Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

______________________________ _______________

Zhenping Wang Date

Two Essays in Financial Economics

By

Zhenping Wang
Doctor of Philosophy

Business

Jay Shanken   Tarun Chordia
Advisor

T. Clifton Green
Committee Member

Vivian Zhanwei Yue
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Two Essays in Financial Economics

By

Zhenping Wang

M.S., Columbia University, 2010

B.A., Harbin Institute of Technology, 2008

Advisor: Jay Shanken, Ph.D., Carnegie-Mellon University, 1983

Advisor: Tarun Chordia, Ph.D., University of California, Los Angeles, 1993

An abstract of

A dissertation submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in Business

2018

Abstract

Two Essays in Financial Economics

By Zhenping Wang

This dissertation explores new aspects of two empirical issues in financial economics: earnings predictability and informed trading. In the first essay (“Forecasting earnings from early announcers: a latent factor approach”), I propose a new method to predict non-announcing firms' earnings news using the cross section of all available early announcers' earnings news, the number of which can be as large as thousands. The method assumes common latent factors driving the earnings news 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 the earnings surprise and the earnings announcement return with both statistical and economic significance. A long-short trading strategy based on the extracted information realizes a 10% alpha annually, indicating a delayed reaction of investors. The return predictability is stronger for small firms or firms with less investor attention. Controlling a series of documented information channels has little impact on the predictive power of this extracted measure. In the second essay (“Option trading leverage and stock returns”), I construct a new measure to capture informed trading in the equity option market. Informed investors choose optimal leverage as a function of risk and cost. In the option market, they establish the chosen exposure using options of low leverage which are more liquid. On the other hand, noise traders prefer options of high leverage which have more lottery-like payoff. Thus, I examine the volume-weighted average of option leverage named Option Trading Leverage Ratio (OTLR). Portfolios with the lowest OTLR underperform portfolios with the highest OTLR by 1.28% monthly or 15% annually. The predictive relationship between OTLR and earnings-related information indicates that better skills in processing public information contribute to the advantage of informed option investors. Empirical findings are consistent with low OTLR capturing negative informed option trading due to the short-sale constraints of the stock market.

Two Essays in Financial Economics

By

Zhenping Wang

M.S., Columbia University, 2010

B.A., Harbin Institute of Technology, 2008

Advisor: Jay Shanken, Ph.D., Carnegie-Mellon University, 1983

Advisor: Tarun Chordia, Ph.D., University of California, Los Angeles, 1993

A dissertation submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in Business

2018

Acknowledgments

I would like to express my deep gratitude to my dissertation committee members: Jay Shanken, Tarun Chordia, Clifton Green and Zhanwei Yue. In particular, I would like to thank my co-chairs Jay and Tarun for their encouragement and support. As outstanding researchers in different fields such as asset pricing, behavior finance, and economics, my committee members have provided me insightful suggestions from the beginning of my papers. In addition, I obtained many invaluable pieces of advice about the job market and the career development in the academia from my committee members. I would also like to thank all the other faculties in the department of finance at Emory University. Especially, I am grateful for all the constructive feedback from Narasimhan Jegadeesh.

I am thankful for the great administrative work of Allison Gilmore and for the data support of Ron Harris. My thanks also go to the Emory alumni who gave me invaluable input on the papers and the job market: Seoyoung Kim, Qing Tong, Quan Wen, Dexin Zhou, Shikha Jaiswal, Kiseo Chung and Badrinath Kottimukkalur. Also, I would like to thank all the Ph.D. students at Emory who provide me with many helpful suggestions on the papers and my presentation skill.

At last, I would like to give my deepest gratitude to my parents: Kefei Wang and Jia Liu. Without their support and encouragement, I will not make it to this dissertation. I get two most wonderful gifts from my parents: love and freedom.

Table of Contents

List of Tables

List of Figures

First Essay: Forecasting Earnings from Early Announcers: A Latent Factor Approach

Zhenping Wang

Emory University

Abstract

I propose a new method to predict non-announcing firms' earnings news using the cross section of all available early announcers' earnings news, the number of which can be as large as thousands. The method assumes common latent factors driving the earnings news 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 the earnings surprise and the earnings announcement return with both statistical and economic significance. A long-short trading strategy based on the extracted information realizes a 10% alpha annually, indicating a delayed reaction of investors. The return predictability is stronger for small firms or firms with less investor attention. Controlling a series of documented information channels has little impact on the predictive power of this extracted measure.

1 Introduction

Thousands of firms release their earnings quarterly for the same operating period but on different dates. How can we update earnings expectations for late announcers by learning from numerous announced earnings of early announcers? The literature usually simplifies all announced earnings to an aggregated measure such as the average value. Instead, I apply a new method to more effectively extract the information relevant with the earnings news of each non-announcing firm from the cross section of early announcers' earnings news. 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 both the earnings surprise and the announcement return with both statistical and economic significance.

Using all the 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 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), attempt to forecast non-announcers' earnings using the average earnings news of early announcers belonging to the same industry. In this setting, the aggregation measure is intended to capture the industry news. The effectiveness of this measure highly depends on the relationship between the forecasting target and the predictors. For example, the launch of 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 makers such as Nokia and BlackBerry. When Apple and Nokia are early announcers, their earnings news may offset each other of which the average is a poor proxy for the industry news. 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 the common information using a wealth of independent variables. In this setting, earnings news 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 average measure, the technology shock is extracted from the cross section of early announcers earnings news given that earnings news will load differently on the factor. As presented in the empirical part of this study, most forecasting targets, or the earnings news of late announcers, are assembled by two latent factors that are common with the earnings news of early announcers. A simple average of all announced earnings news would not completely capture 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 the specific economic links, the factor structure assumed by 3PRF captures the statistical covariance among individual firms earnings. The extracted measure from 3PRF sums up all types of potential economic links. Specifically, I find that the predicted earnings news from 3PRF forecast the earnings surprise with both statistical and economic significance, with a series of stock characteristics controlled. For the 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 the analyst forecast error (FE), another measure of earnings surprise, one standard deviation change in the 3PRF measure predicts a 10% 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 estimate the information using 3PRF with sufficient early announcers, investors may have already incorporated 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 the earnings announcement return. One standard deviation change of the extracted earnings news predicts a change of 23 basis points in the same direction for the earnings announcement return in excess of market return, which is approximately 11% annualized. Furthermore, an event study reveals that investors do not react to the news contained by the 3PRF measure substantially before the earnings announcement. A long-short weekly-rebalanced trading strategy that exploits the return predictability earns a 10% 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 for one third of the sample period from January1981 to December 2016. These findings indicate that the 3PRF measure captures valuable information before the earnings announcement. Also, a further investigation reveals that the return predictability is stronger for small firms, or firms with less investor attention proxied by the institutional ownership, analyst coverage and trading turnover.

A vital question is the degree of the additional contribution by the 3PRF measure that extracts information from the early announcers' earnings news 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 higher 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 announced earnings surprise 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 the 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 the principal component analysis (PCA), which has widespread applications, such as the macroeconomic time series forecasting (Stock and Watson, 2002). The PCA approach selects the 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 the predictors and the target, thus establishing a more connected predictive 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.

[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.

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 the 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 the pooled cross section regression to capture the average relationship between predictors and earnings. For example, Fama and French (2000) examine the earnings predictability using a series of stock characteristics. The pooled cross section regression waives the requirement of a 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 to the regression is easy due to the large sample size. As a result, I use the pooled cross section regression method in this study.

The information spillover literature examines whether investors react to different types of information in a timely and properly manner. If the market is inefficient, investors may have a delayed reaction to public signals or misreact to such information[3]. This paper is closely related to the studies 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 the information diffusion among stocks. Rather than search for various economic links between firms, I take advantage of their statistical links of earnings news, 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 with 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

[2]See Kothari (2001)

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

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.

2 Methodology

This section discusses assumptions, such as the latent factor structure of earnings news, 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. $EA_{i,q}$ 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 unexpected 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 $EA_{j,q} < EA_{i,q}$.

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

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

$\mathbf{F}_q$ 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

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

Thus, predictors' or early announcers' earnings news is expanded by the same latent factors of firm i's earnings news, 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 $UE_{i,q}$[4].

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 news. If such is the case, the power of 3PRF is diminished. The empirical section shows that the 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 fundamentals 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) attempt 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, as in PLS, or a group of new variables motivated by the economics theory. The authors develop two methods to estimate 3PRF. The first is a closed-form solution derived from an equivalent constrained least

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

squares problem under the assumption that only the 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 $EA_{j,q} < EA_{i,q}$, 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

$$UE_{j,t} = \psi_{j,0} + \psi_{j,1} UE_{i,t} + \nu_{j,t} \tag{4}$$

with all the available common histories for $UE_{i,t}$ and $UE_{j,t}$ before quarter q or $t \leq q-1$. I require a minimum length of 20 quarters' history in this regression. $UE_{i,t}$ is used as a proxy for the common latent factor. The estimated loading $\hat{\psi}_{j,1}$ captures the degree of dependence of $UE_{j,t}$ on the common factor.

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

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

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 100 predictors are enforced for each cross section estimation.

In the last step, the time series regression

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

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

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 the 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.

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 the 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 the out-of-sample implementation of 3PRF in that all estimations are calculated on the samples before the release date of the forecast target $UE_{i,q}$. In addition, not all individual firms have correlated earnings given that earnings of individual firms are usually very idiosyncratic; 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 the relevant factors. This filter requires that the estimated loadings of regressing each predictor on the target, as in the first step of one latent factor model, are 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 for different 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 has the lowest Bayesian Information Criterion (BIC) as calculated in Krmer and Sugiyama (2011), in which an unbiased estimate of the Degrees of Freedom of partial least squares is derived. Specifically, for the m-factor model in predicting $UE_{i,q}$, the Degrees of Freedom are

$$D\hat{O}F(m) = 1 + \sum_{k=1}^{m} c_k trace(\mathbf{P}^k) - \sum_{l,k=1}^{m} \mathbf{t}_l^{'}\mathbf{P}^k\mathbf{t}_l + (\mathbf{y} - \hat{\mathbf{y}}_m)^{'} \sum_{k=1}^{m} \mathbf{P}^k\mathbf{v}_k + m \tag{7}$$

where $\mathbf{P} = \mathbf{X}\mathbf{X}^{'}$. $\mathbf{X}$ is the (q-1)×N matrix of predictors, or the history of N early announcers' earnings news. c_k is the kth element of the vector $\mathbf{B}^{-1}\mathbf{T}\mathbf{y}$. $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, ..., \mathbf{t}_m]$ is the (q-1)×m matrix of the estimated m latent factors. $\mathbf{y}$ is the vector of earnings news of firm i before quarter q. $\mathbf{B}$ is the Krylov basis decomposition $(\langle \mathbf{t}_l, \mathbf{P}^k\mathbf{y} \rangle) \in R^{m\times m}$. $\mathbf{v}_k$ is the kth column of the matrix $\mathbf{T}(\mathbf{B}^{-1})$. $\hat{\mathbf{y}}_m$is the vector of fitted values of $\mathbf{y}$. The BIC is

$$BIC(m) = \sum_{s=1}^{q-1} (y_s - \hat{y}_{m,s})/(q-1) + log(q-1)\hat{\sigma}^2 D\hat{O}F(m)/(q-1) \tag{8}$$

where $\hat{\sigma} = \sqrt{\sum_{s=1}^{q-1}(y_s - \hat{y}_{m,s})/(q-1-D\hat{O}F(m))}$.

3 Data and summary statistics

3.1 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 1981 to December 2016[5]. Analyst forecasts are from Institutional Brokers' Estimate System (IBES) provided by Thomson and Reuters. Because of data availability, tests involving analyst forecasts are performed on a sample from January 1985 to December 2016.

Section 2 describes the 3PRF estimation procedures to acquire $\hat{UE}_{i,q}$. $\hat{UE}_{i,q}$ is estimated using all the available early announcers' earnings news at the end of four trading days before the

[5]The quarterly accounting variables of Compustat can be traced back to the 1960s. In the estimation of 3PRF, a minimum length of 20-quarter history is required for the first step, and a minimum number of 100 qualified early announcers are required for the second step. Given such conditions, the estimation can not be performed for a large proportion of the cross section before 1981. As a result, I use the sample from January 1981.

announcement day or at the end of two trading days before the time window of announcement returns. This additional one trading day gap between the estimation date and the announcement return window is enforced because some firms announce earnings after trading hours.

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) find 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[6].

I calculate two measures to capture the earnings surprise. The first is the standardized unexpected earnings or SUE (Jegadeesh, Kim, Krische, and Lee, 2004), defined as

$$SUE_{i,q} = \frac{UE_{i,q}}{Std_{i,q}(UE)} \tag{9}$$

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

$$FE_{i,q} = \frac{E_{i,q} - Med_{i,q}}{P_{i,q}} \tag{10}$$

in which $Med_{i,q}$ is the median of all available analyst earnings forecasts at the end of three trading days before the earnings announcement. Only the most recent forecast from each analyst is included. Analyst forecasts made more than 90 days earlier than the announcement day are considered as stale forecasts and deleted from the sample. $P_{i,q}$ is the stock price of firm i at the end of quarter q.

I derive two measures using the 3PRF prediction to forecast the earnings surprise. The first measure is the predicted SUE (PSUE). For quarter q and firm i, PSUE is defined as

$$PSUE_{i,q} = \frac{\hat{UE}_{i,q}}{Std_{i,q-1}(UE)} \tag{11}$$

Compared to SUE, PSUE replaces UE with the predicted UE from 3PRF. Using data before the announcement of $E_{i,q}$, PSUE is scaled by the standard deviation of UE using observations from

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

$q-8$ to $q-1$. 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 as

$$PFE_{i,q} = \frac{\hat{UE}_{i,q} + E_{i,q-4} - Med_{i,q}}{P_{i,q}} \tag{12}$$

where $E_{i,q}$ is replaced by $\hat{UE}_{i,q} + E_{i,q-4}$ in the definition of $FE_{i,q}$.

To measure the revision of analyst forecasts, define

$$FR_{i,q} = \frac{Med_{i,q,EA_{i,q}-3} - Med_{i,q,q}}{P_{i,q}} \tag{13}$$

where $Med_{i,q,d}$ denotes the median of analysts forecasts calculated using all available data at the end of day d. This variable is the analyst forecasts revision from the fiscial quarter end to the end of three trading days before the earnings announcement. Correspondingly, I define a variable to capture the expected analyst revision using the information captured by 3PRF under the assumption that analysts react to the information:

$$PFR_{i,q} = \frac{\hat{UE}_{i,q} + E_{i,q-4} - Med_{i,q,q}}{P_{i,q}} \tag{14}$$

To measure the market reaction to the earnings announcement, I define the earnings announcement return (AR) as the cumulative return, in excess of the market return, from two trading days before the announcement ($EA_{i,q}$) to two trading days after the announcement:

$$AR_{i,q} = \sum_{t=EA_{i,q}-2}^{EA_{i,q}+2} (r_{i,t} - r_{mkt,t}) \tag{15}$$

3.2 Summary statistics

Panel A of table 1 presents the summary statistics for the key variables in the paper. On average, the sample has 775 observations of SUE with the 3PRF predictions at each quarter from January 1981 to December 2016, while it has 477 observations of FE with 3PRF predictions quarterly from January 1985 to December 2016. An observation from panels A indicates that the distributions of the earnings surprise measures and the announcement return are very similar for samples with or without predictions generated by 3PRF. Figure 2 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 short earnings history of less than 20 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.

Panel B of table 1 summarizes the training sample size and the number of factors used in the 3PRF prediction. For the early sample from 1981 to 2000, $\hat{UE}_{i,q}$ is estimated by training samples of 30 quarterly time series and 233 early announcers on average. For the later sample from 2001 to 2016, the predicted earnings news from 3PRF is estimated using larger training samples. For both subsamples, most predictions are composed of two factors, which demonstrates the necessity of a multi-factor assumption.

4 Empirical tests

In this section, empirical tests are conducted to support the argument that predicted earnings news from 3PRF has strong power in forecasting earnings surprise. In addition, because of the delayed reaction of investors, news captured by 3PRF is not fully reflected in the late announcers' stock prices. The tests in Section 3.2 demonstrate that 3PRF measures forecast earnings announcement return with both statistical and economic significance. Trading strategies exploiting this return predictability gain an annual alpha of greater than 10%. 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. Section 3.3 also shows that large firms, or firms with higher institutional ownership, or firms with more following analysts, or firms with higher turnover incorporate the information captured by 3PRF faster than other firms.

4.1 Prediction of earnings surprise

To examine the predictive power of 3PRF measures, I conduct panel regressions of earnings surprise, SUE or 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 turnover. To control market reactions to the information before the announcement, 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, measured from the quarter end to three trading days before the announcement. 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[7].

Table 2 presents the estimation results of regressing SUE on PSUE derived from the 3PRF prediction. Model 1 shows the regression on PSUE 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 PSUE 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 PSUE are all statistically significant and have very similar values. One standard deviation change in PSUE predicts approximately 20% of one standard deviation change in SUE in the same direction.

Models 6 and 7 of table 2 control for analyst forecasts. I construct a variable $\mathrm{PSUE_a}$ defined as

$$PSUEa_{i,q} = \frac{U\hat{E}a_{i,q}}{Std_{i,q-1}(UE)} \tag{16}$$

where

$$U\hat{E}a_{i,q} = Med_{i,q} - E_{i,q-4} \tag{17}$$

[7]These two measures of abnormal returns 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.

$E_{i,q}$ 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 $q-1$. $Med_{i,q}$ is the median of all available analyst earnings forecasts at the end of three trading days before the announcement. Essentially, $\mathrm{PSUE_a}$ is the predicted SUE using analyst forecasts instead of the 3PRF extracted measure. For stocks covered by analysts, PSUE still have independent information in forecasting SUE.

Table 3 shows analyst reactions to information captured by 3PRF. Analyst revision is defined as the difference between the median of analyst forecasts divided by the fiscal quarter end price during a given time period. Table 3 presents regressions of analyst revisions from the corresponding fiscal quarter end to three trading days before the earnings announcement. The results indicate that the contemporaneous analyst revisions are consistent with the information captured by 3PRF. Table 4 shows the tests of regressing FE on PFE. The strong forecasting power of PFE confirms that analysts fail to utilize all information from early announcers although we do observe consistent analyst revisions in table 3.

4.2 Prediction of market reaction

Table 5 demonstrates that after controlling a series of stock characteristics, one standard deviation change in PSUE predicts a change of the earnings announcement return by 25 basis points in the same direction, which is approximately 12% annualized. Models 6 and 7 of table 5 examine the predictive magnitude of 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 PSUE is slightly larger for stocks with analyst coverage using the Fama-MacBeth method. In general, the regression results prove that PSUE has both statistical and economic significance in terms of predicting the earnings announcement return.

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 PSUE value. As shown in Table 6, the stock characteristics are very similar across all five groups. Stocks with the 3PRF prediction across all groups are slightly larger stocks with higher liquidity and lower volatility. This is not surprising that smaller firms with more idiosyncratic earnings will not have 3PRF predictions since 3PRF captures systematic earnings news. 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.

Table 7 presents the cumulative returns of stock groups for the following two time windows: Q to [EA-3] and [EA-2] to [EA+2]. Q denotes the quarter end of the corresponding fiscal quarter. [EA+n] or [EA-n] stands for n trading days after or before the earnings announcement. I examine stock returns, raw or abnormal, before and during the earnings announcement. For each group, I obtain a quarterly time series of the average cumulative return or the average 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 7 presents the value-weighted average of the time series. Panel A shows the event returns for five stock groups ranked by PSUE value at each quarter. The results reveal that the stock prices do not significantly respond to early announcers before the announcement. We also observe a significant difference of earnings announcement returns between the stocks groups of highest PSUE value and lowest PSUE value. As a result, it seems that the market does not respond to the information captured by 3PRF before the information is released on the earnings announcement day.

Panel B to Panel D of table 7 presents event returns of stock groups ranked by PSUE while controlling for one characteristics. For each quarter, stocks are divided to five groups ranked by the characteristic value. Within each group, stocks are further divided to five groups ranked by the PSUE value. As a result, we end up with five stock groups ranked by PSUE with similar value of the characteristics. For example, the stock group of highest PSUE value will include stocks belonging to the highest PSUE quintile across all five groups formed by the characteristics. Table 6 shows that one-month lag return and momentum exhibit an insignificant increase as the PSUE value becomes higher. Thus, I control for these two variables in Panel B and Panel C. In panel D, I control for the cumulative returns from the fiscal quarter end to three trading days before the earnings announcement. The results are consistent with what we find in panel A. We observe significant return difference only during the earnings announcement, indicating a delayed reaction of investors.

To examine the economic significance of the return predictability, I develop a weekly rebalanced trading strategy. Specifically, the long (short) side buys (shorts) stocks with the highest (lowest) PSUE quintile ranking at the end of the first trading day of each week. The position is liquidated at the end of the first trading day of the next week. PSUE is calculated using all available early announcers at the end of the last trading day of the previous week. A minimum number of 10 stocks are required to maintain the portfolios. Only stocks with a price larger than $1 on the formation day are considered. Table 8 shows the results of this trading strategy. For a sample of approximately 36 years, this trading strategy is active for about one third of the sample period. The average holding weeks per year is 16 weeks. The long minus short strategy gains a 10% value-weighted alpha and a 7% equal-weighted alpha annually. The last four columns of panels B and C present the factor loadings of the long minus short returns.

4.3 Further analysis

4.3.1 Subsample analysis

I perform predictive regressions using two subsamples: one early sample before the year 2000 and one later sample starting from the year 2001. Table 9 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 slightly more predictive information during the early sample. The magnitude of the PSUE in forecasting the earnings announcement return of the later sample is very similar with that of the early sample, indicating that investors do not react differently during the later sample period.

4.3.2 Alternative information channels

Several 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 market 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. 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 these information channels. The first type is calculated as the average of earnings surprises or announcement returns of stock groups that are fundamentally connected with the target or are shown to have 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 the information from of the informed group only. The second type is the average cumulative returns of five trading days preceding [EA-3]. This measure considers the possibility that some stock prices reflect information faster even when they have not yet announced their earnings.

Tables 10 to 12 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 10, the estimated coefficients of PSUE 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[8]. According to the results of model 8 as shown in panel A of table 10, the average SUE of early announcers within the same industry (IND_SUE), and that of early announcers belonging to the customer industries (CUS_SUE) or supplier industries (SUP_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,

[8]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 10 to 12

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 11 shows the regression results of forecasting the analyst forecast error using the predicted FE and additional controls. As presented in model 8 of panel A, only the average FE of early announcers belonging to the same industries and the customer industries offer independent information among all additional controls. Table 12 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 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 derived from 3PRF predictions proves the advantage of this method when facing a large number of predictors.

4.3.3 Variations of return predictability

The information captured by 3PRF is public. Although investors pay a lot of attention to the earnings announcement of individual firms, it seems that they cannot fully transfer the information to related firms. In this section, I study how the return predictability of the 3PRF measure varies in different scenarios. Specifically, I check the magnitude difference of PSUE in forecasting the earnings announcement return for stocks with dissimilar market capitalization, institutional ownership, analyst coverage, turnover, idiosyncratic volatility, and macro economic conditions.

Model 1 in table 13 and 14 shows the magnitude of PSUE in forecasting SUE and AR for different firm size. I_MED equals to PSUE if the market capitalization of the stock is above the 33rd percentile and below or equal to the 67th percentile of the cross section. I_SMALL equals to PSUE if the market capitalization of the stock is less than or equal to the 33rd percentile. Model 1 in table 13 reveals that the magnitude of PSUE in forecasting SUE is weaker for firms that are smaller, since earnings of large firms are less idiosyncratic than small firms. In contrast, model 1 in table 14 shows that the magnitude of PSUE in forecasting AR is larger for small firms. This is consistent with the argument that large stocks incorporate information faster than small stocks.

