The Pennsylvania State University

The Graduate School

Smeal College of Business

INFORMATION CHOICE, UNCERTAINTY, AND EXPECTED RETURNS

A Dissertation in

Business Administration

by

David Gempesaw

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

May 2019

The dissertation of David Gempesaw was reviewed and approved* by the following:

Charles Cao
Professor of Finance
Dissertation Co-Adviser
Co-Chair of Committee

Timothy Simin
Associate Professor of Finance
Dissertation Co-Adviser
Co-Chair of Committee

Stephen Lenkey
Assistant Professor of Finance

Kai Du
Assistant Professor of Accounting

Jeremiah Green
Associate Professor of Accounting (Texas A&M University)
Special Member

Brent Ambrose
Professor of Risk Management
Director of Smeal College of Business Ph.D. Program

*Signatures are on file in the Graduate School.

Abstract

In this dissertation, I investigate the empirical relationship between investors' information choices and the cross-section of risk and return in the equity market. My analysis builds upon the rational expectations equilibrium model of information choice and investment choice developed by Van Nieuwerburgh and Veldkamp (2010). I estimate a variable from the model called the learning index that reflects the theoretical expected benefits of learning about an asset for a rational average investor. Using this measure as a proxy for information flow, I find that stocks with higher values of the learning index have lower expected returns and volatilities in the cross-section on average. I provide support for the interpretation of the learning index through analyses based on short run and long run patterns in returns and volatilities, other measures of information flow, the information environment surrounding earnings announcements, and measures of information processing costs. Taken together, my findings provide evidence in support of the model's predictions and illustrate a new approach to empirically measure investors' information choices and assess the effects of these choices.

Contents

List of Figures vi

List of Tables vii

Acknowledgements x

1 Introduction 1

1.1 Overview 1

1.2 Literature review 7

1.3 Hypothesis development 9

1.4 Illustrative example: Doral Financial Corporation 13

2 Methodology and data 16

2.1 Estimating the learning index 16

2.2 Data sources and variable definitions 18

2.3 Descriptive statistics 20

3 Learning and the cross-section of risk and return 23

3.1 Explaining the cross-section of expected returns 23

3.1.1 Portfolio sorting 23

3.1.2 Cross-sectional regressions 26

3.2 Explaining the cross-section of volatility 29

3.2.1 Portfolio sorting 29

3.2.2 Cross-sectional regressions 31

4 Support for interpretation of the learning index 36

4.1 Contemporaneous impact of learning on price and volatility 36

4.2 Long-term predictability 39

4.3 Relationship with measures of information flow 42

4.3.1 Learning index and analyst coverage 46

4.3.2 Learning index and EDGAR downloads 48
4.4 Learning prior to earnings announcement month 49
4.4.1 Variation in earnings announcement activity 51
4.5 Learning during earnings announcement month 52
4.5.1 Variation in earnings announcement activity 54
4.5.2 Changes in the learning index 55
4.6 Learning costs and return/volatility predictability 57

5 Additional discussion and robustness checks **61**
5.1 Discussion of model assumptions 61
5.2 Alternative asset pricing models 63
5.3 Explaining the cross-section of volatility (Bivariate portfolio sorting) 64
5.4 Explaining the cross-section of implied volatility 65
5.5 Explaining the cross-section of market beta 66
5.6 Additional control variables 68
5.7 Subperiod analysis 69
5.8 Components of the learning index 71
5.9 Alternative test assets 74

6 Conclusion **76**

References **78**

Figures and Tables **83**

Appendix **130**

List of Figures

Figure 1 Illustrative example: Doral Financial Corporation 84
Figure 2 Long-term return predictability: $LI5 - LI1$ portfolio average return . . . 96
Figure 3 Long-term return predictability: $LI5 - LI1$ portfolio risk-adjusted average return 97
Figure 4 Long-term volatility predictability: $LI5 - LI1$ portfolio average abnormal return volatility 98
Figure 5 Long-term volatility predictability: $LI5 - LI1$ portfolio average abnormal systematic volatility 99
Figure 6 Long-term volatility predictability: $LI5 - LI1$ portfolio average abnormal idiosyncratic volatility 100

List of Tables

Table 1 Cross-sectional summary statistics 85
Table 2 Cross-sectional correlations 86
Table 3 Transition probabilities for portfolios sorted by learning index 87
Table 4 Explaining the cross-section of expected returns: Portfolios of stocks sorted by learning index 88
Table 5 Explaining the cross-section of expected returns: Cross-sectional regressions 89
Table 6 Explaining the cross-section of volatility: Portfolios of stocks sorted by learning index 90
Table 7 Explaining the cross-section of total volatility: Cross-sectional regressions 91
Table 8 Explaining the cross-section of systematic volatility: Cross-sectional regressions 92
Table 9 Explaining the cross-section of idiosyncratic volatility: Cross-sectional regressions 93
Table 10 Contemporaneous price impact of learning: Portfolios of stocks sorted by changes in learning index 94
Table 11 Contemporaneous volatility impact of learning: Portfolios of stocks sorted by changes in learning index 95
Table 12 Relationship with measures of information flow: Portfolios of stocks sorted by learning index controlling for firm size 101
Table 13 Learning index and analyst coverage: Cross-sectional regressions 102
Table 14 Learning index and EDGAR downloads: Cross-sectional regressions . . . 103
Table 15 Learning prior to earnings announcement month: Portfolios of stocks sorted by lagged learning index controlling for firm size 104
Table 16 Variation in earnings announcement activity: Portfolios of stocks sorted by lagged learning index controlling for firm size 105
Table 17 Learning during earnings announcement month: Portfolios of stocks sorted by contemporaneous learning index controlling for firm size 106

Table 18 Variation in earnings announcement activity: Portfolios of stocks sorted by contemporaneous learning index controlling for firm size 107
Table 19 Learning during earnings announcement month: Portfolios of stocks sorted by changes in learning index controlling for firm size 108
Table 20 Variation in earnings announcement activity: Portfolios of stocks sorted by changes in learning index controlling for firm size 109
Table 21 Learning costs and return predictability: Portfolios of stocks sorted by number of business segments and learning index controlling for firm size 110
Table 22 Learning costs and volatility predictability: Portfolios of stocks sorted by number of business segments and learning index controlling for firm size 111
Table 23 Learning costs and return predictability: Portfolios of stocks sorted by firm complexity and learning index controlling for firm size 112
Table 24 Learning costs and volatility predictability: Portfolios of stocks sorted by firm complexity and learning index controlling for firm size 113
Table 25 Alternative asset pricing models: Portfolios of stocks sorted by learning index 114
Table 26 Explaining the cross-section of volatility: Portfolios of stocks sorted by learning index controlling for past volatility 115
Table 27 Explaining the cross-section of implied volatility: Portfolios of stocks sorted by learning index 116
Table 28 Explaining the cross-section of implied volatility: Portfolios of stocks sorted by learning index controlling for past implied volatility 117
Table 29 Explaining the cross-section of market beta: Portfolios of stocks sorted by learning index 118
Table 30 Explaining the cross-section of market beta: Portfolios of stocks sorted by learning index controlling for past market beta 119
Table 31 Explaining the cross-section of market beta: Cross-sectional regressions 120
Table 32 Additional control variables: Cross-sectional return regressions 121
Table 33 Additional control variables: Cross-sectional volatility regressions 122
Table 34 Subperiod analysis: Portfolios of stocks sorted by learning index 123
Table 35 Subperiod analysis: Learning index coefficient from cross-sectional regressions 124
Table 36 Components of the learning index: Cross-sectional summary statistics and correlations 125
Table 37 Components of the learning index: Portfolio sorting 126
Table 38 Components of the learning index: Cross-sectional return regressions . . 127

Table 39 Components of the learning index: Cross-sectional volatility regressions 128
Table 40 Alternative test assets: Portfolios sorted by learning index 129
Table A1 Variable definitions 131

Acknowledgements

I would like to thank my dissertation committee co-chairs Charles Cao and Tim Simin for their guidance, mentorship, and advice. I also thank Stephen Lenkey, Kai Du, and Jeremiah Green for sharing their time and insight as members of my dissertation committee. I thank the faculty, staff, and my fellow Ph.D. students at Penn State who have helped me in various ways throughout the course of my doctoral studies. I am grateful to my family for their constant support, particularly my parents for being the best role models I could ask for. Finally, I thank my wife Nicole for her love and sacrifice and for always believing in me.

Chapter 1

Introduction

1.1 Overview

In the US equity market, price discovery is driven by the trading activity of active investors — for every $1 in trades based on a passive index strategy, active stock selectors trade approximately $22.[1] Active asset management involves making decisions not only about portfolio allocation, but also about information acquisition. While it is difficult to observe investors' information choices, a large literature provides insight regarding the impact of these choices on investment performance by analyzing observable outcomes such as portfolio holdings and investment returns, which are functions of an investor's ability and decision to acquire and use information.[2] In contrast, less is known about the impact of information choices on outcomes for the underlying assets.

In this dissertation, I investigate the role of investors' learning decisions in determining the cross-section of risk and expected return. Rather than attempt to directly measure what investors know, I rely on a theory that predicts which assets a rational investor would choose to learn about. Van Nieuwerburgh and Veldkamp (2010) present a rational expectations

[1] "Viewpoint: Index investing supports vibrant capital markets," BlackRock, October 2017.

[2] For example, see Grinblatt and Titman (1989), Wermers (2000), Kosowski, Timmermann, Wermers, and White (2007), Kacperczyk, Sialm, and Zheng (2008), and Cremers and Petajisto (2009).

general equilibrium model in which investors are able to reduce uncertainty about the future payoffs of particular risky assets before making investment decisions. The model generates predictions regarding the relationship between aggregate learning and the risk and return characteristics of the individual assets. First, learning about an asset results in lower uncertainty or risk — an increase in information corresponds to more precise conditional expectations of future payoffs. Second, learning about an asset results in a lower expected return — given an average information signal, risk-averse investors prefer to hold assets that they know more about. Finally, the model delivers a prediction about information choices through a measure called the learning index (LI). The learning index represents the expected benefits of learning about a particular asset and is increasing in the asset's prior squared Sharpe ratio and expected pricing errors. In equilibrium, higher values of the learning index correspond to a greater degree of learning about future unknown payoffs.

The objective of this paper is to test these predictions in the context of the equity market by empirically estimating the learning index. The conclusions from these tests have direct implications for empirical asset pricing. Standard pricing models used by academics and practitioners do not account for the ability of investors to reduce the risk of particular assets by learning. This omission leads to patterns in pricing errors that can be predicted by the learning index. The use of the learning index to measure information choices has a number of additional advantages. For example, estimating the learning index only requires historical return data. As such, this methodology can be applied to any market or set of assets for which return history is available and for which information acquisition is an important component of price determination. Furthermore, the fact that the learning index is derived from theory facilitates interpretation of the measure — the empirical learning index identifies which assets would be most valuable to learn about for the average investor (according to the model) and therefore represents a prediction of cross-sectional variation in information flow. In addition, unlike other theories that rely on untestable assumptions about investors' unobservable information sets, the theory of Van Nieuwerburgh and Veldkamp (2010) and the

information choices predicted by the learning index can be tested with observable variables — the learning index is estimated based on historical returns and is used to predict cross-sectional patterns in realized returns and risk.[3]

I implement a novel methodology to estimate the learning index for individual stocks at the end of each month from 1964 to 2016. The empirical learning index is a cross-sectional ranking ranging from zero to one in each monthly cross-section. The underlying testable expectation is that higher values of the learning index correspond to a greater degree of investor learning and information flow. Using this measure, I first test the predicted relation between learning and expected returns. Univariate portfolio analyses indicate a negative cross-sectional relation between LI and stock returns over the following month. For value-weighted portfolios, the average return spread between the highest and lowest quintile portfolios sorted on LI is -0.50% per month or -6.2% per year. After adjusting portfolio returns for exposure to market, size, value, profitability, investment, and momentum factors, the difference in risk-adjusted return between the extreme quintiles is -0.52% per month or -6.4% per year. As an alternative approach, I use two-stage cross-sectional regressions to examine the explanatory power of LI while controlling for several stock characteristics that are recognized in the literature as important predictors of future stock returns. Coefficient estimates from the multivariate regressions indicate that the difference between the stock with the highest and lowest learning index in an average cross-section is approximately -0.41% per month or -5.0% per year (all else equal). These results support the model's prediction that an increase in information about an asset corresponds to a lower expected return.

Next, I evaluate the relationship between information choices and risk. For each stock-month, I construct a measure of abnormal return volatility in the next month relative to the average level of volatility over the prior 12 months. Using value-weighted quintile portfolios formed based on LI, I find that abnormal return volatility in the following month is 3.98%

[3]Van Nieuwerburgh and Veldkamp (2009) and Veldkamp (2011) propose this property as an advantage of the theoretical analysis of information choice.

lower on average for high LI stocks compared to low LI stocks (for comparison, the abnormal monthly return volatility of the average stock in my sample is 1.27%). I decompose abnormal return volatility into systematic and idiosyncratic components and find that the learning index predicts cross-sectional differences in both components of risk. This result indicates that learning about an asset reduces not only firm-specific uncertainty, but also return co-movement with systematic risk factors. I arrive at similar conclusions using alternative sorting approaches and multivariate cross-sectional regressions of next month systematic, idiosyncratic, or total volatility on LI and a set of control variables. Taken together, these results suggest that the observed negative cross-sectional relation between LI and expected return derives from investors' decisions to reduce risk through learning.

After investigating the explanatory power of the learning index for the cross-section of stock returns and volatilities, I perform a number of analyses intended to provide support for the interpretation of the learning index as a proxy for learning decisions and information flow.

First, I investigate the contemporaneous price impact of learning. If, on average, investor learning results in a lower future return due to a reduction in future risk, then this implies that the current price will increase on average to reflect a lower expected return and the current volatility will increase as investors trade upon their information. I find that stocks with the largest increase in LI tend to have the highest maximum daily and weekly returns during the month. These stocks also have the highest maximum absolute returns during the month, reflecting a higher level of short-term volatility.

Second, I form value-weighted quintile portfolios based on LI and track the differences in expected returns and abnormal risk between the extreme quintiles over a long-term horizon. If investors learn information and trade based on that information, prices should move toward their intrinsic values and not revert in the future. The difference in risk-adjusted monthly returns between extreme value-weighted LI quintiles is negative and significant for up to eight months following portfolio formation. These differences are not reversed during

the subsequent two years, suggesting that the return predictive power of LI is due to investor learning rather than temporary price pressure or mispricing. LI also predicts significant differences in average abnormal volatility between extreme value-weighted quintiles for seven months after portfolio formation.

Third, to test the notion that the learning index is representative of investor learning and information flow, I investigate the relationship between LI and a number of variables reflecting investor attention or information demand. Using a bivariate portfolio sorting approach to control for firm size, I find that stocks with higher values of LI have greater abnormal trading activity, analyst coverage, forecast revisions, improvements in forecast accuracy, EDGAR filing downloads, and news reading activity on Bloomberg terminals.

Fourth, I perform two sets of tests to examine the relationship between the learning index and the information environment surrounding quarterly earnings announcements. If investors learn about a firm *prior* to an earnings announcement, this should reduce the amount of new information revealed in the announcement and result in a smaller market reaction on average. After controlling for size, I find that stocks with a higher learning index in the month before the announcement tend to have smaller market reactions to earnings announcements and a smaller post-earnings announcement drift. High lagged LI stocks also experience a higher degree of abnormal trading activity in the month prior to an earnings announcement. These findings are consistent with more information being acquired for high lagged LI stocks and incorporated into prices before earnings announcements.

On the other hand, if a stock has a high learning index *during* the earnings announcement month, this suggests that information flow is high in that month, potentially because of the information disclosed in the announcement. I find that stocks with higher values of LI in the earnings announcement month tend to have a larger market reaction and greater trading activity in the three-day window around the announcement and a smaller post-earnings announcement drift over the subsequent quarter. The results on learning prior to and during the earnings announcement month are also stronger when there are relatively fewer firms

announcing earnings, consistent with investors' learning efforts being concentrated among fewer firms.

Finally, I examine how variation in information processing cost affects the explanatory power of the learning index for risk and return. Using a measure of firm complexity based on sales concentration across segments within a firm, I find that the learning index predicts larger differences in next month return for less complex firms. However, I find that the explanatory power of LI for abnormal volatility is similar across levels of firm complexity. The evidence suggests that a given proportional reduction in risk corresponds to a greater reduction in expected return for firms that are more transparent and easier to understand. Altogether, the results from these analyses reinforce the idea that the learning index is representative of the information choices of investors and the flow of information about a given firm.

Before proceeding, a note of clarification is in order regarding the perspective of my empirical analysis. This study focuses on information choices at the aggregate level, not at the individual investor level. The empirical learning index serves as a prediction about variation in a firm's information environment resulting from the information acquisition decisions of all investors. While assets with higher values of the learning index are predicted to have lower equilibrium expected returns, an individual investor who learns about these assets makes more informed investment choices and has a higher expected portfolio return. Because the empirical learning index does not directly provide insight into information choices at the individual investor level, I focus only on testing the relationships between learning, risk, and return at the aggregate level.

The rest of the dissertation is organized as follows. In the remainder of this chapter, I provide a review of related literature and highlight the contribution of this paper. I also discuss the model of Van Nieuwerburgh and Veldkamp (2010), describe the learning index, and outline the model's relevant predictions. I then present an illustrative example of the information content of the learning index using a case study. In Chapter 2, I describe the

procedure to empirically estimate the learning index and summarize the data. Chapter 3 presents empirical results on the cross-sectional explanatory power of the learning index for expected returns and volatilities. In Chapter 4, I provide supporting evidence for the interpretation of the learning index. Chapter 5 contains additional discussion of the model and results from various robustness checks. Chapter 6 concludes.

