | (0.107) | (0.122) |
| IVOL | | -0.059** | -0.059 | 0.032 | -0.104*** | 0.130 | -0.010 |
| | | (0.024) | (0.066) | (0.051) | (0.038) | (0.084) | (0.077) |
| N | 149652 | 131767 | 131767 | 131481 | 131767 | 83643 | 83941 |
| Adj. $R^2$ | 0.000 | 0.001 | 0.001 | 0.023 | 0.001 | 0.026 | 0.000 |
| Start | 1979 | 1979 | 1979 | 1979 | 1979 | 1985 | 1985 |
| End | 2016 | 2016 | 2016 | 2016 | 2016 | 2016 | 2016 |
| Fixed effect | No | No | No | F & Q | FM | F & Q | FM |
| Clustered Std Error | No | No | F & Q | F & Q | No | F & Q | No |

Table 8 Stock Characteristics for Stock Groups Formed by PSUE

This table reports quarterly average of stock characteristics (Panel A) and its cross-section percentile (Panel B) for five stock groups formed by $PSUE_f$ at each quarter. Size is market capitalization (in million dollars) of the stock. BM is the ratio of book equity and market capitalization. Return is the stock return of most recent month. MOM is the cumulative stock return of prior one year. Amihud is market adjusted illiquidity as in Amihud (2002) scaled by $10^6$. Turn is average daily turnover in the past six months. IVOL is idiosyncratic volatility using three-factor model. TVOL is total volatility. TA is total accrual scaled by total asset. The sample is from Jan 1979 to Dec 2016.

PanelA: quarterly average of stock characteristics

| $PSUE_f$ | Size | BM | Return | MOM | Amihud | Turn | TVOL | IVOL | TA |
|---|---|---|---|---|---|---|---|---|---|
| Low | 3,052 | 0.72 | 0.88% | 11.71% | 0.67 | 47.07% | 33.95% | 27.43% | 4.29% |
| 2 | 2,988 | 0.78 | 0.97% | 10.92% | 0.82 | 45.38% | 33.63% | 27.23% | 3.45% |
| 3 | 2,970 | 0.81 | 1.14% | 12.58% | 0.76 | 44.85% | 33.42% | 27.04% | 2.43% |
| 4 | 2,817 | 0.82 | 1.19% | 15.32% | 0.81 | 45.18% | 33.92% | 27.45% | 1.72% |
| High | 2,255 | 0.84 | 1.41% | 16.88% | 0.92 | 44.78% | 34.51% | 28.11% | 1.60% |

Panel B: quarterly average of cross-section stock characteristics percentile

| $PSUE_f$ | Size | BM | Return | MOM | Amihud | Turn | TVOL | IVOL | TA |
|---|---|---|---|---|---|---|---|---|---|
| Low | 70 | 49 | 51 | 53 | 30 | 55 | 34 | 33 | 54 |
| 2 | 69 | 53 | 52 | 52 | 31 | 54 | 33 | 32 | 52 |
| 3 | 69 | 55 | 52 | 53 | 31 | 54 | 33 | 32 | 49 |
| 4 | 68 | 55 | 52 | 55 | 32 | 54 | 34 | 33 | 47 |
| High | 67 | 56 | 53 | 56 | 34 | 53 | 35 | 34 | 47 |

Table 9 Market Reaction to Information Captured by $PSUE_f$

This table reports the event returns of stock groups formed by $PSUE_f$ quarterly before and after two time points: calculation day for PSUE_first and earnings announcement day. Event returns are presented in two forms of cumulative returns in percentage: raw return and alpha based on four factor model. Q represents the corresponding fiscal quarter end. [First + n] denotes n trading days after the calculation day of PSUE_first. EA is earnings announcement day. EA-(+)n represents n trading days before(after) EA. Column two presents quarterly average of realized SUE of each stock groups. The sample is from Jan 1979 to Dec 2016.

Panel A: value-weighted returns

| | | Q to [First+1] | | [First+2] to [EA-3] | | [EA-2] to [EA+2] | | [First+2] to [EA+2] | |
|---|---|---|---|---|---|---|---|---|---|
| $PSUE_f$ | SUE | Return | Alpha | Return | Alpha | Return | Alpha | Return | Alpha |
| Low | -0.58 | 0.49 | -0.26 | -0.07 | -0.25 | 0.03 | -0.05 | -0.02 | -0.25 |
| 2 | -0.32 | 0.42 | -0.30 | 0.07 | -0.10 | 0.30 | 0.17 | 0.37 | 0.10 |
| 3 | -0.04 | 0.87 | 0.03 | 0.14 | -0.10 | 0.29 | 0.19 | 0.44 | 0.10 |
| 4 | 0.28 | 0.68 | -0.11 | 0.16 | 0.03 | 0.21 | 0.04 | 0.37 | 0.07 |
| High | 0.49 | 0.75 | -0.17 | 0.11 | -0.04 | 0.21 | 0.09 | 0.33 | 0.06 |
| H-L | 1.07*** | 0.26 | 0.09 | 0.18* | 0.20* | 0.18 | 0.14 | 0.35** | 0.30* |
| T-stat | 11.89 | 1.38 | 0.46 | 1.75 | 1.85 | 1.59 | 1.25 | 2.26 | 1.79 |

Panel B: equal-weighted returns

| | | Q to [First +1] | | [First+2] to [EA-3] | | [EA-2] to [EA+2] | | [First+2] to [EA+2] | |
|---|---|---|---|---|---|---|---|---|---|
| $PSUE_f$ | SUE | Return | Alpha | Return | Alpha | Return | Alpha | Return | Alpha |
| Low | -0.77 | 0.45 | -0.22 | 0.13 | -0.15 | 0.12 | -0.01 | 0.22 | -0.16 |
| 2 | -0.46 | 0.52 | -0.17 | 0.22 | -0.06 | 0.26 | 0.12 | 0.45 | 0.05 |
| 3 | -0.14 | 0.53 | -0.13 | 0.15 | -0.14 | 0.25 | 0.12 | 0.39 | -0.02 |
| 4 | 0.22 | 0.64 | -0.08 | 0.25 | -0.01 | 0.33 | 0.19 | 0.54 | 0.17 |
| High | 0.50 | 0.63 | -0.10 | 0.24 | -0.07 | 0.56 | 0.38 | 0.77 | 0.31 |
| H-L | 1.27*** | 0.18* | 0.12 | 0.12 | 0.09 | 0.45*** | 0.39*** | 0.56*** | 0.47*** |
| T-stat | 22.28 | 1.75 | 1.43 | 1.37 | 1.25 | 6.82 | 5.73 | 4.92 | 4.16 |

Table 10 Market Reaction for Size Subsamples

This table reports the event returns of stock groups formed by $PSUE_f$ quarterly before and after two time points: calculation day for $PSUE_f$ and earnings announcement day. Two subsamples are presented separately. Panel A presents stock groups using stocks with size above the median calculated using NYSE stocks, while panel B shows results using stocks with size below the NYSE median. Event returns are presented in two forms of cumulative returns in percentage: raw return and alpha based on four factor model. [First + n] denotes n trading days after the calculation day of $PSUE_f$. EA is earnings announcement day. EA-(+)n represents n trading days before(after) EA. Column two presents quarterly average of realized SUE of each stock groups. The sample is from Jan 1979 to Dec 2016.