I use three proxies to capture investor attention. First of all, Badrinath et al. (1995) document that stocks with higher institutional ownership (IOR) incorporate information faster than others. Also, Jegadeesh et al. (2004) argue that stocks with more following analysts (ANALYST) reflect information faster than others. Chordia and Swaminathan (2000) find that stocks with higher turnover (TURN) respond to information faster. To find out whether the predictability of PSUE changes for varying investor attention while controlling for the size effect, I construct two variables for each proxy. For each quarterly cross section, stocks are firstly divided into three groups by their size values. Within each size group, they are separated to another three groups according to their proxy values. The variable I_MED_proxy (I_LOW_proxy) equals to PSUE if the proxy value of the stock belongs to the secondly (thirdly) ranked proxy group in any size group.

Model 2 to model 4 of table 13 show how the predictability of PSUE varies with each proxy of investor attention when forecasting SUE. We can see that investor attention do not matter much in this case. Instead, according to model 2 to model 4 of table 14, return predictability of PSUE primarily exists for stocks with lower institutional ownership, less analyst coverage, and lower turnover. This indicates that stocks with more investor attention incorporate the information captured by 3PRF faster than other stocks.

Limits to arbitrage may also play a role to determine the power of return predictability. It is possible that investors are aware of the information but they are unable to take advantage of it due to the arbitrage risk. I use idiosyncratic volatility to measure the arbitrage risk following Stambaugh, Yu, and Yuan (2015). The variables I_MED_IVOL and I_LOW_IVOL are constructed similarly with the investor attention proxies. Model 5 of table 14 reveals that the return predictability of PSUE are primarily driven by stocks with low arbitrage risk. The empirical finding is inconsistent with the argument that low arbitrage risk facilitates information incorporation. Instead, it is consistent with the fact that the 3PRF measure has stronger return predictability for stocks with less idiosyncratic earnings since 3PRF only captures systematic information. The results also indicate that arbitrage risk does not explain why investors react to the information captured by 3PRF slowly.

Loh and Stulz (2018) find that analysts work harder and investors rely more on analysts during the recessions. I use the recessions classified by the National Bureau of Economic Research and construct a variable I_REC which equals to PSUE during the recessions and zero otherwise. As shown in Model 6 of table 13 and 14, I do not find large variations in predictability during recessions.

5 Conclusions

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 release. What is the best way to collect information from hundreds of early announcers, which is relevant with 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 the earnings news, information can be extracted from the early announcers considering the dynamics among both the early announcers and the late announcers. The success of predicting the earnings surprise and return demonstrates the effectiveness of the adopted methodology.

Second Essay: Option Trading Leverage and Stock Returns

Zhenping Wang

Emory University

Abstract

Noise traders prefer options of high leverage which have more lottery-like payoff. On the other hand, informed investors will prefer options of lower leverage which have lower arbitrage risk and trading cost. Such divergent preferences provide an opportunity to detect informed trading in the option market. In this paper I examine volume-weighted average of option leverage named Option Trading Leverage Ratio (OTLR). Portfolios with the lowest OTLR underperform portfolios with the highest OTLR by 1.28% monthly or 15% annually, which is consistent with the assumption that low OTLR captures informed option trading which tends to be on the negative information due to short-sale constraint in the stock market. Predictive relationship between OTLR and earnings-related information indicates that better skills in processing public information contributes to the advantage of informed option investors.

1 Introduction

In a complete market with no trading frictions, options are redundant in the sense that their cash flow can be replicated by underlying assets and risk-free assets. Under such ideal assumptions, the equity option market does not provide any more information than the stock market. However, in reality, investors have to face short sale constraints and limited capital in the stock market, whereas both frictions can be alleviated using equity options. An investor with strong beliefs about stock price but limited capital will expect to achieve leveraged profit by investing in options. In addition, for investors with bearish opinions, instead of taking the considerable effort to short a stock directly, negative exposure to the underlying stock can easily be established using options. Given a stock market with these two trading frictions, the equity option market serves as a better alternative trading venue, rather than a market of redundant derivatives, for both informed and noise traders with convictions about stock values.

Unlike stocks, options with the same underlying asset provide divergent leverage exposures. Thus, an investor of strong belief about future stock returns, informed or uninformed, can establish highly leveraged positions using options in order to achieve maximized profit. However, informed investors, usually more sophisticated than noise traders, will have to consider several factors to determine the optimal leverage exposure. High leverage will not only bring leveraged payoff, but also raises leveraged concern for informed investors. Most informed investors are institutional investors such as mutual fund or hedge fund managers. They invest using other people's capital, and their career success highly depends on investment performance. De Long, Shleifer, Summers, and Waldmann (1990) model the situation in which noise traders drive price away from fundamental and deter arbitrage. Facing such noise trader risk, delegated portfolio managers who trade against noise traders, have to consider possible capital outflow when stock price does not converge to fundamental, or even diverges further away from its fundamental value, as described in Shleifer and Vishny (1997). Thus, compared to noise traders, informed option trading will establish relatively low leveraged positions.

This low leverage preference of informed investors are confirmed in research studies such as Ang, Gorovyy, and van Inwegen (2011). They find that the average leverage ratio of hedge fund is around 2. Such a low leverage preference is also observed by industry practitioners. Reports conducted by government agencies or investment firms reach the same consensus: hedge fund industry is not highly leveraged even though they can. The average leverage was around 2 but fell to 1.5 or even lower after the financial crisis in 2008[9].

To maintain low leverage using options, one investor has two choices. The first way is to invest options of low leverage directly, such as at-the-money (ATM) options. The other one is to invest highly leveraged options using less capital while keep a large amount of cash on hand. However, the second one is not the standard industry practice. Large proportion of cash balance might be interpreted as a signal of delinquent portfolio managers[10]. Indeed, both mutual fund and hedge fund tend to have very low level of cash balances, usually below 10% on average[11]. Another significant concern about using options of high leverage is liquidity. A survey by Tabb Group[12] indicates that the biggest concern of institutional investors facing options is liquidity. Highly leveraged options tend to be more illiquid in terms of trading volumes and bid-ask spread, as shown by Chaudhury (2014). Johnson and So (2012) argues that options of higher leverage will have larger transaction costs compared to ones of lower leverage due to market makers who expect to incur larger loss when sell options of higher leverage to informed investors. As a result, informed investors with low leverage preference will trade more using options with similar leverage level, instead of using highly leveraged option combined with large amount of cash balance.

On the other hand, options of high leverage are very attractive to noise traders such as retail investors, who prefer large positive skewness as documented in the behavior bias literature such as Barber and Odean (2008) and Kumar (2009). The more lottery-like feature, low cost but a small probability of extremely high returns, of highly leveraged options satisfies the gambling appetite of retail investors, as shown in Boyer and Vorkink (2014). Different trading preferences for option leverage between informed and noise investors provide an opportunity to detect informed trading in the option market. In this paper, I develop a measure named Option Trading Leverage Ratio, or OTLR, to capture informed option trading. OTLR is calculated as the monthly average of option leverage, weighted by daily trading volumes.

[9]For example, an article that summarizes the leverage exposure is `http://seekingalpha.com/article/124783-a-graphical-look-at-hedge-fund-leverage`

[10]See `http://www.wsj.com/articles/SB870299784602635500` for example.

[11]See `https://stockcharts.com/articles/decisionpoint/2016/01/interesting-mutual-fund-stats.html` for mutual fund, and `http://www.businessinsider.com/bofa-hedge-funds-are-fully-invested-2013-3` for hedge fund.

[12]Equity options trading 2008: risking out of obscurity. Feb, 2008. Tabb Group.

Low option trading leverage potentially reflects the dominant presence of informed option traders. Consistent with this supposition, I find that low option trading leverage happens more when institutional ownership is high. Informed investors are mostly institutional investors who have better resources to collect private information and also possess better skills in processing public information. OTLR is also negatively correlated with previous proxies for negative information trading, such as the relative volume ratio of option to stock as in Johnson and So (2012). Evidence also supports the assumption that informed option trading captured by OTLR takes advantage of stock mispricing. Consistent with the liquidity concern of sophisticated investors, lower leveraged option trading tends to have a smaller option bid-ask spread but tighter underlying stock liquidity. Overall, contemporaneous correlation analysis highlights the potential that OTLR is an indicator of negatively informed option trading.

The long-short strategy that sells stocks with the lowest OTLR quintile and buys stocks with the highest OTLR quintile earns 1.28% monthly alpha(value-weighted), or 15% annually. This empirical finding is consistent with the assumption that informed option investors trade more negative information than positive information in the option market, as in Johnson and So (2012) and Ge, Lin, and Pearson (2016).

Option leverage, or λ, belongs to the family of standard option Greeks that measures option price sensitivity to certain pricing parameters. λ measures percentage change of option payoff caused by percentage change of underlying asset price, calculated as the multiplication of option delta and stock price divided by option price. Previous papers also use option delta or moneyness to proxy for option leverage, which are potentially very noisy measures of option leverage. For an example, an at-the-money (ATM) option will have constant moneyness and an almost constant delta very close to 0.5. On the contrary, the λ of this option can change drastically when option pricing parameters move, such as implied volatility. As a comparison, I construct Option Trading Delta Ratio (OTDR) and Option Trading Moneyness Ratio (OTMR) similarly with OTLR. Neither OTDR nor OTMR offers as strong predictive power as OTLR.

Whether the option market leads the stock market at the microstructure level has not reached a solid consensus. Previous empirical studies present mixed evidence[13]. Meanwhile, more and more evidence is found that options contain predictive information for stock returns of weekly, monthly

[13] For a detailed literature review, please refer to Muravyev, Pearson, and Broussard (2013).

or even longer horizons. It seems that the stock market reacts rather slowly to information in the option market. Previous literature focuses on two types of information proxy obtained using options: price-related and volume-related. This paper is closely related to the strand of research on volume-related information of options.

High trading volume is an indicator for information asymmetry (Grossman and Stiglitz, 1980; Kyle, 1985) or investor disagreement (Harris and Raviv, 1993; Kandel and Pearson, 1995). Easley, O'hara, and Srinivas (1998) develops a model in which positive (negative) option volumes predict positive (negative) stock returns, due to the high leverage feature of options. They define positive (negative) option volumes as buying (selling) call (put) options or selling (buying) put (call) options. Pan and Poteshman (2006) also studies signed option volumes using a private dataset, not observed by the public, and finds that put-call volume ratio negatively predicts stock returns up to one week using option volumes initiated by buyers to open new positions. Both studies have to infer information from signed option volumes. Roll, Schwartz, and Subrahmanyam (2010) is the first empirical paper attempting to understand relative unsigned option volume scaled by its underlying stock volume, denoted by OS. They collect several pieces of evidence that OS potentially reflects informed option trading. One of the most important pieces of evidence is that OS predicts post-earnings announcement returns.

Johnson and So (2012) looks further into the information of OS. They set up a model in which informed traders face short cost as in reality. In equilibrium, informed investors trade more negative information than positive information using options. Such asymmetrical trading behavior indicates that high OS will predict low stock returns on average. Empirical evidence supports their model's prediction: a strategy that buys stocks with highest decile weekly OS and sells stocks with lowest decile weekly OS earns about 19% annualized risk-adjusted return. Ge et al. (2016) uses signed option volume data and supports the argument of Johnson and So (2012) by showing that there are more components of option volumes with negative information than those with positive information. However, OS is potentially a noisy indicator of informed trading in the option market. Noise traders may also trade their negative belief more in the option market. Furthermore, the high leverage feature of options is very attractive to speculators who have been shown to prefer highly skewed payoffs. Thus, high OS ratio could also indicate intense speculative trading in the option market. OTLR, motivated by the different trading preferences of informed and uninformed investors, will be helpful in filtering noisy information in OS. In addition, OTLR is calculated without the involvement of stock trading volume, and thus eliminates the impact of contemporaneous informed trading in the equity market, as in the OS measure.

Whether OTLR provides additional predictive power to previously documented negative information variables is an empirical question. To answer this question, I compare OTLR with OS as in Johnson and So (2012) and with short interest, scaled by shares outstanding, as in Asquith, Pathak, and Ritter (2005). At the beginning of each month, stocks are grouped into five groups by OS or short interest as of the previous month. Within each quintile group, stocks are further divided into five portfolios by the OTLR of last month. OTLR positively predicts risk-adjusted return for most OS or short interest groups, while the alpha spread is largest for groups with highest OS or short interest. The evidence suggests that OTLR not only helps filter out noise in short interest and OS, but also offers additional information not captured by those two informational signals.

To examine whether OTLR picks up the predictability of stock characteristics, I form 5 by 5 double-sorted portfolios first by stock characteristics and then by OTLR. I find that long-short strategy by OTLR delivers significant risk-adjusted returns for three out of five size groups, among which it earns the largest alpha, 1.72% monthly, among stocks of smallest size. Also, OTLR generates positive alpha spread for each group of stocks classified by book-to-market ratio. Not surprisingly, OTLR has the largest return spread among stocks with the lowest momentum returns, while it still has independent predictability for stocks with the highest momentum returns. Miller (1977) argues that short-sale constraint deters the price adjustment to negative information. I use institutional ownership as a proxy for short-sale constraint as in Nagel (2005). Consistent with Miller (1977) prediction, I find that OTLR long-short strategy earns the largest alpha, 2.25% monthly, among stocks with lowest institutional ownership.

Fama-Macbeth regression helps to investigate information captured by OTLR while controlling for multiple relevant variables. Since OTLR is a proxy for negative information, it may not offer much predictability for the upside. To take care of this potential non-linearity, I set up a variable, OTLR0, that equals to OTLR when OTLR falls below the cross-section median, or equals to zero otherwise. In the regression analysis, I regress one-month ahead stock returns on OTLR, OTLR0 and a series of relevant variables, including OS, short interest, idiosyncratic volatility, Amihud (2002) liquidity, Stambaugh et al. (2015) mispricing score, analyst dispersion as in Diether, Malloy, and Scherbina (2002), and stock characteristics such as size, book-to-market ratio, monthly stock return, and momentum returns. The regression results suggest that, all else being equal, one standard deviation decrease of OTLR0, or OTLR when OTLR is smaller than the cross-section median, predicts 0.50% decrease of monthly stock returns, which is about 6% annually. Furthermore, consistent with the model implication in Johnson and So (2012), the relative volume ratio of option to stock is more informative in predicting future stock returns when option leverage is lower. To check whether OTLR offers long-lasting predictive power, I also use stock returns up to six months ahead of independent variables. After controlling for a series of stock characteristics and option variables, the predictive power of OTLR does not last more than one month.

The strong association of OTLR and future stock returns suggests the possibility that OTLR also predicts future earnings, one of the most important pieces of information to evaluate stock values. Empirically, I find low OTLR predicts negative changes of monthly analyst forecast consensus. Furthermore, although OTLR does not significantly predict earnings surprise based on analyst consensus, OTLR has significant power in forecasting standard unexplained earnings (SUE) and market reaction to earnings announcement. Controlling a series of stock characteristics and option characteristics, one standard deviation decrease of OTLR below median will predict 0.05 standard deviation decrease of CAR(-1,1), which is about 0.47% or about 40% after annualization. Last but not least, OTLR predicts post-earnings announcement returns (PEAD) up to 60 trading days after the earnings announcement. The predictability for earnings-related variables primarily comes from OTLR below the cross-section median, or OTLR0. For example, one standard deviation decrease of OTLR0 predicts a 0.06 standard deviation decrease of CAR(2,60), which is about 1.4% for the cumulative period or 6% after being annualized. The overall evidence indicates that informed option investors trade their information before analysts incorporate the same information into their earnings forecasts. Also, informed option investors take advantage of market underreaction to earnings announcement. As a result, at least part of predictive power of OTLR comes from informed option investors who are better at analyzing public information compared to marginal investors.

Motivated by different trading preferences between informed and uninformed investors, I use option trading leverage ratio, or OTLR, to detect informed trading in the option market. Besides the trading choice of options, OTLR is strongly correlated with option price or implied volatility, as shown in section 2.2. Such association raises the concern that OTLR predicts stock returns because it picks up information from option pricing. Price-related variables of options have been used to forecast stock returns in previous literature. Previous documented information proxy includes volatility spread between implied volatility and realized volatility by Bali and Hovakimian (2009), put-call parity violation by Cremers and Weinbaum (2010), slope of put implied volatility by Xing, Zhang, and Zhao (2010) and changes of call and put implied volatility by An, Ang, Bali, and Cakici (2014). None of them finds a direct relationship between implied volatility and stock returns. Also, price-related variables considered in previous research do not consider the impact of trading choice. They either extract information from all options regardless of whether they are traded, or focus on at-the-money options which tend to be more liquid than others. On the contrary, OTLR captures leverage preference of option traders, calculated using implied volatility on the trading day. The predictive power of OTLR does not fully rely on previous price-correlated measures. However, the positive association between low option leverage and high implied volatility is not surprising, which may arise endogenously. The model in An et al. (2014) shows that demand by informed option traders will cause contemporaneous positive change in implied volatility or option price. As a robustness check, I also examine the predictive power of OTLR using a subsample that deletes 10% of observations with extreme implied volatility. The predictive power of OTLR stays strong in this subsample.

In this paper, section 3 summarizes data and variables used, and also offers an analysis of differences among option leverage, delta and moneyness. Section 3 conducts portfolio analysis and multivariate regression to investigate whether OTLR predicts future stock returns. Section 4 analyzes the predictive power of OTLR about firm cash flows. Section 5 concludes.

2 Data and variables

2.1 Data and summary statistics

Stock-related data (price, return and volume) is obtained from the Center for Research in Securities Prices (CRSP). Option-related data (price, implied volatility, Greeks and option volume) is from OptionMetrics. Short interest, book equity, and earnings data are retrieved from Compustat. Analyst forecast data is from the Institutional Brokers' Estimate System (I/B/E/S) provided by Thomson Reuters. Data for number of shares owned by institutional investors is obtained from

Thomson-Reuters 13F database. The main sample, an intersection of CRSP and OptionMetrics, consists of stocks with share code 10 or 11 and their equity options from 1996 to 2014. Variables are constructed at the end of each month. Stocks with prices above \$5 are included. Following Johnson and So (2012) to avoid measurement errors caused by illiquid options, stocks with options that have less than 100 call and 100 put contracts traded during each month are excluded. Each variable is winsorized by 1% at both sides in the cross section, except cases when some variables such as stock returns are used as predictive dependent variable in regression tests. The main sample has 254,475 firm-month observations, which has about 1,116 firms per month.

U.S. equity options are American-style options, which means the holder of options can exercise his options any time before expiration. Black-Scholes model will not accurately price American-style options with underlying asset paying dividend like stocks. [14] OptionMetrics uses the Cox, Ross, and Rubinstein (1979) binomial tree model to infer implied volatility and option Greeks. They use the most recent dividend payment as the dividend yield input, and assume the underlying stock pays dividends at a specific date, which is either the announced future ex-dividend date or projected ex-dividend date according to their extrapolation algorithm if the actual date is not known.

I obtain end-of-day option price, underlying security price, implied volatility and option greeks from OptionMetrics. For a given equity option at the end of day t, its leverage ratio, or λ_t, is defined as

$$\lambda_t = \frac{\partial O_t}{\partial S_t} \times \frac{S_t}{O_t} = \frac{Delta_t \times S_t}{O_t} \tag{18}$$

in which $Delta_t$ measures the rate of change of option price corresponding to one unit of change in the underlying asset price, the value of which will depend on the option pricing model used. S_t is the underlying stock price at time t, O_t is the average of end-of-day bid and ask price for the option. λ_t is defined as the percentage change in an option price scaled by the percentage change in the underlying price, which measures the actual leveraged ratio of payoff an investor will earn by investing in an option instead of its underlying stock.

Here is an example of option leverage. Assume one has \$100 to buy stocks of company A, which is traded at \$10. One can also buy a call option of the stock with strike equals to \$10, which is

[14] Black and Scholes (1973) model can be extended for European options with the underlying asset paying continuous dividend yield or proportional dividends of the underlying asset price at pre-determined times. However, both cases don't apply for American options with underlying assets paying cash dividends. The main problem is that the Black and Scholes (1973) model does not include the optimal early exercising scenario when the underlying pays dividend.

traded at \$0.5. Thus, by buying the option one can control 200 shares of stock instead of 10 shares by buying the stock directly. Assuming now the stock goes up to \$11, a portfolio of 10 shares of stocks will have a payoff of \$10, whereas a portfolio of 200 call options will have a payoff of \$100 that is ten times leverage of the stock portfolio. [15]

To capture option traders' choice of leverage in month m, I define Option Trading Leverage Ratio, or OTLR, for each firm i in month m as

$$OTLR_{i,m} = \frac{\sum_{d=1}^{D}\sum_{k=1}^{K}(|\lambda_{i,k,d,call}| \times OPVol_{i,k,d,call} + |\lambda_{i,k,d,put}| \times OPVol_{i,k,d,put})}{\sum_{d=1}^{D}\sum_{k=1}^{K}(OPVol_{i,k,d,call} + OPVol_{i,k,d,put})} \tag{19}$$

D is the number of trading days in month m for options with underlying stock i. K is the number of strikes available. Call or put denotes for its corresponding option type. For firm i, $OTLR_{i,m}$ measures the average of absolute option leverage ratio, value-weighted by daily option trading volumes in month m. OTLR measures the trading preference of option leverage for firm i in month m.

At each trading day d, only options with 15 to 45 days to expiration are included. Thus, the sample does not include options with longer expiration, which tend to be more illiquid. Since the paper studies prediction for monthly stock returns, a minimum of 15 days to expiration is required. [16]

Previous papers such as Pan and Poteshman (2006) and Xing et al. (2010) also study information contained in call and put separately. To check whether there is any difference in terms of leverage choice, I also define OTLRC and OTLRP for call and put options respectively, in which only trading of the corresponding option type is included.

Panel A of table 1 shows summary statistics of OTLR in different subsample periods. The average of OTLR is increasing over time, indicating an rising preference for high leverage. This might be due to the increasing population of noise traders or retail investors in the option market as the advancing development of computerized trading system.

Recent papers such as Roll, Schwartz, and Subrahmanyam (2009) use delta as a proxy of option

[15] Delta is assumed to be 0.5 since this option is at the money. If the option has 30 days to expiration with zero dividend yield and zero risk-free rate, the price of \$0.5 is about 44% implied volatility using Black and Scholes (1973) model.

[16] The choice of days to expiration does not significantly affect the results. I also try some other ranges such as from 5 to 35 days to expiration as in Johnson and So (2012). Results are quantitatively similar.

leverage, the absolute value of which is negatively correlated with λ. Some papers like Ge et al. (2016) also use moneyness, which is the ratio of option exercise price over stock price, to infer option leverage. Out-of-the-money (OTM) call and put options are considered as more leveraged investments than at-the-money (ATM) or in-the-money (ITM) options. Both measures are highly correlated with λ. Thus, as alternative measures of option trading leverage, I define Option Trading Delta Ratio (OTDR) as

$$OTDR_{i,m} = \frac{\sum_{d=1}^{D}\sum_{k=1}^{K}(|\Delta_{i,k,d,call}| \times OPVol_{i,k,d,call} + |\Delta_{i,k,d,put}| \times OPVol_{i,k,d,put})}{\sum_{d=1}^{D}\sum_{k=1}^{K}(OPVol_{i,k,d,call} + OPVol_{i,k,d,put})} \tag{20}$$

I also define Option Trading Moneyness Ratio (OTMR) as

$$OTMR_{i,m} = \frac{\sum_{d=1}^{D}\sum_{k=1}^{K}(k/S_{i,d} \times OPVol_{i,k,d,call} + (1 - k/S_{i,d}) \times OPVol_{i,k,d,put})}{\sum_{d=1}^{D}\sum_{k=1}^{K}(OPVol_{i,k,d,call} + OPVol_{i,k,d,put})} \tag{21}$$

$S_{i,d}$ is stock price for firm i at the end of day d. However, section 2.3 shows that delta and moneyness are both very noisy proxies for option leverage. Thus, OTDR and OTMR do not have predictability of future stock returns as strong as OTLR.