1.2 Literature review

This paper contributes to a line of research featuring empirical applications of noisy rational expectations equilibrium models focused on the information content of prices. Biais, Bossaerts, and Spatt (2010) argue that prices contain information that is value-relevant to an uninformed investor and document that a price-contingent portfolio based on ex-ante information outperforms a passive index. Banerjee (2011) presents a model that nests the rational expectations and differences of opinion approaches, each of which delivers contrasting predictions regarding how investors use prices. The author finds empirical evidence suggesting that investors exhibit rational expectations and condition their beliefs on prices. Burlacu, Fontaine, Jiminez-Garcés, and Seasholes (2012) develop a measure of information precision and supply uncertainty based on the work of Admati (1985) and investigate its relationship with expected returns.

The theoretical models underlying the aforementioned literature typically rely on the assumption that information asymmetry is exogenously determined (e.g., all investors receive a private information signal, or a certain fraction of investors are assumed to be informed). In contrast, my empirical analysis is based on a model that treats investors' information sets as endogenous. Kacperczyk, Van Nieuwerburgh, and Veldkamp (2016) construct and test a closely related model of mutual fund managers' attention allocation and portfolio choices. While there are a number of common themes between that model and the model underlying my paper, the empirical focus of the two papers is different. These authors concentrate on

identifying patterns in mutual fund investment and performance that vary with the business cycle, whereas I am interested in directly estimating the learning index at the individual asset level and using it in cross-sectional analyses.

My paper also relates to the literature investigating the empirical relationship between information flow, expected returns, and risk. Botosan (1997) finds that greater voluntary disclosure by firms is associated with a lower cost of equity capital. Using firm age as a proxy for uncertainty about future profitability, Pastor and Veronesi (2003) show that firms with lower uncertainty have lower market-to-book ratios and lower volatilities. Pan, Wang, and Weisbach (2015) find that volatility is decreasing in CEO tenure, arguing that uncertainty is reduced over time as investors learn about CEO ability. Using SEC Form 8-K filing frequency as a measure of information intensity, Zhao (2017) demonstrates that information intensity reduces expected uncertainty and expected return. Each of these studies focus on information flows that are exogenous from the investor's perspective. I provide complementary evidence to this literature by demonstrating a cross-sectional link between information, returns, and risk using a measure intended to reflect investors' endogenous learning decisions.

Prior studies have introduced a number of empirical proxies for investor attention or information acquisition. For example, Barber and Odean (2007) use news coverage, abnormal trading volume, and extreme one-day returns as indirect measures of retail investor attention. Recently, researchers have proposed more direct measures information acquisition, such as download activity of SEC filings (e.g., Li and Sun (2017) and Crane, Crotty, and Umar (2018)) or news reading activity on Bloomberg terminals (Ben-Rephael, Da, and Israelsen (2017)).

Da, Engelberg, and Gao (2011) construct a measure of retail investor attention based on Google search frequency. These authors show that increased search activity on Google is associated with future stock price appreciation. At first glance, this finding may seem contradictory to the evidence of a negative relation between learning and returns presented in this paper. My research focus differs from that of this particular study in a subtle but important way. Da et al. (2011) are interested in examining the price pressure created by

attention-driven buying activity. In contrast, I study the reduction in risk and risk premium associated with an increase in information. Da et al. (2011) show that higher search activity on Google predicts higher stock prices during the subsequent two weeks, followed by reversal within one year. This pattern is consistent with an explanation based on attention-induced price pressure. In contrast, I demonstrate that the learning index predicts lower future monthly returns that are not reversed in the long run. This pattern is inconsistent with the price pressure hypothesis. Instead, it suggests that an increase in information is associated with more efficient prices (as evidenced by the lack of reversal) and a lower risk premium.

I add to the aforementioned literature by introducing a theoretically-motivated prediction of investors' learning behavior. This measure has a number of limitations and advantages relative to the other existing proxies. For instance, the empirical learning index does not depend on direct observation of investor learning. Instead, it identifies which assets rational investors would want to learn about in the context of the model. Furthermore, the learning index cannot be used to differentiate between the information activities of retail investors and institutional investors. Rather, it should be viewed as a prediction of information flow for the average investor. On the other hand, use of the learning index is advantageous because it is less restricted by data availability. Data on EDGAR file download activity, Bloomberg news reading activity, and Google search activity begin in 2003, 2004, and 2010, respectively. Since estimation of the learning index only requires historical return data, it can be used during earlier time periods as well as for other sets of assets.

1.3 Hypothesis development

My empirical analysis is based on the rational expectations general equilibrium model of information choice and investment choice developed by Van Nieuwerburgh and Veldkamp (2010). The authors explore the impact of different assumptions regarding learning technologies and investor preferences on the optimal information acquisition strategy. I focus

on the version of their model with mean-variance preferences and entropy-based learning because these assumptions lead to a prediction of specialized learning (as opposed to generalized learning), which is consistent with the actual behavior of informed investors. Further discussion of the model's assumptions is presented in Section 5.1.

The model contains multiple risky assets and multiple investors. Prior to investing, investors have the ability to acquire information about unknown asset payoffs f, which are assumed to be normally distributed with mean μ and variance Σ. The learning decision involves choosing which assets to learn about and how much to learn about them, subject to a learning capacity constraint. The model assumes independent asset payoffs and independent information signals about these payoffs. If assets are correlated, an eigen decomposition can be used to form independent linear combinations of the correlated assets. These synthetic assets can be interpreted as principal components (PC), risk factors, or Arrow-Debreu securities. Specifically, a non-diagonal covariance matrix Σ can be decomposed into an eigenvector matrix Γ and a diagonal eigenvalue matrix Λ: $\Sigma = \Gamma\Lambda\Gamma'$. The eigenvalue matrix contains the variances of the principal components, while the eigenvector matrix contains the loadings of the correlated assets on the principal components. With these assumptions, the investor's information choice is equivalent to choosing the posterior variance of each principal component.

The model takes place over three periods: information choices are made in period 1, investment choices are made in period 2, and payoffs and utility are realized in period 3. The model is solved using backward induction. The optimal investment choice in period 2 is a diversified portfolio that conditions on an investor's prior beliefs, information signal realizations, and prices: $q^* = \frac{1}{\rho}\hat{\Sigma}^{-1}\left(\hat{\mu} - pr\right)$, where ρ is the coefficient of risk aversion, p is a vector of prices, r is the risk-free rate, and $\hat{\mu}$ and $\hat{\Sigma}$ are the posterior mean and variance of payoffs. Similar to Admati (1985), equilibrium prices are a linear function of payoffs and supply shocks x: $pr = A + Bf + Cx$. The coefficient matrices A, B, and C are functions of the posterior beliefs of the average investor, the level of risk aversion, and the asset supply. The

average investor can be viewed as a representative investor whose posterior mean $\hat{\mu}_a$ is the average of all investors' posterior means and whose posterior variance $\hat{\Sigma}_a$ is the harmonic average of all investors' posterior variances.

In period 1, the optimal information choice is to allocate all learning capacity toward the principal component with the highest value of the learning index. The learning index for PC i is

$$LI_i = \left(\left((I - B)\mu - A \right)' \Gamma_i \right)^2 \Lambda_i^{-1} + (1 - \Lambda_{Bi})^2 + \Lambda_i^{-1} \Lambda_{Ci}^2 \sigma_x^2. \tag{1.1}$$

The first term of (1.1) is equivalent to the prior squared Sharpe ratio of PC i: $\frac{\left(E\left[\Gamma_i'(f - pr)\right]\right)^2}{Var[\Gamma_i' f]}$. Alternatively, this term can be viewed as the product of two terms: $E\left[\Gamma_i'(f - pr)\right]$ and $\frac{E\left[\Gamma_i'(f - pr)\right]}{Var[\Gamma_i' f]}$, which is equivalent to ρ times the expected investment in PC i. These two terms indicate that the value of learning is greater for an asset with a high expected excess return and a high expected portfolio share. Consequently, there are increasing returns to learning — expecting to hold more of an asset makes it more valuable to learn about that asset, while learning more about an asset makes the asset less risky and more attractive to hold.

The second term reflects expected pricing errors related to the informativeness of prices about payoffs. Λ_{Bi} is the i^{th} eigenvalue of B and captures the relationship between payoffs and prices. When Λ_{Bi} is lower, prices covary less with payoffs, making information about payoffs more valuable to learn. The third term reflects expected pricing errors related to the sensitivity of prices to supply shocks. Λ_i and Λ_{Ci} are the i^{th} eigenvalues of the prior covariance matrix Σ and C, respectively. σ_x^2 is the variance of supply shocks, which is assumed to be the same for all PCs. Holding prior uncertainty constant, higher values of Λ_{Ci} indicate that supply shocks have a greater impact on prices, creating pricing errors that can be exploited by an informed investor.

In general equilibrium, ex-ante identical investors specialize by learning about a single factor, but each investor chooses to learn about different factors due to strategic substitutability — investors prefer to learn information that other investors do not know. As more investors learn about a given factor, the expected return on that factor is reduced, which

reduces the value of learning about that factor. The model has a unique equilibrium in which the aggregate learning capacity of all investors determines the number of risk factors that the economy learns about. However, each individual employs a mixed strategy and randomizes over which of these factors to learn about.

The model generates predictions for the relationships between information choices, risk, and expected returns: an increase in information about an asset leads to a reduction in uncertainty and a lower expected return. The model also provides predictions about the impact of learning on systematic risk exposure and prediction errors from a typical asset pricing model such as the CAPM. Similar to Biais et al. (2010) and Banerjee (2011), Van Nieuwerburgh and Veldkamp (2010) derive a conditional CAPM relation in which risk and expected return are measured conditional on information that the average investor knows.[4] In contrast, the standard unconditional CAPM beta is based only on past return information. Predictions of expected returns from the unconditional CAPM do not account for investors' ability to reduce risk through learning. Learning more information about an individual asset reduces the asset's total risk without changing the asset's correlation with the market risk factor. If investors learn more about an asset, the conditional CAPM beta (i.e., the beta conditional on the information learned by investors) will be lower than the unconditional CAPM beta, and the conditional expected return will be lower than the unconditional expected return. Therefore, the model predicts that learning reduces co-movement with systematic risk factors. The discrepancy between the empirically estimated unconditional risk exposure and the unobserved conditional risk exposure leads to cross-sectional variation in factor model pricing errors that is related to investors' learning decisions.

I apply these predictions to the cross-section of domestic equities by estimating the learning index for individual stocks and conducting the following analyses. First, I test the hypothesized relationship between learning and expected returns by examining the cross-sectional explanatory power of LI for future stock returns and risk-adjusted returns.

[4] See Section A.5 of the technical appendix to Van Nieuwerburgh and Veldkamp (2010) for proof.

Second, I test the hypothesized relationship between learning and risk by investigating the predictive power of LI for cross-sectional variation in return volatility, systematic volatility, and idiosyncratic volatility. Third, I test the hypothesis that LI captures information flow through analyses based on short run and long run patterns in returns and volatilities, other measures of information flow, the information environment surrounding earnings announcements, and measures of information processing costs.

1.4 Illustrative example: Doral Financial Corporation

In this section, I examine events occurring in the early 2000s related to Doral Financial Corporation, an NYSE-listed mortgage banking company with operations in Puerto Rico and New York City, in order to illustrate the correspondence between the learning index and investor learning or information flow. During this period, Doral Financial was the leading residential mortgage lender in Puerto Rico. As of the end of 2004, the company's market capitalization was approximately $5.31 billion and its share price was $49.25. The majority of the company's income came in the form of interest-only strips (IOs) from selling or securitizing residential mortgage loans. In order to value its IOs, Doral Financial assumed that interest rates were fixed at the spot rate, or the 90-day LIBOR rate at the end of each reporting period, instead of using an implied forward rate which more accurately reflected market expectations. This methodology significantly overstated the value of the company's portfolio. Senior management at the company became aware of this issue in late 2004.

On January 18, 2005, Doral Financial announced an impairment charge to its portfolio in an earnings press release for the fourth quarter of 2004, triggering a 12% drop in its stock price on the following day. On March 15, 2005, Doral Financial publicly disclosed its improper use of the spot rate methodology in its Form 10-K for the fiscal year 2004. The company's stock price dropped 44% over the next three days. According to restated financial statements, the company's pre-tax income between 2000 and 2004 was overstated by a

cumulative amount of approximately $921 million, $283.1 million of which was attributable to improper IO valuation. As a result of this valuation issue and additional accounting and disclosure irregularities, the total decline in Doral Financial's stock price ended up being over 80% of its value at the end of 2004, corresponding to a reduction in equity market value of more than $4 billion.

Figure 1 plots monthly values of the learning index for Doral Financial Corporation, the number of downloads of the company's SEC filings from the EDGAR database, the company's stock price, and its monthly return volatility from January 2004 through September 2005. During 2004, both the learning index and the number of EDGAR downloads stayed at relatively low levels. The average monthly value of the learning index during this time was 0.06 and the average number of EDGAR downloads in a month was 220. Over this year, the monthly volatility of the company's stock returns stayed consistent as the stock price increased from $32 to almost $50. In January 2005, the company's stock price declined 12% as a result of the announcement regarding impairment. During this month, return volatility increased from 16% to 47%, the learning index increased from 0.01 to 0.30, and the number of EDGAR downloads increased from 144 to 412. In March 2005, the stock price declined almost 45% due to the disclosures contained in the annual report. In this month, return volatility increased from 33% to 113%, the learning index increased from 0.09 to 0.85, and the number of EDGAR downloads increased from 317 to 1125. Over the next six months, the stock price continued to fall 40% of its value at the end of March. By September, monthly return volatility was 25%, the learning index was 0.26, and the number of EDGAR downloads was 249.

These patterns serve as an example of the relationship between predicted information flow, as measured by the learning index, and one existing measure of actual information flow. Since the theoretical predictions of reductions in expected return and risk are cross-sectional in nature, they cannot be directly applied to a single event. Rather, these predictions should be expected to hold on average across assets. This example also shows how information

acquisition is not always associated with a subsequent purchase of the asset — the direction of the investment decision depends on whether the observed information signal contains good or bad news. In this situation, investors learned about accounting issues that carried significant negative implications for the value of the company. As investors learned this new information, they updated their expectations about the future payoff of this particular stock, and the resulting trading activity moved the price closer to intrinsic value.

Chapter 2

Methodology and data

2.1 Estimating the learning index

My objective is to measure the learning index at the end of each month for each stock in the sample. The estimation procedure generally follows the approach described in Van Nieuwerburgh and Veldkamp (2009) and Veldkamp (2011). I use a two-year rolling window of weekly returns to construct prices, payoffs, and an estimate of the payoff covariance matrix. I use weekly returns instead of monthly or daily returns in order to increase the number of observations within the window while avoiding the effects of non-synchronous trading. Following convention in the literature, weekly returns are measured from Wednesday close to the next Wednesday close (or the closing price on the latest trading day within this period in case of a holiday). The following steps are performed at each month-end.

Step 1: Construct price (p) and payoff (f) time-series for each stock. The price of each stock is set equal to one in the first week. Stock prices then evolve according to the respective weekly return series. Because prices are assumed to be log-normally distributed, I use log prices to be consistent with the model's assumptions. The stock price in the following week is used as a proxy for the stock's payoff, and returns are calculated as $f - pr$. To avoid look-ahead bias in the empirical tests, estimation is only based on information available at

the end of the current month. Therefore, the final payoff observation in each window is the price at the end of the last full week in the current month.

Step 2: Convert the cross-section of correlated stocks to a set of uncorrelated assets. Estimate the prior covariance matrix Σ of payoffs from Step 1. To account for heteroskedasticity across individual assets, payoffs are standardized to have zero mean and unit variance prior to computing the covariance matrix and performing the eigen decomposition. This approach is equivalent to the maximum explanatory component analysis of Xu (2007) and avoids overweighting stocks with high idiosyncratic volatility when extracting the principal components. Decompose Σ into a diagonal eigenvalue matrix Λ and an eigenvector matrix Γ: $\Sigma = \Gamma\Lambda\Gamma'$. Construct principal component prices $\left(\Gamma' p\right)$, payoffs $\left(\Gamma' f\right)$, and returns $\left(\Gamma'(f - pr)\right)$.

Step 3: Estimate the learning index for principal components. The first term of the learning index is estimated by dividing squared average return by the variance of payoffs. The second and third term require estimation of the equilibrium price equation at the principal component level: $\Gamma' pr = \Gamma' A + \Gamma' Bf + \Gamma' Cx$. Since principal components are uncorrelated, this is equivalent to estimating a separate regression for each principal component of its price on a constant and its payoff. This step involves a time-series regression of two non-stationary variables. The underlying theory suggests that in equilibrium, there exists a linear combination of these variables that is stationary. As such, these variables are said to be cointegrated, and the cointegrating vector can be consistently estimated using OLS. In untabulated analysis, I verify the stationarity of the residuals from this regression. The payoff coefficient Λ_B and the regression R^2 are used to compute the second and third term.[1]

Step 4: Estimate the learning index for stocks. Pre-multiply the principal component learning index vector by the eigenvector matrix: $\Gamma\left(LI^{PC}\right)$. The learning index for a given stock

[1]Estimating $(1 - \Lambda_{Bi})^2$: If prices follow the pricing equation $pr = A + Bf + Cx$, then OLS can be used to directly estimate B. The OLS estimate is $\Sigma^{-1}\Sigma B = B$. Since assets are assumed to be independent, B is a diagonal covariance matrix and the eigenvalues of B are the diagonal elements of the matrix. For PC i, the OLS coefficient is a direct estimate of Λ_{Bi}.