Panel A: value-weighted returns for firms of size above NYSE median

| | | Q to [First+1] | | [First+2] to [EA-3] | | [EA-2] to [EA+2] | | [First+2] to [EA+2] | |
|---|---|---|---|---|---|---|---|---|---|
| $PSUE_f$ | SUE | Return | FF4 Alpha | Return | FF4 Alpha | Return | FF4 Alpha | Return | FF4 Alpha |
| Low | -0.56 | 0.45 | -0.30 | -0.09 | -0.25 | 0.03 | -0.02 | -0.04 | -0.22 |
| 2 | -0.32 | 0.43 | -0.32 | 0.09 | -0.12 | 0.35 | 0.18 | 0.43 | 0.09 |
| 3 | -0.06 | 0.97 | 0.09 | 0.13 | -0.06 | 0.21 | 0.07 | 0.34 | 0.00 |
| 4 | 0.24 | 0.72 | -0.12 | 0.13 | 0.00 | 0.22 | 0.09 | 0.36 | 0.10 |
| High | 0.48 | 0.80 | -0.11 | 0.11 | -0.04 | 0.22 | 0.11 | 0.33 | 0.08 |
| H-L | 1.04*** | 0.35* | 0.18 | 0.20* | 0.22** | 0.20* | 0.12 | 0.38*** | 0.30* |
| T-stat | 10.83 | 1.71 | 0.80 | 1.78 | 1.91 | 1.74 | 1.05 | 2.57 | 1.85 |

Panel B: value-weighted returns for firms of size below NYSE median

| | | Q to [First+1] | | [First+2] to [EA-3] | | [EA-2] to [EA+2] | | [First+2] to [EA+2] | |
|---|---|---|---|---|---|---|---|---|---|
| $PSUE_f$ | SUE | Return | FF4 Alpha | Return | FF4 Alpha | Return | FF4 Alpha | Return | FF4 Alpha |
| Low | -0.69 | 0.34 | -0.26 | 0.25 | -0.10 | 0.00 | -0.13 | 0.25 | -0.22 |
| 2 | -0.40 | 0.40 | -0.16 | 0.17 | -0.15 | 0.21 | 0.05 | 0.39 | -0.09 |
| 3 | -0.06 | 0.41 | -0.18 | 0.05 | -0.26 | 0.07 | 0.01 | 0.15 | -0.23 |
| 4 | 0.27 | 0.49 | -0.16 | 0.23 | -0.06 | 0.29 | 0.17 | 0.53 | 0.12 |
| High | 0.52 | 0.55 | -0.11 | 0.30 | -0.01 | 0.44 | 0.25 | 0.73 | 0.23 |
| H-L | 1.20*** | 0.21 | 0.15 | 0.05 | 0.08 | 0.43*** | 0.38*** | 0.47*** | 0.44*** |
| T-stat | 19.28 | 1.46 | 1.14 | 0.33 | 0.72 | 3.77 | 3.73 | 2.81 | 3.09 |

Table 11 Trading Strategies on Information Spillover

This table reports daily abnormal returns in percentage to trading strategies that long(short) stocks with high(low) $PSUE_f$ at the end of one trading day after the calculation day of $PSUE_f$, and liquidate assets at the end of one trading day after the earnings announcement day. The long(short) side requires at least ten stocks to maintain the portfolio, or it will be liquidated in case of less than ten stocks. Require stock price $\geq$ \$5 on formation date. Panel A reports the average number of trading days for which one stock stays in one trading side. Panel B (Panel C) reports daily portfolio returns in excess of market returns and daily alphas using CAPM, Fama and French (1993) three factor model and Carhart (1997) four factor model, both value weighted (equal weighted). The first column indicates the model used. Column 2 to column 4 reports return or alpha of long side, short side and long minus short. The sample is from Jan 1979 to Dec 2016. The remaining columns report factor loadings of long minus short trading returns. Newey and West (1987) standard errors are used. The notations ***, ** and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Panel A: trading days

| | Summary of holding days for each stock | | | | | | |
|---|---|---|---|---|---|---|---|
| N of trading | Mean | STD | P5 | P25 | Median | P75 | P95 |
| 3,399 | 11 | 6 | 4 | 6 | 10 | 14 | 23 |

Panel B: value-weighted returns

| | Long | Short | L-S | Mktrf | SMB | HML | UMD |
|---|---|---|---|---|---|---|---|
| Ret-Mkt | 0.02 | -0.04 | 0.06*** | | | | |
| | (1.47) | (-2.51) | (3.03) | | | | |
| CAPM | 0.02 | -0.04 | 0.07*** | -0.03 | | | |
| | (1.58) | (-2.45) | (3.07) | (-1.00) | | | |
| FF3 | 0.02 | -0.04 | 0.07*** | -0.02 | 0.07 | 0.02 | |
| | (1.48) | (-2.55) | (3.05) | (-0.70) | (1.25) | (0.29) | |
| Carhart4 | 0.02 | -0.04 | 0.06*** | -0.01 | 0.07 | 0.08 | 0.11*** |
| | (1.46) | (-2.51) | (2.99) | (-0.23) | (1.21) | (1.24) | (2.85) |

Panel C: equal-weighted returns

| | Long | Short | L-S | Mktrf | SMB | HML | UMD |
|---|---|---|---|---|---|---|---|
| Ret-Mkt | 0.03 | -0.02 | 0.05*** | | | | |
| | (2.44) | (-1.63) | (3.93) | | | | |
| CAPM | 0.04 | -0.01 | 0.05*** | -0.03* | | | |
| | (3.23) | (-0.93) | (4.02) | (-1.93) | | | |
| FF3 | 0.03 | -0.01 | 0.05*** | -0.03 | 0.02 | 0.01 | |
| | (3.61) | (-1.45) | (4.01) | (-1.62) | (0.80) | (0.36) | |
| Carhart4 | 0.03 | -0.01 | 0.05*** | -0.02 | 0.02 | 0.05 | 0.06*** |
| | (3.64) | (-1.35) | (3.95) | (-1.12) | (0.79) | (1.40) | (2.86) |