Two competing measures for informed trading with negative information are short interest ratio (Short), which is short interest divided by shares outstanding, as in Asquith et al. (2005), and relative volume ratio of option to stock as in Johnson and So (2012), which is defined as

$$OS_{i,m} = \frac{\sum_{d=1}^{D}\sum_{k=1}^{K} OPVol_{i,k,d,call} + OPVol_{i,k,d,put}}{\sum_{d=1}^{D} StkVol_{i,d}} \tag{22}$$

where $StkVol_{i,d}$ is the trading volume of underlying stock i on day d in month m.

To test whether OTLR predicts stock returns, at the beginning of each month I form quintile portfolios by ranking OTLR as of the previous month. The time series average of cross-section options and underlying stock characteristics of OTLR portfolios are shown in panel B of table 1. OTLR ranges from 7 to 21 for quintile portfolios. OTLRC and OTLRP changes monotonically with OTLR portfolios, confirming the previous finding that there is no opposite information between unsigned call volume and unsigned put volume. OTDR and OTMR decrease monotonically with OTLR, consistent with the fact that smaller absolute delta and out-of-the-money options have higher leverage exposure. The portfolio with the lowest OTLR has an Option Trading Delta Ratio close to 0.5, indicating option leverage of this portfolio is similar with a portfolio of at-the-money options.

In this paper, I define at-the-money (ATM) options as options of moneyness between 0.9 and 1. ATM option volatility premium (Prem), which is the difference between ATM implied volatility (ATM_IV) and historical volatility, increases with OTLR according to panel B of table 1. Positive option volatility premium happens when investors pay more than the fair value of the option. ATM_IV for firm i in month m is calculated as the monthly average of end-of-day implied volatilities of ATM call and put in month m. The historical volatility is calculated as the standard deviation of daily returns in the same month. The negative correlation of OTLR and Prem indicates that informed traders act when options are cheaper. ATM option bid-ask spread, or OBA, is the monthly average of end-of-day ask minus bid divided by the midpoint of ATM call and put options. The empirical fact that OBA also decreases with OTLR getting lower shows that informed traders are sophisticated enough to trade when transaction costs are low.

Examining underlying stock characteristics, I find that for portfolios with the lowest OTLR, the underlying assets are small stocks with relatively higher book-to-market ratio and with high prior momentum returns. Informed traders have a higher chance to discover profitable investment opportunities among small stocks that tend to have lower institutional ownership, as discussed by Bennett, Sias, and Starks (2003). Also, it seems that informed option trading takes advantage of overpricing caused by naive momentum chasing behavior. Hong and Stein (1999) discusses the possibility of overreaction caused by trend-chasing investment.

Panel B of table 2 presents correlations among option and stock variables, which are standardized to mean 0 and standard deviation 1 in each monthly cross section. Consistent with portfolio level analysis, OTDR and OTMR are both negatively correlated with OTLR. However, the correlation is not very large. Section 2.2 explains the discrepancy between option leverage and its proxies including delta and moneyness. Given the fact that both delta and moneyness are very noisy proxies for option leverage, OTDR and OTMR do not have predictive power of future stock returns as strong as OTLR. OS has very low correlation with OTLR, indicating OTLR may provide additional power in detecting informed trading in option markets. OTLRC and OTLRP are highly correlated with OTLR. As a result, option type does not matter in interpreting option trading leverage. Short is negatively correlated with OTLR, indicating some option traders share the same information with short sellers in the stock market. OTLR also negatively correlates with mispricing score (Misp), the average of ranking percentiles on 11 anomaly variables as in Stambaugh et al. (2015), and analyst dispersion (Disp), the standard deviation of analyst annual earnings forecasts scaled by the forecast mean as in Diether et al. (2002). Higher values of Misp and Disp indicate higher overpricing. Obviously, informed traders trade against stock overpricing using options. Lower OTLR is associated with higher idiosyncratic volatility (Ivol), which is the standard deviation of residuals from regressing daily stock returns on Fama and French (1993) three-factor model in the same month. Ivol is considered as proxy for arbitrage risk, which impedes arbitrage activity, as in Stambaugh et al. (2015). The negative correlation between OTLR and Ivol shows that facing higher arbitrage risk, informed traders tend to be more risk- averse and choose lower leverage in option trading.

Analysis using portfolio characteristics or stock-level variable correlations does not control for all relevant information. To have a better understanding of OTLR, in section 2.3, I use (Fama and MacBeth, 1973) regression to study what stock or option characteristics are associated with OTLR contemporaneously.

2.2 Option leverage, delta and moneyness

Many previous studies, such as Ge et al. (2016) and Roll et al. (2010), have been using delta or moneyness as a proxy for option leverage. To see how these two measures differ from option leverage, or λ, I use Black and Scholes (1973)s model for a simplified analysis. Although Black and Scholes (1973)s model cannot be used to accurately price American options with an underlying asset that pays dividends, it can be used to analyze options when the underlying stock does not pay any dividend. In figure 4, I plot how absolute delta and option leverage, both calculated using Black and Scholes (1973)s model, change with moneyness and implied volatility for call and put options with 30 days to expiration. The underlying stock has zero dividend yield. The risk-free rate is assumed to be zero. In each subfigure, three moneyness ratios, or the ratio of strike over underlying price, are assumed: 0.8, 1 and 1.2.

From figure 4a, absolute delta does not change much with implied volatility for the call option with strike equals to underlying price. Delta of out-of-the-money (in-the-money) calls with moneyness equal to 1.2 (0.8) increases (decreases) almost linearly with implied volatility except for volatility lower than 30%. Delta can be interpreted as the probability that the option will be in-the-money at expiration. As implied volatility increases, OTM (ITM) call has a higher chance to expire in (out of) the money and thus has higher (lower) delta. According to figure 4b, option leverage is a convex function of implied volatility for ATM call and OTM call, where the latter has higher positive convexity. For ATM and OTM call option, option leverage is very sensitive to volatility change when volatility is below about 50%, and becomes more insensitive to volatility change for higher implied volatility. The option leverage of ITM call does not change significantly with implied volatility. From figure 4c and figure 4d, OTM, ATM and ITM put options have similar absolute delta and option leverage properties with the corresponding call options.

Figure 4 shows that delta and moneyness can be misleading indicators for option leverage. In a portfolio of one ATM call options, different option leverage will be established depending on the implied volatility. However, the portfolio delta will be very close to 0.5 and the portfolio moneyness will be constantly 1 regardless of high or low implied volatility. The difference arises from different sensitivity of delta and option price to implied volatility. From eq 19, option leverage is a multiplier of delta and the ratio of stock price to option price. Implied volatility will affect the value of delta and option price differently, resulting in the discrepancy between delta and option leverage.

In sum, delta and moneyness are both noisy proxies for option leverage. In section 3.1, we will see that OTDR and OTMR do not predict stock returns as well as OTLR does. Also, since implied volatility plays an important role in deciding option leverage, I perform additional tests using samples that delete stocks with top and bottom 10% implied volatility in section 3.1, to alleviate concern that the predictability of OTLR is driven by extreme implied volatility.

2.3 What affects option trading leverage?

In this section, I use Fama and MacBeth (1973) regression to study what stock and option characteristics are associated with option traders' choice of leverage. For each month, OTLR is regressed on a series of concurrent stock characteristics and option characteristics. The average of estimated coefficients is reported in table 3, with standard deviation calculated using Newey and West (1987) method to adjust for serial correlation. Although the test cannot be used to infer any causal relationship, it can be used to examine contemporaneous correlation between OTLR and one variable while controlling for other relevant information.

For each regression in table 3, I control for stock characteristics including size, book-to-market, monthly stock return, momentum return of prior six month and cumulative return from prior 12th month to prior 7th month. In regression (1), I add two variables that have been used as informed trading signals of negative information: relative volume of option to stock (OS), and short interest ratio (Short). The estimated coefficients show that low OTLR is associated with high OS and high Short. When option traders choose low leverage exposure, they tend to trade more and their trading is positively correlated with short sellers' trading. Thus it is not surprising that OTLR also detects informed trading with negative information.

In regression (2), I add institutional ownership (IOR), which is the ratio of institution-owned shares over shares outstanding, to the examination. The result shows that when option traders choose low leverage exposure, the underlying stock tends to have higher institutional ownership. Institutional investors are considered to be sophisticated investors with more resources to discover new information and with better skills to process information. This result supports the assumption that institutional investors drive the low leverage exposure in option trading.

Sophisticated investors will also take liquidity into consideration carefully. In regression (3), I find that low leverage option trading happens more to options with lower trading costs, measured by at-the-money option bid-ask spread. Option volatility premium, measured by the difference between implied volatility and historical volatility of at-the-money options, seems not to be significantly correlated with OTLR. However, when I add all relevant variables into the regression as in regression (6), option volatility premium has a significant negative correlation with OTLR. The change of estimate is due to omitted variable bias. Low leverage option trading happens more when options are priced with higher premiums. This could be driven by market makers who learn from the orders of informed option traders and update option price accordingly, as modeled by An et al. (2014). Another possible reason is that the abnormal demand for low leverage options drives option price up, according to the demand-based option pricing literature such as Bollen and Whaley (2004) and Garleanu, Pedersen, and Poteshman (2009). In either case, lower leverage option trading is associated with more expensive options, highlighting informed traders' expectation of a higher payoff to offset the option premium.

Idiosyncratic risk adds uncertainty for arbitrage, which is a primary source of limits to arbitrage. With the presence of noise traders, stock price may not converge to fundamental value for a long time, which will further deter arbitrage as described in De Long et al. (1990) and Shleifer and Vishny (1997). It is important for informed traders to control leverage when trading against noise investors. Regression (4) shows that when idiosyncratic volatility is higher, option trading leverage is lower. If the stock market is relatively illiquid, informed traders will trade more information in option markets. In regression (6), low leverage option trading is associated with illiquidity in the stock market, measured by Amihud (2002) liquidity proxy, which is the average of absolute price change over trading volume for the previous 12 months.

Previous papers show that the advantage of sophisticated investors may not come from their ability to collect private information but their better skills in processing public information, such as Engelberg, Reed, and Ringgenberg (2012). Thus, an important source of the predictability of OTLR could be stock mispricing based on public information. In regression (5), OTLR is showed to be negatively correlated with mispricing score and analyst dispersion, both of which well-documented proxies for overpricing such as in Diether et al. (2002) and Stambaugh et al. (2015).

Regression (6) includes all relevant variables, which alleviates the concern of omitted variable bias. Low leverage option trading happens more to small stocks and value stocks, which is similar with the conclusions in section 2.1. However, after controlling for other variables, low leverage trading is associated with low momentum stocks. OS, option bid-ask spread, and mispricing score become less correlated with OTLR but still significant, while Short is no longer significantly correlated with OTLR. IOR, option volatility premium, idiosyncratic volatility and Amihud liquidity are more correlated with OTLR in the same direction with estimates from previous regressions. In sum, the evidence is consistent with the assumption that low leverage option trading is primarily driven by informed investors who establish optimal leverage, which is determined by risk and costs, using options of low leverage due to liquidity concern.

3 Option trading leverage ratio and stock returns

3.1 Portfolios performance

To examine the predictability of OTLR, at the beginning of each month, stocks are grouped to five portfolios by their last month OTLR. Value-weighted portfolio returns are presented in panel A of Table 4. In terms of monthly excess returns, the portfolio with the lowest OTLR outperforms the portfolio with the highest OTLR by 0.90%. For risk-adjusted return, or alpha, using Fama and French (1993) three factor model, the portfolio with the lowest OTLR earns -1.36% while the portfolio with the highest OTLR earns 0.21%. A strategy that buys the highest OTLR portfolio and sells the lowest OTLR portfolio will generate 1.57% abnormal returns on average. For alpha using the four factor model with additional momentum factor, the portfolio with lowest OTLR earns -1.14% while the portfolio with highest OTLR earns 0.14%. The long-short strategy generates 1.28% abnormal returns on average. About 90% of abnormal returns come from the short side, which is consistent with Stambaugh, Yu, and Yuan (2012) that short sale constraint in the stock market exerts overpricing. Figure 7 presents risk-adjusted returns of the long leg and short leg after portfolio formation. Alpha earned by the short leg gradually decreases over time. The evidence indicates that the stock price incorporates information of OTLR rather slowly.

Section 2.2 presents evidence that implied volatility is an important determinant of option leverage. Previous literature like Bali and Hovakimian (2009) uses measures derived from implied volatility to predict stock returns. To alleviate the concern that predictability of OTLR is driven by implied volatility, I form a subsample for which stocks with top and bottom 10% at-the-money implied volatility in each cross section are deleted. Quintile portfolios are grouped by OTLR using stocks in this subsample, whose performance is presented in panel B of table 4. The risk-adjusted return of long-short strategy using three-factor or four-factor models is still statistically significant, which is about three-fourth of abnormal returns using the full sample. It is not surprising that information captured by low leverage option trading is closely associated with high implied volatility. On one hand, high implied volatility means options are expensive and thus drives down option leverage. On the other hand, pricey options could be an outcome of informed trading when market makers learn information from coming orders as described in An et al. (2014), or an outcome of the excessive systematic demand for low leverage options as modeled by Garleanu et al. (2009). Although previous literature uses some proxies derived from implied volatility to predict stock returns, implied volatility alone does not offer strong predictability as in Bali and Hovakimian (2009). Also, OTLR uses daily implied volatility that is value-weighted by trading volumes. OTLR captures trading choice of option leverage, instead of a month-end snapshot of option pricing as in previous research.

Section 2.2 shows that delta and moneyness are very noisy proxies for option leverage. To compare OTLR with alternative leverage measures, delta and moneyness, used in previous literature, I also form quintile portfolios by Option Trading Delta Ratio (OTDR) and Option Trading Moneyness Ratio (OTMR), both of which are negatively correlated with OTLR. From panel A of table 5, the long-short strategy of OTDR does not earn significant abnormal returns using three-factor or four-factor models. The portfolio with highest OTMR underperforms the portfolio with lowest OTMR by 0.42% in terms of risk-adjusted return using the four-factor model, about 30% of alpha spread generated by OTLR. Thus, option leverage ratio constructed by delta or moneyness does not serve as a good predictor of informed trading.

3.2 Portfolio returns controlling for stock characteristics

Panel B of table 1 presents stock characteristics of portfolios formed by OTLR. Fama-Macbeth regression results in Table 3 confirm that low OTLR stocks tend be small in market capitalization, tend to be value firms, and have lower momentum returns. In this section, I examine portfolio performance by controlling stock characteristics. Each month, stocks are assigned to one of five portfolios by size, book-to-market ratio, momentum returns and residual institutional ownership as of previous month end, and then within each characteristic quintile, stocks are further grouped into five portfolios based on OTLR as of previous month. At the beginning of each month, I form 25 portfolios and hold for one month. Table 6 presents average risk-adjusted returns, or alpha, earned by the formed portfolios, calculated using the four-factor model.

From panel A of table 6, the long-short strategy of OTLR earns positive abnormal returns within each size rank. The abnormal return is largest for the smallest size rank, which is about 1.72% monthly alpha. The evidence goes against the argument that OTLR strategy earns positive alpha because it picks up the size premium.

From panel B of table 6, the long-short strategy of OTLR earns positive abnormal returns within each book-to-market (BM) rank. The abnormal return is largest for largest BM rank, which is about 2.25% monthly alpha. The long-short strategy of OTLR earns significant positive alpha in three groups formed by BM, which suggests that OTLR has independent information for future returns.

Panel C of table 6 shows that the long-short strategy of OTLR earns positive alpha for each momentum rank, in which OTLR strategy earns the largest risk-adjusted return in the ranking of lowest momentum returns. It is not surprising that informed traders who drive low option leverage take advantage of low momentum stocks by shorting them in the option market. However, in the highest momentum stocks, stocks with lowest OTLR earn significant negative abnormal returns, -0.84% monthly, suggesting that informed traders also possess information other than low momentum.

Miller (1977) argues that in a market with short sale constraints, stock price will reflect the opinion of marginally optimistic investors. As a result, mispricing will be primarily driven by overpricing, which has been confirmed by the empirical studies such as Stambaugh et al. (2012) and Stambaugh et al. (2015). If low OTLR captures negative information that has not been incorporated into the stock market due to short sale constraints, the long-short strategy will earn the largest alpha among stocks that are most difficulty to short. Previous papers such as Asquith et al. (2005) and Nagel (2005) argue that institutional ownership (IOR) is a proxy for short sale constraint, since institutional investors are the main supplier of shorting. Following his study, I determine residual institutional ownership by regressing logit IOR to log size and squared log size in order to control for the positive correlation between IOR and size. Performance of double-sorted portfolios by residual IOR and OTLR is presented in panel D of table 6. Consistent with Miller (1977)s argument, the long-short strategy of OTLR earns highest alpha in the rank of lowest residual IOR, or among stocks most difficulty to be shorted. The strategy of OTLR earns positive risk-adjusted return across all five ranks of residual IOR, suggesting that even with favorable shorting conditions in the stock market, informed traders still prefer option markets due to leveraged payoff and convenience to establish negative exposure[17].

3.3 Portfolio returns controlling for other informed trading proxy

Previous literature has discovered several proxies to detect informed traders with negative information. Among all, two are most closely related to OTLR. One is short interest ratio (Short), which is the ratio of short shares outstanding divided by total shares outstanding. Asquith et al. (2005) presents evidence that Short negatively predicts stock returns, especially in short-sale constrained

[17] DAvolio (2002) documents that most stock loans can be terminated any time by lenders, named recall risk faced by short sellers. Once recalled, the mean time to re-establish short positions is more than 20 trading days. As a result, even if the cost of shorting is low and shorting supply is abundant, the short seller might still experience difficulty in keeping negative exposure for a long period.

stocks. High short interest on highly short-sale constrained stocks means short sellers maintain negative exposure facing high shorting cost, which indicates they believe those stocks will have stronger negative returns in the future. Another one is relative volume ratio of option to stock (OS). Johnson and So (2012) shows that OS is a negative predictor of underlying stock returns. They develop a model in which informed traders, facing high shorting cost in the stock market, will trade negative information more than positive information in the option market, and conclude that the ratio of aggregate option volume over stock volume negatively predicts underlying stock returns.

To examine whether OTLR has information beyond Short and OS, I first form five portfolios by Short or OS. Within each quintile of Short or OS portfolios, stocks are then sorted into five portfolios by OTLR. 25 portfolios are created at the beginning of each month, and are held for the month. Risk adjusted returns are shown in Table 7, using the four-factor model.

From panel A of table 7, we can see that the strategy that buys stocks with highest OTLR and sells stocks with lowest OTLR earns higher alpha for stocks with higher ranking of OS. When informed traders possess negative information, they tend to trade more in options in order to extract maximize profit, while choosing low leveraged options to alleviate risk and liquidity concerns. Even in the lowest rank of OS, OTLR strategy earns 0.61% monthly alpha, showing that OTLR offers additional information about informed trading.

Panel B of table 7 presents abnormal returns of double-sorted portfolios by Short and OTLR. OTLR strategy earns significant positive alpha for stocks with the highest two rankings of Short. Obviously, informed traders in the option market share the same information as short sellers. Furthermore, OTLR helps to filter out uninformed Short. For example, portfolios with high Short but high OTLR, indicating dominant presence of noisy traders, do not earn significant negative abnormal returns. In section 3.4, I use Fama-Macbeth regression to control all relevant stock characteristics, results of which confirm that OTLR offers additional information beyond Short.

To compare profitability among these three informed signals of negative information, I calculate information ratios for long-short strategy using OTLR, OS and Short respectively, and for strategy using double-sorted portfolios by the combination of these three signals. Information ratio is the ratio of abnormal return, using the four-factor model as the benchmark, divided by the standard deviation of residual, which is annualized by multiplying the square root of 12 because monthly returns are used. Table 8 presents the results. In general, the information ratio for equal-weighted strategy returns is larger than value weighted returns. Small firms are more likely to be short-sale constrained, and are more likely to be overpriced. For value-weighted portfolio returns, OTLR strategy has the largest information ratio about 1, compared with that of strategy by OS or Short, which is 0.46 and 0.31 respectively. Combining OS and OTLR delivers the highest information ratio using equal-weighted portfolio returns. Adding OTLR to filter noise in OS or Short delivers a higher information ratio, using value-weighted portfolio returns, than the strategy combining OS and Short. In sum, OTLR has orthogonal information to predict stock returns in addition to OS and Short.

Furthermore, OTLR also has longer predictability than Short and OS. In figure 5, I plot risk-adjusted returns, for long-short strategy using OTLR, Short and OS respectively, up to one year after portfolio formation. In general, average alpha of OTLR strategy is larger than that of OS or Short up to one year after portfolio formation. The difference can be better presented using cumulative alpha as shown in figure 6. The difference of cumulative alpha for a one year holding period is about 4% (5%) between OTLR and Short (OS).

Figure 8 shows alphas of double-sorted portfolios after portfolio formation. Consistent with the conclusion from information ratio shown in table 8, a strategy that combines Short or OS with OTLR performs better than a strategy that combines Short and OS up to one year after portfolio formation.

3.4 Multivariate regression analysis

Double sorted portfolios can only give an incomplete picture about the independent information of OTLR. Since only two variables are considered, OTLR may offer additional predictability due to its correlation with other variables not included. In this section, I use Fama-MacBeth regression to study information in OTLR by regressing one-month ahead stock returns on all relevant variables. Regression results are reported in table 9. To facilitate interpretation of economic significance, I standardize each independent variable to have mean 0 and standard deviation 1 in the cross section, while dependent variables are stock returns in percentage.

Since OTLR, OS and Short are signals for negative information trading, it is possible that their information comes from the short side and exhibits a non-linear relationship with stock returns.

To check this possibility, I form decile portfolios by OTLR, OS and Short respectively, and plot time-series average returns for each decile portfolio as in figure 9. The plot provides some evidence that OTLR predicts stock return approximately in a linear way for OTLR below median, while Short and OS predict stock returns in a linear relationship for values above median. To take care of the possible non-linearity, I add additional variables in the Fama-MacBeth regressions:

$$x0_{i,m} = \begin{cases} x_{i,m} & : x_{i,m} < Median(x_{i,m}) \\ 0 & : x_{i,m} \geq Median(x_{i,m}) \end{cases} \tag{23}$$

for firm i in month m. $Median(x_{i,m})$ denotes median of $x_{i,m}$ for each cross section at month m. x can be OTLR, OS or SR. When both x and $x0$ are added to regressions, the coefficient of x measures the linear relationship between the dependent variable and x above median, while the sum of coefficients of x and $x0$ measures linear relationship between the dependent and x below median.

In the model of Johnson and So (2012), the equilibrium results indicate that option volume is more informed when the option has lower leverage exposure. In their model, options of higher leverage exposure will have higher bid-ask spread to compensate the expected loss of market makers given certain probability of trading with informed investors. As a result, when option volume is observed, it has higher chance that the option is traded by informed investors when the corresponding option has lower leverage. To test whether OS is more informative when leverage is low, Iadd the multiplication of OTLR and OS to each regression of one-month ahead stock returns. If the estimated coefficient of the interaction variable is positive, it shows that the negative relationship between OS and future stock returns gets stronger as option leverage becomes lower.

In regression (1) and (2) in table 9, I consider three informed signals only. Regression (1) shows that the average relationship between short interest and future stock returns is the only statistical significant one. In regression (2), the estimated coefficient of OTLR0 is about ten time the one of OTLR in regression (1), which indicates the information of OTLR comes from OTLR below median. Surprisingly, the result in regression (2) indicates the information of OS also comes from OS below median after controlling for OTLR and Short. As argued by Boehmer, Huszar, and Jordan (2010), no negative signal could be good news as informed traders stay away from such stocks, and the market seems to ignore such information. Controlling OS and OTLR, no significant nonlinearity is detected for Short.