Estimating $\Lambda_i^{-1}\Lambda_{Ci}^2\sigma_x^2$: First, compute the unconditional variance of prices: $Var(p) = Var(A + Bf + Cx) = B\Sigma B' + CC'\sigma_x^2$. This expression gives us the total sum of squares of prices. Because the asset supply shocks are assumed to be the regression residual, $CC'\sigma_x^2$ is the unexplained sum of squares and $B\Sigma B'$ is the explained sum of squares. Then $\frac{1-R^2}{R^2}$ corresponds to $\left(B\Sigma B'\right)^{-1}CC'\sigma_x^2$. That is, for asset i, $\Lambda_i^{-1}\Lambda_{Ci}^2\sigma_x^2 = \frac{1-R^2}{R^2}\Lambda_{Bi}^2$.

is a weighted sum of PC learning indexes where the weights are based on the contribution of the stock to each PC. A well-known practical issue involved in eigen decomposition is that the sign of an eigenvector is arbitrary. While this does not make a difference theoretically, it poses an empirical problem. To resolve this issue, I use the square of the normalized eigenvector elements as weights in calculating the stock learning index. This excludes the possibility of a stock having a negative learning index, which has no theoretical interpretation. Because the eigenvectors are standardized to unit length (i.e., the sum of squares for every eigenvector is one), an eigenvector element squared represents the contribution of the stock to the corresponding principal component. Therefore, a stock's learning index can be interpreted as a weighted sum of principal component learning indexes, where the weights are proportional to the stock's contribution to each principal component.

Because the number of stocks in each monthly cross-section varies over time, I rank-transform the stock learning index and its components to the interval [0,1] to facilitate interpretation and comparability across cross-sections. This transformation would only potentially affect the cross-sectional regression analyses and would have no impact on the portfolio sorting results. The main conclusions regarding the explanatory power of the learning index for risk and return are qualitatively similar without this transformation. Further discussion of the learning index and its components is presented in Section 5.8.

2.2 Data sources and variable definitions

I obtain daily and monthly data for US common stocks listed on the NYSE, AMEX, and NASDAQ from the Center for Research in Security Prices (CRSP) during the period from July 1962 to December 2016. Stock returns are adjusted for delisting following Beaver, McNichols, and Price (2007). To reduce the impact of microstructure issues and the influence of microcaps on the results, I require stocks to have a price greater than $5 and market capitalization above the 20^{th} NYSE percentile in order to be included in the sample at each month-end. Data for

market, size, value, profitability, investment, and momentum risk factors are obtained from Kenneth French's website.[2] Additional data sources include Compustat, Thomson Reuters Institutional Holdings, Institutional Brokers' Estimate System (I/B/E/S), SEC Electronic Data Gathering, Analysis, and Retrieval (EDGAR) Log Files, Bloomberg, Datastream, and OptionMetrics. The learning index is estimated over the period July 1964 to December 2016, but certain analyses are limited to a subset of this period based on data availability.

For each stock-month, I construct the following characteristics which have been identified in prior studies as important cross-sectional return predictors. Market beta $\left(\beta^{MKT}\right)$ is calculated from a regression of excess stock returns on excess market returns using daily data from the past year. To account for biases due to infrequent trading, I follow Dimson (1979) by including lagged and lead market returns in this regression. The market beta is the sum of the coefficient estimates for the lagged, current, and lead market return. *SIZE* is the natural logarithm of market value of equity. Book-to-market ratio (BM) is the book value of equity in the latest fiscal year ending in the prior calendar year divided by the market value of equity at the end of December of the prior calendar year. Profitability ($PROF$) is annual revenues minus cost of goods sold, interest expense, and selling, general, and administrative expenses divided by book equity for the latest fiscal year ending in the prior calendar year. Investment (INV) is the annual percentage change in total assets. Momentum (MOM) is the cumulative return from month $t-11$ to month $t-1$.

Illiquidity ($ILLIQ$) is the absolute monthly stock return divided by the respective monthly trading volume in dollars, scaled by 10^5. Short-term reversal (STR) is the monthly return of the stock over the past month. Long-term reversal (LTR) is the cumulative return from month $t-59$ to month $t-12$. Idiosyncratic volatility ($IVOL$) is the standard deviation of daily residuals within a month from estimation of the Fama and French (2018) six-factor model, which includes market, size, value, profitability, investment, and momentum risk factors.[3] I also compute total return volatility ($RVOL$) as the standard deviation of daily

[2] mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

[3] Results are robust to the use of alternative factor models to estimate systematic and idiosyncratic volatility.

excess returns within a month, and the systematic component of volatility ($SVOL$) as the square root of the difference between $RVOL^2$ and $IVOL^2$, although these two variables are not used as cross-sectional return predictors.

In addition to the aforementioned variables, I construct the following characteristics which have been identified in prior studies as important predictors for the cross-section of stock volatility. Return on equity (ROE) is earnings before extraordinary items as of the most recent fiscal quarter end divided by common shareholders' equity as of the end of the previous quarter and multiplied by 100. Volatility of return on equity ($ROEVOL$) is the standard deviation of return on equity over the prior 12 fiscal quarters. Firm age (AGE) is the number of years the firm has existed on CRSP. $DIVD$ is a dividend dummy equal to 1 if the firm paid dividends during the most recent fiscal quarter, and 0 otherwise. Leverage (LEV) is total liabilities scaled by the market value of equity as of the most recent fiscal quarter end. $INVPRC$ is the inverse of the stock price, scaled by 100. R is the monthly stock return in percent. All variable definitions are listed in Table A1.

2.3 Descriptive statistics

Table 1 presents time-series averages of monthly cross-sectional summary statistics for the aforementioned stock characteristics. Summary statistics are not presented for LI as it is uniformly distributed between 0 and 1 within each cross-section. In the average month, the average stock in the sample has a market beta of 1.06, market capitalization of $3.62 billion (untabulated), and book-to-market ratio of 0.71. The last row in the table reports time-series summary statistics for the number of stocks in the sample per month. The average (median) number of stocks in the sample in a given month is 1,627 (1,654).

Table 2 presents average cross-sectional correlations between key variables. I include only the characteristics used as return predictors for brevity. On average, stocks with high LI have a lower market beta, lower market capitalization, higher book-to-market ratio, lower

profitability, lower investment, higher illiquidity, lower past returns over short, intermediate, and long horizons, and higher idiosyncratic volatility. These correlations are generally small, indicating that a substantial component of cross-sectional variation in LI is orthogonal to these characteristics.

Table 3 reports transition probabilities for LI-sorted quintile portfolios over 1-month, 6-month, 12-month, 24-month, and 36-month periods. The extreme LI quintiles exhibit a relatively high level of persistence over shorter monthly horizons. Approximately 71.3% (70.3%) of the stocks in the lowest (highest) LI quintile remain in the same quintile in the next month. To a certain extent, this result can be expected given the high degree of overlap in the data used to calculate LI_t and LI_{t+1}. In untabulated analysis, I find that the average (median) first-order correlation coefficient for the learning index is 0.78 (0.83).

As the length of time between the initial month and the final month increases (and the degree of overlap in the data decreases), the level of persistence in LI declines. Panel D of Table 3 contains transition probabilities based on values of LI that are computed using consecutive non-overlapping two year windows. In this panel, the probabilities of transitioning among quintiles are all close to 20%, indicating that the learning index is not a persistent stock characteristic over the long run. I find a similar pattern in Panel E using values of LI that are 36 months apart.

In theory, the learning index is equal across assets in equilibrium and investors are indifferent between learning about all assets. At first glance, the fact that the empirical learning index varies across assets may seem contradictory to the theoretical equilibrium outcome. In reality, learning is not an instantaneous and frictionless process. It may take time for investors to acquire and process information, update their expectations, and incorporate the information into prices. There may also be limits on the amount an investor can reduce uncertainty about a particular asset or limits on the aggregate learning capacity of all investors. While my empirical approach requires investors to follow the predicted information choices of the learning index to at least some extent, the measure is obviously

not a perfect descriptor of actual learning decisions. All of these factors would lead to the observation of cross-sectional dispersion in the empirical learning index at a single point in time or over short horizons. Importantly, however, this dispersion diminishes over longer horizons as illustrated by the patterns in the transition matrices. This outcome is consistent with the theoretical expectation that the value of learning about a particular asset declines as more investors learn about it.

Chapter 3

Learning and the cross-section of risk and return

3.1 Explaining the cross-section of expected returns

In this section, I investigate the ability of the learning index to predict future stock returns using portfolio sorting analyses and two-stage cross-sectional regressions.

3.1.1 Portfolio sorting

At the end of each month, stocks are sorted into quintiles based on LI. For each quintile-month, I calculate value-weighted and equal-weighted average portfolio returns in excess of the risk-free rate $\left(R_{p,t} - R_{f,t}\right)$ in the following month as well as the difference in average returns between the extreme quintiles $(5-1)$. Next, I calculate the time-series average return for each of the portfolios. I also measure risk-adjusted excess returns for each portfolio as the alpha (α) from a time-series regression of portfolio excess returns on nested versions of the six-factor model proposed by Fama and French (2018). The six-factor model includes the market $\left(R_{M,t} - R_{f,t}\right)$, size ($SMB$), and value ($HML$) factors of Fama and French (1993), profitability (RMW) and investment (CMA) factors of Fama and French (2015), and a momentum (UMD)

factor. Specifically, I estimate time-series regressions for each portfolio p using the six-factor model as well as the nested three-factor and five-factor specifications:

$$\begin{aligned} R_{p,t} - R_{f,t} = \alpha_p + \beta_{1,p}\left(R_{M,t} - R_{f,t}\right) + \beta_{2,p} SMB_t + \beta_{3,p} HML_t \\ + \beta_{4,p} RMW_t + \beta_{5,p} CMA_t + \beta_{6,p} UMD_t + \varepsilon_{p,t}. \end{aligned} \tag{3.1}$$

Table 4 presents average excess returns and risk-adjusted excess returns for value-weighted (Panel A) and equal-weighted (Panel B) portfolios. I report Newey and West (1987) t-statistics with a maximum lag order of 12 months to account for potential autocorrelation and heteroskedasticity. In Panel A, the highest LI quintile has an average excess return of 0.638% in the month following portfolio formation, while the lowest LI quintile has an average excess monthly return of 1.138%. The difference in excess returns between these quintiles is −0.500% per month (−6.2% per year) and is significant at the 1% level. These results indicate that expected returns are lower on average for high LI stocks compared to low LI stocks.

The next four columns report risk-adjusted returns estimated using various factor models. After controlling for exposure to market, size, and value risk factors, the value-weighted risk-adjusted return of each quintile is reduced by almost 1%. However, the risk-adjusted return of the $5-1$ portfolio remains economically and statistically significant: the monthly three-factor alpha spread is −0.495% with a t-statistic of −4.29. I find qualitatively similar results after adding the profitability, investment, and momentum factors. The five-factor (six-factor) alpha difference between the extreme LI quintiles is −0.672% (−0.517%) per month or −8.4% (−6.4%) per year. Each of these estimates is significant at the 1% level.

Table 4, Panel B reports results using the returns of equal-weighted portfolios. Quintile 5 has an average excess return of 0.825% and quintile 1 has an average excess return of 1.396% per month. The average monthly return of the $5-1$ portfolio is −0.571%. The average differences in three-factor, five-factor, and six-factor alphas between the extreme quintiles are −0.511%, −0.618%, and −0.551% per month (−6.3%, −7.7%, and −6.8% per

year), respectively.

Overall, the results of portfolio sorting indicate that high LI stocks tend to have lower future returns relative to low LI stocks. These results support the prediction that learning is associated with lower expected returns and risk-adjusted returns. The $5-1$ spreads in equal-weighted and value-weighted returns are economically and statistically significant, even after controlling for exposure to several sources of systematic risk. The return differences are not driven solely by stocks in any particular quintile. Rather, average returns and alphas tend to decrease monotonically as LI increases across quintiles. Throughout the remainder of the paper, I use the Fama and French (2018) six-factor model for risk adjustment and volatility decomposition, although conclusions based on alternative factor model specifications are qualitatively similar.

It is useful to distinguish between the $5-1$ portfolio formed based on values of LI in Table 4 and the hypothetical portfolio of an investor that chooses to learn information. The objective of the analysis in Table 4 is to identify whether there is a difference in expected returns between high LI and low LI stocks, not to evaluate the expected portfolio return of a learning investor. Suppose that an investor learns the most about high LI stocks and the least about low LI stocks. This information choice does not imply that she will take a long position in high LI stocks and a short position in LI stocks. Rather, her investment choice for each asset will depend on her posterior information set. The investor uses her information to buy the assets that she expects to have high payoffs and sell the assets that she expects to have low payoffs. Since learning more about an asset makes these expectations more accurate, the investor's expected portfolio return is increasing in her learning capacity. Therefore, while high LI assets have lower equilibrium expected returns compared to low LI assets, an individual investor who learns about these assets has a higher expected portfolio return compared to an uninformed investor.

3.1.2 Cross-sectional regressions

In this section, I use two-stage cross-sectional regressions to examine the relation between the learning index and expected returns while controlling for other determinants of returns. This approach is appropriate for cross-sectional analysis as it accounts for a time effect in the data (i.e., residuals in a given month are correlated across firms). In the first stage, I estimate monthly cross-sectional regressions of excess stock returns in month $t+1$ on values of LI and a set of ten control variables measured in month t. Of the ten stock characteristics used as controls, the first six are associated with exposure to one of the factors used for risk adjustment in the portfolio sorting analysis. Following the prior literature, I also control for the effects of illiquidity, short-term and long-term return reversals, and idiosyncratic volatility. The full cross-sectional model estimated at the end of each month is

$$\begin{aligned} R_{i,t+1} - R_{f,t+1} = \lambda_{0,t} + \lambda_{1,t} LI_{i,t} + \lambda_{2,t}\beta_{i,t}^{MKT} + \lambda_{3,t} SIZE_{i,t} + \lambda_{4,t} BM_{i,t} \\ + \lambda_{5,t} PROF_{i,t} + \lambda_{6,t} INV_{i,t} + \lambda_{7,t} MOM_{i,t} + \lambda_{8,t} ILLIQ_{i,t} \\ + \lambda_{9,t} STR_{i,t} + \lambda_{10,t} LTR_{i,t} + \lambda_{11,t} IVOL_{i,t} + \varepsilon_{i,t+1}. \end{aligned} \tag{3.2}$$

In the second stage, I calculate the time-series averages of the cross-sectional regression coefficient estimates. As an alternative approach to deal with potential errors-in-variables bias, I also compute precision-weighted time-series averages as in Litzenberger and Ramaswamy (1979), where the weights are inversely proportional to the standard error of the estimates from the first stage.

Table 5 reports equal-weighted average (Panel A) and precision-weighted average (Panel B) slope coefficients, Newey and West (1987) t-statistics in parentheses, and the average adjusted R^2 for each specification. I begin with a univariate regression of excess return on LI in Column 1. The average slope coefficient is -0.691 with a t-statistic of -4.38. Since values of LI range from zero to one, the reported univariate coefficient estimate for LI can be interpreted as the average return difference between the stock with the highest and lowest

value of LI in an average month. As a benchmark, I then estimate a regression of excess return on only the control variables in Column 2. Column 3 presents results from the full regression specification. After controlling for several stock characteristics, the magnitude of the coefficient on LI is slightly reduced (-0.416) relative to the univariate specification, but remains economically and statistically significant.

Panel B presents precision-weighted average slope coefficients from a similar set of three regressions. In this setting, I continue to find a negative and significant relation between LI and subsequent returns. In the univariate regression, the precision-weighted average coefficient on LI is -0.633 and is significant at the 1% level. Using the full multivariate specification in Column 6, the coefficient of interest is -0.405 with a t-statistic of -4.91. In untabulated analyses, I find that the results are robust to the inclusion of additional cross-sectional return predictors as controls, including return volatility, skewness, co-skewness, kurtosis, maximum daily return in the past month, share turnover, institutional ownership, number of institutional owners, number of analyst forecasts, the call-put option implied volatility spread, and the Stambaugh, Yu, and Yuan (2015) mispricing measure. These results reinforce the conclusion that learning is associated with a decrease in expected return, even after controlling for other return predictors and assigning more weight to more precise cross-sectional coefficient estimates.

With respect to the control variables, the signs of the coefficient estimates are generally in accordance with the findings of past studies. The significant precision-weighted average coefficient estimates in Panel B indicate that stocks with lower size, higher book-to-market ratios, higher profitability, lower investment, higher momentum, lower past short-term and long-term returns, and lower idiosyncratic volatility are all associated with higher expected returns. The coefficient estimates for market beta are insignificant in both panels. The precision-weighted average coefficient indicates a negative and significant relation between illiquidity and expected returns. While theory suggests a positive relation between these two variables, Bali, Engle, and Murray (2016) show that the empirical relation between illiquidity

and future stock returns becomes negative within stock samples that exclude extremely small or illiquid stocks.

Coefficient estimates reported in Table 5 can be combined with the cross-sectional summary statistics in Table 1 to get a sense of the relative economic importance of each of the explanatory variables. Based on the precision-weighted average coefficient estimates, current monthly returns (STR) carry the strongest explanatory power for next month returns. An increase of one cross-sectional standard deviation in STR results in a cross-sectional decrease in next month return of $10.00 \times 0.033 \approx 0.33\%$ on average, all else equal. The explanatory power of the learning index for next month returns is comparable to that of momentum, investment, firm size, and idiosyncratic volatility. Increases of one standard deviation in MOM, INV, $SIZE$, $IVOL$, and LI are associated with average cross-sectional differences in expected monthly return of 0.14%, -0.14%, -0.13%, -0.13%, and -0.12% respectively, holding all other variables constant.

In recent years, a number of papers have expressed concerns about data mining or p-hacking in the empirical asset pricing literature. Using a replicated set of 447 anomaly variables, Hou, Xue, and Zhang (2017) find that the explanatory power of 286 (64%) of anomaly variables become insignificant at the 5% level after controlling for microcaps (stocks below the 20^{th} NYSE percentile) and using value-weighted portfolios instead of equal-weighted portfolios. In order to deal with the bias introduced by multiple testing, Harvey, Liu, and Zhu (2016) suggest that researchers use a t-statistic of 3.0 as a hurdle for assessing the significance of a new anomaly. With these concerns in mind, I note that the estimates of interest from value-weighted portfolio sorting in Table 4 and multivariate regressions in Table 5 are based on a sample that excludes microcap stocks and have t-statistics exceeding the higher cutoff of 3.0.