Table 12 Trading Strategies on Earnings Announcement

This table reports the daily abnormal returns in percentage of trading strategies that long(short) stocks with high(low) $PSUE_l$ at the end of three trading day before earnings announcement day, and liquidate assets at the end of one trading day after the earnings announcement day. The long(short) side requires at least ten stocks to maintain the portfolio, or it will be liquidated in case of less than ten stocks. Require stock price higher than $5 on formation date. Panel A reports the average number of trading days for which one stock stays in one trading side. Panel B (Panel C) reports value weighted (equal weighted) daily portfolio returns in excess of market returns and daily alphas using CAPM, Fama and French (1993) three factor model and Carhart (1997) four factor model. The first column indicates the model used. Column 2 to column 4 reports return or alpha of long side, short side and long minus short. The sample is from Jan 1979 to Dec 2016. The remaining columns report factor loadings of long minus short trading returns. Newey and West (1987) standard errors are used. The notations ***, ** and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Panel A: trading days

| | Summary of holding days for each stock | | | | | | |
|---|---|---|---|---|---|---|---|
| N of trading | Mean | STD | P5 | P25 | Median | P75 | P95 |
| 2,607 | 4 | 1 | 2 | 4 | 4 | 4 | 4 |

Panel B: value-weighted returns

| | Long | Short | L-S | Mktrf | SMB | HML | UMD |
|---|---|---|---|---|---|---|---|
| Ret-Mkt | 0.05 | -0.03 | 0.08** | | | | |
| | (2.07) | (-1.01) | (2.17) | | | | |
| CAPM | 0.05 | -0.03 | 0.08** | -0.01 | | | |
| | (2.09) | (-1.01) | (2.19) | (-0.22) | | | |
| FF3 | 0.05 | -0.03 | 0.09** | -0.01 | 0.09 | -0.11 | |
| | (2.11) | (-1.11) | (2.28) | (-0.26) | (1.18) | (-1.09) | |
| Carhart4 | 0.05 | -0.03 | 0.08** | 0.02 | 0.08 | 0.02 | 0.25*** |
| | (2.08) | (-1.02) | (2.20) | (0.52) | (0.97) | (0.23) | (3.64) |

Panel C: equal-weighted returns

| | Long | Short | L-S | Mktrf | SMB | HML | UMD |
|---|---|---|---|---|---|---|---|
| Ret-Mkt | 0.09 | -0.01 | 0.11*** | | | | |
| | (4.70) | (-0.74) | (4.61) | | | | |
| CAPM | 0.10 | -0.01 | 0.11*** | -0.01 | | | |
| | (4.92) | (-0.50) | (4.62) | (-0.58) | | | |
| FF3 | 0.10 | -0.01 | 0.11*** | -0.02 | -0.01 | -0.11** | |
| | (5.53) | (-0.60) | (4.72) | (-0.85) | (-0.13) | (-2.17) | |
| Carhart4 | 0.10 | -0.01 | 0.11*** | 0.00 | -0.02 | -0.02 | 0.16*** |
| | (5.53) | (-0.49) | (4.70) | (0.17) | (-0.37) | (-0.44) | (4.67) |

Table 13 Trading Costs

This table reports daily performance in percentage of long minus short trading strategies with trading costs. Strategy 1 longs(shorts) stocks with high(low) $\mathrm{PSUE_f}$ at the end of one trading day after the calculation day of $\mathrm{PSUE_f}$, and liquidate assets at the end of one trading day after the earnings announcement day. Strategy 2 longs(shorts) stocks with high(low) $\mathrm{PSUE_l}$ at the end of three trading day before earnings announcement day, and liquidate assets at the end of one trading day after the earnings announcement day. The long(short) side requires at least ten stocks to maintain the portfolio, or it will be liquidated in case of less than ten stocks. Require stock price $\geq$ \$5 on formation date. Column weight indicates value-weighted returns (VW) or equal-weighted returns (EW). Column Cost listed trading costs assumption in which the number is the amount of costs in basis points for a round-trip trading. The remaining columns report daily strategy returns in excess of market returns and daily alphas using CAPM, Fama and French (1993) three factor model and Carhart (1997) four factor model. The sample is from Jan 1980 to Sep 2016. Newey and West (1987) standard errors are used. The notations ***, ** and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

| Strategy | Weight | Cost | Ret-Mkt | CAPM | FF3 | Carhart4 |
|---|---|---|---|---|---|---|
| 1 | VW | 0 | 0.06*** | 0.07*** | 0.07*** | 0.06*** |
| | | | (3.03) | (3.07) | (3.05) | (2.99) |
| 1 | VW | 5 | 0.05** | 0.05** | 0.05** | 0.05** |
| | | | (2.44) | (2.47) | (2.46) | (2.40) |
| 1 | VW | 10 | 0.04* | 0.04* | 0.04* | 0.04* |
| | | | (1.84) | (1.88) | (1.86) | (1.80) |
| 1 | EW | 0 | 0.05*** | 0.05*** | 0.05*** | 0.05*** |
| | | | (3.93) | (4.02) | (4.01) | (3.95) |
| 1 | EW | 5 | 0.04*** | 0.04*** | 0.04*** | 0.04*** |
| | | | (3.05) | (3.13) | (3.12) | (3.06) |
| 1 | EW | 10 | 0.03** | 0.03** | 0.03** | 0.03** |
| | | | (2.16) | (2.24) | (2.23) | (2.17) |
| 2 | VW | 0 | 0.08** | 0.08** | 0.09** | 0.09** |
| | | | (2.17) | (2.19) | (2.28) | (2.28) |
| 2 | VW | 5 | 0.05 | 0.05 | 0.06 | 0.05 |
| | | | (1.38) | (1.39) | (1.48) | (1.40) |
| 2 | VW | 10 | 0.02 | 0.02 | 0.03 | 0.02 |
| | | | (0.59) | (0.60) | (0.68) | (0.59) |
| 2 | EW | 0 | 0.11*** | 0.11*** | 0.11*** | 0.11*** |
| | | | (4.61) | (4.62) | (4.72) | (4.70) |
| 2 | EW | 5 | 0.08*** | 0.08*** | 0.08*** | 0.08*** |
| | | | (3.33) | (3.34) | (3.44) | (3.40) |
| 2 | EW | 10 | 0.05** | 0.05** | 0.05** | 0.05** |
| | | | (2.05) | (2.07) | (2.16) | (2.10) |