In regression (3) and (4), I add firm size, book-to-market ratio, one-month lag returns and momentum returns of prior one year as controls. The estimated coefficients of OTLR and OTLR0 do not change much. In regression (5), I examine whether limits to arbitrage in the stock market explains the predictability of OTLR. Idiosyncratic volatility has been considered as proxy for arbitrage risk in the literature such as Stambaugh et al. (2015). Also, low liquidity in the stock market might drive informed traders to use options instead of stocks, for which I use Amihud (2002)s liquidity measure. The result in regression (5) shows a decrease of OTLR0 coefficient, from 0.51 to 0.35, but it is still positive and significant. In regression (6), I control for known mispricing measures documented by previous literature: mispricing score as in Stambaugh et al. (2015) and analyst dispersion as in Diether et al. (2002). The coefficient of OTLR0 does not change significantly. In regression (7) I include all variables in regard to the regression. The sum of estimated coefficients of OTLR0 and OTLR is about 0.51, which means that one standard deviation decrease of OTLR below median will predict 0.51% lower monthly stock returns, all else being equal, which is about 6% annually. Consistent with the prediction in Johnson and So (2012), OS is more informative when OTLR is low, as shown by the positive estimated coefficient of the interaction term in each regression of Table 9.

According to portfolio analysis as in section 3.3, portfolios formed by OTLR exhibit persistent profitability up to one year after portfolio formation. Thus, I regress monthly returns up to six months in the future as shown in table 10. Regression (n) in table 10 uses n-month ahead stock returns as a dependent variable. In general, OTLR does not contribute in predicting stocks returns further ahead in the future, after controlling for a series of stock and option characteristics.

4 Option trading leverage ratio and earnings

Earning is one of the most important pieces of information for stock valuation. If OTLR strongly predicts stock returns, it has a high chance to predict earnings information as well. In this section, I examine whether OTLR predicts analyst earnings forecast revision, earnings surprise and post earnings announcement returns (PEAD).

Following Jegadeesh et al. (2004), I measure monthly forecast revision (Est_rev) as changes of monthly analyst forecast average of quarterly earnings, scaled by lagged month end stock price. I use three measures to capture earnings surprise. The first measures earnings surprise using analyst forecast as the consensus, which is the difference between actual earnings and average analyst forecast just before the announcement, scaled by last fiscal quarter end price as in Livnat and Mendenhall (2006). The second measure is called standardized unexplained earnings or SUE in previous literature, calculated as the difference between current quarterly earnings and earnings of four-quarter ago, divided by standard deviation of this difference using previous two-year observations, as in Jegadeesh et al. (2004). "Street" earnings are used as suggested by Bradshaw and Sloan (2002), which is Compustat reported EPS minus special items and then times the difference by 65%. The third measure is about market reaction to quarterly earnings announcement, measured by CAR(-1,1) as cumulative market-adjusted return, the difference between stock return and market value-weighted return, from one day before the announcement to one day after the announcement.

Fama-MacBeth cross-section regressions are run every month for forecast revision or every calendar quarter for earnings surprise on variables observed as of previous month relative to the corresponding dependent variable. The time-series average of estimated coefficients and Newey and West (1987) adjusted t-statistics are reported in table 12. To facilitate interpretation of economic significance, both dependent and independent variables are standardized to mean 0 and standard deviation 1 in the cross section. Table 11 reports summary statistics for dependent variables studied, all in percentage except SUE.

Regression (1) in table 12 shows that after controlling a series of stock characteristics, low OTLR predicts negative analyst forecast revision, indicating analysts learn the same information possessed by informed option traders over time. Regression (2) shows that OTLR below median is the primary source of predictability. The estimated coefficients of interactive term between OTLR and OS is significantly negative, which contradicts the model implication in Johnson and So (2012).

For earnings surprise using analyst forecast consensus, according to regression (3) and (4), low OTLR below median predicts negative earnings surprise, with a linear coefficient about 0.023, while the predictive relationship is not statistically significant. Surprisingly, OTLR above median has a small negative association with earnings surprise. The predictive relationship between OS and earnings surprise using analyst forecast consensus is opposite with the one between OS and stock returns, which is not consistent with the model prediction in Johnson and So (2012). The multiplication of OTLR and OS has no additional information for earnings surprise. Regression (5) and (6) presents evidence that low OTLR is also associated with low SUE, with one standard deviation decrease of OTLR predicting 0.03 standard deviation decrease of SUE, all else being equal. CAR(-1,1) measures market reaction to the earnings announcement, reflecting how surprised the market actually is. According to regression (8), one standard deviation decrease of OTLR below median will predict 0.05 standard deviation decrease of CAR(-1,1), which is about 0.47% or about 40% after annualization. In sum, OTLR predicts analyst forecast revision, meaning that informed option investors trade their information before analysts incorporate the same information into their earnings forecasts. Although OTLR does not predict earnings surprise based on analyst forecast consensus, it predicts market reaction to earnings announcement. As a result, at least part of predictive power of OTLR comes from better skills in processing public information.

CAR(X,Y) denotes cumulative market adjusted return from X trading days to Y trading days relative to earnings announcement day. To examine whether OTLR predicts post earnings announcement returns, I regress CAR(2,10), CAR(2,20), CAR(2,40) and CAR(2,60) on OTLR and relevant stock characteristics as of month previous to earnings announcement. In all four cases, OTLR0 shows strong predictability as presented in Table 13. For example, one standard deviation decrease of OTLR0 predicts 0.06 standard deviation decrease of CAR(2,60), which is about 1.40% or 6% after annualized transformation. OS does not significantly predict post earnings announcement returns, while the multiplication of OTLR and OS shows that OTLR is more informative when OS is larger. The evidence indicates that informed option traders take advantage of market underreaction to earnings announcement.

Overall, I find OTLR possesses information to predict analyst forecast revision, earnings surprise and post earnings announcement returns. The evidence found is consistent with the assumption that OTLR reflects informed trading in the option market with negative information about underlying stocks. At least part of predictive power of OTLR comes from informed option investors who are better at analyzing public information compared to marginal investors.

5 Conclusions

In this paper I examine information in Option Trading Leverage Ratio (OTLR), calculated as the volume-weighted monthly average of option leverage. Sophisticated investors choose optimal leveraged positions as a function of risk and cost. Previous empirical evidence indicates a low level of leverage among institutional investors. For example, Ang et al. (2011) finds that the average leverage of hedge fund is about 2. In terms of option trading. informed investors use options of low leverage, such as at-the-money options, to establish his optimal exposures due to their avoidance of large cash balance and liquidity concern. On the other hand, noise traders such as retail investors are attracted to options of high leverage which have more lottery-like features. Such different trading preferences cause a divergent favored level of option leverage. Under the assumption that informed investors trade more negative information than positive information using options due to short-sale constraint in the stock market as argued by Johnson and So (2012), low option trading leverage should indicate the dominant presence of negatively informed investors, and predict negative future stock returns. Consistent empirical evidence is found to support the close connection between option trading leverage and negatively informed trading.

First, I find that OTLR is contemporaneously correlated with variables consistent with the informed trading assumption. Low OTLR happens more when short interest and relative volume ratio of option to stock are both high, where the latter two variables are documented signals for negatively informed trading. Institutional investors constitute the major population of informed traders. High institutional ownership indicates dominant presence of informed investors. Empirically low OTLR is associated with high institutional ownership of underlying equity. Also, low OTLR detects informed option traders who take advantage of stock mispricing. Facing high arbitrage risk and high illiquidity in the stock market, informed traders choose relatively low option leverage. When OTLR is low, option trading costs also tend to be low, reflecting that sophisticated investors care about option liquidity in terms of transaction costs.

Second, the quintile portfolio with lowest OTLR underperforms the quintile portfolio with highest OTLR by 1.28% monthly or 15% annually. Using Fama-Macbeth regression to control a series of previously documented predictive variables, OTLR has significant predictive power for one-month ahead stock returns. OTLR also forecasts firm earnings-related information, such as analyst forecast revision, earnings surprise and post-earnings announcement returns. The overall evidence is consistent with the assumption that OTLR reflects informed trading in the option market with negative information about underlying stocks. Also, at least part of predictive power of OTLR comes from informed option investors who are better at analyzing public information compared to marginal investors.

This paper provides a new tool, OTLR, to detect informed trading in the option market. OTLR utilizes the rich dynamics of options that divergent leverage exposures can be established using options with the same underlying equity. The predictive power of OTLR comes from different trading preferences between informed and noise investors motivated by previous literature. Derivatives such as options are considered redundant assets in a complete market. However, in reality, trading frictions of the underlying asset might induce asymmetric informed trading between the two markets. This paper contributes to the understanding of information discovery that happens in a multi-market environment.

References

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

An, B.-J., Ang, A., Bali, T. G., Cakici, N., 2014. The Joint Cross Section of Stocks and Options: The Joint Cross Section of Stocks and Options. The Journal of Finance 69, 2279–2337.

Ang, A., Gorovyy, S., van Inwegen, G. B., 2011. Hedge fund leverage. Journal of Financial Economics 102, 102–126.

Asquith, P., Pathak, P. A., Ritter, J. R., 2005. Short interest, institutional ownership, and stock returns. Journal of Financial Economics 78, 243–276.

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.

Bali, T. G., Hovakimian, A., 2009. Volatility Spreads and Expected Stock Returns. Management Science 55, 1797–1812.

Barber, B. M., Odean, T., 2008. All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. Review of Financial Studies 21, 785–818.

Bennett, J. A., Sias, R. W., Starks, L. T., 2003. Greener Pastures and the Impact of Dynamic Institutional Preferences. Review of Financial Studies 16, 1203–1238.

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. The journal of political economy pp. 637–654.

Boehmer, E., Huszar, Z. R., Jordan, B. D., 2010. The good news in short interest. Journal of Financial Economics 96, 80–97.

Bollen, N. P., Whaley, R. E., 2004. Does net buying pressure affect the shape of implied volatility functions? Journal of Finance pp. 711–753.

Boyer, B. H., Vorkink, K., 2014. Stock Options as Lotteries. The Journal of Finance 69, 1485–1527.

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.

Chaudhury, M., 2014. Option Bid-Ask Spread and Liquidity. SSRN Scholarly Paper ID 1769126, Social Science Research Network, Rochester, NY.

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.

Cox, J. C., Ross, S. A., Rubinstein, M., 1979. Option pricing: A simplified approach. Journal of financial Economics 7, 229–263.

Cremers, M., Weinbaum, D., 2010. Deviations from Put-Call Parity and Stock Return Predictability. Journal of Financial and Quantitative Analysis 45, 335–367.

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.

DAvolio, G., 2002. The market for borrowing stock. Journal of Financial Economics 66, 271–306.

De Long, J. B., Shleifer, A., Summers, L. H., Waldmann, R. J., 1990. Noise Trader Risk in Financial Markets. Journal of Political Economy 98, 703–738.

Diether, K. B., Malloy, C. J., Scherbina, A., 2002. Differences of opinion and the cross section of stock returns. The Journal of Finance 57, 2113–2141.

Easley, D., O'hara, M., Srinivas, P. S., 1998. Option volume and stock prices: Evidence on where informed traders trade. The Journal of Finance 53, 431–465.

Engelberg, J. E., Reed, A. V., Ringgenberg, M. C., 2012. How are shorts informed?: Short sellers, news, and information processing. Journal of Financial Economics 105, 260–278.

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.

Fama, E. F., MacBeth, J. D., 1973. Risk, Return, and Equilibrium: Empirical Tests. Journal of Political Economy 81, 607–36.

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

Garleanu, N., Pedersen, L. H., Poteshman, A. M., 2009. Demand-Based Option Pricing. Review of Financial Studies 22, 4259–4299.

Ge, L., Lin, T.-C., Pearson, N. D., 2016. Why does the option to stock volume ratio predict stock returns? Journal of Financial Economics 120, 601–622.

Grossman, S. J., Stiglitz, J. E., 1980. On the Impossibility of Informationally Efficient Markets. American Economic Review 70, 393–408.

Harris, M., Raviv, A., 1993. Differences of Opinion Make a Horse Race. Review of Financial Studies 6, 473–506.

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., 1999. A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets. The Journal of Finance 54, 2143–2184.

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

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.

Johnson, T. L., So, E. C., 2012. The option to stock volume ratio and future returns. Journal of Financial Economics 106, 262–286.

Kandel, E., Pearson, N. D., 1995. Differential Interpretation of Public Signals and Trade in Speculative Markets. Journal of Political Economy 103, 831–72.

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.

Krmer, N., Sugiyama, M., 2011. The Degrees of Freedom of Partial Least Squares Regression. Journal of the American Statistical Association 106, 697–705.

Kumar, A., 2009. Who Gambles in the Stock Market? The Journal of Finance 64, 1889–1933.

Kyle, A. S., 1985. Continuous Auctions and Insider Trading. Econometrica 53, 1315–35.

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.

Loh, R. K., Stulz, R. M., 2018. Is Sell-Side Research More Valuable in Bad Times?*. The Journal of Finance 0.

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.

Muravyev, D., Pearson, N. D., Broussard, J. P., 2013. Is there price discovery in equity options? Journal of Financial Economics 107, 259–283.

Nagel, S., 2005. Short sales, institutional investors and the cross-section of stock returns. Journal of Financial Economics 78, 277–309.

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

Pan, J., Poteshman, A. M., 2006. The Information in Option Volume for Future Stock Prices. Review of Financial Studies 19, 871–908.

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.

Roll, R., Schwartz, E., Subrahmanyam, A., 2009. Options trading activity and firm valuation. Journal of Financial Economics 94, 345–360.

Roll, R., Schwartz, E., Subrahmanyam, A., 2010. O/S: The relative trading activity in options and stock. Journal of Financial Economics 96, 1–17.

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

Stambaugh, R. F., Yu, J., Yuan, Y., 2012. The short of it: Investor sentiment and anomalies. Journal of Financial Economics 104, 288–302.

Stambaugh, R. F., Yu, J., Yuan, Y., 2015. Arbitrage Asymmetry and the Idiosyncratic Volatility Puzzle. The Journal of Finance 70, 1903–1948.

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.

Xing, Y., Zhang, X., Zhao, R., 2010. What Does the Individual Option Volatility Smirk Tell Us About Future Equity Returns? Journal of Financial and Quantitative Analysis 45, 641–662.

Table 1 Summary Statistics

Panel A reports quarterly average of summary statistics for variables listed in the first column. SUE_p (SUE_{np}) is the standardized unexpected earnings for the sample with (without) 3PRF predictions. $\hat{U}E$ is the predicted earnings news from 3PRF, estimated using all available early announcers at the end of four trading days before the earnings announcement. PSUE is the predicted SUE derived from $\hat{U}E$. FE_p (FE_{np}) is analyst forecast error using analyst forecasts up to three trading days before the earnings announcement date for the sample with (without) 3PRF predictions. PFE is the predicted analyst forecast error derived from $\hat{U}E$. AR_p (AR_{np}) is the earnings announcement return adjusted by market return. All variables except AR are winzorized by 1%. The sample for SUE and PSUE is from Jan 1981 to Dec 2016, while the sample for FE and PFE is from Jan 1985 to Dec 2016. Panel B reports summarized information about $\hat{U}E$, with subsamples from Jan, 1981 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 the 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 statistics

Variable	N	Mean	STD	Median	Skew	Kurt
SUE_p	775	0.15	1.28	0.11	0.33	0.67
SUE_{np}	1288	0.35	1.37	0.25	0.43	0.63
PSUE	775	0.19	1.03	0.11	0.91	3.16
FE_p	477	-0.47%	2.53%	-0.01%	-4.24	29.11
FE_{np}	1304	-0.39%	2.36%	0.00%	-5.16	41.58
PFE	477	-0.37%	4.11%	-0.04%	-2.39	21.90
AR_p	809	0.13%	8.87%	-0.15%	1.12	17.26
AR_{np}	1282	0.35%	9.40%	0.02%	1.73	26.06

Panel B: summary of $\hat{U}E$

	Predictors		Factors		
Sample	T	N	One	Two	Three
1981 to 2000	30	233	27	357	22
2001 to 2016	36	397	126	1134	70

Table 2 Forecast of Quarterly Earnings Surprise

This table presents the regression results in forecasting standardized unexpected earnings (SUE). The leftmost column shows regressors. PSUE is the predicted SUE by 3PRF, estimated using all available early announcers at the end of four trading days before the earnings announcement. $PSUE_a$ is predicted SUE using analyst forecasts available at the end of three trading days before the earnings announcement. 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 three trading day before the earnings announcement. 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. Turn is the share turnover calculated as the average daily share turnover over the prior 6 month. Daily turnover is the number of shares traded divided by the shares outstanding. IVOL is the idiosyncratic volatility using three-factor model. 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	0.454***	0.213***	0.213***	0.204***	0.218***	0.069***	0.145***
	(0.003)	(0.004)	(0.012)	(0.012)	(0.009)	(0.007)	(0.033)
$PSUE_a$						0.706***	0.486***
						(0.017)	(0.021)
Lag(SUE)		0.333***	0.333***	0.259***	0.324***	0.126***	0.134**
		(0.003)	(0.010)	(0.009)	(0.008)	(0.007)	(0.055)
Log(Size)		0.056***	0.056***	-0.052***	0.061***	-0.104***	0.092
		(0.004)	(0.007)	(0.017)	(0.006)	(0.022)	(0.062)
Log(BM)		-0.042***	-0.042***	-0.086***	-0.047***	-0.082***	-0.068**
		(0.003)	(0.005)	(0.009)	(0.006)	(0.011)	(0.028)
Cum_alpha		0.064***	0.064***	0.059***	0.080***	0.048***	0.095***
		(0.003)	(0.004)	(0.004)	(0.005)	(0.004)	(0.021)
Hist_alpha		0.128***	0.128***	0.144***	0.148***	0.085***	0.219***
		(0.003)	(0.005)	(0.005)	(0.006)	(0.006)	(0.063)
TA		-0.028***	-0.028***	-0.040***	-0.029***	-0.010**	0.007
		(0.003)	(0.004)	(0.004)	(0.005)	(0.004)	(0.023)
Log(Turn)		-0.034***	-0.034***	-0.070***	-0.034***	-0.045***	-0.104
		(0.003)	(0.006)	(0.007)	(0.005)	(0.010)	(0.066)
IVOL		-0.052***	-0.052***	-0.013**	-0.034***	-0.031***	0.021
		(0.003)	(0.009)	(0.005)	(0.006)	(0.008)	(0.062)
N	110810	102462	102462	102070	102462	58454	58889
adj. R-sq	0.182	0.301	0.301	0.322	0.300	0.581	0.486
Start	1981	1981	1981	1981	1981	1981	1981
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 3 Reaction of Analysts to Information Captured by 3PRF

This table reports the contemporaneous regressions of analyst revisions on the predicted analyst revisions derived from ÛE, which is the predicted earnings news from 3PRF, estimated using all available early announcers at the end of four trading days before the earnings announcement. The dependent variable is the analyst revision from the corresponding fiscal quarter end 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 the earnings announcement day. EA-n represents n trading days before EA. Analyst revision is the difference between the median of analyst forecast consensus divided by the fiscal quarter end price. PFR is the predicted revision derived from ÛE. 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 three trading day before the earnings announcement. 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. Turn is the share turnover calculated as the average daily share turnover over the prior 6 month. Daily turnover is the number of shares traded divided by the shares outstanding. IVOL is the idiosyncratic volatility using three-factor model. Intercept is not reported for abbreviation. 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
PFR	0.095***	0.053***	0.053***	0.051***	0.075*
	(0.005)	(0.006)	(0.012)	(0.011)	(0.041)
Lag(FE)		0.002	0.002	-0.005	0.084**
		(0.006)	(0.016)	(0.014)	(0.038)
Log(Size)		0.056***	0.056***	0.064	0.100***
		(0.008)	(0.013)	(0.043)	(0.023)
Log(BM)		-0.045***	-0.045***	-0.034*	-0.049***
		(0.006)	(0.009)	(0.018)	(0.012)
Cum_alpha		0.161***	0.161***	0.140***	0.142***
		(0.006)	(0.014)	(0.012)	(0.022)
Hist_alpha		0.144***	0.144***	0.134***	0.150***
		(0.006)	(0.012)	(0.011)	(0.034)
TA		0.011*	0.011*	0.007	0.031
		(0.006)	(0.006)	(0.006)	(0.032)
Log(Turn)		0.029***	0.029***	0.010	0.023
		(0.008)	(0.011)	(0.022)	(0.021)
IVOL		-0.211***	-0.211***	-0.140***	-0.201***
		(0.009)	(0.024)	(0.024)	(0.032)
N	35251	32026	32026	31428	32026
adj. R-sq	0.009	0.077	0.077	0.161	0.071
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 4 Forecast of analyst forecast errors

This table reports regressions of analyst forecast errors on the predicted forecast errors 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 three trading days before the earnings announcement. PFE is the predicted analyst forecast error by 3PRF, estimated using all available early announcers at the end of four trading days before the earnings announcement. 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 three trading day before the earnings announcement. 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. Turn is the share turnover calculated as the average daily share turnover over the prior 6 month. Daily turnover is the number of shares traded divided by the shares outstanding. IVOL is the idiosyncratic volatility using three-factor model. Intercept is not reported for abbreviation. 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	0.194***	0.107***	0.107***	0.071***	0.109***
	(0.004)	(0.005)	(0.017)	(0.017)	(0.020)
Lag(FE)		0.164***	0.164***	0.069***	0.161***
		(0.005)	(0.018)	(0.017)	(0.024)
Log(Size)		-0.014***	-0.014	0.099*	0.021**
		(0.006)	(0.015)	(0.050)	(0.010)
Log(BM)		-0.109***	-0.109***	-0.177***	-0.075***
		(0.004)	(0.013)	(0.024)	(0.010)
Cum_alpha		0.063***	0.063***	0.056***	0.064***
		(0.004)	(0.008)	(0.008)	(0.009)
Hist_alpha		0.159***	0.159***	0.147***	0.155***
		(0.004)	(0.013)	(0.015)	(0.014)
TA		0.005	0.005	0.001	0.020***
		(0.004)	(0.005)	(0.006)	(0.007)
Log(Turn)		-0.074***	-0.074***	-0.132***	-0.074***
		(0.005)	(0.012)	(0.018)	(0.011)
IVOL		-0.260***	-0.260***	-0.221***	-0.169***
		(0.006)	(0.038)	(0.032)	(0.018)
N	68251	56944	56944	56494	56944
adj. R-sq	0.035	0.150	0.150	0.210	0.147
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 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 is the predicted SUE by 3PRF, estimated using all available early announcers at the end of four trading days before the earnings announcement. $PSUE_a$ is predicted SUE using analyst forecasts available at the end of three trading days before the earnings announcement. 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 three trading day before the earnings announcement. 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. Turn is the share turnover calculated as the average daily share turnover over the prior 6 month. Daily turnover is the number of shares traded divided by the shares outstanding. IVOL is the idiosyncratic volatility using three-factor model. 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
PUSE	0.167***	0.231***	0.231***	0.232***	0.245***	0.123**	0.298**
	(0.031)	(0.041)	(0.056)	(0.053)	(0.065)	(0.062)	(0.129)
$PSUE_a$						0.153**	0.215**
						(0.060)	(0.095)
Lag(SUE)		-0.081**	-0.081	-0.185***	0.040	-0.328***	-0.091
		(0.040)	(0.060)	(0.051)	(0.056)	(0.064)	(0.142)
Log(Size)		0.267***	0.267***	-2.302***	0.078	-2.367***	-0.054
		(0.042)	(0.085)	(0.248)	(0.056)	(0.379)	(0.155)
Log(BM)		0.210***	0.210***	0.042	0.207***	-0.079	0.200
		(0.033)	(0.067)	(0.095)	(0.052)	(0.113)	(0.164)
Cum_alpha		-0.326***	-0.326***	-0.457***	-0.528***	-0.284***	-0.401***
		(0.032)	(0.058)	(0.054)	(0.056)	(0.080)	(0.128)
Hist_alpha		0.121***	0.121*	-0.177**	0.138**	-0.285***	0.162
		(0.034)	(0.072)	(0.074)	(0.060)	(0.102)	(0.109)
TA		-0.038	-0.038	-0.100**	-0.150***	-0.024	-0.240*
		(0.033)	(0.047)	(0.043)	(0.053)	(0.053)	(0.125)
Log(Turn)		-0.468***	-0.468***	-0.609***	-0.399***	-0.554***	-0.102
		(0.037)	(0.092)	(0.103)	(0.070)	(0.162)	(0.180)
IVOL		0.064	0.064	0.164	-0.084	0.154	-0.131
		(0.039)	(0.151)	(0.123)	(0.076)	(0.250)	(0.175)
N	115205	102797	102797	102403	102797	58820	59250
adj. R-sq	0.000	0.004	0.004	0.029	0.003	0.031	0.001
Start	1981	1981	1981	1981	1981	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 6 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, which is the predicted SUE by 3PRF, estimated using all available early announcers at the end of four trading days before the earnings announcement. Size is market capitalization (in million dollars) of the stock. BM is the ratio of book equity and market capitalization. SRET is the stock return of the most recent month. MOM is the cumulative stock return of 11 months prior to the most recent month. 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 1981 to Dec 2016.