3.2 Explaining the cross-section of volatility

In the context of the model by Van Nieuwerburgh and Veldkamp (2010), learning about an asset leads to a reduction in the posterior variance of the asset's payoff. In this section, I investigate the cross-sectional relationship between the learning index and return volatility.

3.2.1 Portfolio sorting

I first conduct a univariate portfolio sorting analysis using quintiles sorted on LI. Because return volatility is serially correlated, I use a measure of abnormal return volatility as the dependent variable in the sorting analysis. This measure can be viewed as a proxy for a stock's posterior variance relative to its prior variance. My objective in this section is to determine whether there is a difference in average abnormal volatility between the extreme LI quintiles. My expectation is that the abnormal volatility of high LI stocks should be lower on average compared to the abnormal volatility of low LI stocks. As is often the case in portfolio analyses, I am not directly interested in the level of the dependent variable (abnormal volatility) for any particular quintile over the sample period. The decision to learn about certain stocks does not imply that I should empirically observe negative abnormal volatility on average for these stocks. In the theoretical model where there is only one period and uncertainty only changes due to information acquisition, I would indeed expect assets that are learned about to experience a decrease in volatility relative to prior volatility levels. In reality, stock volatility may change over time for reasons unrelated to investor learning. The model does not directly lead to a time-series prediction about whether volatility is increasing or decreasing on average for any given quintile. It only provides a cross-sectional prediction regarding the comparison of abnormal volatility for stocks subject to a high degree of learning relative to that of stocks subject to a lower degree of learning.

I measure abnormal return volatility ($ARVOL$) as the difference between next month return volatility and average return volatility in the prior 12 months, scaled by average

return volatility in the prior 12 months and multiplied by 100. As the model predicts that learning also reduces the systematic component of risk, I also construct measures of abnormal systematic volatility and abnormal idiosyncratic volatility. $ASVOL$ is monthly systematic volatility divided by average monthly systematic volatility over the previous 12 months, minus one and multiplied by 100. Similarly, $AIVOL$ is monthly idiosyncratic volatility divided by average monthly idiosyncratic volatility over the previous 12 months, minus one and multiplied by 100.

I sort stocks based on LI into quintiles each month and examine the pattern in time-series means of portfolio average abnormal volatility across quintiles. Table 6 presents value-weighted (Panel A) and equal-weighted (Panel B) portfolio average abnormal volatility. In Panel A, abnormal return volatility is 3.975% lower on average for high LI stocks relative to low LI stocks. This difference is significant at the 1% level. The results in the next two columns suggest that the information choices of investors predict cross-sectional differences in both systematic and idiosyncratic volatility. On average, abnormal systematic (idiosyncratic) volatility in the month following portfolio formation is 4.639% (2.773%) lower for high LI stocks compared to low LI stocks, with a t-statistic of -5.79 (-5.13). I arrive at similar conclusions if abnormal volatility is weighted equally within each portfolio. On average, the differences in $ARVOL$, $ASVOL$, and $AIVOL$ between extreme equal-weighted portfolios is -3.595%, -4.123%, and -3.015%, respectively. Each of these estimates is significant at the 1% level.

Altogether, the results from these sorting analyses indicate that learning is associated with a cross-sectional reduction in both the firm-specific and systematic components of risk. The findings based on the systematic component of volatility do not necessarily imply that the choice to learn about a stock involves the discovery of market-wide or macroeconomic information. Rather, the results support the idea that learning news about a firm can reduce not only firm-specific uncertainty, but also uncertainty arising from co-movement with the market or other common risk factors. For robustness, I consider defining abnormal

volatility as the next month volatility relative to current month volatility or relative to average volatility over the prior 3 or 6 months. In addition, I consider using the standard deviation of monthly volatility as the denominator as well as calculating absolute differences instead of relative differences (i.e., using just the numerator) in volatility compared to the prior 1, 3, 6, or 12 months. I also estimate systematic and idiosyncratic volatility using alternative factor model specifications. Finally, instead of examining abnormal volatility, I use a bivariate portfolio sorting approach to examine the relationship between LI and the level of volatility in the following month while controlling for the past 12-month average volatility level. The results of this particular analysis are presented in Section 5.3. My conclusions are qualitatively similar under each of these robustness checks.

3.2.2 Cross-sectional regressions

Next, I use two-stage cross-sectional regressions to examine the cross-sectional relationships between the learning index and total return volatility, systematic volatility, and idiosyncratic volatility in a multivariate setting. In the first stage, I estimate monthly cross-sectional regressions of a measure of volatility in month $t+1$ on values of LI and a set of control variables. In the second stage, I calculate the time-series averages and Litzenberger and Ramaswamy (1979) precision-weighted time-series averages of the cross-sectional regression coefficient estimates. The full cross-sectional model estimated at the end of each month is

$$\begin{aligned} VOL_{i,t+1} = {} & \lambda_{0,t} + \lambda_{1,t} LI_{i,t} + \lambda_{2,t} ROE_{i,t} + \lambda_{3,t} ROEVOL_{i,t} + \lambda_{4,t} AGE_{i,t} \\ & + \lambda_{5,t} DIVD_{i,t} + \lambda_{6,t} LEV_{i,t} + \lambda_{7,t} INVPRC_{i,t} + \lambda_{8,t} SIZE_{i,t} \\ & + \lambda_{9,t} BM_{i,t} + \lambda_{10,t} MOM_{i,t} + \lambda_{11,t} STR_{i,t} + \lambda_{12,t} R_{i,t+1} \\ & + \sum_{j=0}^{11} \gamma_{j,t} VOL_{i,t-j} + \varepsilon_{i,t+1} \end{aligned} \tag{3.3}$$

where VOL is one of total return volatility ($RVOL$), systematic volatility ($SVOL$), or idiosyncratic volatility ($IVOL$). Pastor and Veronesi (2003) find that stock return volatility is higher for less profitable firms, firms with more volatile profitability, younger firms, and firms that do not pay dividends. Based on this, I include return on equity (ROE), the volatility of return on equity ($ROEVOL$), firm age (AGE), and a dividend dummy ($DIVD$) as controls. Prior studies also show that stock return volatility increases after stock prices fall due to a leverage effect (Christie (1982); Cheung and Ng (1992)), while Duffee (1995) documents a contemporaneous relation between return and volatility. As such, I include financial leverage (LEV), the inverse of stock price ($INVPRC$), and the stock return in the next month (R) as control variables. I also include $SIZE$, BM, MOM, and STR to account for the impact of well-known sources of risk. Finally, I control for 12 lagged monthly values of the respective volatility measure in all specifications since volatility is highly persistent over time. The coefficient estimates on lagged volatilities and the intercept term are not reported in the tables for brevity. Based on the availability of data for the explanatory variables, the sample period for this analysis is December 1974 to December 2016.

Table 7 reports the regression results for total return volatility. Equal-weighted coefficient averages are presented in Panel A. In the first column, I estimate monthly cross-sectional regressions of return volatility in the next month on LI while controlling for lagged monthly volatilities over the past year. With this specification, I find a negative relation between LI and volatility. The coefficient on LI is -1.369 and is significant at the 1% level. This finding supports the prediction that learning is associated with lower uncertainty in the cross-section. In Column 2, I estimate a regression of next month return volatility on only the control variables (including lagged volatility) as a benchmark. Consistent with Pastor and Veronesi (2003), firms with lower return on equity, firms with higher volatility of return on equity, younger firms, and non-dividend-paying firms are all associated with higher stock return volatility. In addition, I find a positive and significant contemporaneous relation between return and volatility as well as between the inverse price level and volatility.

Column 3 of Table 7 presents results from the full regression specification. After controlling for a number of characteristics known to have cross-sectional explanatory power for volatility, I continue to find a negative and significant relation between the learning index and volatility. The coefficient on LI in the full specification is -1.355 with a t-statistic of -7.22. This result suggests that, in the average month, the next month return volatility of the stock with the highest value of LI is 1.355 percentage points lower on average than the stock with the lowest value of LI, holding all other variables constant. Panel B of Table 7 presents precision-weighted averages of the cross-sectional coefficient estimates from three similar specifications. The coefficient estimates for LI in Panel B are comparable in magnitude and significance to those in Panel A. In Column 6, the coefficient on LI is -1.190 and is significant at the 1% level. To obtain an approximation of the relative impact of learning on return volatility, I compare these coefficient estimates to the sample average return volatility reported in Table 1. For the average stock in an average cross-section, an increase in the value of LI from zero to one (all else being equal) is associated with a cross-sectional difference in return volatility of $\frac{-1.355}{34.20} \approx -3.96\%$ based on the equal-weighted average LI coefficient, or $\frac{-1.190}{34.20} \approx -3.48\%$ based on the precision-weighted average LI coefficient.

Based on precision-weighted average coefficient estimates in Column 6, variation in the next month return, current month return, and stock price have the largest impact on return volatility in the following month. All else equal, increases of one cross-sectional standard deviation in R, STR, and $INVPRC$ are associated with average cross-sectional differences of 0.87, -0.86, and 0.68 percentage points in next month return volatility. The explanatory power of the learning index for future monthly return volatility is comparable to that of the book-to-market ratio, dividend dummy, and return on equity. Increases of one standard deviation in LI, BM, $DIVD$, and ROE correspond to cross-sectional decreases in return volatility in the following month of 0.35, 0.33, 0.30, and 0.21 percentage points on average, holding all other variables constant.

In the next two tables, I focus on explaining the systematic and idiosyncratic components

of return volatility using a similar cross-sectional multivariate analysis. Table 8 presents equal-weighted averages (Panel A) and precision-weighted averages (Panel B) of coefficient estimates from the systematic volatility regressions. In the first column, I regress $SVOL$ in the next month on LI while controlling for lagged monthly values of $SVOL$ over the past year. The results indicate a negative and significant cross-sectional relation between learning and systematic risk in the following month. The coefficient on LI is -1.052 and is significant at the 1% level. Column 2 reports estimates from a benchmark specification that includes lagged values of $SVOL$ and all explanatory variables besides LI. The coefficient estimates in these columns are consistent with those in Table 7 with respect to sign and significance, with a few exceptions. Firm size and momentum are not significantly related to total return volatility but are positively related to the systematic component of volatility in this specification.

The third column of Table 8 reports results from the regression of next month $SVOL$ on the full set of explanatory variables. After controlling for various stock characteristics associated with volatility, I find that the learning index continues to carry negative and significant explanatory power for cross-sectional variation in systematic volatility during the following month. In Column 3, the average coefficient estimate on LI is -0.793 (t-statistic $= -5.54$). To evaluate the impact of LI on a relative basis, I compare the LI coefficient estimates from the full specifications to the sample average value of $SVOL$ reported in Table 1. For the average stock in an average cross-section, an increase in the value of LI from zero to one holding all other variables constant is associated with a difference in $SVOL$ of $\frac{-0.793}{22.74} \approx -3.48\%$. In terms of economic significance, the explanatory power of LI for the systematic component of next month volatility is comparable to that of the dividend dummy and book-to-market ratio. I arrive at similar conclusions based on the precision-weighted coefficient averages reported in Panel B.

In Table 9, I repeat the cross-sectional regression analyses using idiosyncratic volatility as the dependent variable and lagged values of idiosyncratic volatility as controls. In Column

1, I regress $IVOL$ in the following month on LI in the current month and 12 lagged monthly values of $IVOL$. The equal-weighted average coefficient estimate on LI is -0.702 (t-statistic $= -3.83$). Column 2 reports equal-weighted average coefficient estimates from the benchmark specification. Consistent with the findings in the previous two tables, the control variables exhibit significant explanatory power for cross-sectional variation in next month idiosyncratic volatility. I find that firm size and momentum are negatively related to $IVOL$. Thus, it appears that combining the negative effects of these variables on $IVOL$ with their positive effects on $SVOL$ results in the insignificant relations with total volatility reported in Table 7.

After controlling for a number of other stock characteristics, I find that the explanatory power of LI for cross-sectional variation in $IVOL$ becomes stronger. The coefficient estimate from Column 3 indicates that, in the average month, the next month idiosyncratic volatility of the stock with the highest value of LI is 0.934 percentage points lower on average than the stock with the lowest value of LI, all else equal (t-statistic $= -6.92$). For the average stock in an average cross-section, this estimate corresponds to a cross-sectional decrease in idiosyncratic volatility of $\frac{0.934}{24.46} \approx 3.82\%$. In terms of economic significance, the cross-sectional explanatory power of LI for next month $IVOL$ is comparable to that of the dividend dummy and book-to-market ratio. The results based on precision-weighted coefficient averages in Panel B are qualitatively similar.

In total, the analyses in this section support the hypothesis that investor learning leads to a cross-sectional reduction in volatility. When combined with the findings in Section 3.1, the results suggest that this cross-sectional reduction in risk corresponds to a cross-sectional reduction in risk premium or expected return.

Chapter 4

Support for interpretation of the learning index

4.1 Contemporaneous impact of learning on price and volatility

The central hypothesis in this paper is that assets that are subject to a greater degree of learning have lower expected returns and risk on average in the cross-section. This conclusion has two important implications. First, holding future payoffs fixed, a lower expected return on average implies an increase in the current price on average. Second, while future volatility is expected to decline in the cross-section for assets subject to a greater degree of learning, the trading activity resulting from investor learning should correspond to an increase in volatility in the short run as prices adjust to reflect new information. In this section, I present analyses aimed at identifying these effects.

There are a number of challenges involved in using the learning index to identify the short run impacts of learning on prices and volatility. For example, because the empirical learning index is measured at a relatively low frequency, it is not feasible to conduct an event study analysis using this measure to pinpoint the exact timing of information flow and

price impact. Furthermore, the learning index only represents a prediction of information choices and does not rely on direct observation of actual learning activity. Thus, while it is theoretically possible to estimate the learning index at a daily or weekly frequency, there is no guarantee that this approach would precisely identify the timing of the learning activity. One potential approach would be to examine monthly returns and volatility during the same month that the learning index is calculated. This approach is also problematic due to the persistence in the learning index from one month to the next, particularly in the extreme quintiles (as shown in Table 3). If a given stock has a high learning index in the current month, there is a good chance that the stock had a high learning index value in the prior month. The empirical evidence in Section 3.1 would then suggest that this stock will have a lower return and lower abnormal volatility relative to the cross-section during the current month.

Given these challenges and the persistence in the learning index, I instead examine changes in this measure over 1-month and 3-month horizons in order to investigate the short run impact of learning. I expect that stocks with large (more positive) changes in LI relative to past values are more likely to experience large increases in information flow relative to past values. If this is true, then these stocks are more likely to exhibit a large increase in price on average at some point during this horizon to reflect increased investor demand for assets that are now less risky. Focusing on instances where the benefits of learning increase by a large amount in a short period of time improves my ability to identify contemporaneous price and volatility impact.

At the end of each month, I sort stocks into quintiles based on the change in LI over either a 1-month or 3-month horizon. The 1-month (3-month) change in LI is computed as the value of LI at the end of the portfolio formation month t minus the value of LI at the end of month $t-1$ ($t-3$). To measure short run price impact, I use the maximum daily ($MAXDRET$) or weekly ($MAXWRET$) return over the 1-month and 3-month horizon. To measure short run volatility impact, I use the maximum absolute daily ($MAX|DRET|$) or

weekly ($MAX|WRET|$) return over these horizons.

For reference, I first discuss average values of the sorting variable within each quintile (results are untabulated). Since the learning index is defined as a cross-sectional ranking from zero to one, average changes in LI exhibit a symmetric distribution centered around zero in the third quintile. The fourth quintile has an average 1-month (3-month) change in LI of approximately 0.07 (0.10), while the fifth quintile has an average 1-month (3-month) change in LI of approximately 0.21 (0.30). The first and second quintiles exhibit an almost identical pattern in average changes in LI as the fourth and fifth quintiles, except the average changes are negative rather than positive.

Panel A of Table 10 presents value-weighted average maximum returns for each quintile of changes in LI. On average, stocks with the largest (most positive) change in LI compared to the prior month have an average maximum daily return during the month of 3.943%. In comparison, stocks with the smallest (most negative) change in LI compared to the prior month have a value-weighted average maximum daily return of 3.676%. The average difference in $MAXDRET$ between the extreme quintiles based on $LI_t - LI_{t-1}$ is 0.267% (t-statistic = 7.90). Thus, when a stock experiences a large upward movement in the learning index over a particular month, it is more likely to have experienced a larger maximum daily return during that month. When examining changes in LI and maximum returns over a 3-month horizon, the average values of $MAXDRET$ within each quintile are all larger than their counterparts based on a 1-month horizon, but the spread in average values of $MAXDRET$ between extreme quintiles is similar: 0.264% with a t-statistic of 3.65. I arrive at similar conclusions based on maximum weekly returns. Over a 1-month (3-month) horizon, the average value of $MAXWRET$ for the $5-1$ portfolio is 0.490% (0.403%). Each of these estimates is significant at the 1% level. My conclusions are also qualitatively similar when analyzing equal-weighted portfolios (results reported in Panel B).

In Table 11, I examine the contemporaneous relationship between changes in LI and short run volatility as measured by the maximum absolute daily or weekly return during

the specified horizon. Results based on value-weighted portfolios are presented in Panel A. Stocks with the largest (smallest) 1-month change in LI have an average maximum absolute daily return of 4.600% (4.155%) and an average maximum absolute weekly return of 6.419% (5.581%). The average difference in $MAX|DRET|$ between extreme value-weighted quintiles is 0.446%, and the average difference in $MAX|WRET|$ is 0.838% (t-statistics of 10.35 and 13.88, respectively). In the lower half of the table, I analyze changes in LI and maximum absolute returns during 3-month horizons. Similar to the results in the previous table, the average short run volatility for each quintile is larger when measured over the longer horizon, but the spread between the highest and lowest quintiles remains consistent.