Table 14 Compare first and last predicted measures by 3PRF

This table reports predictive regressions of SUE, FE and AR on 3PRF measures. The first row indicates the dependent variable for the regression results shown in each column. SUE is the standardized unexpected earning. FE is analyst forecast error. AR is the cumulative return in excess of market return from [EA-2] to [EA+2]. EA is earnings announcement day. EA-n (EA+n) represents n trading days before (after) EA. The second row lists the type of 3PRF measures used in the regression. First (Last) is the first (last) predicted measure by 3PRF. PSUE is the predicted SUE by 3PRF. PFE is the predicted analyst forecast error by 3PRF. Size is the market capitalization of the stock. BM is the ratio of book equity and market capitalization. Cum_alpha is the cumulative alpha, calculated using four factor model, from most recent fiscal quarter end to one trading day after calculate day of $PSUE_f$. Hist_alpha is the historical alpha of daily returns for one year before the most recent fiscal quarter end, calculated using three-factor model. TA is the ratio of total accrual scaled by total asset. IVOL is the idiosyncratic volatility using three-factor model. Amihud is the market adjusted illiquidity as in Amihud (2002). Intercept is not reported for abbreviation. Standard errors are reported in parentheses. All variables except AR are winzorized by 1% and standardized to mean 0 and standard deviation 1. The notations ***, ** and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

| | SUE | | FE | | AR | |
|---|---|---|---|---|---|---|
| | First | Last | First | Last | First | Last |
| | 1 | 2 | 3 | 4 | 5 | 6 |
| PSUE | 0.094*** | 0.234*** | | | 0.103*** | 0.174*** |
| | (0.009) | (0.012) | | | (0.029) | (0.046) |
| PFE | | | 0.049*** | 0.082*** | | |
| | | | (0.010) | (0.012) | | |
| Lag(SUE) | 0.368*** | 0.229*** | | | -0.021 | -0.091* |
| | (0.009) | (0.010) | | | (0.036) | (0.054) |
| Lag(FE) | | | 0.093*** | 0.076*** | | |
| | | | (0.013) | (0.016) | | |
| Log(Size) | -0.023 | -0.043*** | -0.024 | -0.016 | -1.670*** | -1.863*** |
| | (0.014) | (0.016) | (0.030) | (0.037) | (0.146) | (0.179) |
| Log(BM) | -0.089*** | -0.100*** | -0.157*** | -0.182*** | 0.025 | 0.049 |
| | (0.008) | (0.010) | (0.018) | (0.020) | (0.062) | (0.087) |
| Cum_alpha | 0.040*** | 0.060*** | 0.045*** | 0.068*** | -0.154*** | -0.285*** |
| | (0.004) | (0.004) | (0.007) | (0.008) | (0.038) | (0.047) |
| Hist_alpha | 0.152*** | 0.151*** | 0.123*** | 0.124*** | -0.123** | -0.104 |
| | (0.005) | (0.005) | (0.008) | (0.008) | (0.047) | (0.065) |
| TA | -0.020*** | -0.028*** | 0.001 | -0.002 | -0.077** | -0.085** |
| | (0.004) | (0.004) | (0.005) | (0.006) | (0.031) | (0.039) |
| Log(Amihud) | -0.063*** | -0.070*** | -0.110*** | -0.151*** | -0.316*** | -0.271*** |
| | (0.008) | (0.009) | (0.017) | (0.018) | (0.074) | (0.095) |
| IVOL | -0.023*** | -0.022*** | -0.097*** | -0.105*** | 0.032 | -0.028 |
| | (0.004) | (0.005) | (0.011) | (0.015) | (0.051) | (0.065) |
| N | 131058 | 76384 | 80734 | 45115 | 131481 | 76651 |
| Adj. $R^2$ | 0.372 | 0.352 | 0.151 | 0.165 | 0.023 | 0.026 |
| Start | 1979 | 1979 | 1985 | 1985 | 1979 | 1979 |
| End | 2016 | 2016 | 2016 | 2016 | 2016 | 2016 |
| Fixed effect | Yes | Yes | Yes | Yes | Yes | Yes |
| Clustered Std Error | Yes | Yes | Yes | Yes | Yes | Yes |

Table 15 Subsample tests

This table presents regression results for two subsamples. The early sample is from Jan 1979 to Dec 2000 for the forecasting of standardized unexpected earnings(SUE) and earnings announcement returns(AR). The early sample starts from Jan 1986 for the forecasting of analyst forecast errors (FE). The later sample is from Jan 2001 to Dec 2016. $PSUE_f(PFE_f)$ is the firstly predicted SUE(FE) generated by 3PRF with at least 30 qualified early announcers. Size is the market capitalization of the stock. BM is the ratio of book equity and market capitalization. Cum_alpha is the cumulative alpha, calculated using four factor model, from most recent fiscal quarter end to one trading day after calculate day of $PSUE_f$. Hist_alpha is the historical alpha for one year before the most recent fiscal quarter end, calculated using three-factor model. TA is the ratio of total accrual divided by total asset. IVOL is idiosyncratic volatility using three-factor model. Amihud is market adjusted illiquidity as in Amihud (2002). Standard errors are reported in parentheses. All variables except AR are winzorized by 1% for each calendar quarter. All variables except AR are winzorized by 1% and standardized to mean 0 and standard deviation 1. AR is in percentage. The notations ***, ** and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