PanelA: quarterly average of stock characteristics

PSUE	Size	BM	SRET	MOM	Amihud	Turn	TVOL	IVOL	TA
1	1234	0.93	0.13%	-1.12%	3.27	44.29%	46.12%	38.59%	2.40%
2	1466	0.91	0.59%	5.88%	2.89	42.77%	43.91%	36.69%	1.63%
3	1549	0.92	0.79%	13.05%	2.68	42.53%	42.39%	35.30%	0.46%
4	1633	0.88	0.96%	20.23%	2.63	42.92%	41.76%	34.66%	0.53%
5	2229	0.74	1.26%	26.68%	1.64	47.15%	39.45%	32.53%	2.55%

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

PSUE	Size	BM	SRET	MOM	Amihud	Turn	TVOL	IVOL	TA
1	54	59	48	42	45	51	45	45	49
2	56	58	49	47	43	51	42	42	47
3	57	58	50	52	42	51	41	41	45
4	58	56	51	57	41	51	41	40	45
5	63	49	51	61	36	54	39	38	51

Table 7 Market Reaction to Information Captured by $PSUE_f$

This table reports the value-weighted event returns of stock groups. [Q,EA-3] is the period from the fiscal quarter end to three trading days before the earnings announcement. [EA-2,EA+2] is the period from two trading days before the announcement to two trading days after the announcement. Event returns are presented in two forms of cumulative returns in percentage: raw return and alpha based on the four factor model. Panel A reports the five stock groups sorted by PSUE. To control for certain characteristics (SRET in panel B, MOM in panel C, and CumRet[Q,EA-3] in panel D), 5 by 5 dependent double sorting is performed firstly on the characteristics and then on PSUE. SRET is the previous month return. MOM is the cumulative stock return of 11 months prior to the previous month. CumRet[Q,EA-3] is the cumulative returns from the fiscal quarter end to three trading days before the earnings announcement. Panel B to Panel D report the stock groups formed by the PSUE rank in each group formed by the characteristics.

		[Q,EA-3]		[EA-2,EA+2]	
Panel A: stock groups formed by PSUE					
PSUE	SUE	Return	Alpha	Return	Alpha
Low	-0.41	0.69	-0.53	0.17	-0.16
2	-0.13	0.71	-0.34	0.29	-0.10
3	0.19	0.61	-0.37	0.60	0.20
4	0.51	0.69	-0.25	0.47	0.08
High	1.34	0.93	-0.11	0.77	0.38
H-L	1.75***	0.24	0.42*	0.60***	0.54***
T-stat	19.47	0.99	1.73	2.74	3.24
Panel B: control for SRET					
PSUE	SUE	Return	Alpha	Return	Alpha
Low	-0.38	0.83	-0.43	0.23	-0.14
2	-0.13	0.46	-0.51	0.28	-0.11
3	0.19	0.80	-0.28	0.52	0.19
4	0.51	0.72	-0.19	0.55	0.12
High	1.33	0.89	-0.08	0.81	0.41
H-L	1.71***	0.06	0.36	0.58***	0.55***
T-stat	19.38	0.24	1.34	2.81	3.2
Panel C: control for MOM					
PSUE	SUE	Return	Alpha	Return	Alpha
Low	-0.33	0.58	-0.53	0.18	-0.10
2	-0.05	0.65	-0.31	0.40	-0.04
3	0.19	0.32	-0.69	0.74	0.33
4	0.56	1.04	0.07	0.47	0.10
High	1.29	0.94	-0.13	0.79	0.38
H-L	1.62***	0.36*	0.40*	0.61***	0.48***
T-stat	19.49	1.82	1.66	2.64	2.66
Panel D: control for CumRet[Q, EA-3]					
PSUE	SUE	Return	Alpha	Return	Alpha
Low	-0.38	0.83	-0.42	0.28	-0.09
2	-0.12	0.82	-0.27	0.29	-0.04
3	0.17	0.82	-0.09	0.46	0.06
4	0.54	0.56	-0.41	0.52	0.12
High	1.33	0.76	-0.23	0.84	0.40
H-L	1.71***	-0.07	0.19	0.56***	0.48***
T-stat	19.76	-0.33	0.79	2.88	3.36

Table 8 Trading Strategies on Information Spillover

This table reports weekly returns in percentage of trading strategies that long(short) stocks with highest (lowest) PSUE quintile. The portfolios are formed at the end of the first trading day of each week, and liquidate at the end of the first trading day of the next week. PSUE is calculated using all available early announcers up to the last trading day of the previous week. To calculate the 3PRF prediction, require at least 100 qualified early announcers. The long(short) side requires at least ten stocks to maintain the portfolio. Require stock price $\geq$ \$1 on formation date. Panel A reports the summary of the number of trading weeks each year from 1981 to 2016. Panel B (Panel C) reports value weighted weekly portfolio returns in excess of risk-free return and value weighted weekly 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 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 weeks

Mean	STD	P5	P25	Median	P75	P95
16	5	10	13	16	19	24

Panel B: value-weighted returns

	Long	Short	L-S	Mktrf	SMB	HML	UMD
Ret-Rf	0.54	0.05	0.49***				
	(4.16)	(0.36)	(4.00)				
CAPM	0.27	-0.24	0.51***	-0.06			
	(2.64)	(-2.66)	(4.18)	(-0.86)			
FF3	0.26	-0.30	0.55***	-0.08	-0.12	-0.31***	
	(2.51)	(-3.24)	(4.40)	(-1.35)	(-1.08)	(-3.00)	
Carhart4	0.25	-0.28	0.53***	-0.05	-0.15	-0.20*	0.21**
	(2.45)	(-3.20)	(4.28)	(-0.63)	(-1.32)	(-1.85)	(2.57)

Panel C: equal-weighted returns

	Long	Short	L-S	Mktrf	SMB	HML	UMD
Ret-Rf	0.54	0.12	0.42***				
	(4.46)	(0.93)	(5.12)				
CAPM	0.27	-0.14	0.41***	0.04			
	(3.16)	(-1.68)	(5.10)	(0.85)			
FF3	0.26	-0.17	0.43***	0.03	0.00	-0.14*	
	(3.83)	(-2.60)	(5.41)	(0.62)	(0.02)	(-1.89)	
Carhart4	0.25	-0.16	0.41***	0.07*	-0.04	-0.02	0.26***
	(3.78)	(-2.47)	(5.40)	(1.86)	(-0.50)	(-0.24)	(5.74)

Table 9 Subsample tests

This table presents regression results for two subsamples. The early sample is from Jan 1981 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. $\hat{U}E$ is the predicted earnings news from 3PRF, estimated using all available early announcers at the end of four trading days before the earnings announcement. PSUE is the predicted SUE derived from $\hat{U}E$. PFE is the predicted analyst forecast error derived from $\hat{U}E$. 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 three trading day before the earnings announcement. 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. Turn is the share turnover calculated as the average daily share turnover over the prior 6 month. Daily turnover is the number of shares traded divided by the shares outstanding. IVOL is the idiosyncratic volatility using three-factor model. 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. 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	0.191***	0.205***			0.275***	0.232***
	(0.017)	(0.014)			(0.076)	(0.066)
PFE			0.088**	0.064***		
			(0.036)	(0.019)		
Lag(SUE)	0.272***	0.239***			0.039	-0.301***
	(0.013)	(0.010)			(0.075)	(0.061)
Lag(FE)			-0.049	0.076***		
			(0.030)	(0.019)		
Log(Size)	-0.142***	-0.072***	0.134*	0.107	-2.855***	-2.970***
	(0.037)	(0.023)	(0.072)	(0.066)	(0.363)	(0.317)
Log(BM)	-0.112***	-0.092***	-0.123***	-0.193***	0.067	-0.037
	(0.019)	(0.011)	(0.033)	(0.029)	(0.170)	(0.117)
Cum_alpha	0.084***	0.051***	0.076***	0.054***	-0.652***	-0.433***
	(0.008)	(0.004)	(0.015)	(0.009)	(0.104)	(0.063)
Hist_alpha	0.197***	0.137***	0.186***	0.142***	-0.086	-0.232**
	(0.010)	(0.006)	(0.023)	(0.017)	(0.087)	(0.091)
TA	-0.064***	-0.039***	0.021	-0.003	-0.314***	-0.064
	(0.009)	(0.005)	(0.014)	(0.006)	(0.107)	(0.047)
Log(Turn)	-0.096***	-0.078***	-0.058*	-0.148***	-0.572***	-0.741***
	(0.015)	(0.009)	(0.032)	(0.021)	(0.148)	(0.130)
IVOL	-0.030**	-0.004	-0.130***	-0.229***	0.184	0.143
	(0.011)	(0.006)	(0.034)	(0.036)	(0.142)	(0.157)
N	25521	76409	9200	47144	25720	76537
adj. R-sq	0.364	0.311	0.185	0.217	0.046	0.031
Start	1981	2001	1985	2001	1981	2001
End	2000	2016	2000	2016	2000	2016
Fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Clustered Std Error	Yes	Yes	Yes	Yes	Yes	Yes

Table 10 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 [EA-7, EA-3] of different informed group. EA-n denotes n trading days prior to the earnings announcement. IND is a group of stocks of the same industry with the forecast target as defined by Fama and French (1997) 48 industries. CUS(SUP) 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. X (IOR, TURN or Analyst) is the group of stocks with the highest ranking of X when stocks are firstly sorted to four groups by size and then sorted to four groups by X within each size group. IOR denotes institutional ownership, TURN is the stock turnover, and Analyst is the number of analyst that cover the stock. Other controls are lag(SUE), log(size), log(BM), log(Cum_alpha), log(Hist_alpha), TA, IVOL, log(TURN). The sample is from Jan 1981 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 SUE

	1	2	3	4	5	6	7	8
PSUE	0.204***	0.201***	0.201***	0.204***	0.205***	0.204***	0.200***	0.197***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.013)
IND_SUE		0.062***						0.041***
		(0.005)						(0.006)
CUS_SUE			0.052***					0.034***
			(0.005)					(0.006)
SUP_SUE			0.017***					0.014**
			(0.006)					(0.006)
BIG_SUE				-0.072				-0.069
				(0.042)				(0.041)
IOR_SUE					-0.059			-0.038
					(0.032)			(0.028)
TURN_SUE						-0.040**		-0.004
						(0.017)		(0.021)
Analyst_SUE							0.010	0.020
							(0.032)	(0.034)

Panel B: controls of value-weighted average of five-day cumulative returns from [EA-7, EA-3]

	1	2	3	4	5	6	7	8
PSUE	0.204***	0.204***	0.203***	0.204***	0.205***	0.204***	0.200***	0.203***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
IND_RET		-0.000						0.001
		(0.004)						(0.006)
CUS_RET			-0.001					-0.002
			(0.005)					(0.007)
SUP_RET			0.002					0.002
			(0.005)					(0.007)
BIG_RET				-0.000				-0.010
				(0.004)				(0.010)
IOR_RET					0.001			
					(0.004)			
TURN_RET						0.002		0.011
						(0.004)		(0.009)
Analyst_RET							-0.004	
							(0.009)	

Table 11 Robustness Tests of FE Forecast

This table presents forecasting regressions of FE on the 3PRF measure controlling additional information channels. 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 [EA-7, EA-3] of different informed group. EA-n denotes n trading days prior to the earnings announcement. IND is a group of stocks of the same industry with the forecast target as defined by Fama and French (1997) 48 industries. CUS (SUP) 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. X (IOR, TURN or Analyst) is the group of stocks with the highest ranking of X when stocks are firstly sorted to four groups by size and then sorted to four groups by X within each size group. IOR denotes institutional ownership, TURN is the stock turnover, and Analyst is the number of analyst that cover the stock. Other controls are lag(SUE), log(size), log(BM), log(Cum_alpha), log(Hist_alpha), TA, IVOL, log(TURN). 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 FE

	1	2	3	4	5	6	7	8
PFE	0.071***	0.070***	0.068***	0.071***	0.067***	0.071***	0.071***	0.065***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.018)	(0.017)	(0.017)	(0.018)
IND_FE		0.057***						0.024**
		(0.013)						(0.010)
CUS_FE			0.076***					0.062***
			(0.018)					(0.018)
SUP_FE			-0.005					-0.010
			(0.013)					(0.014)
BIG_FE				0.022				-0.052
				(0.034)				(0.036)
IOR_FE					0.039			0.045
					(0.028)			(0.028)
TURN_FE						0.031		0.040
						(0.032)		(0.030)
Analyst_FE							-0.045	
							(0.076)	

Panel B: controls of value-weighted average of five-day cumulative returns from [EA-7, EA-3]

	1	2	3	4	5	6	7	8
PFE	0.071***	0.073***	0.071***	0.071***	0.067***	0.071***	0.071***	0.072***
	(0.017)	(0.018)	(0.018)	(0.017)	(0.018)	(0.017)	(0.017)	(0.018)
IND_RET		0.008						0.008
		(0.007)						(0.012)
CUS_RET			0.007					0.005
			(0.009)					(0.013)
SUP_RET			0.000					0.010
			(0.009)					(0.012)
BIG_RET				-0.001				-0.021
				(0.006)				(0.016)
IOR_RET					0.000			
					(0.006)			
TURN_RET						-0.001		-0.000
						(0.006)		(0.012)
Analyst_RET							0.008	
							(0.022)	

Table 12 Robustness Tests of AR Forecast

This table presents forecasting regressions of AR (in percentage) on the 3PRF measure controlling additional information channels. 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 [EA-7, EA-3] of different informed group. EA-n denotes n trading days prior to the earnings announcement. IND is a group of stocks of the same industry with the forecast target as defined by Fama and French (1997) 48 industries. CUS (SUP) 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. X (IOR, TURN or Analyst) is the group of stocks with the highest ranking of X when stocks are firstly sorted to four groups by size and then sorted to four groups by X within each size group. IOR denotes institutional ownership, TURN is the stock turnover, and Analyst is the number of analyst that cover the stock. Other controls are lag(SUE), log(size), log(BM), log(Cum_alpha), log(Hist_alpha), TA, IVOL, log(TURN). The sample is from Jan 1985 to Dec 2016. All independent variables are winzorized by 1% and standardized to mean 0 and standard deviation 1.

Panel A: controls of value-weighted average of AR

	1	2	3	4	5	6	7	8
PSUE	0.232***	0.227***	0.227***	0.232***	0.224***	0.231***	0.226***	0.202***
	(0.053)	(0.053)	(0.055)	(0.054)	(0.055)	(0.054)	(0.055)	(0.059)
IND_AR		0.042						0.027
		(0.044)						(0.039)
CUS_AR			0.049					0.037
			(0.054)					(0.056)
SUP_AR			0.007					0.050
			(0.053)					(0.057)
BIG_AR				-0.090				0.036
				(0.179)				(0.171)
IOR_AR					-0.312			-0.270
					(0.211)			(0.151)
TURN_AR						-0.329		-0.253
						(0.201)		(0.193)
Analyst_AR							0.133	0.070
							(0.172)	(0.173)

Panel B: controls of value-weighted average of five-day cumulative returns from [EA-7, EA-3]

	1	2	3	4	5	6	7	8
PSUE	0.232***	0.227***	0.227***	0.232***	0.225***	0.232***	0.227***	0.222***
	(0.053)	(0.053)	(0.055)	(0.054)	(0.055)	(0.054)	(0.055)	(0.055)
IND_RET		0.031						0.075
		(0.084)						(0.064)
CUS_RET			0.042					-0.024
			(0.066)					(0.073)
SUP_RET			-0.039					-0.127
			(0.094)					(0.088)
BIG_RET				0.040				0.117
				(0.113)				(0.120)
IOR_RET					-0.004			
					(0.122)			
TURN_RET						-0.007		-0.029
						(0.113)		(0.117)
Analyst_RET							-0.101	
							(0.105)	

Table 13 Forecast of Quarterly Earnings Surprise

This table presents the regression results in forecasting standardized unexpected earnings (SUE). The leftmost column shows regressors. PSUE is the predicted SUE by 3PRF, estimated using all available early announcers at the end of four trading days before the earnings announcement. I_MED (I_SMALL) equals to PSUE if the stock belongs to the size percentile [34,67] ([1,33]) in the quarterly cross section, and equals to 0 otherwise. IOR denotes institutional ownership. Analyst is the number of analyst that cover the stock. TURN is the stock turnover. IVOL is the idiosyncratic volatility. Stocks are double sorted to three groups firstly by size, and then sorted to three groups by a characteristics X (IOR, ANALYST, TURN or IVOL) within each size group. I_MED_X (I_LOW_X) equals to PSUE if the stock belongs to the secondly (thirdly) ranked group based on X, and equals to 0 otherwise. I_REC equals to PSUE if the fiscal quarter of SUE overlaps with the recession periods and equals to 0 otherwise. Other controls are lag(SUE), log(size), log(BM), log(Cum_alpha), log(Hist_alpha), TA, IVOL, log(TURN). The sample is from Jan 1981 to Dec 2016.

	1	2	3	4	5	6
PSUE	0.272***	0.285***	0.253***	0.286***	0.261***	0.282***
	(0.014)	(0.016)	(0.017)	(0.016)	(0.015)	(0.014)
I_MED	-0.092***	-0.091***	-0.084***	-0.091***	-0.092***	-0.092***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
I_SMALL	-0.135***	-0.135***	-0.125***	-0.135***	-0.136***	-0.135***
	(0.012)	(0.012)	(0.013)	(0.013)	(0.012)	(0.013)
I_MED_IOR		0.000				
		(0.009)				
I_LOW_IOR		-0.044***				
		(0.011)				
I_MED_ANALYST			0.010			
			(0.010)			
I_LOW_ANALYST			0.017			
			(0.011)			
I_MED_TURN				-0.018*		
				(0.010)		
I_LOW_TURN				-0.025**		
				(0.012)		
I_MED_IVOL					0.020**	
					(0.009)	
I_LOW_IVOL					0.011	
					(0.010)	
I_REC						-0.048**
						(0.020)
N	102070	102070	95421	102070	102070	102070
Start	1981	1981	1981	1981	1981	1981
End	2016	2016	2016	2016	2016	2016

Table 14 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 is the predicted SUE by 3PRF, estimated using all available early announcers at the end of four trading days before the earnings announcement. I_MED (I_SMALL) equals to PSUE if the stock belongs to the size percentile [34,67] ([1,33]) in the quarterly cross section, and equals to 0 otherwise. IOR denotes institutional ownership. Analyst is the number of analyst that cover the stock. TURN is the stock turnover. IVOL is the idiosyncratic volatility. Stocks are double sorted to three groups firstly by size, and then sorted to three groups by a characteristics X (IOR, ANALYST, TURN or IVOL) within each size group. I_MED_X (I_LOW_X) equals to PSUE if the stock belongs to the secondly (thirdly) ranked group based on X, and equals to 0 otherwise. I_REC equals to PSUE if the fiscal quarter of SUE overlaps with the recession periods and equals to 0 otherwise. Other controls are lag(SUE), log(size), log(BM), log(Cum_alpha), log(Hist_alpha), TA, IVOL, log(TURN). The sample is from Jan 1981 to Dec 2016.

	1	2	3	4	5	6
PSUE	0.175***	0.038	0.007	0.055	0.115	0.205***
	(0.056)	(0.083)	(0.099)	(0.088)	(0.091)	(0.055)
I_MED	0.000	0.002	0.030	-0.003	-0.010	0.001
	(0.073)	(0.073)	(0.075)	(0.073)	(0.073)	(0.073)
I_SMALL	0.221**	0.229**	0.222**	0.221**	0.208**	0.220**
	(0.104)	(0.103)	(0.106)	(0.104)	(0.105)	(0.104)
I_MED_IOR		0.235**				
		(0.100)				
I_LOW_IOR		0.187**				
		(0.093)				
I_MED_ANALYST			0.249**			
			(0.111)			
I_LOW_ANALYST			0.167			
			(0.125)			
I_MED_TURN				0.150		
				(0.094)		
I_LOW_TURN				0.214**		
				(0.099)		
I_MED_IVOL					-0.033	
					(0.097)	
I_LOW_IVOL					0.236***	
					(0.087)	
I_REC						-0.150
						(0.138)
N	102403	102403	95754	102403	102403	102403
adj. R-sq	0.029	0.030	0.030	0.029	0.030	0.029

Table 15

Sample Summary of OTLR and portfolios

Panel A shows time series average of cross section summary statistics for option trading leverage ratio (OTLR) by subsamples. All variables are obtained at monthly frequency. OTLR is calculated for each underlying stock, which is the weighted monthly average of option absolute leverage, using trading volumes as weights. Panel B shows time-series average of characteristics of quintile portfolios formed by OTLR. OTLRC (OTLRP) is option trading leverage ratio calculated using call (put) options only. OTDR is monthly average of option absolute delta, value-weighted by trading volumes. OTMR is monthly average of option strike to underlying price for call, and one minus this ratio for put, value-weighted by trading volumes. At-the-money option premium (Prem) is the difference between average of implied volatilities of at-the-money (ATM) options and historical volatility of the underlying stock. ATM options include call and put options with moneyness between 0.9 and 1.1. Historical volatility is annualized standard deviation of daily stock returns in one month. ATM option bid-ask spread (OBA) is the average of bid ask spreads of ATM options. Size is market capitalization in million dollars. NYSE size group (NYSE) is the average of size decile formed using NYSE based market capitalization deciles. Book-to-market (BM) is the ratio of book equity and market capitalization. Momentum (Mom) is the cumulative stock return of prior six months. The sample spans from 1996 to 2014. Option variables are calculated using options with 15 to 45 days to expiration. Stocks with price less than $5 are excluded. Each monthly cross section is winzorized by 1% at each tail.