The findings in this section indicate that stocks with the largest increase in the learning index over a 1-month or 3-month horizon tend to have the largest maximum daily or weekly return as well as the most extreme (positive or negative) daily or weekly return during this period. As a final test, I re-perform the analyses in this section using levels of LI rather than changes in LI (untabulated). Using this approach, I fail to find a robust relationship between LI and maximum daily/weekly returns or maximum absolute daily/weekly returns during the portfolio formation month. This result illustrates the difficulty in using levels of LI to examine short run price and volatility impact, and it also highlights the advantage of focusing on large changes in LI to identify these effects. In addition, this result serves as evidence that the learning index is not merely capturing the effects of lottery demand identified by Bali, Cakici, and Whitelaw (2011).

4.2 Long-term predictability

In this section, I examine the cross-sectional explanatory power of the learning index for subsequent months up to three years. To the extent that the learning index reflects investors learning fundamental information and incorporating this information into prices, I expect that prices move toward fundamental value and do not reverse in the long run. Alternatively,

if the explanatory power of the learning index derives from temporary price movements away from intrinsic value, I expect this mispricing to be eventually corrected over time.

At the end of each month t, I sort stocks into quintiles based on LI and track the difference in average returns between the highest LI quintile and the lowest LI quintile $(5-1)$ in each of the 36 months after portfolio formation. Figure 2 presents average monthly returns for the spread portfolio. The average return of the value-weighted $5-1$ portfolio is most negative in the month immediately following portfolio formation and subsequently moves toward zero. By month $t+5$, the negative average return spread is no longer significant at the 10% level. The spread based on equal-weighted portfolios remains significant until month $t+7$.

Figure 3 presents the $LI5-LI1$ portfolio risk-adjusted average return in each of the next 36 months. On a risk-adjusted basis, the return spread between the highest and lowest LI quintiles is negative and significant until month $t+8$ using either value-weighted or equal-weighted portfolios. With one exception, the risk-adjusted returns of both the value-weighted and equal-weighted spread portfolio are not statistically different from zero in any of the subsequent months. The results indicate that the explanatory power of LI for returns continues in a declining manner over a period of several months. Furthermore, after adjusting for co-movement with systematic risk factors, the monthly return differences between extreme LI quintiles are not reversed over the subsequent three year period. This finding supports the notion that the cross-sectional explanatory power of LI for returns reflects the effects of prices moving closer to (rather than further from) their fundamental values as investors learn and trade upon new information.

Jegadeesh and Titman (1993) find that portfolios sorted on past performance exhibit momentum during the next 12 months, followed by reversal over the subsequent one to two years. To ensure that my findings are not attributable to this effect, I repeat this long-term predictability analysis using portfolios sorted based on MOM. Consistent with the findings of Jegadeesh and Titman (1993), there is evidence of short-term momentum followed by a longer-term reversal using the raw returns of either equal-weighted or value-

weighted portfolios. As expected, controlling for exposure to the momentum factor completely eliminates these patterns. The fact that the learning index has explanatory power for long run returns even after controlling for the momentum factor suggests that LI is not merely capturing cross-sectional variation in momentum returns.

Next, I repeat the portfolio sorting analysis and track the difference in value-weighted and equal-weighted average abnormal volatility between the extreme LI quintiles over the subsequent 36 months. In Figure 4, the average spread in $ARVOL$ is negative and significant for seven months after portfolio formation using value-weighted portfolios and eight months after portfolio formation using equal-weighted portfolios. Beyond this point, all values of $ARVOL$ are not statistically different from zero. This result suggests that the cross-sectional relation between the learning index and risk is not attributable to temporary decreases in volatility.

When I decompose abnormal volatility into systematic and idiosyncratic components, I find that the two volatility components exhibit different patterns over the long run. Figure 5 shows that the differences in abnormal systematic volatility predicted by LI tend to reverse to a certain extent over the long run. The spread in average $ASVOL$ is negative and significant until month $t+5$ ($t+7$) for value-weighted (equal-weighted) portfolios, but turns positive and significant in some of the subsequent months. On the other hand, average values of $AIVOL$ for the $5-1$ portfolio presented in Figure 6 are negative and significant until month $t+13$ ($t+11$) for value-weighted (equal-weighted) portfolios and are not statistically different from zero for the next two years.

The findings based on long-term volatility predictability suggest that the effects of learning (as measured by the learning index) are more permanent for the idiosyncratic component of risk than for the systematic component of risk. While learning appears to reduce return co-movement with systematic risk factors over the short run, this effect is partially reversed over time. Combined with the results on long-term return predictability, the patterns in long-term volatility predictability suggest that the observed reversal in raw returns in Figure

2 is associated with the reversal in systematic risk. I find no evidence of a reversal in risk-adjusted returns and idiosyncratic risk. In untabulated analyses, I arrive at similar conclusions using alternative factor model specifications for risk adjustment and volatility decomposition. In aggregate, the results in this section support the interpretation of the learning index by demonstrating that the cross-sectional differences in risk and risk-adjusted returns predicted by this measure are generally long-lasting.

4.3 Relationship with measures of information flow

In this section, I examine the cross-sectional relation between LI and a number of proxies for investor attention or information demand. I consider measures related to trading activity, analyst coverage, forecast revision and accuracy, SEC filing download activity on EDGAR, and Bloomberg news reading activity. In practice, information acquisition efforts are likely to be constrained by the fact that smaller firms may be less visible to investors, less informationally transparent, or may have less information available for acquisition. As such, for this analysis I use a bivariate dependent portfolio sorting approach based on size and LI. This approach allows me to investigate the outcomes of differences in information choices across firms while controlling for the impact of firm visibility, informational transparency, or the amount of acquirable information (as captured by firm size).

At the end of each month, I sort stocks into quintiles based on firm size. Then, within each size quintile, I sort stocks based on LI. Each LI subquintile is combined across size quintiles into a single quintile. This procedure creates portfolios of stocks with differences in LI but similar distributions of size. To verify that my prior conclusions regarding the explanatory power of LI for risk and return are robust to this bivariate sorting approach, I use the same approach to examine patterns in returns and abnormal volatility across LI quintiles while controlling for firm size. The results from these untabulated analyses are qualitatively similar to those presented in Table 4 and Table 6.

Table 12 reports portfolio average values of six different proxies of information flow as well as the respective sample period over which each analysis is performed. The first proxy is abnormal trading activity. According to Barber and Odean (2007), trading activity is likely to increase as investors learn new information about a firm. I first measure trading activity as monthly share turnover, or the total number of shares traded within a month divided by shares outstanding. I then measure abnormal turnover ($ATURN$) as turnover during the current month divided by average monthly turnover over the previous 12 months, minus one and multiplied by 100. Data for this variable are available from CRSP for the full sample period (July 1964 to December 2016). On average, high (low) LI stocks experience a 8.547% (3.676%) increase in monthly share turnover relative to average monthly turnover during the past year. The difference in abnormal turnover between the extreme quintiles is 4.871% with a t-statistic of 5.78. This difference suggests a greater level of abnormal trading activity among stocks expected to be subject to a greater degree of investor learning.

The next three proxies relate to analyst coverage, forecast revisions, and forecast accuracy. I expect greater analyst coverage to be associated with an increase in the information available about a firm. Consistent with this idea, Hong, Lim, and Stein (2000) use analyst coverage as a measure of the rate of information flow. The arrival of new information about a firm should also correspond to a revision of analysts' expectations and more accurate forecasts. Zhang (2008) finds that timely analyst forecast revisions improve market efficiency. Harford, Jiang, Wang, and Xie (2018) show that greater effort by analysts in acquiring information is associated with more frequent forecast revisions and more accurate forecasts. Beginning in July 1984, I measure analyst coverage ($nFCST$) each month as the number of analyst forecasts of earnings per share (EPS) recorded by I/B/E/S for the nearest fiscal quarter. I also measure the number of analyst forecast revisions since the last month ($nREV$). In addition, I construct a measure of the change in forecast accuracy (ΔFA) from one month to the next. First, I calculate the error in the mean forecast for the nearest fiscal quarter as the absolute value of the difference between the mean EPS forecast and the actual EPS as a percentage

of the actual EPS. I subtract this error from one to measure forecast accuracy. Finally, I compute the monthly change in forecast accuracy (ΔFA) as forecast accuracy in the current month minus forecast accuracy in the prior month, multiplied by 100. Larger values of (ΔFA) represent increases in forecast accuracy. This variable is computed by firm within a given forecast period so that forecast errors are not compared across different forecast periods.

The evidence indicates a positive association between the learning index and analysts' decisions to follow firms and update forecasts. After controlling for the effects of size, stocks with the highest (lowest) values of LI are covered by an average of 8.621 (7.483) analysts. The difference in coverage is approximately one analyst with a t-statistic of 6.76. On average, 2.505 (2.076) analysts covering a high (low) LI stock revise their forecasts from the prior month. The average difference in the number of forecast revisions is 0.428 and is significant at the 1% level. In an average monthly cross-section in my sample, the mean (median) firm has 8.05 (6.70) analyst forecasts and 2.31 (1.30) forecast revisions. As such, the estimates of cross-sectional spreads in analyst coverage and forecast revisions between extreme LI quintiles are also economically significant.

In Column 4 of Table 12, I investigate the relationship between the learning index and changes in forecast accuracy. For all LI quintiles, the average monthly percentage change in forecast accuracy is positive. This pattern implies that on average, the mean forecast estimate become more accurate (relative to the actual realized value) as the fiscal quarter end approaches. Stocks in the highest (lowest) LI quintile have an average improvement in forecast accuracy of 4.319% (2.294%). The difference in ΔFA between the extreme LI quintiles is 2.025% on average (t-statistic = 4.63). Therefore, while the EPS forecasts for all stocks in the sample tend to move closer on average to the actual realized EPS, the monthly increase in accuracy is greater for stocks with higher values of the learning index. In untabulated analyses, I find qualitatively similar results when I compute forecast errors using the median rather than the mean analyst forecast.

The fifth proxy is based on downloads of company filings from the SEC EDGAR database.

Although the data are available beginning in January 2003, there are significant known issues with the data due to lost or corrupted log files prior to March 2003 and between September 24, 2005 and May 10, 2006. As such, I begin my analysis in March 2003 and drop months with partial coverage from my sample for this analysis. Using this data, Crane et al. (2018) provide evidence on the value of this information by showing that hedge fund usage of publicly-available SEC filings predicts fund performance. Following the methodology of Ryans (2017) to screen out algorithmic download activity, I measure $EDGAR$ as the number of human downloads of a company's SEC filings during the month.[1] After controlling for the size of the firm, I find that the filings of firms with the highest (lowest) values of LI are downloaded approximately 872 (742) times within a month on average. The difference in average EDGAR downloads between these quintiles is approximately 130 with a t-statistic of 3.54. This result supports the notion that investors are more likely to gather information for stocks with higher values of the learning index.

The sixth proxy is based on a measure of Bloomberg news reading activity proposed by Ben-Rephael et al. (2017). Bloomberg provides a variable called "News Heat - Daily Max Readership" that measures readership interest in a company relative to the past 30 days. The variable ranges from 0 to 4, with 0 indicating relatively low interest and 4 indicating unusually high interest. The data for this variable are available beginning February 17, 2010, although historical data are missing for periods between December 2010 and January 2011 as well as between August 2011 and November 2011. As such, I begin my analysis in March 2010 and drop any months with partial coverage from my sample for this analysis. Following Ben-Rephael et al. (2017), I measure abnormal attention at the daily frequency using a dummy variable that is equal to 1 if the Bloomberg daily maximum is a 3 or 4, and 0 otherwise. I then aggregate this measure to the monthly frequency by computing the total number of days with abnormal attention within a month (BBG). After controlling for size, high (low) LI stocks receive abnormal investor attention during 3.166 (2.805) days within

[1] I obtain summarized EDGAR log file data from James Ryans' website: http://www.jamesryans.com/.

a month on average. The difference in abnormal attention days between high and low LI stocks is 0.361 with a t-statistic of 5.63. Overall, the patterns documented in Table 12 serve as supporting evidence of a relationship between the learning index and information flow.

4.3.1 Learning index and analyst coverage

In this section, I use multivariate regressions to further examine the relationship between the learning index and analyst coverage. I consider a set of control variables that is similar to those used by Hong et al. (2000) in their analysis of the cross-sectional determinants of analyst coverage: firm size ($SIZE$), a dummy variable equal to one for stocks traded on the NASDAQ Stock Exchange ($NASDAQ$), book-to-market ratio (BM), market beta (β^{MKT}), inverse of stock price ($INVPRC$), return volatility ($RVOL$), momentum (MOM), average monthly turnover over the past 12 months ($\overline{TURN}$), and an interaction term between $NASDAQ$ and $\overline{TURN}$. For certain specifications, I also include industry dummy variables based on two-digit Standard Industrial Classification (SIC) code.[2] The full cross-sectional model estimated at the end of each month is

$$\begin{aligned}\ln(1+nFCST)_{i,t} = {} & \lambda_{0,t} + \lambda_{1,t}LI_{i,t} + \lambda_{2,t}SIZE_{i,t} + \lambda_{3,t}NASDAQ_{i,t} + \lambda_{4,t}BM_{i,t} \\ & + \lambda_{5,t}\beta^{MKT}_{i,t} + \lambda_{6,t}INVPRC_{i,t} + \lambda_{7,t}RVOL_{i,t} + \lambda_{8,t}MOM_{i,t} \\ & + \lambda_{9,t}\overline{TURN}_{i,t} + \lambda_{10,t}\left(NASDAQ^{*}\overline{TURN}_{i,t}\right) \\ & + \sum_{j=1}^{15}\gamma_{j,t}INDUSTRY_{j,t} + \varepsilon_{i,t}\end{aligned} \tag{4.1}$$

where the dependent variable is the natural logarithm of one plus the number of analyst forecasts of quarterly earnings during the month. The sample period for this analysis begins in July 1984 and ends in December 2016.

Table 13, Panel A reports equal-weighted coefficient averages from monthly cross-sectional

[2]Following Hong et al. (2000), the industry dummy variables correspond to the following groups of two-digit SIC codes: 01 – 09; 10 – 14; 15 – 19; 20 – 21; 22 – 23; 24 – 27; 28 – 32; 33 – 34; 35 – 39; 40 – 48; 49; 50 – 52; 53 – 59; 60 – 69; and 70 – 79.

regressions of log number of analyst forecasts. In Column 1, I include only the learning index in a univariate specification. Contrary to the expected relation, the average coefficient estimate is negative and significant. This estimate implies that, when no other variables are controlled for, stocks with higher values of the learning index tend to be covered by a lower number of analysts. In Column 2, I investigate how the relationship between analyst coverage and the learning index changes after controlling for firm size. Using this specification, I find that firm size and analyst coverage are positively related. More importantly, I find that the coefficient on LI is positive and significant. This result suggests that the combination of a negative relation between LI and firm size and a positive relation between firm size and analyst coverage leads to a negative omitted variable bias on the LI coefficient in Column 1. It also demonstrates the importance of using bivariate portfolio sorting in Table 12 to control for firm size. After controlling for firm size, the average coefficient on LI is 0.175 with a t-statistic of 8.23.

Column 3 of Table 13 reports results from a specification which includes firm size, a NASDAQ dummy, book-to-market ratio, market beta, inverse stock price, return volatility, momentum, average monthly turnover, and a NASDAQ-turnover interaction term. I find that firm size is the strongest predictor of analyst coverage, followed by average monthly turnover. After controlling for these additional variables, the coefficient on LI is 0.119 (t-statistic = 10.84). In Column 4, I include a set of dummy variables to control for differences in analyst coverage across industries. With this specification, the average coefficient on LI remains positive and significant at the 1% level. Based on these estimates, there is a $e^{0.094} - 1 \approx 9.9\%$ difference in the number of analyst forecasts of quarterly earnings between the highest and lowest LI stocks in a given month on average, holding all other variables constant. Conclusions based on precision-weighted coefficient averages in Panel B are qualitatively similar. The results of this analysis are consistent with the notion that higher values of the learning index are associated with greater information flow as evidenced by increased analyst coverage.

4.3.2 Learning index and EDGAR downloads

In this section, I use multivariate regressions to investigate the relationship between the learning index and EDGAR downloads. I use a number of control variables that are considered in Li and Sun (2017): firm size ($SIZE$), the natural logarithm of one plus the number of analyst forecasts ($\ln(1+nFCST)$), average monthly turnover over the past 12 months ($\overline{TURN}$), idiosyncratic volatility ($IVOL$), momentum (MOM), book-to-market ratio (BM), institutional ownership (IO), a dummy variable equal to one for stocks included in the S&P 500 index ($SP500$), and a dummy variable equal to one if the firm announces quarterly earnings during the month (EAM). The full cross-sectional model estimated at the end of each month is

$$\begin{aligned}\ln(1+EDGAR)_{i,t} = \lambda_{0,t} + \lambda_{1,t}LI_{i,t} + \lambda_{2,t}SIZE_{i,t} + \lambda_{3,t}\ln(1+nFCST)_{i,t} \\ + \lambda_{4,t}\overline{TURN}_{i,t} + \lambda_{5,t}IVOL_{i,t} + \lambda_{6,t}MOM_{i,t} + \lambda_{7,t}BM_{i,t} \\ + \lambda_{8,t}IO_{i,t} + \lambda_{9,t}SP500_{i,t} + \lambda_{10,t}EAM_{i,t} + \varepsilon_{i,t}\end{aligned} \tag{4.2}$$

where the dependent variable is the natural logarithm of one plus the number of EDGAR downloads during the month. The sample period for this analysis is March 2003 to December 2016.

Panel A of Table 14 presents equal-weighted coefficient averages from monthly cross-sectional regressions of log number of EDGAR downloads. In the first column, I include only the learning index in a univariate specification. The average coefficient on LI is -0.105, indicating that higher values of the learning index are associated with a lower number of downloads during the month without controlling for any other variables. In the second column, I add firm size as a control variable. Using this specification, I find a positive and significant relation between firm size and the number of EDGAR downloads, indicating that investors tend to acquire more information via EDGAR for larger firms. After controlling for size, the coefficient on the learning index is 0.181 (t-statistic = 3.56). Similarly to the

previous section, the results in the first two columns of the table suggest the presence of a negative omitted variable bias on the LI coefficient in Column 1.