| | SUE | | FE | | AR | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| $PSUE_f$ | 0.084*** | 0.091*** | | | 0.052* | 0.126*** |
| | (0.010) | (0.012) | | | (0.028) | (0.040) |
| $PFE_f$ | | | 0.025* | 0.050*** | | |
| | | | (0.013) | (0.012) | | |
| Lag(SUE) | 0.369*** | 0.343*** | | | 0.093** | -0.122** |
| | (0.011) | (0.011) | | | (0.043) | (0.047) |
| Lag(FE) | | | 0.019 | 0.098*** | | |
| | | | (0.017) | (0.016) | | |
| Log(Size) | -0.102*** | -0.024 | -0.004 | -0.036 | -1.845*** | -2.577*** |
| | (0.031) | (0.023) | (0.052) | (0.046) | (0.243) | (0.259) |
| Log(BM) | -0.111*** | -0.104*** | -0.107*** | -0.193*** | 0.144 | -0.104 |
| | (0.015) | (0.012) | (0.022) | (0.024) | (0.102) | (0.091) |
| Cum_alpha | 0.054*** | 0.033*** | 0.056*** | 0.042*** | -0.161*** | -0.160*** |
| | (0.005) | (0.006) | (0.009) | (0.009) | (0.041) | (0.053) |
| Hist_alpha | 0.186*** | 0.143*** | 0.152*** | 0.114*** | -0.010 | -0.202*** |
| | (0.007) | (0.006) | (0.013) | (0.010) | (0.054) | (0.064) |
| TA | -0.031*** | -0.022*** | 0.014 | -0.002 | -0.112** | -0.069* |
| | (0.006) | (0.004) | (0.010) | (0.006) | (0.050) | (0.039) |
| Log(Amihud) | -0.097*** | -0.080*** | -0.063** | -0.144*** | -0.354*** | -0.443*** |
| | (0.013) | (0.011) | (0.028) | (0.020) | (0.089) | (0.115) |
| IVOL | -0.039*** | -0.013** | -0.090*** | -0.095*** | -0.051 | 0.116 |
| | (0.007) | (0.005) | (0.013) | (0.014) | (0.060) | (0.074) |
| N | 45879 | 85039 | 20934 | 59675 | 46222 | 85112 |
| Adj. $R_2$ | 0.422 | 0.349 | 0.142 | 0.160 | 0.042 | 0.024 |
| Start | 1979 | 1979 | 1985 | 1985 | 1979 | 1979 |
| End | 2016 | 2016 | 2016 | 2016 | 2016 | 2016 |
| Fixed effect | Yes | Yes | Yes | Yes | Yes | Yes |
| Clustered Std Error | Yes | Yes | Yes | Yes | Yes | Yes |

Table 16 Robustness Tests of SUE Forecast

This table presents forecasting regressions of SUE on the 3PRF measure controlling additional information channels as listed in column 1. Panel A uses the value-weighted average of early announcers SUEs belonging to different informed groups. Panel B uses the value-weighted average of five-day cumulative returns from [First-3, First+1] of different informed group. First -(+) n denotes n trading days prior (post) to the estimation day of $PSUE_f$. CUSIND (SUPIND) is a group of stocks that belong to the customer(supplier) industries of the forecasting target as defined in Menzly and Ozbas (2010). BIG is a group of stocks with size above 66th percentile among NYSE stocks. IND is a group of stocks within same industry with the forecast target as defined by Fama and French (1997) 48 industries. Analyst is the group of stocks covered by analysts. IOR is the group of stocks with institutional ownership above 66th percentile among NYSE stocks. TURN is the group of stocks with turnover above 66th percentile among NYSE stocks. Other controls are lag(SUE), log(size), log(BM), log(Cum_alpha), log(Hist_alpha), TA, IVOL, log(Amihud). The sample is from Jan 1979 to Dec 2016. All variables are winzorized by 1% and standardized to mean 0 and standard deviation 1.

Panel A: controls of value-weighted average of SUEs

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $PSUE_f$ | 0.094*** | 0.092*** | 0.092*** | 0.094*** | 0.093*** | 0.094*** | 0.088*** | 0.091*** |
| | (0.009) | (0.009) | (0.009) | (0.009) | (0.009) | (0.009) | (0.009) | (0.009) |
| IND_SUE | | 0.058*** | | | | | | 0.034*** |
| | | (0.006) | | | | | | (0.006) |
| CUSIND_SUE | | | 0.054*** | | | | | 0.036*** |
| | | | (0.005) | | | | | (0.006) |
| SUPIND_SUE | | | 0.016** | | | | | 0.009 |
| | | | (0.008) | | | | | (0.008) |
| BIG_SUE | | | | 0.040* | | | | 0.033 |
| | | | | (0.023) | | | | (0.034) |
| IOR_SUE | | | | | 0.010 | | | -0.003 |
| | | | | | (0.019) | | | (0.028) |
| TURN_SUE | | | | | | -0.001 | | -0.022 |
| | | | | | | (0.025) | | (0.026) |
| Analyst_SUE | | | | | | | 0.038* | |
| | | | | | | | (0.023) | |

Panel B: controls of value-weighted average of five-day cumulative returns from [First-3, First+1]

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $PSUE_f$ | 0.094*** | 0.094*** | 0.094*** | 0.094*** | 0.093*** | 0.094*** | 0.088*** | 0.094*** |
| | (0.009) | (0.009) | (0.009) | (0.009) | (0.009) | (0.009) | (0.009) | (0.009) |
| IND_Ret | | -0.013** | | | | | | -0.011* |
| | | (0.006) | | | | | | (0.006) |
| CUSIND_Ret | | | -0.012** | | | | | -0.006 |
| | | | (0.006) | | | | | (0.007) |
| SUPIND_Ret | | | 0.000 | | | | | -0.001 |
| | | | (0.008) | | | | | (0.011) |
| BIG_Ret | | | | -0.004 | | | | 0.008 |
| | | | | (0.006) | | | | (0.011) |
| IOR_Ret | | | | | -0.005 | | | |
| | | | | | (0.006) | | | |
| TURN_Ret | | | | | | -0.004 | | |
| | | | | | | (0.006) | | |
| Analyst_Ret | | | | | | | -0.005 | |
| | | | | | | | (0.005) | |

Table 17 Robustness Tests of FE Forecast

This table presents forecasting regressions of FE on the 3PRF measure controlling additional information channels as listed in column 1. Panel A uses the value-weighted average of early announcers FEs belonging to different informed groups. Panel B uses the value-weighted average of five-day cumulative returns from [First-3, First+1] of different informed group. First -(+) n denotes n trading days prior (post) to the estimation day of $PFE_f$. CUSIND (SUPIND) is a group of stocks that belong to the customer(supplier) industries of the forecasting target as defined in Menzly and Ozbas (2010). BIG is a group of stocks with size above 66th percentile among NYSE stocks. IND is a group of stocks within same industry with the forecast target as defined by Fama and French (1997) 48 industries. Analyst is the group of stocks covered by analysts. IOR is the group of stocks with institutional ownership above 66th percentile among NYSE stocks. TURN is the group of stocks with turnover above 66th percentile among NYSE stocks. Other controls are lag(SUE), log(size), log(BM), log(Cum_alpha), log(Hist_alpha), TA, IVOL, log(Amihud). The sample is from Jan 1985 to Dec 2016. All variables are winzorized by 1% and standardized to mean 0 and standard deviation 1.