Panel A: Summary statistics of option trading leverage ratio (OTLR) by subsamples

Start	End	Firms	Mean	STD	P25	Median	P75	Skew
1996	2000	927	9.64	4.02	6.61	8.54	12.01	0.96
2001	2005	1,021	12.50	5.20	8.52	11.43	15.54	0.85
2006	2010	1,329	13.37	5.24	9.67	12.41	16.04	0.97
2010	2014	1,204	16.81	7.39	11.31	15.50	20.99	0.93

Panel B: Characteristics of portfolios formed by OTLR

	OTLR	OTLRC	OTLRP	OTDR	OTMR	Prem	OBA	Size	NYSE	BM	Mom
Low	6.78	7.52	6.34	0.48	0.65	1.87%	0.24	1,566	4.01	0.48	13.38%
2	9.47	10.11	8.81	0.45	0.66	1.90%	0.21	2,706	5.28	0.46	13.22%
3	11.81	12.48	11.17	0.44	0.64	2.21%	0.21	5,126	6.48	0.46	12.61%
4	14.95	15.67	14.21	0.43	0.63	2.42%	0.22	9,853	7.70	0.45	10.45%
High	21.40	22.27	20.00	0.41	0.63	2.32%	0.27	20,414	8.87	0.44	8.95%

Table 16

Summary statistics and correlations

Panel A includes time series average of cross section summary statistics of monthly variables other than OTLR. OTLRC (OTLRP) is option trading leverage ratio calculated using call (put) options only. OTDR is monthly average of option absolute delta, value-weighted by trading volumes. OTMR is monthly average of option strike to underlying price for call, and one minus this ratio for put, value-weighted by trading volumes. Relative option to stock volume ratio (OS) is sum of option volumes divided by its underlying stock volumes. Short interest (Short) is short interest divided by shares outstanding. At-the-money option premium (Prem) is the difference between average of implied volatilities of at-the-money (ATM) options and historical volatility of the underlying stock. ATM options include call and put options with moneyness between 0.9 and 1.1. Historical volatility is annualized standard deviation of daily stock returns in previous month. ATM option bid-ask spread (OBA) is the average of bid ask spreads of ATM options. Size is market capitalization in million dollars. Book-to-market (BM) is the ratio of book equity and market capitalization. Stock return (Return) is monthly stock return. Momentum (Mom) is the cumulative stock return of prior six months. Idiosyncratic volatility (Ivol) is annualized standard deviation of residuals from regressing daily stock returns on Fama and French (1993) three-factor model in previous month. Amihud illiquidity (Amihud) is the average of absolute daily return divided by daily trading volume, adjusted by the market average, using data of past 12 months, as in Amihud (2002). Mispricing score (Misp) is the average of ranking percentiles on 11 anomaly variables as in Stambaugh et al. (2015). Analyst dispersion (Disp) is the standard deviation of analyst annual earnings forecasts divided by the mean of the forecasts as in Diether et al. (2002). Panel B reports correlations of standardized variables with mean 0 and standard deviation 1 in each monthly cross section. The sample spans from 1996 to 2014, except for Misp which ends in 2013. Option variables are calculated using options with 15 to 45 days to expiration. Stocks with price less than \$5 are excluded. Each cross section is winzorized by 1% at each tail.

Panel A: Summary statistics

Variables	Firms	Mean	STD	P25	Median	P75	Skew
OTLRC	1,116	13.61	5.69	9.43	12.42	16.71	0.98
OTLRP	1,116	12.11	5.53	8.00	11.07	15.28	0.85
OTDR	1,116	0.44	0.09	0.39	0.43	0.49	0.56
OTMR	1,116	0.64	0.20	0.53	0.66	0.78	-0.69
OS	1,116	0.04	0.05	0.01	0.03	0.05	2.46
Short	938	0.06	0.05	0.02	0.04	0.07	1.74
Prem	1,114	2.14%	15.03%	-3.72%	3.54%	10.23%	-1.13
OBA	1,114	0.23	0.15	0.13	0.20	0.30	1.37
Size	1,116	7,932	13,803	890.49	2423.63	7702.00	2.80
BM	1,089	0.46	0.34	0.22	0.37	0.60	1.68
Return	1,116	1.47%	12.80%	-6.20%	0.78%	8.13%	0.49
Mom	1,104	11.70%	36.44%	-10.96%	6.37%	26.79%	1.12
Ivol	1,116	35.83%	20.27%	21.37%	30.97%	44.76%	1.58
Amihud	1,095	0.30%	1.48%	0.02%	0.06%	0.21%	14.29
Misp	1,047	49.49	13.44	39.66	48.62	58.70	0.25
Disp	970	15.58%	38.43%	1.81%	4.42%	12.54%	5.28

Table 16-continued

Panel B: Correlations

	OTLR	OTLRC	OTLRP	OTDR	OTMR	OS	Short	Prem	OBA	Return	Size	BM	Mom	Ivol	Amihud	Misp	Disp
OTLR	1.00																
OTLRC	0.95	1.00															
OTLRP	0.90	0.78	1.00														
OTDR	-0.26	-0.26	-0.13	1.00													
OTMR	-0.05	-0.14	-0.06	-0.05	1.00												
OS	-0.10	-0.09	-0.07	-0.12	0.03	1.00											
Short	-0.40	-0.37	-0.40	0.02	-0.01	0.18	1.00										
Prem	-0.01	0.00	-0.02	-0.05	0.01	0.07	0.06	1.00									
OBA	0.15	0.14	0.15	0.20	-0.07	-0.38	-0.09	-0.01	1.00								
Return	0.01	-0.05	0.09	0.12	0.10	0.02	-0.01	-0.09	0.02	1.00							
Size	0.51	0.48	0.50	-0.08	-0.03	0.20	-0.36	-0.02	-0.17	0.01	1.00						
BM	-0.02	-0.01	-0.01	0.09	-0.02	-0.13	0.02	-0.01	0.20	0.03	-0.07	1.00					
Mom	0.00	-0.03	0.05	0.01	0.05	0.06	-0.04	0.00	-0.08	0.00	0.00	-0.15	1.00				
Ivol	-0.59	-0.57	-0.58	0.05	0.12	0.07	0.31	-0.57	-0.10	0.05	-0.33	-0.04	0.00	1.00			
Amihud	-0.23	-0.23	-0.23	0.06	0.05	0.00	0.06	0.02	0.11	0.05	-0.15	-0.02	0.16	0.26	1.00		
Misp	-0.31	-0.29	-0.32	0.02	0.09	0.00	0.24	0.03	0.02	-0.01	-0.22	0.24	-0.22	0.25	0.09	1.00	
Disp	-0.21	-0.20	-0.21	0.03	0.05	0.02	0.14	0.00	0.01	-0.01	-0.10	0.12	-0.03	0.17	0.07	0.19	1.00

Table 17

Fama-MacBeth regressions of OTLR

Fama and MacBeth (1973) cross-section regressions are implemented every month. OS is relative volume ratio of option to stock. Short is short interest divided by shares outstanding. IOR is institutional owned shares divided by shares outstanding in most recent quarter. ATM options include options with moneyness within 0.9 and 1.1. OBA is the average of bid ask spreads of ATM options. Prem is the difference between ATM_IV and historical volatility. ATM_IV is the average of implied volatilities of ATM options. Ivol is annualized standard deviation of residuals of regressing daily stock returns on Fama and French (1993) three-factor model. Amihud is market adjusted illiquidity as in Amihud (2002). Misp is mispricing score as in Stambaugh et al. (2015). Disp is analyst dispersion of annual earnings forecasts. Size is market capitalization. BM is the ratio of book equity and market capitalization. Return is monthly stock return. Mom(X,Y) is the cumulative stock return from lagged month Y to lagged month X. The sample is from 1996 to 2014, except for Misp which ends in 2013. Sample includes options with 15 to 45 days to expiration and stocks with price above $5. Each cross section is winzorized by 1% at each tail and is standardized to mean 0 and standard deviation 1. Newey and West (1987) t-statistics are reported in parentheses. The notations ***, ** and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.0707***	0.0122***	0.0124***	0.0067***	0.0039	-0.0612***
	(3.93)	(6.44)	(7.31)	(5.32)	(1.46)	(-3.60)
OS	-0.1588***					-0.0657***
	(-15.16)					(-14.57)
Short	-0.0749***					-0.0099
	(-7.42)					(-0.85)
IOR		-0.0152***				-0.0950***
		(-2.90)				(-12.85)
OBA			0.2688***			0.1234***
			(21.47)			(7.27)
Prem			0.0080			-0.3891***
			(1.14)			(-11.71)
Ivol				-0.3105***		-0.6727***
				(-33.73)		(-11.82)
Log(Amihud)				0.0065		-0.3618***
				(0.30)		(-8.81)
Misp					-0.0915***	-0.0181**
					(-7.17)	(-2.50)
Log(Disp)					-0.1882***	-0.1011***
					(-15.42)	(-13.90)
Log(Size)	0.6095***	0.6258***	0.6837***	0.4866***	0.5557***	0.0546
	(38.06)	(40.39)	(65.04)	(14.86)	(32.12)	(1.34)
Log(BM)	-0.0048	0.0560***	0.0123	0.0272**	0.0883***	-0.0249***
	(-0.35)	(3.32)	(0.89)	(2.28)	(5.77)	(-3.41)
Return	-0.0037	-0.0135	-0.0230**	0.0050	-0.0196*	0.0010
	(-0.28)	(-1.20)	(-2.48)	(0.58)	(-1.85)	(0.11)
Mom(1,6)	0.0010	-0.0067	0.0011	-0.0116	-0.0385**	0.0314*
	(0.04)	(-0.30)	(0.06)	(-0.64)	(-2.02)	(1.69)
Mom(7,12)	-0.0320*	-0.0287*	-0.0124	-0.0202	-0.0490***	-0.0040
	(-1.70)	(-1.69)	(-0.89)	(-1.49)	(-3.55)	(-0.38)
Adj. R^2	47.91%	48.76%	55.37%	55.13%	52.85%	63.21%
N Obs	204,473	241,335	240,954	241,229	192,611	165,242

Table 18

Factor regression results by portfolio returns formed by OTLR

This table reports average excess return of quintile portfolios and time-series regression results of Fama and French (1993) three-factor (MKTRF, SMB and HML) model and four factor (Fama-French three factors and UMD) model. Quintile portfolios are formed at the beginning of each month by last month option trading leverage (OTLR). Value-weighted portfolio return in percentage is used in the regression. Panel A uses all qualified observations, while Panel B excludes stocks with top and bottom 10% ATM_IV in each monthly cross section. ATM_IV is the average of implied volatilities of ATM options, for which ATM options include options with moneyness within 0.9 and 1.1. The sample is from 1996 to 2014. Sample includes options with 15 to 45 days to expiration and stocks with price above $5. Newey and West (1987) t-statistics are reported in parentheses. For the row that reports results of portfolios that long highest OTLR and short lowest OTLR, the notations ***, ** and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Panel A: Factor regressions for all observations

	Ex Ret	Three-factor				Four-factor				
	Average	Alpha	MKTRF	SMB	HML	Alpha	MKTRF	SMB	HML	UMD
Low 1	-0.27	-1.36	1.80	0.83	-0.57	-1.14	1.67	0.89	-0.67	-0.33
	(-0.35)	(-6.28)	(14.65)	(8.06)	(-3.02)	(-4.66)	(18.29)	(7.26)	(-4.89)	(-4.53)
2	0.52	-0.38	1.53	0.64	-0.54	-0.29	1.48	0.66	-0.58	-0.13
	(0.82)	(-1.80)	(18.11)	(6.87)	(-3.20)	(-1.33)	(21.23)	(6.72)	(-3.69)	(-2.26)
3	0.67	-0.10	1.36	0.30	-0.38	-0.08	1.36	0.31	-0.39	-0.02
	(1.27)	(-0.61)	(33.57)	(7.17)	(-4.87)	(-0.59)	(27.07)	(7.26)	(-5.78)	(-0.39)
4	0.55	-0.07	1.14	-0.03	-0.13	-0.08	1.15	-0.03	-0.12	0.02
	(1.35)	(-0.79)	(56.44)	(-1.17)	(-4.42)	(-1.06)	(59.28)	(-1.35)	(-4.70)	(1.56)
High 5	0.62	0.21	0.80	-0.29	0.07	0.14	0.84	-0.31	0.11	0.10
	(2.08)	(3.38)	(20.56)	(-12.34)	(1.19)	(1.93)	(29.08)	(-11.30)	(2.09)	(4.43)
High-Low	0.90	1.57***	-1.00***	-1.12***	0.64***	1.28***	-0.83***	-1.20***	0.78***	0.43***
	(1.59)	(6.24)	(-6.36)	(-9.24)	(2.61)	(4.34)	(-7.22)	(-8.21)	(4.29)	(4.86)

Panel B: Factor regressions for sample with 10% extreme ATM_IV deleted at both sides

	Ex Ret	Three-factor				Four-factor				
	Average	Alpha	MKTRF	SMB	HML	Alpha	MKTRF	SMB	HML	UMD
Low 1	0.01	-0.99	1.67	0.75	-0.54	-0.84	1.58	0.79	-0.61	-0.21
	(0.02)	(-3.79)	(13.79)	(5.91)	(-2.62)	(-3.37)	(15.79)	(5.88)	(-3.44)	(-2.72)
2	0.89	0.02	1.49	0.60	-0.53	0.10	1.45	0.62	-0.57	-0.11
	(1.42)	(0.09)	(21.31)	(6.89)	(-3.99)	(0.47)	(24.71)	(6.65)	(-4.60)	(-2.25)
3	0.63	-0.15	1.39	0.31	-0.37	-0.13	1.38	0.32	-0.37	-0.03
	(1.16)	(-0.93)	(36.61)	(6.85)	(-4.59)	(-0.91)	(29.74)	(6.51)	(-5.35)	(-0.52)
4	0.58	-0.07	1.21	0.02	-0.24	-0.08	1.22	0.02	-0.23	0.02
	(1.27)	(-0.54)	(58.61)	(0.70)	(-6.12)	(-0.67)	(54.07)	(0.61)	(-5.61)	(0.80)
High 5	0.74	0.19	1.01	-0.21	0.02	0.15	1.04	-0.22	0.04	0.07
	(2.07)	(2.75)	(33.66)	(-6.37)	(0.25)	(1.77)	(43.33)	(-6.58)	(0.62)	(2.99)
High-Low	0.72	1.18***	-0.66***	-0.96***	0.55**	0.99***	-0.54***	-1.01***	0.65***	0.28***
	(1.51)	(4.35)	(-4.47)	(-6.53)	(2.10)	(3.59)	(-4.65)	(-6.46)	(2.80)	(3.14)

Table 19

Factor regression results by portfolio returns formed by alternative leverage measures

This table reports excess return of quintile portfolios and time-series regression results of Fama and French (1993) three-factor (MKTRF, SMB and HML) model and four factor (Fama-French three factors and UMD) model. In Panel A, quintile portfolios are formed at the beginning of each month by last month option trading delta ratio (OTDR). In Panel B, quintile portfolios are formed at the beginning of each month by last month option trading moneyness ratio (OTMR). Both panels use value-weighted excess returns in percentage. The sample is from 1996 to 2014. Sample includes options with 15 to 45 days to expiration and stocks with price above $5. Newey and West (1987) t-statistics are reported in parentheses. For the row that reports results of portfolios that long highest OTDR or OTMR and short lowest OTDR or OTMR, the notations ***, ** and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Panel A: Factor regressions for quintile portfolios formed by OTDR

	Ex Ret	Three-factor				Four-factor				
	Average	Alpha	MKTRF	SMB	HML	Alpha	MKTRF	SMB	HML	UMD
Low 1	0.65	0.09	1.03	-0.11	-0.09	0.06	1.05	-0.12	-0.07	0.04
	(1.58)	(1.17)	(60.46)	(-4.19)	(-3.62)	(0.78)	(69.31)	(-4.28)	(-3.11)	(2.75)
2	0.59	0.04	1.07	-0.13	-0.20	0.03	1.08	-0.13	-0.19	0.01
	(1.36)	(0.43)	(51.58)	(-3.72)	(-4.66)	(0.33)	(46.69)	(-3.90)	(-4.63)	(0.76)
3	0.50	-0.09	1.06	-0.07	-0.05	-0.10	1.06	-0.07	-0.04	0.01
	(1.35)	(-1.02)	(63.01)	(-1.61)	(-1.40)	(-1.17)	(57.82)	(-1.74)	(-1.39)	(0.87)
4	0.60	0.02	0.98	0.00	0.02	0.04	0.97	0.01	0.02	-0.02
	(1.76)	(0.27)	(43.96)	(0.06)	(0.76)	(0.35)	(30.56)	(0.17)	(0.65)	(-0.52)
High 5	0.65	0.05	0.93	0.10	0.22	0.06	0.92	0.10	0.21	-0.02
	(2.16)	(0.48)	(28.57)	(3.26)	(3.17)	(0.55)	(24.81)	(3.37)	(3.21)	(-0.73)
High-Low	0.01	-0.05	-0.11***	0.20***	0.30***	-0.00	-0.13***	0.22***	0.28***	-0.06*
	(0.03)	(-0.29)	(-2.97)	(4.28)	(4.11)	(-0.02)	(-3.35)	(4.35)	(3.77)	(-1.97)

Panel B: Factor regressions for quintile portfolios formed by OTMR

	Ex Ret	Three-factor				Four-factor				
	Average	Alpha	MKTRF	SMB	HML	Alpha	MKTRF	SMB	HML	UMD
Low 1	0.78	0.14	1.02	0.01	0.24	0.24	0.96	0.04	0.19	-0.15
	(2.27)	(1.02)	(46.54)	(0.28)	(3.76)	(1.62)	(28.86)	(1.02)	(4.60)	(-2.85)
2	0.58	0.04	1.00	-0.11	-0.03	0.02	1.01	-0.12	-0.02	0.03
	(1.52)	(0.41)	(52.01)	(-3.53)	(-0.70)	(0.21)	(43.13)	(-3.87)	(-0.52)	(1.62)
3	0.64	0.10	1.05	-0.11	-0.19	0.06	1.07	-0.12	-0.17	0.06
	(1.65)	(1.42)	(63.61)	(-3.56)	(-6.05)	(0.81)	(57.95)	(-4.88)	(-5.94)	(3.07)
4	0.35	-0.21	1.04	-0.01	-0.15	-0.25	1.06	-0.02	-0.13	0.06
	(0.87)	(-2.36)	(47.22)	(-0.16)	(-3.48)	(-2.68)	(40.70)	(-0.41)	(-3.38)	(2.06)
High 5	0.38	-0.22	1.07	0.01	-0.08	-0.18	1.05	0.02	-0.10	-0.06
	(0.96)	(-1.90)	(47.84)	(0.11)	(-1.37)	(-1.74)	(38.51)	(0.39)	(-2.06)	(-2.09)
High-Low	-0.40	-0.36*	0.06*	-0.00	-0.32***	-0.42*	0.10**	-0.02	-0.29***	0.09
	(-1.64)	(-1.78)	(1.96)	(-0.09)	(-4.12)	(-1.93)	(2.03)	(-0.37)	(-4.26)	(1.44)

Table 20

Double sorted portfolios: stock characteristics and OTLR

This table presents risk-adjusted returns of time-series regression using four-factor (MKTRF, SMB, HML and UMD) model. 25 portfolios are formed at the beginning of every month by sorting stocks on previous month end stock characteristics (size in Panel A, BM in Panel B, Momentum in Panel C, residual IOR in Panel D) and option trading leverage ratio (OTLR). Stocks are firstly sorted into five groups based on stock characteristics. Within each characteristic group, stocks are sorted into five groups based on OTLR. Size is market capitalization. BM is the ratio of book equity and market capitalization. Momentum is the cumulative stock return of prior six month. Res IOR is the residual institutional ownership (IOR) obtained by regressing IOR on size in each cross section, as in Nagel (2005). Portfolio excess returns are value weighted returns in percentage. The sample is from 1996 to 2014. Sample includes options with 15 to 45 days to expiration and stocks with price above $5. Newey and West (1987) t-statistics are reported in parentheses. For the last column that reports results of portfolios that long highest OTLR and short lowest OTLR within each characteristic group, the notations ***, ** and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Panel A: Portfolios formed by size and OTLR

Size	OTLR					
	Low	2	3	4	High	H-L
Low	-1.59	-0.73	-0.93	0.18	0.13	1.72***
	(-4.69)	(-2.51)	(-3.62)	(0.61)	(0.89)	(4.20)
2	-0.79	-0.36	-0.49	-0.29	0.31	1.10***
	(-3.36)	(-1.33)	(-2.47)	(-1.40)	(1.90)	(3.13)
3	-0.53	-0.04	0.09	-0.10	0.12	0.65
	(-1.74)	(-0.19)	(0.51)	(-0.62)	(0.61)	(1.57)
4	-0.50	0.10	0.20	0.22	0.40	0.90**
	(-1.83)	(0.57)	(1.14)	(1.57)	(2.85)	(2.37)
High	-0.31	-0.24	0.10	0.12	0.19	0.50
	(-1.27)	(-1.97)	(0.94)	(1.07)	(2.00)	(1.61)

Panel B: Portfolios formed by book-to-market ratio and OTLR

BM	OTLR					
	Low	2	3	4	High	H-L
Low	-0.71	-0.36	-0.24	0.35	0.09	0.80
	(-1.63)	(-1.10)	(-0.89)	(2.08)	(0.66)	(1.59)
2	-0.97	-0.07	-0.09	-0.14	0.13	1.09**
	(-2.02)	(-0.28)	(-0.50)	(-0.98)	(1.21)	(2.09)
3	-0.59	-0.42	0.02	-0.11	0.23	0.81
	(-1.27)	(-1.41)	(0.08)	(-0.95)	(2.03)	(1.65)
4	-0.72	-0.40	0.07	-0.17	0.04	0.75**
	(-2.33)	(-1.60)	(0.27)	(-1.28)	(0.31)	(2.06)
High	-2.09	-0.50	-0.03	-0.34	0.16	2.25***
	(-4.55)	(-1.57)	(-0.12)	(-1.76)	(1.31)	(4.72)

Table 20-continued
Double sorted portfolios: stock characteristics and OTLR

Panel C: Portfolios formed by momentum and OTLR

Momentum	OTLR					
	Low	2	3	4	High	H-L
Low	-1.86	-1.15	-0.28	-0.33	0.38	2.24***
	(-3.84)	(-2.67)	(-1.11)	(-1.03)	(1.97)	(4.00)
2	-0.89	-0.24	-0.19	-0.07	0.35	1.24***
	(-3.43)	(-1.06)	(-0.86)	(-0.37)	(2.38)	(3.93)
3	-0.27	-0.32	0.00	0.20	0.39	0.66*
	(-1.17)	(-1.36)	(0.01)	(1.25)	(2.76)	(1.94)
4	-1.11	-0.26	-0.27	-0.26	-0.12	0.99***
	(-4.20)	(-1.25)	(-1.19)	(-1.74)	(-0.87)	(3.08)
High	-0.84	-0.30	0.03	0.14	0.10	0.94*
	(-1.85)	(-0.73)	(0.08)	(0.69)	(0.43)	(1.81)

Panel D: Portfolios formed by residual IOR and OTLR

Res IOR	OTLR					
	Low	2	3	4	High	H-L
Low	-2.24	-0.95	-0.28	-0.07	0.01	2.25***
	(-3.97)	(-2.33)	(-0.77)	(-0.31)	(0.07)	(4.03)
2	-0.60	-0.15	0.26	0.19	0.25	0.86**
	(-1.79)	(-0.73)	(1.48)	(1.16)	(1.47)	(2.06)
3	-0.42	0.08	-0.02	0.17	0.21	0.63*
	(-1.58)	(0.42)	(-0.08)	(1.28)	(1.53)	(1.80)
4	-0.42	-0.44	0.15	-0.32	0.14	0.56**
	(-2.15)	(-1.77)	(0.41)	(-1.34)	(1.47)	(2.40)
High	-0.48	-0.61	0.10	-0.27	0.02	0.50
	(-1.49)	(-1.48)	(0.32)	(-0.99)	(0.21)	(1.48)