In the third column of Table 14, I include the eight aforementioned control variables. Stocks with higher share turnover, higher idiosyncratic volatility, lower past returns, and higher book-to-market ratios all exhibit higher levels of EDGAR download activity. I find a positive but insignificant relation on average between IO and the number of EDGAR downloads (this relation is significant at the 1% level when computing precision-weighted coefficient averages). The results also suggest a higher number of downloads for stocks in the S&P 500 and for stocks with earnings announcements during the month. Firm size carries the strongest explanatory power for EDGAR download activity out of all independent variables considered.

After controlling for these additional variables, the coefficient on LI is 0.137 and is significant at the 1% level. This estimate indicates that, in an average month, there is a $e^{0.137} - 1 \approx 14.7\%$ difference in the number of EDGAR downloads between the highest and lowest LI stocks on average, holding all other variables constant. I arrive at similar conclusions using precision-weighted coefficient averages in Panel B. Thus, the evidence of a positive relation between LI and EDGAR download activity in a multivariate setting provides further support for the interpretation of the learning index as a proxy for information flow.

4.4 Learning prior to earnings announcement month

In this section, I examine the relationship between the learning index and the information environment prior to quarterly earnings announcements. If a stock has a high learning index in the month prior to an earnings announcement, I expect that investors are learning more about a firm and incorporating information into prices before the announcement. If this is true, then the average market reactions to earnings announcements of stocks with high lagged values of LI should be smaller in magnitude than those of low LI stocks. I also expect

to observe a higher level of abnormal trading activity in the month prior to the earnings announcement as investors trade upon their information.

I measure the market reaction to earnings announcements using the magnitude of cumulative abnormal returns. Absolute returns can also be interpreted as a simple measure of volatility. Daily abnormal returns are calculated as the difference between the daily stock return and the daily return on a portfolio of firms matched on size and book-to-market ratio. I measure the absolute value of the cumulative abnormal return on the day of the announcement $\left(|CAR|_d\right)$ and during a three-day window around the announcement $\left(|CAR|_{d-1,d+1}\right)$. I also examine the magnitude of the post-earnings announcement drift, measured as the absolute value of the cumulative abnormal return during the period from two days after the announcement through one day after the following quarterly announcement $\left(|CAR|_{q+1}\right)$. In addition to these measures of market reaction, I use abnormal monthly turnover in the month prior to the announcement ($ATURN_{t-1}$) to capture abnormal levels of trading activity.

As in the previous section, I use a bivariate dependent sorting approach to control for the effects of firm size. At the end of each month, all stocks with a quarterly earnings announcement during the month are sorted into quintiles based on lagged market capitalization, and then based on lagged values of LI within each size quintile. Each LI subquintile is then combined across the size quintiles. Due to data availability, the sample period for this analysis is October 1971 to December 2016.

Table 15 reports average values and associated t-statistics of the market reaction and trading activity for each quintile. After controlling for firm size, stocks with high values of LI in the prior month tend to have smaller market reactions to quarterly earnings announcements. On average, the $|CAR|$ on the event date is 0.119% smaller for high lagged LI stocks compared to low lagged LI stocks. Over a three-day window, the difference in market reaction between the extreme LI quintiles is -0.214%. These estimates are significant at the 5% and 1% level, respectively. The results also suggest that the absolute magnitude of the

drift in abnormal returns over the quarter following the earnings announcement tends to be smaller for stocks with high lagged values of LI. The average spread in $|CAR|_{q+1}$ between the high and low lagged LI quintiles is -0.338% (t-statistic = -1.65).

The last column in Table 15 indicates a higher degree of abnormal trading activity during the month prior to an earnings announcement for stocks with high lagged LI relative to low lagged LI stocks. On average, high (low) lagged LI stocks experience a 1.838% (-0.409%) change in share turnover during the month prior to a quarterly earnings announcement relative to all months in the past year. The difference in abnormal monthly turnover between the extreme quintiles is 2.247% with a t-statistic of 2.45.

In sum, the evidence in this section indicates that stocks with higher values of LI in the month prior to an earnings announcement tend to have smaller abnormal market reactions to the announcement. These stocks also exhibit a greater level of abnormal trading activity in the month before the earnings announcement. The findings support the idea that the learning index reflects learning decisions and the flow of information prior to earnings announcements.

4.4.1 Variation in earnings announcement activity

In this section, I examine how the conclusions from the prior section vary with the level of earnings announcement activity during the month. If there are fewer earnings announcements during the month, investors' learning efforts may be more concentrated among these firms. Conversely, during months when many firms are reporting earnings, the learning efforts of investors may be spread out over many firms, and the predictions of the learning index may be less powerful. In order to test this hypothesis, I count the number of earnings announcements in each month and split the sample period into two groups: months with high earnings announcement activity (above the yearly median) and months with low earnings announcement activity (below the yearly median). The rest of the testing procedure remains the same as in the prior section.

For each of the two sample period groups, Table 16 reports time-series means of quintile averages of market reactions and abnormal trading activity. Results for months with high activity are presented in Panel A, and results for months with low activity are presented in Panel B. In Panel A, the differences in the three measures of market reaction between extreme lagged LI quintiles are all negative but insignificant. In Panel B, each of these coefficients are negative and significant. During months with relatively low earnings announcement activity, the differences in $|CAR|_d$, $|CAR|_{d-1,d+1}$, and $|CAR|_{q+1}$ between the highest and lowest lagged LI quintiles are -0.169%, -0.295%, and -0.608%, with t-statistics of -2.32, -2.83, and -2.19, respectively.

The fourth column of Table 16 presents differences in abnormal trading activity in the month prior to an earnings announcement. The difference in $ATURN$ between extreme lagged LI quintiles is 2.117% during high activity months and 2.377% during low activity months, but only the first of these two estimates is significant. The evidence from this analysis suggests that the ability of the learning index to predict information flow prior to earnings announcements is stronger when learning is concentrated among a lower number of firms.

4.5 Learning during earnings announcement month

In this section, I examine the relationship between the learning index and the information environment during the earnings announcement month. The analyses in this section are similar to those in the previous section but feature one methodological difference that corresponds to a significantly different prediction. Here, I consider values of the learning index during the earnings announcement month instead of prior to the announcement month. If a stock has a high learning index in the month of an earnings announcement, I expect that a high level of information flow occurred during that month, likely due to the earnings announcement disclosure. If this is true, then the average market reactions to

earnings announcements of stocks with high contemporaneous values of LI should be larger in magnitude than those of low LI stocks. To the extent that the information flow results from the earnings announcement, I also expect a higher level of abnormal trading activity surrounding the earnings announcement date.

To test this hypothesis, I use the same bivariate dependent sorting approach controlling for firm size as before, but I use contemporaneous values of the learning index rather than lagged values. At the end of each month, all stocks with a quarterly earnings announcement during the month are sorted into quintiles based on firm size and then based on contemporaneous values of LI within each size quintile. Each LI subquintile is then combined across the size quintiles. I use the same three measures of market reaction ($|CAR|_d$, $|CAR|_{d-1,d+1}$, and $|CAR|_{q+1}$) as before. Similar to Lerman, Livnat, and Mendenhall (2008), I measure abnormal trading activity around the earnings announcement date ($ATURN_{d-1,d+1}$) as average daily turnover during the three-day period around the announcement date d divided by average daily turnover during a non-event period of days $d-63$ through $d-8$, minus one and multiplied by 100.

Table 17 reports average market reaction and abnormal trading activity for each LI quintile. After controlling for firm size, stocks with high values of LI in the earnings announcement month tend to have larger market reactions to these informational events. On average, the $|CAR|$ on the event date is 0.139% larger for high LI stocks compared to low LI stocks. Over a three-day window, the difference between the extreme LI quintiles is 0.349%. These estimates are each significant at the 1% level. In the third column, I find that a greater level of information flow during the earnings announcement month is followed by a smaller return drift in absolute value over the subsequent quarter. Thus, for high LI stocks, a greater degree of investor learning occurring around the earnings announcement translates into a smaller post-earnings announcement drift. The average spread in $|CAR|_{q+1}$ between the high and low LI quintiles is -0.381% (t-statistic = -1.68).

The next column in Table 17 indicates a higher degree of abnormal trading activity

during the three-day period surrounding an earnings announcement for high LI stocks relative to low LI stocks. On average, high (low) LI stocks experience a 66.748% (52.331%) increase in daily share turnover around the earnings announcement relative to the prior non-event period. The difference in abnormal daily turnover between the extreme quintiles is 14.417% with a t-statistic of 8.95. Overall, I find that higher values of LI during the earnings announcement month are associated with greater levels of information flow during that month. The evidence suggests that this increased information flow occurs around the earnings announcement date. Stocks with high contemporaneous values of LI tend to have larger market reactions and greater abnormal trading activity during the three-day period around the announcement. These stocks also have a smaller subsequent post-earnings announcement drift.

4.5.1 Variation in earnings announcement activity

In this section, I examine how the conclusions from the previous section vary with the level of earnings announcement activity during the month. The logic for these tests is similar to that in Section 4.4.1. When there are fewer earnings announcements during the month, investors' learning efforts may be more concentrated among fewer firms. If this is true, I expect the results from the previous tests in Table 17 to be stronger when there is less announcement activity. To test this hypothesis, I separately analyze months with high earnings announcement activity (above the yearly median) and months with low earnings announcement activity (below the yearly median). The rest of the testing procedure remains the same.

For each of the two sample period groups, Table 18 reports time-series means of quintile averages of market reactions and abnormal trading activity around the earnings announcement date. Panel A (Panel B) presents results for months with high (low) earnings announcement activity. In Panel A, the differences in $|CAR|_d$ and $|CAR|_{d-1,d+1}$ between extreme LI quintiles are 0.116% and 0.330%, with t-statistics of 1.98 and 2.84. In Panel

B, the differences in $|CAR|_d$ and $|CAR|_{d-1,d+1}$ between extreme LI quintiles are 0.163% and 0.369%, with t-statistics of 2.54 and 3.48. In the third column of the table, I find that the spread in $|CAR|_{q+1}$ between high and low LI quintiles is -0.745 (t-statistic = -2.41) on average during low activity months. In contrast, this spread is not significantly different from zero during high activity months.

In the last column of Table 18, I find that the difference in abnormal daily turnover around the earnings announcement date between extreme LI quintiles is larger during months with fewer announcements. The difference in ($ATURN_{d-1,d+1}$) between the highest and lowest LI quintiles is 12.235% during high activity months and 16.607% during low activity months. Each estimate is significant at the 1% level. The results from these tests suggest that the predictive power of the learning index around earnings announcements is dependent on the level of announcement activity. When there are fewer firms releasing earnings during a month, the difference in market reactions and abnormal trading activity between high and low LI stocks is larger. This effect corresponds to a greater reduction in post-earnings announcement drift for high LI stocks compared to low LI stocks.

4.5.2 Changes in the learning index

In this section, I continue my examination of the relationship between the learning index and the information environment during the earnings announcement month, focusing now on changes in the learning index relative to the previous month. Given the persistence in the learning index from one month to the next, stocks with high values of the learning index in the earnings announcement month are likely to have had high values in the previous month. If the learning index is representative of information flow, then I expect market reactions to earnings announcements to be largest for those stocks with low lagged values of the learning index and high contemporaneous values of the learning index. To test this hypothesis, I use bivariate dependent sorting based on the change in the learning index during the month ($LI_t - LI_{t-1}$) while controlling for firm size. At the end of each month, all

stocks with a quarterly earnings announcement during the month are sorted into quintiles based on firm size and then based on $LI_t - LI_{t-1}$ within each size quintile. Each $LI_t - LI_{t-1}$ subquintile is then combined across the size quintiles.

Table 19 presents quintile average market reactions and abnormal trading activity around the announcement date. After controlling for firm size, I find that stocks with larger increases in the learning index tend to have larger market reactions to earnings announcements occurring during the month. On average, the $|CAR|$ on the event date is 0.428% larger for high $LI_t - LI_{t-1}$ stocks compared to low $LI_t - LI_{t-1}$ stocks (t-statistic = 10.55). Over a three-day window around the announcement date, the difference in average $|CAR|$ between extreme quintiles is 0.885%, with a t-statistic of 12.64.

In the third column of Table 19, I find a U-shaped pattern in average values of $|CAR|_{q+1}$ across $LI_t - LI_{t-1}$ quintiles. The difference in average $|CAR|_{q+1}$ between extreme quintiles is not significantly different from zero. Given that the extreme $LI_t - LI_{t-1}$ quintiles are more likely to contain stocks with a higher learning index in either the previous month or the current month, the U-shaped pattern is consistent with the results in Table 15 and Table 17 and indicates that stocks with high lagged LI or high contemporaneous LI tend to exhibit smaller post-earnings announcement drifts.

The last column of Table 19 presents average values of abnormal daily turnover around the earnings announcement date. After controlling for size, I find that the difference in $ATURN_{d-1,d+1}$ between extreme quintiles is 20.795% (t-statistic = 12.04). Thus, the evidence in this table supports the idea that the relationship between the contemporaneous learning index and information flow during an earnings announcement month becomes even stronger after accounting for the degree of information flow prior to the announcement month. For a given stock with a high degree of information flow during an earnings announcement (as measured by a high contemporaneous learning index value), the resulting market reaction and abnormal trading activity tend to be larger if the stock also had a low degree of information flow prior to the earnings announcement month (as measured by a low lagged

learning index value).

In Table 20, I examine how variation in earnings announcement activity affects my conclusions regarding learning index changes during the earnings announcement month. In Panel A, the differences in $|CAR|_d$, $|CAR|_{d-1,d+1}$, and $ATURN_{d-1,d+1}$ between extreme $LI_t - LI_{t-1}$ quintiles are 0.341%, 0.768%, and 19.902%. In Panel B, the differences in these three variables between extreme quintiles are 0.516%, 1.003%, and 21.692%. Each of these six estimates is significant at the 1% level. In both panels, I find an insignificant relationship between the change in the learning index and $|CAR|_{q+1}$ during both high and low earnings announcement activity months. Consistent with the conclusions from the previous related analyses, the results indicate that the explanatory power of the learning index for measures of information flow is stronger when there are fewer earnings announcements during the month.

4.6 Learning costs and return/volatility predictability

In this section, I examine how differences in firm complexity affect the relationship between the learning index and cross-sectional variation in risk and return. Investors may have more difficulty learning about firms with complicated organizational structures. If this is true, then I expect the learning index to carry greater explanatory power among stocks with lower information processing costs. I consider two measures of firm complexity using data from Compustat. The first measure, $nSEG$, is the number of business segments in different industries (defined based on 4-digit SIC code). The second measure reflects the concentration of sales from various industry segments within the firm ($COMPLEX$), computed as $1-\sum_{j=1}^{J} s_j^2$, where s_j is the fraction of the firm's total sales generated by industry segment j. Higher values of $COMPLEX$ are associated with greater complexity due to lower sales concentration within the firm. In an average month in my sample, the average (median) firm has 2.47 (1.92) business segments and a $COMPLEX$ value of 0.29 (0.24). For each

stock-month, I assign values of $nSEG$ and $COMPLEX$ based on the segment information disclosed in the annual report for the corresponding fiscal year. On occasion, firms issue restated financial statements for prior fiscal years due to a subsequent reorganization of its segments. In order to avoid improperly matching stock-months to organizational structures that were not yet established as of that month, I only use segment information from the year of the original annual report.

Since firms with more segments tend to be larger on average, I use trivariate portfolio sorting to examine the predictive ability of the learning index for risk and return across different levels of firm complexity while controlling for the relationship between complexity and firm size. At the end of each month, I sort stocks into quintiles based on firm size. Within each size quintile, I then sort stocks into terciles based on one of the two measures of complexity and group stocks together across size quintiles within each complexity tercile. This procedure results in three groups of stocks that exhibit differences in complexity but similar distributions of firm size. To verify the effectiveness of this sorting procedure, I examine the mean and median values of $nSEG$, $COMPLEX$, and $SIZE$ across these newly formed terciles. The difference in mean (median) $nSEG$ between extreme terciles is 2.403 (2.254), while the difference in mean (median) $COMPLEX$ between extreme terciles is 0.523 (0.532). For both measures of firm complexity, the difference in mean and median firm size between extreme terciles is not significantly different from zero, indicating that the complexity terciles are relatively balanced in terms of firm size. Within each tercile, I then sort stocks into quintiles based on the learning index.

Panel A of Table 21 presents next month excess return and risk-adjusted excess return for the resulting 15 value-weighted portfolios as well as for the three $LI5-LI1$ spread portfolios within each complexity tercile. The $LI5-LI1$ return spread is -0.805% for stocks in the lowest $nSEG$ tercile and -0.508% for stocks in the highest $nSEG$ tercile. After adjusting for risk exposure, I find that the difference in these spreads widens: the FF6 alpha of the $LI5-LI1$ portfolio is -1.079% for stocks with the lowest number of industry segments and

−0.316% for stocks with the highest number of industry segments (t-statistics of −4.40 and −1.80, respectively). Conclusions are qualitatively similar using equal-weighted portfolios in Panel B.

Table 22 reports next month abnormal volatilities for each of the 15 portfolios. In Panel A, I find that the spreads in value-weighted abnormal volatility between high LI and low LI stocks are similar within the extreme $nSEG$ groups. Within the low $nSEG$ tercile, the spread in $ARVOL$ between extreme LI quintiles is −4.028. Within the high $nSEG$ tercile, the $LI5-LI1$ spread in $ARVOL$ is −4.206. Each estimate is significant at the 1% level.