Panel A: controls of value-weighted average of FEs

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $PFE_f$ | 0.049*** | 0.048*** | 0.048*** | 0.049*** | 0.048*** | 0.048*** | 0.049*** | 0.048*** |
| | (0.010) | (0.010) | (0.010) | (0.010) | (0.010) | (0.010) | (0.010) | (0.010) |
| IND_FE | | 0.047*** | | | | | | 0.014 |
| | | (0.009) | | | | | | (0.009) |
| CUSIND_FE | | | 0.062*** | | | | | 0.054*** |
| | | | (0.013) | | | | | (0.014) |
| SUPIND_FE | | | 0.005 | | | | | 0.001 |
| | | | (0.010) | | | | | (0.011) |
| BIG_FE | | | | -0.003 | | | | 0.006 |
| | | | | (0.033) | | | | (0.039) |
| IOR_FE | | | | | 0.076 | | | 0.067 |
| | | | | | (0.084) | | | (0.088) |
| TURN_FE | | | | | | -0.015 | | -0.049 |
| | | | | | | (0.052) | | (0.067) |
| Analyst_FE | | | | | | | -0.005 | |
| | | | | | | | (0.041) | |

Panel B: controls of value-weighted average of five-day cumulative returns from [First-3, First+1]

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $PFE_f$ | 0.049*** | 0.049*** | 0.050*** | 0.048*** | 0.048*** | 0.048*** | 0.048*** | 0.050*** |
| | (0.010) | (0.010) | (0.010) | (0.010) | (0.010) | (0.010) | (0.010) | (0.010) |
| IND_Ret | | 0.006 | | | | | | -0.010 |
| | | (0.010) | | | | | | (0.009) |
| CUSIND_Ret | | | 0.021 | | | | | 0.027** |
| | | | (0.014) | | | | | (0.014) |
| SUPIND_Ret | | | -0.013 | | | | | -0.011 |
| | | | (0.012) | | | | | (0.020) |
| BIG_Ret | | | | 0.006 | | | | 0.002 |
| | | | | (0.014) | | | | (0.024) |
| IOR_Ret | | | | | 0.006 | | | |
| | | | | | (0.014) | | | |
| TURN_Ret | | | | | | 0.003 | | |
| | | | | | | (0.015) | | |
| Analyst_Ret | | | | | | | 0.006 | |
| | | | | | | | (0.014) | |

Table 18 Robustness Tests of AR Forecast

This table presents forecasting regressions of AR on the 3PRF measure controlling additional information channels as listed in column 1. Panel A uses the value-weighted average of early announcers ARs belonging to different informed groups. Panel B uses the value-weighted average of five-day cumulative returns from [First-3, First+1] of different informed group. First -(+) n denotes n trading days prior (post) to the estimation day of $PSUE_f$. CUSIND (SUPIND) is a group of stocks that belong to the customer(supplier) industries of the forecasting target as defined in Menzly and Ozbas (2010). BIG is a group of stocks with size above 66th percentile among NYSE stocks. IND is a group of stocks within same industry with the forecast target as defined by Fama and French (1997) 48 industries. Analyst is the group of stocks covered by analysts. IOR is the group of stocks with institutional ownership above 66th percentile among NYSE stocks. TURN is the group of stocks with turnover above 66th percentile among NYSE stocks. Other controls are lag(SUE), log(size), log(BM), log(Cum_alpha), log(Hist_alpha), TA, IVOL, log(Amihud). The sample is from Jan 1979 to Dec 2016. All variables are winzorized by 1% and standardized to mean 0 and standard deviation 1.

Panel A: controls of value-weighted average of ARs

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $PSUE_f$ | 0.103*** | 0.101*** | 0.107*** | 0.103*** | 0.107*** | 0.104*** | 0.102*** | 0.110*** |
| | (0.029) | (0.029) | (0.030) | (0.029) | (0.030) | (0.029) | (0.030) | (0.030) |
| IND_AR | | 0.072** | | | | | | 0.040 |
| | | (0.028) | | | | | | (0.033) |
| CUSIND_AR | | | 0.072** | | | | | 0.049 |
| | | | (0.033) | | | | | (0.038) |
| SUPIND_AR | | | 0.019 | | | | | 0.012 |
| | | | (0.042) | | | | | (0.042) |
| BIG_AR | | | | -0.018 | | | | -0.100 |
| | | | | (0.085) | | | | (0.096) |
| IOR_AR | | | | | -0.119 | | | -0.111 |
| | | | | | (0.144) | | | (0.150) |
| TURN_AR | | | | | | 0.182* | | 0.200* |
| | | | | | | (0.096) | | (0.108) |
| Analyst_AR | | | | | | | 13.068 | |
| | | | | | | | (26.188) | |

Panel B: controls of value-weighted average of five-day cumulative returns from [First-3, First+1]

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| PSUEf | 0.103*** | 0.101*** | 0.107*** | 0.104*** | 0.107*** | 0.104*** | 0.102*** | 0.105*** |
| | (0.029) | (0.029) | (0.030) | (0.029) | (0.030) | (0.029) | (0.030) | (0.030) |
| IND_Ret | | 0.003 | | | | | | 0.006 |
| | | (0.052) | | | | | | (0.077) |
| CUSIND_Ret | | | 0.035 | | | | | 0.032 |
| | | | (0.061) | | | | | (0.077) |
| SUPIND_Ret | | | -0.057 | | | | | -0.056 |
| | | | (0.066) | | | | | (0.082) |
| BIG_Ret | | | | -0.017 | | | | -0.002 |
| | | | | (0.055) | | | | (0.101) |
| IOR_Ret | | | | | -0.006 | | | |
| | | | | | (0.054) | | | |
| TURN_Ret | | | | | | -0.032 | | |
| | | | | | | (0.053) | | |
| Analyst_Ret | | | | | | | -0.015 | |
| | | | | | | | (0.059) | |