Table 21

Double sorted portfolios: informed short and OTLR

This table presents risk-adjusted returns of time-series regression using four-factor (MKTRF, SMB, HML and UMD) model. 25 portfolios are formed at the beginning of every month by sorting stocks on previous month informed short variable (OS in Panel A, Short in Panel B) and option trading leverage ratio (OTLR). Stocks are firstly sorted into five groups based on informed short variable. Within each formed group, stocks are sorted into five groups based on OTLR. OS is relative volume ratio of option to stock. Short is short interest divided by shares outstanding. Portfolio excess returns are value weighted returns in percentage. The sample is from 1996 to 2014. Sample includes options with 15 to 45 days to expiration and stocks with price above $5. Newey and West (1987) t-statistics are reported in parentheses. For the last column that reports results of portfolios that long highest OTLR and short lowest OTLR within each informed short group, the notations ***, ** and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Panel A: Portfolios formed by OS then OTLR

OS	OTLR					
	Low	2	3	4	High	H-L
Low	-0.30	0.13	0.51	0.21	0.31	0.61*
	(-1.06)	(0.75)	(2.54)	(1.57)	(2.60)	(1.87)
2	-0.42	-0.67	-0.08	0.04	0.29	0.71*
	(-1.19)	(-2.65)	(-0.36)	(0.23)	(2.46)	(1.68)
3	-0.61	-0.16	0.19	-0.07	0.14	0.75**
	(-2.07)	(-0.58)	(0.85)	(-0.38)	(1.08)	(2.22)
4	-1.55	-0.25	-0.27	-0.14	0.20	1.75***
	(-3.85)	(-0.85)	(-1.08)	(-0.77)	(2.04)	(4.16)
High	-1.31	-0.66	-0.07	-0.23	-0.05	1.26***
	(-3.74)	(-1.59)	(-0.25)	(-1.04)	(-0.51)	(3.32)

Panel B: Portfolios formed by Short then OTLR

Short	OTLR					
	Low	2	3	4	High	H-L
Low	-0.07	-0.02	-0.10	0.26	0.27	0.35
	(-0.27)	(-0.08)	(-0.78)	(2.12)	(2.04)	(0.98)
2	-0.34	-0.15	0.11	0.05	0.09	0.43
	(-0.98)	(-0.71)	(0.59)	(0.38)	(0.61)	(1.02)
3	-0.50	-0.49	-0.01	0.18	-0.13	0.37
	(-1.53)	(-2.35)	(-0.07)	(1.03)	(-0.87)	(0.94)
4	-1.40	-0.53	-0.12	-0.15	-0.08	1.32***
	(-3.60)	(-2.12)	(-0.40)	(-0.63)	(-0.46)	(3.15)
High	-1.26	-0.60	-0.13	-0.38	-0.11	1.15**
	(-3.33)	(-2.29)	(-0.39)	(-1.24)	(-0.48)	(2.59)

Table 22

Information ratios of long-short trading strategies

This table shows annualized information ratio of long-short trading strategies, which is calculated as monthly risk-adjusted return divided by standard deviation of residuals, using four-factor (MKTRF, SMB, HML and UMD) model as the benchmark, and then annualized by multiplying $\sqrt{12}$. Portfolios are formed at the beginning of every month by trading signals as of previous month, and are held for one month. The second column indicates type of portfolio returns used: equal-weighted or value-weighted. The last column presents details of long-short strategy used. For the first three strategy (OTLR, OS, and Short), quintile portfolios are formed by the corresponding variable. For the last three strategy, 5×5 portfolios are formed by double-sorting on two variables. OTLR is calculated for each underlying stock, which is the weighted average of option leverage, using trading volumes as weights. OS is relative volume ratio of option to stock. Short is short interest divided by shares outstanding. The sample is from 1996 to 2014. Sample includes options with 15 to 45 days to expiration and stocks with price above $5.

Strategy	Returns	Info Ratio	Details
OTLR	Equal	1.20	Long high OTLR, short low OTLR
	Value	1.03	
OS	Equal	0.47	Long low OS, short high OS
	Value	0.46	
Short	Equal	1.30	Long low Short, short high Short
	Value	0.31	
OS/OTLR	Equal	1.71	Long low OS and high OTLR, short high OS and low OTLR
	Value	0.86	
Short/OTLR	Equal	1.12	Long low Short and high OTLR, short high Short and low OTLR
	Value	0.81	
Short/OS	Equal	1.39	Long low Short and low OS, short high Short and high OS
	Value	0.30	

Table 23

Fama-MacBeth regressions of one month ahead stock returns

Fama and MacBeth (1973) cross-section regressions are implemented every month. One-month ahead stock return in percentage is used. OTLR is option trading leverage ratio. OS is relative volume ratio of option to stock. Short is short interest divided by shares outstanding. Var0 (Var can be OTLR, OS and Short) equals to Var if Var is below the cross section median or equals to 0 otherwise. Ivol is idiosyncratic volatility using three-factor model. Amihud is market adjusted illiquidity as in Amihud (2002). Misp is mispricing score as in Stambaugh et al. (2015). Disp is analyst dispersion of earnings forecasts. Size is market capitalization. BM is the ratio of book equity and market capitalization. Return is monthly stock return. Mom(X,Y) is the cumulative stock return from lagged month Y to lagged month X. The sample is from 1996 to 2014, except for Misp which ends in 2013. Sample includes stocks with price above $5. Each cross section of independent variables are winzorized by 1% at both tails, and are standardized to mean 0 with standard deviation 1. Newey and West (1987) t-statistics are reported in parentheses. The notations ***, ** and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Intercept	0.8647**	0.9424**	0.8960**	0.9429**	0.7854*	0.8788**	0.7260*
	(2.35)	(2.41)	(2.41)	(2.38)	(1.95)	(2.17)	(1.80)
OTLR	0.0669	-0.1383	0.1220	-0.0386	-0.0798	-0.0422	-0.0832
	(0.57)	(-1.42)	(1.12)	(-0.39)	(-0.90)	(-0.44)	(-0.91)
OTLR0		0.6422***		0.5129***	0.3531*	0.6063***	0.5079***
		(3.06)		(2.69)	(1.96)	(3.69)	(3.11)
OS	-0.0719	0.0350	-0.0497	0.0194	0.0006	0.0281	0.0026
	(-1.10)	(0.56)	(-0.94)	(0.33)	(0.01)	(0.45)	(0.04)
OS0		-0.4746***		-0.3173**	-0.3381**	-0.2987**	-0.3208**
		(-2.65)		(-2.24)	(-2.33)	(-2.36)	(-2.49)
Short	-0.2195***	-0.2104**	-0.2247***	-0.1921**	-0.2428***	-0.1119	-0.1606*
	(-2.91)	(-2.22)	(-3.35)	(-2.23)	(-2.95)	(-1.28)	(-1.90)
Short0		-0.0288		-0.1534	-0.2884*	-0.1765	-0.3225**
		(-0.17)		(-0.94)	(-1.67)	(-1.13)	(-1.99)
OTLR*OS	0.1321***	0.1160***	0.1342***	0.1270***	0.1215***	0.1366***	0.1298***
	(3.65)	(3.11)	(3.92)	(3.64)	(3.53)	(3.73)	(3.48)
Ivol					-0.1705**		-0.1295*
					(-2.16)		(-1.67)
Log(Amihud)					-0.7243***		-0.7463***
					(-3.60)		(-3.49)
Misp						-0.1784**	-0.1506**
						(-2.36)	(-2.01)
Log(Disp)						0.0002	0.0044
						(0.00)	(0.08)
Log(Size)			-0.1467	-0.1691	-0.8612***	-0.2078**	-0.8866***
			(-1.28)	(-1.55)	(-3.57)	(-2.08)	(-3.74)
Log(BM)			0.0245	0.0108	-0.0180	0.0587	0.0283
			(0.36)	(0.16)	(-0.27)	(0.78)	(0.38)
Return			-0.1739	-0.1698	-0.0863	-0.3035***	-0.2177**
			(-1.44)	(-1.39)	(-0.67)	(-2.87)	(-2.10)
Mom(1,6)			0.0336	0.0402	0.1944	0.0274	0.1874
			(0.24)	(0.29)	(1.42)	(0.16)	(1.06)
Mom(7,12)			0.0091	0.0202	0.0547	0.0298	0.0571
			(0.07)	(0.15)	(0.40)	(0.19)	(0.36)
Adj-R^2	3.70%	4.36%	7.67%	8.06%	8.62%	9.06%	9.63%
N Obs	213,774	213,774	204,339	204,339	204,248	165,480	165,460

Table 24

Fama-MacBeth regressions of future monthly returns

N-month ahead stock monthly return in percentage is used as dependent variable, where n is reported by the first row in parentheses. OTLR is option trading leverage ratio. OS is relative volume ratio of option to stock. Short is short interest divided by shares outstanding. Var0 (Var can be OTLR, OS and Short) equals to Var if Var is below the cross section median or equals to 0 otherwise. Ivol is idiosyncratic volatility. Amihud is market adjusted illiquidity as in Amihud (2002). Misp is mispricing score as in Stambaugh et al. (2015). Disp is analyst dispersion of annual forecasts. Size is market capitalization. BM is the ratio of book equity and market capitalization. Return is monthly stock return. Mom(X,Y) is the cumulative stock return from lagged month Y to lagged month X. The sample is from 1996 to 2013. Sample includes stocks with price above $5. Independent variables of each cross section are winzorized by 1% at both tails. Independent variables are standardized to mean 0 and standard deviation 1 in each cross section. Newey and West (1987) t-statistics are reported in parentheses. The notations ***, ** and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.7260	0.7139	0.8509*	0.9355**	0.7866*	0.9842**
	(1.63)	(1.56)	(1.86)	(2.06)	(1.67)	(2.21)
OTLR	-0.0832	-0.1196	-0.1176	-0.0479	-0.1111	-0.0983
	(-0.84)	(-1.30)	(-1.21)	(-0.53)	(-1.35)	(-1.23)
OTLR0	0.5079***	0.2203	0.0250	0.0527	0.1008	0.0643
	(3.15)	(1.17)	(0.12)	(0.30)	(0.46)	(0.36)
OS	0.0026	0.0277	-0.0109	0.0561	0.1270**	0.0354
	(0.04)	(0.60)	(-0.18)	(0.82)	(2.42)	(0.63)
OS0	-0.3208**	-0.2306*	-0.1065	-0.0242	-0.2078	-0.1771
	(-2.52)	(-1.92)	(-0.86)	(-0.19)	(-1.43)	(-1.52)
Short	-0.1606**	-0.2101***	-0.2519***	-0.2808***	-0.2749***	-0.3046***
	(-2.10)	(-2.72)	(-3.46)	(-3.70)	(-3.69)	(-4.33)
Short0	-0.3225**	-0.1854	0.0545	0.2115	0.1115	0.2830**
	(-2.17)	(-1.26)	(0.38)	(1.32)	(0.92)	(1.99)
OTLR*OS	0.1298***	0.0453	0.0521	0.0009	-0.0283	0.0305
	(2.93)	(1.58)	(1.41)	(0.02)	(-0.67)	(0.93)
Ivol	-0.1295	-0.1353	-0.0645	-0.0187	-0.1668*	0.0532
	(-1.61)	(-1.52)	(-0.71)	(-0.21)	(-1.68)	(0.58)
Log(Amihud)	-0.7463***	-0.5170**	-0.4625*	-0.2247	-0.4172*	-0.3111
	(-3.46)	(-2.08)	(-1.77)	(-1.06)	(-1.92)	(-1.54)
Misp	-0.1506**	-0.1651**	-0.2225***	-0.2191***	-0.1606***	-0.1870***
	(-2.31)	(-2.59)	(-3.14)	(-3.12)	(-2.73)	(-3.41)
Log(Disp)	0.0044	-0.0130	-0.0334	-0.0816	-0.0267	-0.0131
	(0.08)	(-0.23)	(-0.59)	(-1.35)	(-0.48)	(-0.24)
Log(Size)	-0.8866***	-0.5649**	-0.4357*	-0.3129	-0.4517**	-0.3362
	(-3.69)	(-2.38)	(-1.81)	(-1.52)	(-2.01)	(-1.54)
Log(BM)	0.0283	0.0810	0.1330*	0.1146	0.0660	0.0693
	(0.37)	(1.08)	(1.81)	(1.28)	(0.85)	(0.83)
Return	-0.2177**	-0.0503	0.0890	-0.0676	0.1321	0.2599**
	(-2.05)	(-0.46)	(0.97)	(-0.75)	(1.32)	(2.59)
Mom(1,6)	0.1874	0.1368	0.2237	0.2855*	0.2932**	0.1868
	(1.13)	(0.88)	(1.31)	(1.84)	(2.02)	(1.38)
Mom(7,12)	0.0571	-0.0264	0.0010	-0.0312	-0.2146	-0.2619**
	(0.38)	(-0.17)	(0.01)	(-0.27)	(-1.42)	(-2.24)
Adj-R^2	9.63%	9.13%	8.90%	8.51%	8.44%	7.67%
N Obs	165,460	165,416	165,330	165,206	165,030	164,795

Table 25

Summary statistics of forecast revision, earnings surprise and PEAD

This table reports summary statistics of analyst forecast revision, earnings surprise and PEAD. Est_rev is monthly changes of average analyst forecasts of quarterly earnings divided by previous month end price, as in Jegadeesh et al. (2004). Surprise is the difference of quarterly announced earnings and analyst forecast average just before the announcement, scaled by last fiscal quarter end price, as in Livnat and Mendenhall (2006). SUE is standardized unexplained quarterly earnings, calculated as the difference between current quarterly earnings and earnings of four-quarter ago, divided by its standard deviation using last eight observations , as in Jegadeesh et al. (2004). Street earnings are used which is Compustat reported actual EPS minus special items and then times 65%, as in Bradshaw and Sloan (2002). CAR(X, Y) is the cumulative market-adjusted return from day X to day Y relative to the earnings announcement day, where CRSP value-weighted market return is used to adjust cumulative returns. Except for SUE, all other variables are reported in terms of percentage. Each column reports time series average of summary statistics of each monthly (quarterly) cross section for Est_rev (variables except Est_rev). The sample is from 1996 to 2014. Sample includes stocks with price above $5.

Variable	Firms	Mean	STD	P25	Median	P75	Skew
Est_rev	886	-0.09	0.76	-0.09	-0.01	0.02	-3.68
Surprise	994	0.43	15.39	-0.02	0.05	0.17	-4.38
SUE	944	0.53	1.63	-0.46	0.34	1.36	1.09
CAR(-1,1)	994	-0.28	9.44	-4.10	0.24	4.55	-1.49
CAR(2,5)	994	-0.25	5.61	-2.90	-0.11	2.62	-0.71
CAR(2,10)	994	-0.32	8.07	-4.07	0.02	3.92	-1.26
CAR(2,20)	994	-0.41	11.48	-5.53	0.33	5.72	-1.39
CAR(2,40)	993	-1.46	18.43	-8.49	0.60	8.44	-2.35
CAR(2,60)	991	-2.13	23.63	-10.51	0.96	10.63	-2.61

Table 26

Fama-MacBeth regressions of forecast revisions and earnings surprise

Fama and MacBeth (1973) regressions are run every month for Est_rev and every quarter for earnings surprise. Est_rev is monthly changes of average analyst forecasts divided by previous month end price. Surprise is the difference of quarterly announced earnings and analyst forecast. SUE is standardized unexplained quarterly earnings. CAR(X, Y) is the cumulative market-adjusted return from day X to day Y relative to the announcement. OTLR is option trading leverage ratio. OS is relative volume ratio of option to stock. Short is short interest divided by shares outstanding. Var0 equals to Var if Var is below the cross section median or equals to 0 otherwise. Lag(Dep) is lagged dependent variable. Ltg is mean of analyst long-term growth forecasts. The sample is from 1996 to 2013. All variables are standardized to mean 0 and standard deviation 1 in each cross section. Newey and West (1987) t-statistics are reported in parentheses.

	Est_rev		Surprise		SUE		CAR(-1,1)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Intercept	-0.0826***	-0.0737***	-0.0103	0.0041	-0.1860***	-0.1742***	0.0180**	0.0542***
	(-7.15)	(-5.57)	(-0.75)	(0.20)	(-6.84)	(-7.41)	(2.54)	(4.16)
OTLR	0.0071*	0.0000	-0.0077	-0.0182**	0.0342***	0.0297***	-0.0025	-0.0162**
	(1.79)	(0.01)	(-0.84)	(-2.61)	(6.45)	(3.82)	(-0.48)	(-2.57)
OTLR0		0.0289*		0.0416		0.0126		0.0707***
		(1.85)		(1.46)		(0.54)		(2.94)
OS	-0.0071***	-0.0036	0.0052	0.0084**	0.0025	0.0022	-0.0076	-0.0045
	(-3.29)	(-1.55)	(1.23)	(2.08)	(0.44)	(0.33)	(-1.38)	(-0.84)
OS0		-0.0152*		-0.0161		-0.0049		-0.0040
		(-1.76)		(-0.96)		(-0.32)		(-0.24)
Short	-0.0148***	-0.0144***	-0.0227**	-0.0209*	-0.0078	-0.0113***	-0.0355***	-0.0396***
	(-3.43)	(-3.26)	(-2.00)	(-1.78)	(-1.20)	(-2.89)	(-4.62)	(-5.22)
Short0		0.0087		0.0162		0.0321		0.0378***
		(0.94)		(0.76)		(1.11)		(2.98)
OTLR*OS	-0.0046**	-0.0062***	0.0002	-0.0020	0.0132**	0.0113**	0.0063*	0.0041
	(-2.06)	(-3.01)	(0.07)	(-0.84)	(2.53)	(2.13)	(1.71)	(1.16)
Lag(Dep)	0.2049***	0.2052***	0.2025***	0.2003***	0.3552***	0.3547***	-0.0015	-0.0014
	(4.06)	(4.02)	(3.57)	(3.56)	(28.01)	(27.84)	(-1.64)	(-1.62)
Ltg	0.0158***	0.0159***	-0.0032	-0.0054	0.0676***	0.0669***	0.0053	0.0030
	(3.14)	(3.00)	(-0.26)	(-0.46)	(4.97)	(5.16)	(0.73)	(0.46)
Misp	-0.0020	-0.0020	-0.0109**	-0.0109***	0.0162**	0.0160**	-0.0016	0.0008
	(-0.79)	(-0.81)	(-2.36)	(-2.87)	(2.56)	(2.60)	(-0.27)	(0.13)
Log(Disp)	-0.0173***	-0.0164***	-0.0066	-0.0071	-0.0804***	-0.0813***	-0.0118*	-0.0104
	(-5.86)	(-5.97)	(-0.93)	(-1.17)	(-8.45)	(-8.67)	(-1.84)	(-1.55)
Log(Amihud)	0.0544***	0.0622***	-0.0287	-0.0188	0.0288	0.0366	-0.0530**	-0.0393
	(3.01)	(3.60)	(-1.00)	(-0.66)	(1.22)	(1.63)	(-2.11)	(-1.55)
Log(Size)	0.0565***	0.0644***	-0.0345	-0.0207	0.0606**	0.0704***	-0.0480**	-0.0349
	(3.45)	(4.04)	(-1.50)	(-0.98)	(2.57)	(2.79)	(-2.01)	(-1.47)
Log(BM)	-0.0078	-0.0080*	0.0208**	0.0201**	-0.0539***	-0.0541***	-0.0004	-0.0014
	(-1.63)	(-1.67)	(2.11)	(2.03)	(-5.26)	(-5.23)	(-0.06)	(-0.26)
Return	0.0474***	0.0461***	0.0628***	0.0598***	0.0915***	0.0906***	0.0222**	0.0214**
	(5.17)	(5.13)	(8.63)	(9.42)	(7.01)	(6.80)	(2.37)	(2.10)
Mom(1,6)	0.0351***	0.0334***	0.0649***	0.0652***	0.1447***	0.1416***	0.0106	0.0101
	(6.11)	(6.19)	(5.95)	(5.97)	(16.10)	(16.29)	(1.15)	(0.99)
Mom(7,12)	0.0032	0.0022	0.0317***	0.0312***	0.0530***	0.0534***	0.0134	0.0150
	(0.50)	(0.34)	(3.22)	(3.41)	(4.34)	(4.46)	(1.57)	(1.66)
Adj-R^2	13.92%	14.10%	9.97%	10.29%	48.38%	48.37%	1.73%	1.66%
N Obs	116,735	116,735	38,238	38,238	37,694	37,694	38,166	38,166

Table 27

Fama-MacBeth regressions of PEAD

Fama and MacBeth (1973) cross-section regressions are run every calendar quarter for PEAD from 1996 to 2003. CAR(X, Y) is the cumulative market-adjusted return from day X to day Y relative to the earnings announcement. OTLR is option trading leverage ratio. OS is relative volume ratio of option to stock. Short is short interest divided by shares outstanding. Var0 equals to Var if Var is below the cross section median or equals to 0 otherwise. Lag(Dep) is lagged dependent variable. Ltg is mean of long-term growth forecasts. Misp is Stambaugh et al. (2015) mispricing score. Disp is analyst dispersion of annual earnings forecasts. Amihud is Amihud (2002) market adjusted illiquidity. Size is market capitalization. BM is the ratio of book equity and size. Return is monthly stock return. Mom(X,Y) is the cumulative stock return from lagged month Y to month X. Variables are standardized to mean 0 and standard deviation 1 in cross section. Newey and West (1987) t-statistics are reported in parentheses. Sample includes options with 15 to 45 days to expiration and stocks with price above $5. Each cross section of independent variables are winzorized by 1% at both tails. ***, ** and * indicate statistical significance at the 1%, 5%, and 10%.