In the last three columns of Table 22, I examine levels of volatility across LI quintiles within each $nSEG$ tercile. For these analyses, I use a bivariate dependent sorting approach when defining the LI quintiles to control for past levels of volatility. I find that stocks with higher number of industry segments have lower levels of volatility on average, possibly resulting from the diversification of segments within these firms. While the difference in $ARVOL$ between extreme LI quintiles is comparable across complexity groups, the difference in $RVOL$ between extreme LI quintiles is smaller in magnitude for high complexity stocks. Within the low $nSEG$ tercile, the spread in $RVOL$ between extreme LI quintiles is −3.128 (t-statistic = −2.94). Within the high $nSEG$ tercile, the spread in $RVOL$ between extreme LI quintiles is −1.376 (t-statistic = −2.65). I arrive at similar conclusions based on results using equal-weighted portfolios in Panel B as well as based on untabulated results examining spreads in systematic and idiosyncratic components of volatility.

In Table 23 and Table 24, I use the second measure of firm complexity based on sales concentration to examine variation in the explanatory power of the learning index. All of the conclusions are comparable to those from the previous two tables. The risk-adjusted return of the $LI5-LI1$ value-weighted portfolio is −1.020% (t-statistic = −4.03) among stocks with the highest values of $COMPLEX$ (lowest sales concentration) and −0.318% (t-statistic = −1.85) among stocks with the lowest values of $COMPLEX$ (highest sales concentration). The difference in $ARVOL$ between extreme LI quintiles is −4.134 in the low $COMPLEX$

tercile and -4.133 in the high $COMPLEX$ tercile, whereas the differences in $RVOL$ between extreme LI quintiles is -3.143 and -1.712 for each $COMPLEX$ tercile, respectively. Thus, the evidence in this section indicates that the learning index predicts larger cross-sectional differences in expected return and the level of volatility for firms with lower information processing costs.

Chapter 5

Additional discussion and robustness checks

5.1 Discussion of model assumptions

This section provides additional detail on the model of Van Nieuwerburgh and Veldkamp (2010). In particular, I discuss the properties of various learning technologies considered in the model, the implications of these technologies for the costs of learning, and the rationale for choosing to focus on the version of the model with mean-variance preferences and entropy learning.

There are two aspects of information acquisition cost to consider: how much learning capacity to acquire and how to allocate that learning capacity across assets. Van Nieuwerburgh and Veldkamp (2010) concentrate on the second of these two aspects. Since the capacity allocation decision depends only on the level of capacity acquired, the authors assume that the optimal level of capacity acquired is given exogenously by an unspecified utility cost function. Therefore, the cost of learning in their model is defined by the amount of learning capacity required for a given information signal. This cost is determined by the assumption about the nature of learning.

With an additive learning technology, learning can be characterized as a sequence of independent draws where the cost of a signal is equal to the incremental precision it provides to the investor's posterior beliefs. The total learning cost can be measured by the sum of signal precisions across assets. Under additive learning, a unit increase in precision is equally costly across assets regardless of prior uncertainty. With entropy-based learning costs, learning can be characterized as an increasingly refined search where each new signal depends on prior signals. The total learning cost can be measured by the product of the signal precisions across assets. Under entropy learning, a unit increase in precision is more costly for assets with higher prior uncertainty. Because of this property, investors choose to deepen their knowledge rather than broaden it.

Van Nieuwerburgh and Veldkamp (2010) argue that the entropy-based learning technology is preferable to an additive technology for two reasons. First, it is scale neutral, which means that learning costs are unaffected by the definition of one share of an asset. Second, it leads to a prediction of specialized learning (learning about one asset or risk factor) rather than generalized learning (learning about multiple assets). Combining the entropy technology with constant absolute risk aversion (CARA) preferences leads to a prediction of indifference between any allocation of learning capacity. On the other hand, an investor with mean-variance preferences and entropy learning costs chooses specialization in learning. Under these assumptions, the investor's optimization problem is to maximize a weighted sum of posterior precisions (the weights are given by the learning index), subject to a constraint on the product of those precisions. This yields a prediction of specialized learning where investors allocate all learning capacity toward the asset with the highest learning index in order to maximize the sum while keeping the product low.

The prediction of specialized learning behavior is supported by the empirical observation that concentrated portfolios outperform diversified ones (e.g., Kacperczyk, Sialm, and Zheng (2005) and Ivković, Sialm, and Weisbenner (2008)). Assuming that greater learning capacity is correlated with better investment performance, this observation implies that investors

with greater learning capacity choose to specialize in their information and portfolio choices. My analysis focuses on the version of the model with mean-variance preferences and entropy learning as it leads to a prediction that more closely matches existing empirical evidence.

5.2 Alternative asset pricing models

In this section, I investigate the robustness of the cross-sectional relation between the learning index and expected returns by repeating the portfolio sorting analyses from Section 3.1.1 using four alternative factor model specifications for risk adjustment. I obtain liquidity factor data from Lubos Pastor's website,[1] short-term and long-term reversal factor data from Kenneth French's website,[2] and data for the Stambaugh and Yuan (2017) mispricing factors from Robert Stambaugh's website.[3] Data for the Hou, Xue, and Zhang (2015) q-factor model are provided by Kewei Hou.

I first augment the Fama and French (2018) six-factor model by adding the Pastor and Stambaugh (2003) liquidity factor. I consider a further extension of the previous model by adding a short-term reversal factor and a long-term reversal factor. In addition to these two specifications, I also consider the Stambaugh and Yuan (2017) factor model, which contains market, size, and two mispricing factors, and the Hou et al. (2015) q-factor model, which contains market, size, profitability, and investment factors. The two mispricing factors ($MGMT$ and $PERF$) of Stambaugh and Yuan (2017) capture overpricing or underpricing across 11 well-known anomalies. The $MGMT$ factor is constructed based on six anomaly variables that can be directly affected by the decisions of a firm's management: net stock issues, composite equity issues, accruals, net operating assets, asset growth, and investment to assets. The $PERF$ factor is constructed based on five anomaly variables related to performance: distress, O-score, momentum, gross profitability, and return on assets.

Table 25 reports risk-adjusted excess returns for value-weighted (Panel A) and equal-

[1] faculty.chicagobooth.edu/lubos.pastor/research
[2] mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html
[3] finance.wharton.upenn.edu/~stambaug

weighted (Panel B) quintile portfolios sorted by LI. Using value-weighted portfolio returns, the difference in alpha between extreme LI quintiles ranges from -0.506% per month (-6.2% per year) based on the Stambaugh and Yuan (2017) factor model to -0.635% per month (-7.9% per year) based on the Hou et al. (2015) factor model. Results based on equal-weighted portfolio returns are qualitatively similar. Across all alternative factor model specifications considered, I find that the negative cross-sectional relation between the LI and risk-adjusted returns is robust.

5.3 Explaining the cross-section of volatility (Bivariate portfolio sorting)

In this section, I re-examine the relationship between the learning index and risk using an alternative testing approach. For these tests, I use the level of volatility instead of abnormal volatility as the dependent variable. To control for the prior level of volatility, I use dependent and independent bivariate portfolio sorting. With dependent sorting, I sort stocks into quintiles at the end of each month based on average volatility over the past 12 months. Within each volatility quintile, I sort stocks based on values of LI and then combine each LI subquintile across volatility quintiles. With independent sorting, I sort stocks independently into quintiles at the end of each month based on LI and prior 12-month average volatility. I compute average values of next month volatility within each of the resulting 25 portfolios and then average these values across volatility quintiles for each LI quintile.

Table 26 presents value-weighted (Panel A) and equal-weighted (Panel B) quintile averages of next month return volatility, systematic volatility, and idiosyncratic volatility, controlling for the respective average level of volatility in the prior 12 months. The difference in next month average volatility between high and low LI stocks is negative in all cases and significant in all but one instance. Using bivariate independent sorting, the average spread in next $RVOL$, $SVOL$, and $IVOL$ between extreme LI quintiles is -1.550, -1.383,

and −0.613 percentage points based on value-weighted portfolios, and −1.179, −1.183, and −0.459 percentage points based on equal-weighted portfolios. This serves as complementary evidence of the explanatory power of LI for the cross-section of risk.

5.4 Explaining the cross-section of implied volatility

In this section, I re-examine the relationship between the learning index and risk using option-implied volatility as a proxy for posterior variance. Option-implied volatility can be viewed as a measure of the market's expectation of an asset's volatility over the remaining life of the option. I obtain data beginning in 1996 from the OptionMetrics volatility surface for month-end implied volatilities of at-the-money calls and puts (deltas of 0.5 and -0.5, respectively) with 30 days to maturity. Given that implied volatility is a forward-looking measure, I use portfolio sorting to investigate the contemporaneous relation between the learning index and abnormal implied volatilities of calls ($CVOL$) and puts ($PVOL$). $ACVOL$ is the difference between current month $CVOL$ and average $CVOL$ in the prior 12 months, scaled by average $CVOL$ in the prior 12 months and multiplied by 100. I measure abnormal put-implied volatility ($APVOL$) in a similar manner.

Table 27 reports value-weighted (Panel A) and equal-weighted (Panel B) averages of $ACVOL$ and $APVOL$ for LI-sorted portfolios. I find a negative cross-sectional relation between the learning index and the market's contemporaneous expectation of volatility in the next month. Higher values of LI are associated with lower value-weighted average and equal-weighted averages of $ACVOL$ and $APVOL$. The difference in $ACVOL$ ($APVOL$) between high and low LI quintiles is −2.577% (−2.469%) based on value-weighted portfolios and −2.325% (−2.239%) based on equal-weighted portfolios.

For robustness, I use bivariate dependent and independent portfolio sorting approaches in Table 28 to examine the explanatory power of the learning index for the current level of implied volatility while controlling for the past 12-month average level of implied volatility.

I continue to find that the learning index is negatively related to expectations of future volatility. Using bivariate dependent sorting, the spread in $CVOL$ ($PVOL$) between extreme LI quintiles is -3.789 (-3.806) percentage points based on value-weighted portfolios and -0.850 (-0.804) percentage points based on equal-weighted portfolios.

The results in this section are qualitatively similar if I use next month (instead of current month) abnormal implied volatility as the dependent variable in univariate sorting, or next month levels of implied volatility as the dependent variable in bivariate sorting. Altogether, my findings in this section imply that the learning index carries cross-sectional explanatory power not only for future realized volatility, but also for the market's current expectation of future volatility.

5.5 Explaining the cross-section of market beta

According to Van Nieuwerburgh and Veldkamp (2010), assets that investors learn more about should have returns that are lower than what is predicted by a standard asset pricing model such as the CAPM. My conclusions based on analyses of risk-adjusted returns in Tables 4 and 25 support this idea. The model predicts that learning about an asset leads to a lower conditional covariance with the market. In this section, I investigate this prediction by examining the cross-sectional relation between the learning index and CAPM beta.

In Table 29, I form quintile portfolios based on the learning index and examine abnormal levels of market risk exposure in the next month. I estimate one measure of beta using daily data from the past year (β^{MKT}) and another measure of beta using daily data within the month (β_m^{MKT}). The first measure carries the benefit of more data used in estimation, and the second measure allows for more variation in estimated risk exposure over time. I define abnormal levels of each beta measure ($A\beta^{MKT}$ and $A\beta_m^{MKT}$) as beta in the month following portfolio formation divided by average monthly beta over the previous 12 months, minus one and multiplied by 100.

Table 29 reports the next month value-weighted (Panel A) and equal-weighted (Panel B) quintile average abnormal beta. Consistent with the theoretical prediction that learning reduces co-movement with the market, I find that higher values of LI are associated with lower abnormal levels of market beta. The difference in value-weighted average $A\beta^{MKT}$ between high and low LI stocks is -6.322%, with a t-statistic of -4.81. The result in the next column is based on value-weighted averages of abnormal beta measured using data within a month. The difference in $A\beta_m^{MKT}$ between extreme value-weighted quintiles is negative but insignificant. In Panel B, the $5-1$ spread in equal-weighted average $A\beta^{MKT}$ ($A\beta_m^{MKT}$) is -7.157% (-5.469) with a t-statistic of -7.43 (-3.02).

For robustness, I use bivariate dependent and independent portfolio sorting approaches in Table 30 to examine the relationship between the learning index and next month beta while controlling for the past 12-month average beta. Across all combinations of the various alternatives for portfolio sorting, portfolio weighting, and beta estimation, the relation between LI and next month beta is negative and significant at the 1% level. Using bivariate independent sorting, the spread in β^{MKT} (β_m^{MKT}) between extreme LI quintiles is -0.108 (-0.137) based on value-weighted portfolios and -0.109 (-0.151) based on equal-weighted portfolios.

To further investigate this relation, I estimate multivariate cross-sectional regressions similar to those used in Section 3.2. Instead of using volatility measures, I use next month beta (β_m^{MKT}) as the dependent variable and lagged values of beta as controls. Table 31 presents the results. After controlling for various determinants of risk, I continue to find a negative and significant cross-sectional relation between LI and CAPM beta. Based on precision-weighted coefficient averages, the difference in next month beta between the stocks with the highest and lowest learning index is -0.045 on average, holding all other variables constant.

These results reinforce the prior analyses illustrating the explanatory power of the learning index for dispersion in systematic volatility and factor model pricing errors. Taken

together, my findings support the theoretical prediction that learning about an asset results in a lower conditional covariance with the market.

5.6 Additional control variables

In this section, I use cross-sectional return and volatility regressions to investigate whether the explanatory power of the learning index derives from its relationship with the various other dependent variables studied in this paper. This analysis can indicate the extent to which the learning index provides information content about the cross-section of risk and return that is not already captured by other variables associated with information flow.

For the return regressions, I use the full set of variables from Equation 3.2 as a benchmark specification. I then augment this specification with one of the following variables of interest: maximum daily return during the month ($MAXDRET$), maximum absolute daily return during the month ($MAX|DRET|$),[4] abnormal monthly share turnover ($ATURN$), number of analyst forecasts for the nearest fiscal quarter ($nFCST$), number of analyst forecast revisions ($nREV$), change in forecast accuracy (ΔFA), and number of EDGAR downloads ($EDGAR$).[5] For the volatility regressions, I augment the benchmark specification given by Equation 3.3 with one of the seven aforementioned variables of interest. For brevity, I only report precision-weighted coefficient averages for the learning index and the control variable of interest, and I leave the coefficients for the benchmark control variables untabulated.

Table 32 reports precision-weighted coefficient averages from cross-sectional return regressions. In Column 1, I estimate the benchmark specification over the full sample period (equivalent to Column 6 of Table 5). In the next three columns, I find that controlling for maximum daily return, maximum absolute daily return, or abnormal share turnover has very little impact on the coefficient estimate for LI.[6] Column 5 presents results from estimating

[4]Results from regressions including $MAXWRET$ and $MAX|WRET|$ are qualitatively similar to those based on including $MAXDRET$ and $MAX|DRET|$. These results are left untabulated for brevity.

[5]I forego analysis using of Bloomberg news reading activity (BBG) as a control variable due to limited data availability.

[6]In Column 2 of Table 32, the average coefficient on $MAXDRET$ is positive and significant. This result

Equation 3.2 using data beginning in July 1984. This column serves as a benchmark for the next three specifications which use analyst data. After controlling for analyst coverage, number of forecast revisions, and improvements in forecast accuracy, I continue to find a negative and significant coefficient on LI. Finally, in Column 9 I estimate a benchmark regression using data beginning in March 2003. After controlling for EDGAR download activity, the coefficient on the learning index is qualitatively unchanged.

Using a similar approach, I present precision-weighted coefficient averages from cross-sectional volatility regressions in Table 33. As in the previous table, I find that controlling for each of the additional variables has little impact on the average coefficient estimate for the learning index. Thus, while some of the variables considered in this section carry significant explanatory power for future returns or volatilities, the explanatory power of the learning index is robust to the inclusion of these variables as controls in cross-sectional regressions.

5.7 Subperiod analysis

To examine how the cross-sectional explanatory power of the learning index for risk and return varies over time, I repeat the portfolio sorting analyses over the subperiods July 1964 to December 1989 and January 1990 to December 2016. Table 34 presents value-weighted (Panel A) and equal-weighted (Panel B) portfolio excess return, Fama and French (2018) six-factor risk-adjusted return, and abnormal total, systematic, and idiosyncratic return volatility. While there is slight variation in the magnitude and significance of the estimates of interest over time, the negative cross-sectional relation between LI and the measures of risk and return are evident in both subperiods.

In the earlier sample period, LI strongly predicts cross-sectional variation in raw excess returns. The average return of the $5-1$ value-weighted portfolio is -0.545% (t-statistic

is contrary to the findings of Bali et al. (2011) that stocks with high maximum daily returns during the month command lower expected returns. In untabulated analysis, I find that the coefficient on $MAXDRET$ is negative and significant using various simpler regression specifications, but turns positive and significant after controlling for the combination of market beta, short-term reversal, and idiosyncratic volatility relative to the Fama and French (2018) six-factor model.

= −3.30). The predictive ability of LI for risk-adjusted returns and abnormal idiosyncratic volatility is negative but insignificant during this time, suggesting that a larger portion of the explanatory power of LI for excess returns is attributable to reductions in the systematic component of risk. In the later part of the sample, the spreads in excess return and alpha between extreme value-weighted portfolios are −0.458% and −0.483%. Each estimate is significant at the 5% level. Cross-sectional differences in abnormal volatility between extreme LI portfolios tend to be larger in the later part of the sample. On average, the differences in $ARVOL$, $ASVOL$, and $AIVOL$ between extreme value-weighted portfolios from 1964 to 1989 are −2.489%, −3.236%, and −1.168%, respectively, with corresponding t-statistics of −2.88, −3.12, and −1.62. From 1990 to 2016, these differences increase to −5.198%, −5.776%, and −4.099%. Each of these three estimates is significant at the 1% level. The results using equal-weighted portfolios are qualitatively similar to those based on value-weighted portfolios.