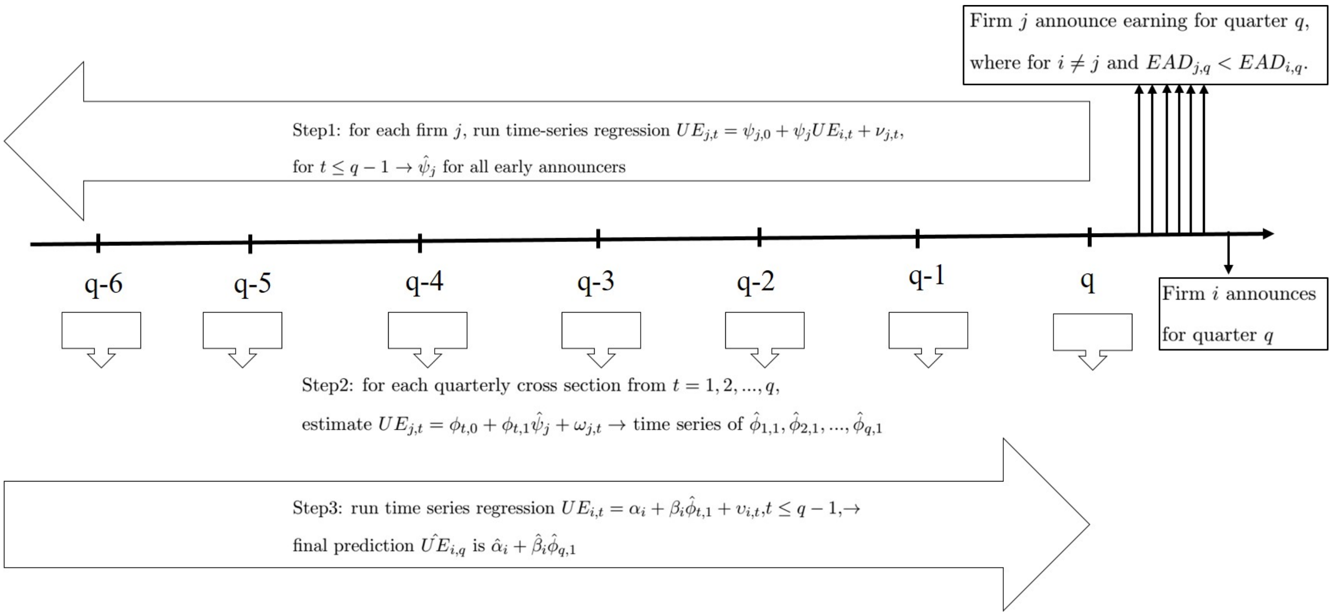


Figure 1 3PRF estimations for one factor assumption. This figures shows three-step OLS estimations of 3PRF in the prediction of unexpected earning(UE) for firm $i$, quarter $q$. Assume one common latent factor for UEs of firm $i$ and other firms that announce before firm $i$. For each firm and quarter, such procedures are performed to get out-of-sample forecasts of the corresponding predicted UEs.

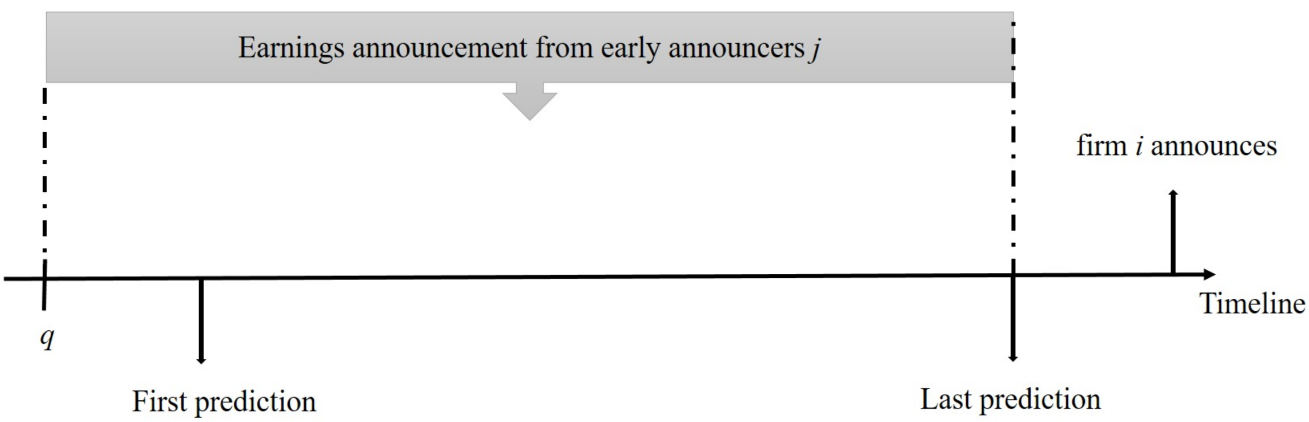


Figure 2 Timeline of 3PRF estimations for quarter $q$ earning of firm $i$. This figure presents a typical estimation timeline for firm $i$ from its fiscal quarter end $q$ to the corresponding earnings announcement day. First prediction is the earliest date on which more than 30 qualified firms have announced. Last prediction is four trading days before the announcement day for firm $i$.