	CAR(2,10)		CAR(2,20)		CAR(2,40)		CAR(2,60)	
Intercept	0.0303**	0.0365	0.0173	0.0376	0.0171	0.0261	0.0044	0.0210
	(2.01)	(1.53)	(1.56)	(1.60)	(1.46)	(1.11)	(0.36)	(1.08)
OTLR	0.0163***	0.0052	0.0203**	0.0045	0.0139*	-0.0011	0.0048	-0.0080
	(2.99)	(0.77)	(2.51)	(0.53)	(1.77)	(-0.14)	(0.47)	(-0.78)
OTLR0		0.0473**		0.0718**		0.0593**		0.0594***
		(2.12)		(2.58)		(2.52)		(2.82)
OS	-0.0124	-0.0126	-0.0023	-0.0036	-0.0032	0.0006	-0.0055	-0.0042
	(-1.37)	(-1.33)	(-0.45)	(-0.56)	(-0.36)	(0.06)	(-0.72)	(-0.48)
OS0		0.0141		0.0237		-0.0047		0.0077
		(0.83)		(0.82)		(-0.15)		(0.30)
Short	-0.0210**	-0.0117	-0.0269**	-0.0171	-0.0184*	-0.0079	-0.0321***	-0.0261**
	(-2.22)	(-0.89)	(-2.52)	(-1.24)	(-1.93)	(-0.65)	(-3.38)	(-2.41)
Short0		-0.0336		-0.0149		-0.0227		-0.0120
		(-1.35)		(-0.70)		(-0.65)		(-0.39)
OTLR*OS	0.0041	0.0014	0.0105**	0.0062	0.0137**	0.0110**	0.0119**	0.0088**
	(0.59)	(0.24)	(2.08)	(1.39)	(2.61)	(2.62)	(2.42)	(2.02)
Lag(Dep)	0.0054***	0.0052***	0.0037***	0.0035***	0.0011	0.0007	0.0008	0.0008
	(3.58)	(3.68)	(2.97)	(2.69)	(1.57)	(0.94)	(1.03)	(1.00)
Ltg	0.0141	0.0110	0.0250	0.0194	0.0072	0.0054	0.0071	0.0058
	(1.28)	(1.00)	(1.57)	(1.37)	(0.44)	(0.32)	(0.70)	(0.57)
Misp	-0.0074	-0.0094	-0.0180**	-0.0229**	-0.0180**	-0.0243**	-0.0266***	-0.0308***
	(-0.93)	(-1.15)	(-2.39)	(-2.18)	(-2.03)	(-2.64)	(-3.91)	(-4.20)
Log(Disp)	-0.0157**	-0.0152**	-0.0191**	-0.0213**	-0.0115**	-0.0123**	-0.0014	-0.0019
	(-2.37)	(-2.40)	(-2.18)	(-2.27)	(-2.31)	(-2.13)	(-0.22)	(-0.28)
Log(Amihud)	-0.0835***	-0.0780**	-0.0772***	-0.0653**	-0.0187	-0.0117	-0.0304	-0.0186
	(-2.78)	(-2.50)	(-3.42)	(-2.43)	(-0.73)	(-0.41)	(-1.23)	(-0.80)
Log(Size)	-0.0833***	-0.0841***	-0.0936***	-0.0869***	-0.0399*	-0.0361	-0.0291	-0.0227
	(-3.55)	(-3.14)	(-4.19)	(-3.18)	(-1.77)	(-1.35)	(-1.09)	(-0.83)
Log(BM)	0.0137	0.0127	0.0162*	0.0169	0.0153	0.0157*	0.0258**	0.0270**
	(1.42)	(1.45)	(1.76)	(1.60)	(1.55)	(1.68)	(2.28)	(2.48)
Return	-0.0354***	-0.0357***	-0.0335**	-0.0377***	-0.0438***	-0.0421***	-0.0268**	-0.0300***
	(-4.87)	(-5.25)	(-2.31)	(-2.66)	(-3.47)	(-3.28)	(-2.39)	(-2.79)
Mom(1,6)	-0.0274***	-0.0271***	-0.0222*	-0.0201*	-0.0078	-0.0061	0.0111	0.0100
	(-4.01)	(-4.25)	(-1.89)	(-1.74)	(-0.44)	(-0.34)	(0.55)	(0.50)
Mom(7,12)	-0.0062	-0.0104	-0.0026	-0.0065	-0.0071	-0.0106	-0.0095	-0.0132
	(-0.43)	(-0.78)	(-0.16)	(-0.40)	(-0.48)	(-0.77)	(-0.66)	(-0.93)
Adj-R^2	6.15%	6.24%	7.31%	7.81%	9.10%	9.63%	9.92%	10.40%
N Obs	38,165	38,165	38,165	38,165	38,160	38,160	38,139	38,139

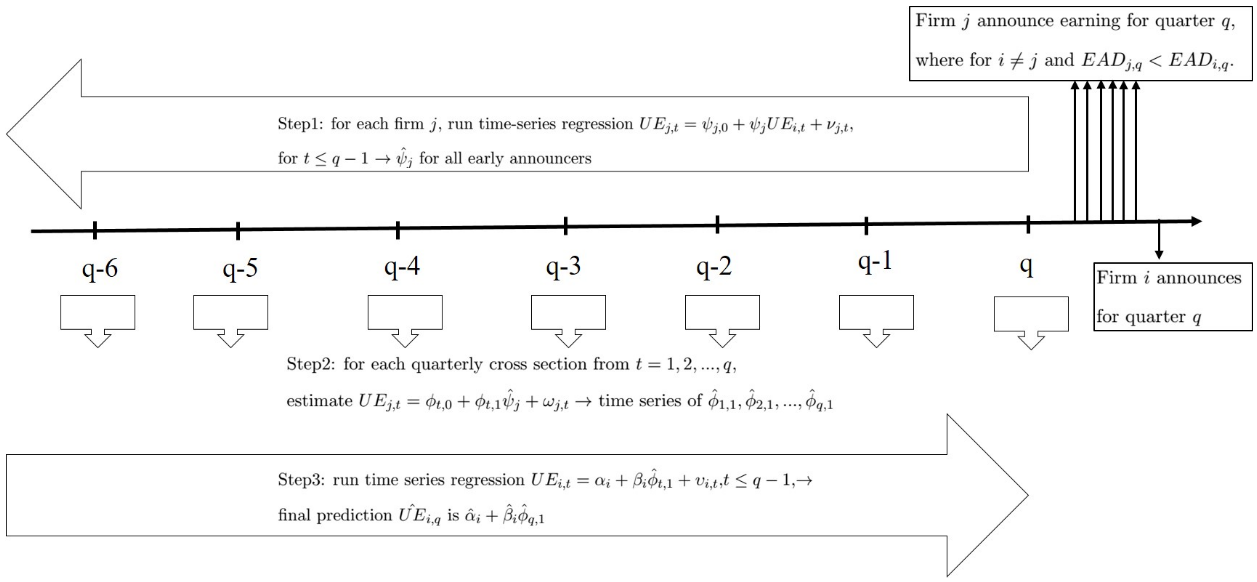


Figure 1 3PRF estimations for one factor assumption. This figures shows three-step OLS estimations of 3PRF in the prediction of unexpected earning(UE) for firm i, quarter q. Assume one common latent factor for UEs of firm i and other firms that announce before firm i. For each firm and quarter, such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

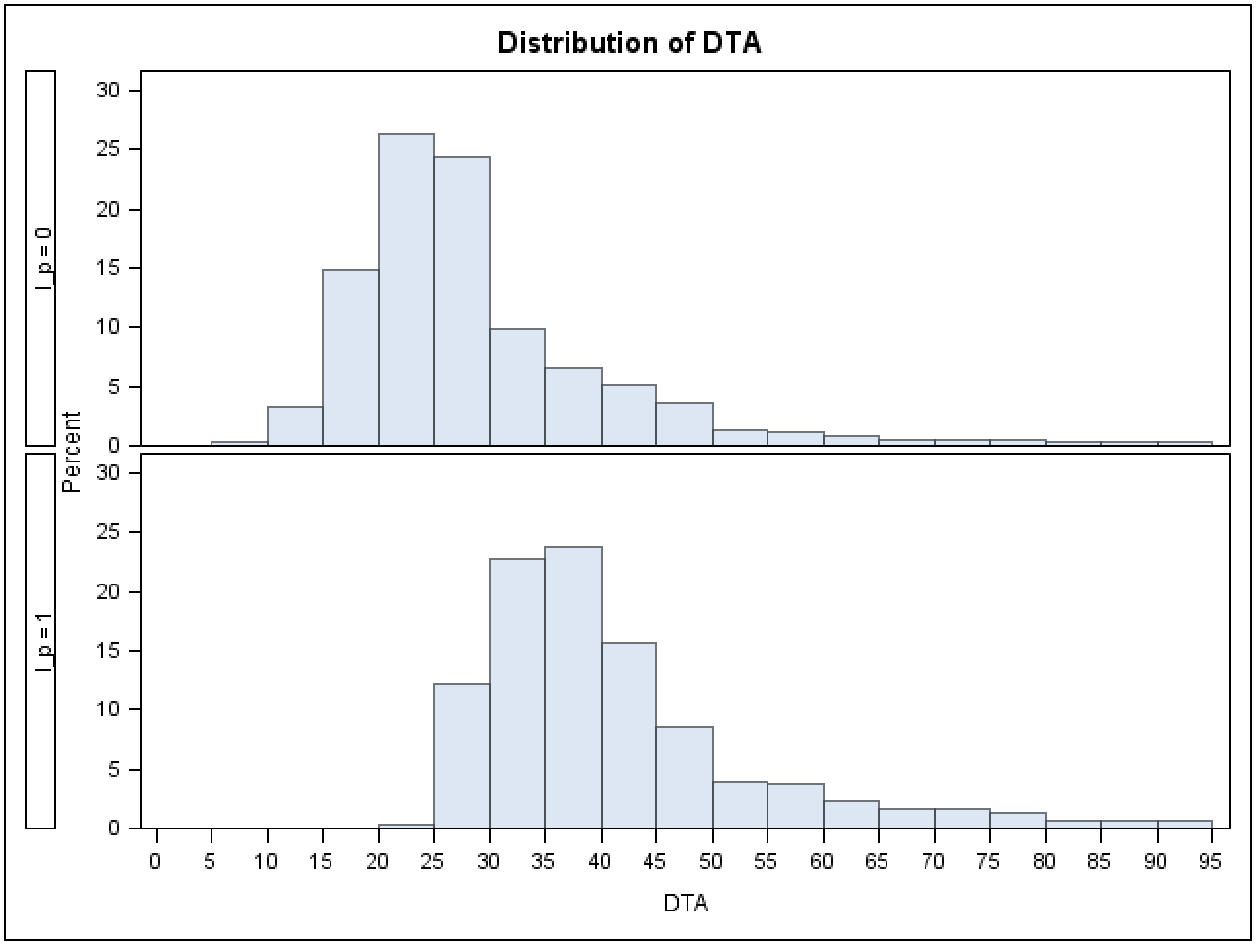


Figure 2 Days to announcement. This figure plots histograms of day-gaps between fiscal quarter end and earnings announcement day. The upper plot is the histogram for firms with no predictions by 3PRF, while the lower graph is the one for firms with 3PRF predictions of UEs.

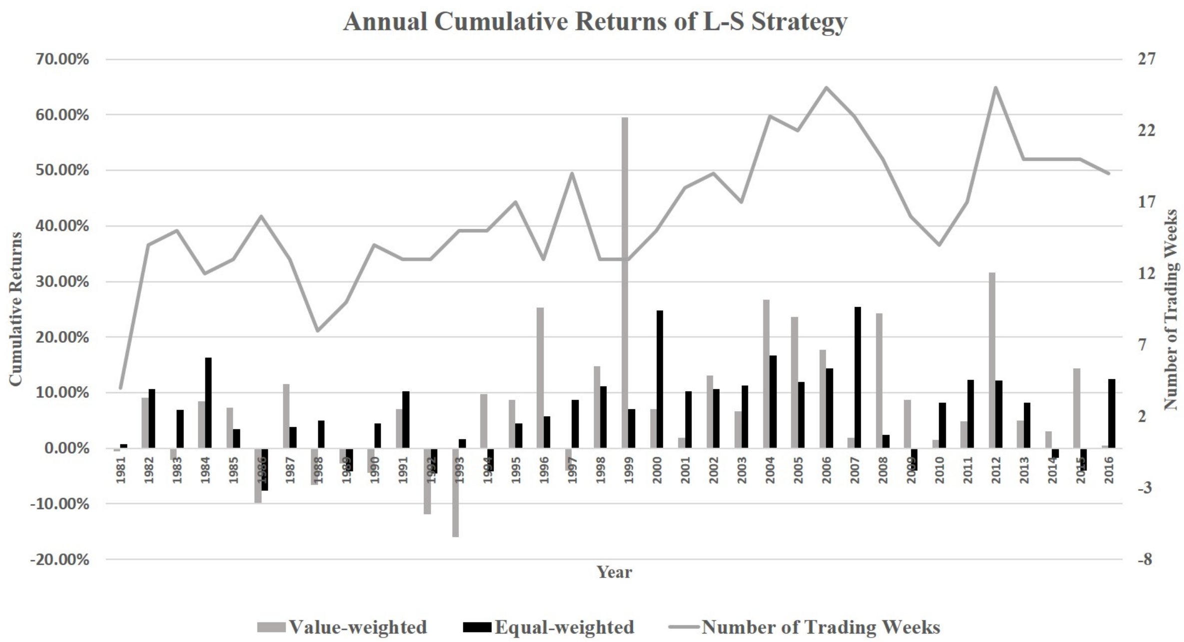
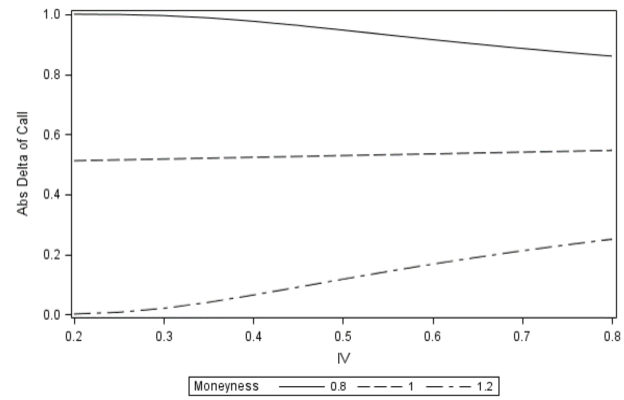
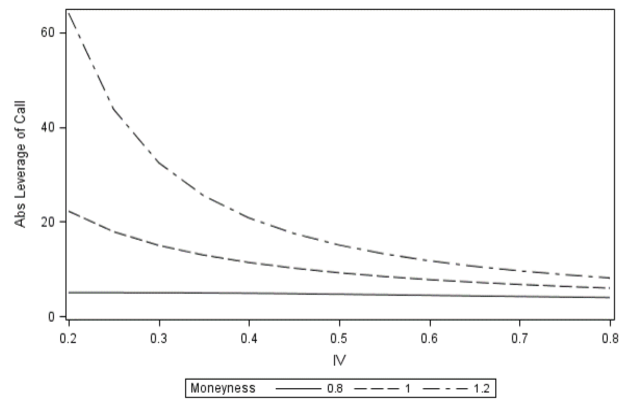


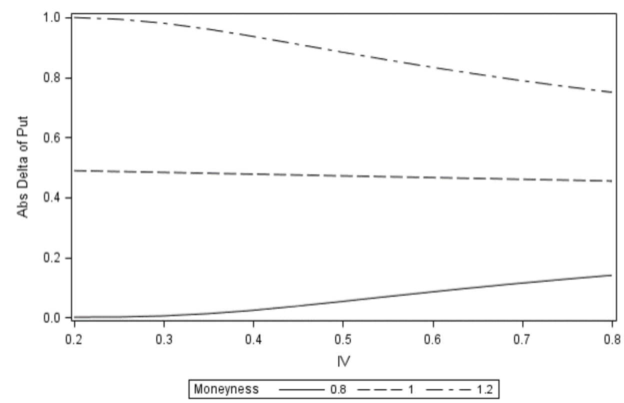
Figure 3 Annual cumulative returns of long-short strategy



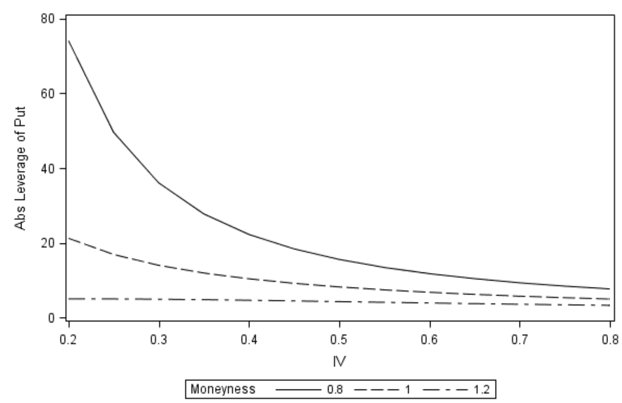
(a) Absolute delta of call options



(b) Absolute leverage of call options



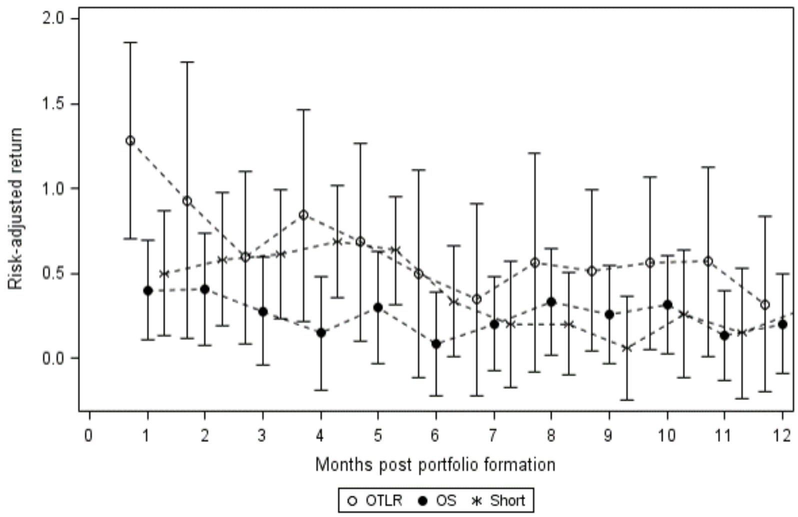
(c) Absolute delta of put options



(d) Absolute leverage of put options

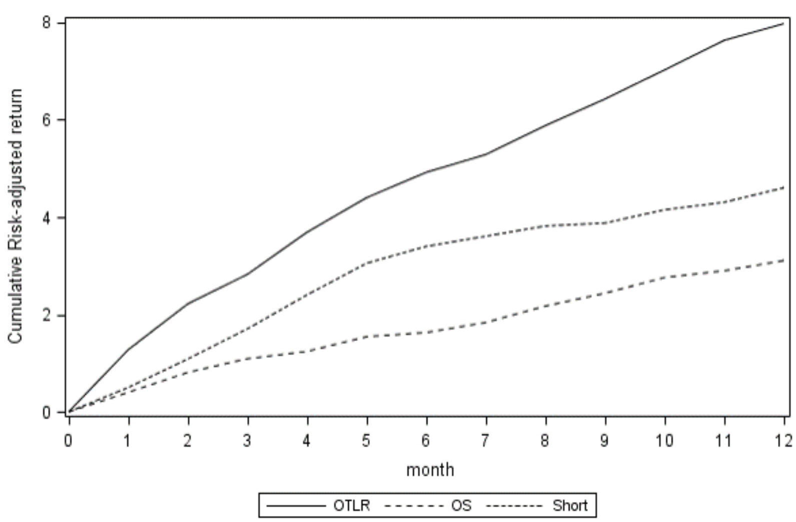
Figure a and b (c and d) shows how absolute delta and option leverage change with implied volatility for a call (put) option with moneyness equals to 0.8, 1, and 1.2. The option has 30 days to expiration and no dividend yield before expiration. Risk free rate is set to 0.

Figure 4 : Delta and leverage



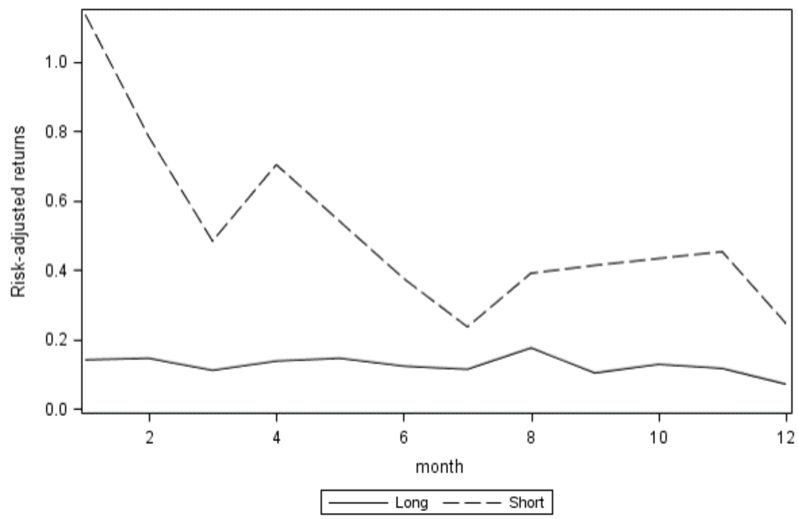
This figure presents time series average of long-short strategy risk-adjusted returns (value-weighted) over time. Quintile portfolios are formed by OTLR, OS and Short respectively at month 0. Y axis is risk-adjusted return in percentage using four-factor (MKTRF, SMB, HML and UMD) model.

Figure 5 : Risk-adjusted returns of long-short strategies



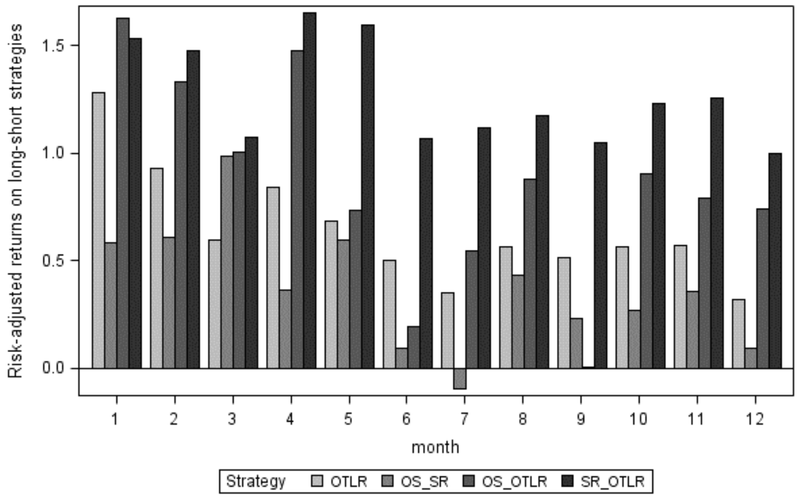
This figure presents time series average of long-short strategy cumulative risk-adjusted returns (value-weighted) over time. Quintile portfolios are formed by OTLR, OS and Short respectively at month 0. Y axis is cumulative risk-adjusted return in percentage using four-factor (MKTRF, SMB, HML and UMD) model.

Figure 6 : Cumulative risk-adjusted returns of long-short strategies



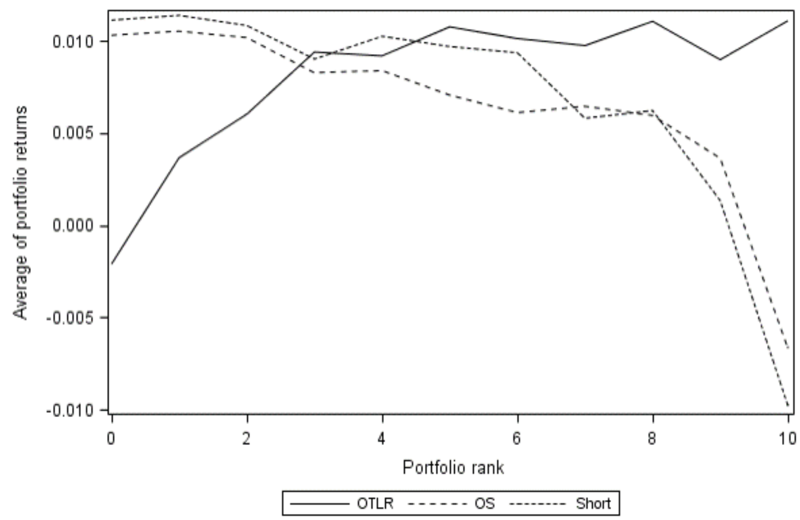
This figure shows time series average of the long leg and short leg risk-adjusted returns (value-weighted) of strategy based on OTLR. Quintile portfolios are formed by OTLR at month 0. Y axis is risk-adjusted return in percentage using four-factor (MKTRF, SMB, HML and UMD) model.

Figure 7 : Convergence of risk-adjusted portfolio returns



This figure presents risk-adjusted returns of long-short strategies post portfolio formation. Strategy OTLR is to long portfolios with highest OTLR and to short portfolios with lowest OTLR, for which quintile portfolios are formed based on OTLR. Strategy OS_SR is to long portfolios with lowest OS and Short (SR) and to short portfolios with highest OS and SR, for which 5 × 5 portfolios are formed by double-sorting on OS and SR. Strategy OS (SR)_OTLR is to long portfolios with lowest OS (SR) and highest OTLR and to short portfolios with highest OS (SR) and lowest OTLR, for which 5 × 5 portfolios are formed by double-sorting on OS (SR) and OTLR. Y axis is risk-adjusted return in percentage using four-factor (MKTRF, SMB, HML and UMD) model.

Figure 8 : Risk-adjusted return over time



This figure shows the time series average of deciles portfolio raw returns formed on OTLR, OS and Short respectively. X axis is the deciles ranking of portfolios.

Figure 9 : Raw returns of decile portfolios