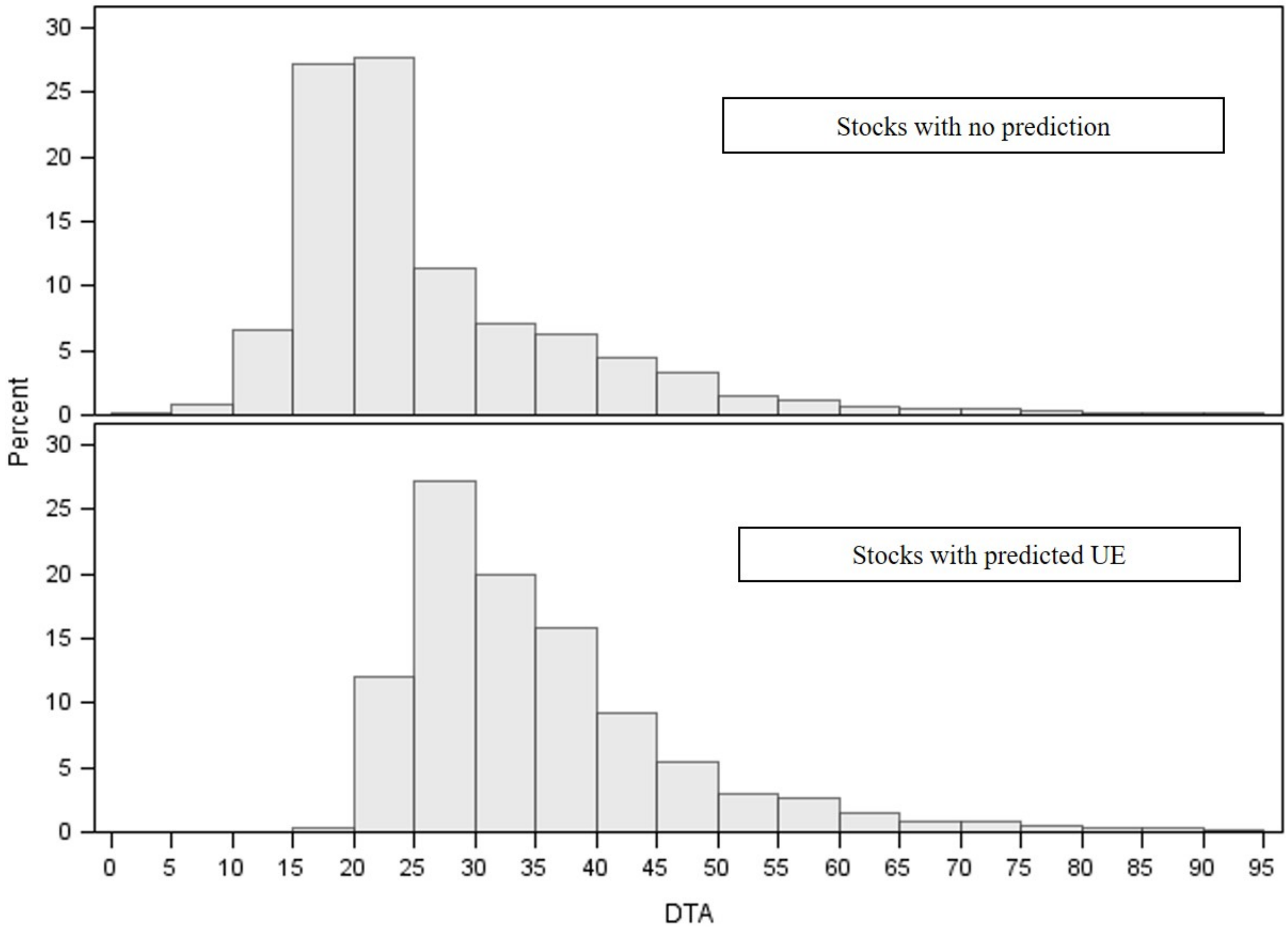


Figure 3 Days to announcement. This figure plots histograms of day-gaps between fiscal quarter end and earnings announcement day. The upper plot is the histogram for firms with no predictions by 3PRF, while the lower graph is the one for firms with 3PRF predictions of UEs.

# References


Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. Journal of Financial Markets 5, 31–56.

Badrinath, S. G., Kale, J. R., Noe, T. H., 1995. Of Shepherds, Sheep, and the Cross-autocorrelations in Equity Returns. The Review of Financial Studies 8, 401–430.

Bradshaw, M. T., Sloan, R. G., 2002. GAAP versus The Street: An Empirical Assessment of Two Alternative Definitions of Earnings. Journal of Accounting Research 40, 41–66.

Brennan, M. J., Jegadeesh, N., Swaminathan, B., 1993. Investment Analysis and the Adjustment of Stock Prices to Common Information. The Review of Financial Studies 6, 799–824.

Carhart, M. M., 1997. On Persistence in Mutual Fund Performance. Journal of Finance 52, 57–82.

Chordia, T., Swaminathan, B., 2000. Trading Volume and Cross-Autocorrelations in Stock Returns. The Journal of Finance 55, 913–935.

Cohen, L., Frazzini, A., 2008. Economic Links and Predictable Returns. The Journal of Finance 63, 1977–2011.

Da, Z., Warachka, M. C., 2009. Cashflow risk, systematic earnings revisions, and the cross-section of stock returns. Journal of Financial Economics 94, 448–468.

Fama, E. F., French, K. R., 1993. Common risk factors in the returns on stocks and bonds. Journal of financial economics 33, 3–56.

Fama, E. F., French, K. R., 1997. Industry costs of equity. Journal of financial economics 43, 153–193.

Fama, E. F., French, K. R., 2000. Forecasting Profitability and Earnings. The Journal of Business 73, 161–175.

Foster, G., 1981. Intra-industry information transfers associated with earnings releases. Journal of accounting and economics 3, 201–232.

Hirshleifer, D., Lim, S. S., Teoh, S. H., 2009. Driven to Distraction: Extraneous Events and Underreaction to Earnings News. The Journal of Finance 64, 2289–2325.

Hong, H., Stein, J. C., Yu, J., 2007. Simple Forecasts and Paradigm Shifts. The Journal of Finance 62, 1207–1242.

Hou, K., Xue, C., Zhang, L., 2015. Digesting Anomalies: An Investment Approach. Review of Financial Studies 28, 650–705.

Jegadeesh, N., Kim, J., Krische, S. D., Lee, C., 2004. Analyzing the analysts: When do recommendations add value? The journal of finance 59, 1083–1124.

Kelly, B., Pruitt, S., 2013. Market expectations in the cross-section of present values. The Journal of Finance 68, 1721–1756.

Kelly, B., Pruitt, S., 2015. The three-pass regression filter: A new approach to forecasting using many predictors. Journal of Econometrics 186, 294–316.

Kothari, S. P., 2001. Capital markets research in accounting. Journal of Accounting and Economics 31, 105–231.

Light, N., Maslov, D., Rytchkov, O., 2017. Aggregation of Information About the Cross Section of Stock Returns: A Latent Variable Approach. Review of Financial Studies p. hhw102.

Livnat, J., Mendenhall, R. R., 2006. Comparing the postearnings announcement drift for surprises calculated from analyst and time series forecasts. Journal of Accounting Research 44, 177–205.

Lo, A. W., MacKinlay, A. C., 1990. Data-Snooping Biases in Tests of Financial Asset Pricing Models. The Review of Financial Studies 3, 431–467.

Menzly, L., Ozbas, O., 2010. Market Segmentation and Cross-predictability of Returns. The Journal of Finance 65, 1555–1580.

Miller, E. M., 1977. Risk, Uncertainty, and Divergence of Opinion. The Journal of Finance 32, 1151–1168.

Newey, W. K., West, K. D., 1987. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. Econometrica 55, 703–708.

Peng, L., Xiong, W., 2006. Investor attention, overconfidence and category learning. Journal of Financial Economics 80, 563–602.

Petersen, M. A., 2009. Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches. The Review of Financial Studies 22, 435–480.

Shleifer, A., Vishny, R. W., 1997. The Limits of Arbitrage. The Journal of Finance 52, 35–55.

Stock, J. H., Watson, M. W., 2002. Forecasting Using Principal Components From a Large Number of Predictors. Journal of the American Statistical Association 97, 1167–1179.

Tetlock, P. C., Saar-Tsechansky, M., Macskassy, S., 2008. More Than Words: Quantifying Language to Measure Firms' Fundamentals. The Journal of Finance 63, 1437–1467.