In Table 35, I estimate two-stage cross-sectional regressions of excess return and volatility over two subperiods. For these analyses, I use the full set of control variables from Equation 3.2 and Equation 3.3, but report only the average LI coefficient estimate for brevity. The first row of the table presents the equal-weighted (Panel A) and precision-weighted (Panel B) average LI coefficient during the first period of July 1966 to December 1989 and the second period of January 1990 to December 2016. In Panel A, the average coefficient in the first (second) period is −0.422 (−0.397). In Panel B, the average coefficient in the first (second) period is −0.411 (−0.411). Each of these four estimates is significant at the 1% level. Thus, in a multivariate setting, the predictive power of LI for next month excess returns is consistent throughout the sample.

The next three rows of Table 35 report equal-weighted and precision-weighted average LI coefficient estimates from cross-sectional regressions of systematic, idiosyncratic, and total return volatility. Due to data availability, these regressions begin in December 1974. As such, I define the first period as December 1974 to December 1995 and the second period as January

1996 to December 2016. My conclusions are similar if I split the sample at December 1989, although a lower number of observations prior to this date reduces statistical power. For each measure of volatility, both the equal-weighted average LI coefficient and precision-weighted average LI coefficient are negative and significant in each subperiod, although the coefficients are larger in absolute value and more significant in the second period than in the first period. This is in accordance with the conclusions from portfolio sorting in Table 34. Combined with the results from portfolio sorting, these findings indicate that the observed relationships between learning, risk, and expected return are not entirely driven by a particular time period within the sample.

5.8 Components of the learning index

In this section, I examine the individual components of the learning index and the relationship of each component to the cross-section of risk and return. In the theoretical model, each individual term is expected to be positively associated with the expected benefits of learning. This analysis can potentially indicate whether the explanatory power of LI is driven by one or two components in particular.

In Panel A of Table 36, I present time-series average cross-sectional summary statistics for each of the three non-rank-transformed components of the learning index as well as the non-rank-transformed sum of the three components. I find that the first and third component of the learning index tend to constitute a larger proportion of the sum. For both the average and the median stock in an average month, the first and third component account for over 87% of the sum. Panel B reports time-series average Pearson and Spearman cross-sectional correlations between these four variables. I find a high degree of cross-sectional correlation between the components, with the monotonic relationships among the components being stronger than the linear relationships. The high correlations may indicate that a large portion of the information content contained in each component is attributable to the weights

on principal components given by the eigenvector matrix, as these weights will be the same across the three components for a given stock. Because of this property, it is difficult to assess the relative importance and independent effect of each component. In any case, I provide results on the explanatory power of each individual component for robustness and to verify that my conclusions are qualitatively unchanged by rank-transformation.

In Table 37, I perform three separate portfolio sorting analyses based on each learning index component. The results indicate that each component carries a comparable level of explanatory power for the various measures of risk and return using either value-weighted or equal-weighted portfolios. The FF6 risk-adjusted returns of the three $5-1$ value-weighted portfolios formed on each learning index component are -0.501%, -0.471%, and -0.490%, with t-statistics of -3.68, -3.24, and -3.32. The differences in $ARVOL$ between extreme value-weighted quintiles formed on each of the three components are -2.772%, -4.342%, and -4.427%, with t-statistics of -4.60, -6.31, and -6.42.

In Table 38, I estimate cross-sectional return regressions using the full set of control variables from Equation 3.2 and replacing LI with the individual components. For brevity, I only report precision-weighted coefficient averages for the learning index components and leave the coefficients for the control variables untabulated. In the first three columns, I estimate univariate cross-sectional regressions based on one of the three learning index components. All three coefficient estimates indicate a negative relationship between each component and next month return and are significant at the 1% level. In Columns 4, 5, and 6, I include control variables in the regressions. The coefficients on the learning index components remain negative and significant at the 1% level.

In Column 7 of Table 38, I estimate a specification including all three components together with the control variables. Each of the three coefficients of interest are not significantly different from zero. This result is likely a result of a multicollinearity problem, as evidenced by the variance inflation factors (VIF) in the last column in of the Table. The variance inflation factor quantifies how much higher the standard error of a coefficient is, relative

to if that variable were uncorrelated with all other explanatory variables. VIFs have a lower bound of one. In untabulated results, I find that the maximum VIF across all control variables in Columns 1 through 7 is approximately 1.513 and the maximum VIF across each component of the learning index in Columns 4 through 6 is 1.135. While there is no universal rule that dictates how high a VIF needs to be in order to signal a problem, the simultaneous inclusion of all learning index components generates significantly higher VIFs of 4.097, 20.421, and 27.838 for each component, respectively.

In Table 39, I estimate cross-sectional regressions of next month return volatility on the full set of control variables from Equation 3.3, again replacing LI with its individual components. Columns 1, 2, and 3 are based on specifications that include one component of the learning index and lagged values of return volatility as controls. Additional control variables are included in the regressions in Columns 4, 5, and 6. Across all six columns, the coefficients of interest from these regressions are all negative and significant.

Column 7 of Table 39 presents results from a specification that includes all three learning index components and all control variables. I find that the coefficient on LI_3 is negative and significant, while the coefficients on the first two components are not significantly different from zero. The last column in the table reports the variance inflation factors associated with the estimates in Column 7. The conclusion from the VIFs regarding a multicollinearity problem is similar to that in the previous table. Across all of the non-volatility control variables in Columns 1 through 7, the maximum VIF is approximately 1.881. In Columns 4 through 6, the maximum VIF across the learning index components is 1.174. Including all three components in the regression specification results in VIFs of 4.356, 21.604, and 29.505 for these three variables.

The results from cross-sectional regressions indicate that multicollinearity between the individual components of the learning index prevents accurate measurement of each variable's relative importance in explaining expected returns and risk. Nevertheless, each component alone carries comparable explanatory power for the cross-section of risk and

return.

5.9 Alternative test assets

In this section, I apply the learning index estimation procedure as described in Section 2.1 to two alternative sets of test assets. I first examine the 49 Fama-French industry portfolios, which are formed based on four-digit SIC codes. Data for value-weighted industry portfolios are obtained from Kenneth French's website. I adjust the returns of industry portfolios for risk using the Fama and French (2018) six-factor model, and focus on the same sample period as the primary analyses (July 1964 to December 2016). I also analyze a set of international equity indexes from developed and emerging markets.[7] Return and market value data for these indexes are obtained from Datastream. I adjust index returns for risk using the Asness, Moskowitz, and Pedersen (2013) global three-factor model, which consists of a global market factor (the MSCI World Index) and two factors capturing value and momentum strategy returns everywhere. Daily data for the global three-factor model is unavailable, so abnormal volatility cannot be decomposed into systematic and idiosyncratic components. The sample spans the period January 1973 to June 2018 and includes a minimum of 17 indexes and a maximum of 68 indexes.

For these analyses, I sort assets (industry portfolios) based on LI into terciles rather than quintiles due to the low number of assets within each cross-section. Table 40 reports next month value-weighted (Panel A) and equal-weighted (Panel B) portfolio excess return, risk-adjusted return, and abnormal return volatility. The difference in return (risk-adjusted return) between the extreme LI terciles is -0.268% (-0.259%) based on value-weighted

[7]The set of developed equity markets consists of Australia, Germany, Belgium, Canada, Denmark, Spain, Finland, France, Hong Kong, Ireland, Israel, Italy, Japan, Korea, Luxembourg, Netherlands, Norway, New Zealand, Austria, Portugal, Sweden, Singapore, Switzerland, United Kingdom, and United States. The set of emerging equity markets consists of United Arab Emirates, Argentina, Bahrain, Bulgaria, Brazil, Colombia, China, Chile, Cyprus, Croatia, Sri Lanka, Czech Republic, Estonia, Egypt, Greece, Hungary, Indonesia, India, Jordan, Kuwait, Lithuania, Malta, Morocco, Mexico, Malaysia, Nigeria, Oman, Peru, Philippines, Pakistan, Poland, Qatar, Romania, Russia, South Africa, Saudi Arabia, Slovenia, Slovakia, Taiwan, Thailand, Turkey, Venezuela, and Vietnam.

terciles. These estimates are significant at the 5% level. I also find a negative and significant relation between LI and abnormal volatility. On average, industries in the high LI tercile experience abnormal monthly volatility that is 1.227% lower than that of the industries within the lowest LI tercile. The results based on equal-weighted portfolio averages are qualitatively similar. These findings are supportive of my conclusions based on individual stocks and are consistent with the idea that greater information flow for particular industries corresponds to a cross-sectional reduction in expected risk and return.

The bottom half of Table 40 presents results from sorting international equity indexes based on values of LI. In Panel A, I find that LI has significant cross-sectional explanatory power for expected returns and risk across equity markets. The tercile portfolio containing markets with the highest (lowest) values of LI has an average monthly excess return of -0.594% (1.511%). The average return difference between these portfolios is -0.917% per month (-11.6% per year) and is significant at the 1% level. After adjusting returns for exposure to global risk factors, the spread in risk-adjusted return is -0.627% per month (-7.8% per year) and is significant at the 5% level. In the last column, the average spread in $ARVOL$ between extreme LI quintiles is -3.376% (t-statistic = -1.97). Results in Panel B using equal-weighted terciles are weaker in comparison to those based on value-weighted terciles. Thus, the explanatory power of LI appears to be concentrated among international equity indexes with greater market value. Overall, the results from using alternative test assets are consistent with the central hypothesized relationship between learning, risk, and expected returns. Furthermore, they demonstrate the applicability of the empirical approach to measuring information flow employed in this paper.

Chapter 6

Conclusion

In this dissertation, I examine the importance of information choice in determining the cross-section of risk and expected return. Much of the asset pricing literature treats an investor's information set as fixed or exogenously determined. In reality, investors have the choice to learn about assets prior to investing. The model of Van Nieuwerburgh and Veldkamp (2010) accounts for this choice, generating predictions for the optimal learning decisions of a rational investor and the resulting impact on risk and risk premiums across assets.

In order to test these predictions, I estimate the learning index from the model for individual stocks. This measure reflects the expected benefits of learning about a particular asset and serves as a prediction of information flow. Consistent with the model's predictions, I find that the empirical learning index is negatively related to both future return and future volatility in the cross-section.

I provide a number of analyses to support the interpretation of the learning index. First, I show that large changes in the learning index are contemporaneously associated with increases in price and volatility. These results illustrate the price appreciation that corresponds to lower expected returns and the heightened return volatility that corresponds to the incorporation of information into prices. I also show that the differences in risk and

risk-adjusted return predicted by the learning index are persistent and do not reverse in the long run, indicating that the explanatory power of this measure is not caused by temporary price pressure or mispricing. In addition, I provide evidence of a contemporaneous relationship between the learning index and abnormal trading activity, analyst coverage, forecast revisions, improvements in forecast accuracy, EDGAR download activity, and Bloomberg news reading activity.

I then demonstrate two distinct patterns in the information environment around quarterly earnings announcements that illustrate the importance of the timing of learning. Higher values of the learning index prior to an earnings announcement are associated with greater abnormal trading activity before the announcement and smaller market reaction to the announcement. On the other hand, higher values of the learning index during an earnings announcement month are associated with greater abnormal trading activity surrounding the announcement and larger market reaction to the announcement. Each of these patterns becomes stronger during months with fewer firms announcing earnings. Finally I explore the role of cross-sectional variation in information processing costs as reflected by firm complexity. I find that the explanatory power of the learning index for expected return and the level of volatility is stronger among firms with less complicated organizational structures.

In aggregate, my findings support the theoretical predictions of Van Nieuwerburgh and Veldkamp (2010) and demonstrate a connection between investors' learning decisions and cross-sectional variation in risk and return. This paper illustrates a new empirical approach for predicting information flow that can be used in a variety of other settings to further investigate the role of information choice in asset pricing.

References

Admati, A. (1985). A noisy rational expectations equilibrium for multi-asset securities markets. *Econometrica 53*, 629–658.

Asness, C., T. Moskowitz, and L. H. Pedersen (2013). Value and momentum everywhere. *Journal of Finance 68*, 929–985.

Bali, T., N. Cakici, and R. Whitelaw (2011). Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics 99*, 427–446.

Bali, T., R. Engle, and S. Murray (2016). *Empirical Asset Pricing: The Cross Section of Stock Returns*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Banerjee, S. (2011). Learning from prices and the dispersion in beliefs. *Review of Financial Studies 24*, 3025–3068.

Barber, B. and T. Odean (2007). 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.

Beaver, W., M. McNichols, and R. Price (2007). Delisting returns and their effect on accounting-based market anomalies. *Journal of Accounting and Economics 2007*, 341–368.

Ben-Rephael, A., Z. Da, and R. Israelsen (2017). It depends on where you search: Institutional investor attention and underreaction to news. *Review of Financial Studies 30*, 3009–3047.

Biais, B., P. Bossaerts, and C. Spatt (2010). Equilibrium asset pricing and portfolio choice under asymmetric information. *Review of Financial Studies 23*, 1503–1543.

Botosan, C. (1997). Disclosure level and the cost of equity capital. *The Accounting Review 72*, 323–349.

Burlacu, R., P. Fontaine, S. Jiminez-Garcés, and M. Seasholes (2012). Risk and the cross section of stock returns. *Journal of Financial Economics 2012*, 511–522.

Cheung, Y.-W. and L. Ng (1992). Stock price dynamics and firm size: An empirical investigation. *Journal of Finance 47*, 1985–1997.

Christie, A. (1982). The stochastic behavior of common stock variances. *Journal of Financial Economics 10*, 407–432.

Crane, A., K. Crotty, and T. Umar (2018). Do hedge funds profit from public information? *Working paper, Rice University*.

Cremers, M. and A. Petajisto (2009). How active is your fund manager? A new measure that predicts performance. *Review of Financial Studies 22*, 3329–3365.

Da, Z., J. Engelberg, and P. Gao (2011). In search of attention. *Journal of Finance 66*, 1461–1499.

Dimson, E. (1979). Risk measurement when shares are subject to infrequent trading. *Journal of Financial Economics 7*, 197–226.

Duffee, G. (1995). Stock returns and volatility: A firm-level analysis. *Journal of Financial Economics 37*, 399–420.

Fama, E. and K. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*, 3–56.

Fama, E. and K. French (2015). A five-factor asset pricing model. *Journal of Financial Economics 116*, 1–22.

Fama, E. and K. French (2018). Choosing factors. *Journal of Financial Economics 128*, 234–252.

Grinblatt, M. and S. Titman (1989). Mutual fund performance: An analysis of quarterly portfolio holdings. *Journal of Business 62*, 393–416.

Harford, J., F. Jiang, R. Wang, and F. Xie (2018). Analyst career concerns, effort allocation, and firms' information environment. *Working paper, University of Washington, University at Buffalo, Singapore Management University, and University of Delaware*.

Harvey, C., Y. Liu, and H. Zhu (2016). ... and the cross-section of expected returns. *Review of Financial Studies 29*, 5–68.

Hong, H., T. Lim, and J. Stein (2000). Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies. *Journal of Finance 55*, 265–295.

Hou, K., C. Xue, and L. Zhang (2015). Digesting anomalies: An investment approach. *Review of Financial Studies 28*, 650–705.

Hou, K., C. Xue, and L. Zhang (2017). Replicating anomalies. *Working paper, Ohio State University and University of Cincinnati*.

Ivković, Z., C. Sialm, and S. Weisbenner (2008). Portfolio concentration and the performance of individual investors. *Journal of Financial and Quantitative Analysis 43*, 613–655.

Jegadeesh, N. and S. Titman (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance 48*, 65–91.

Kacperczyk, M., C. Sialm, and L. Zheng (2005). On the industry concentration of actively managed equity mutual funds. *Journal of Finance 60*, 1983–2011.

Kacperczyk, M., C. Sialm, and L. Zheng (2008). Unobserved actions of mutual funds. *Review of Financial Studies 21*, 2379–2416.

Kacperczyk, M., S. Van Nieuwerburgh, and L. Veldkamp (2016). A rational theory of mutual funds' attention allocation. *Econometrica 84*, 571–626.

Kosowski, R., A. Timmermann, R. Wermers, and H. White (2007). Can mutual fund "stars" really pick stocks? New evidence from a bootstrap analysis. *Journal of Finance 61*, 2551–2595.

Lerman, A., J. Livnat, and R. Mendenhall (2008). The high-volume return premium and post-earnings announcement drift. *Working paper, New York University and University of Notre Dame*.

Li, F. W. and C. Sun (2017). Information acquisition and expected returns: Evidence from EDGAR search traffic. *Working paper, Singapore Management University and Hong Kong University of Science and Technology*.

Litzenberger, R. and K. Ramaswamy (1979). The effect of personal taxes and dividends on capital asset prices: Theory and empirical evidence. *Journal of Financial Economics 7*, 163–195.

Newey, W. and K. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*, 703–708.

Pan, Y., T. Wang, and M. Weisbach (2015). Learning about CEO ability and stock return volatility. *Review of Financial Studies 28*, 1623–1666.

Pastor, L. and R. Stambaugh (2003). Liquidity risk and expected stock returns. *Journal of Political Economy 111*, 642–685.

Pastor, L. and P. Veronesi (2003). Stock valuation and learning about profitability. *Journal of Finance 58*, 1749–1789.

Ryans, J. (2017). Using the EDGAR log file data set. *Working paper, London Business School*.

Stambaugh, R., J. Yu, and Y. Yuan (2015). Arbitrage asymmetry and the idiosyncratic volatility puzzle. *Journal of Finance 70*, 1903–1948.

Stambaugh, R. and Y. Yuan (2017). Mispricing factors. *Review of Financial Studies 30*, 1270–1315.

Van Nieuwerburgh, S. and L. Veldkamp (2009). Information immobility and the home bias puzzle. *Journal of Finance 64*, 1187–1215.

Van Nieuwerburgh, S. and L. Veldkamp (2010). Information acquisition and under-diversification. *Review of Economic Studies 77*, 779–805.

Veldkamp, L. (2011). *Information Choice in Macroeconomics and Finance*. Princeton, New Jersey: Princeton University Press.

Wermers, R. (2000). Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses. *Journal of Finance 55*, 1655–1695.

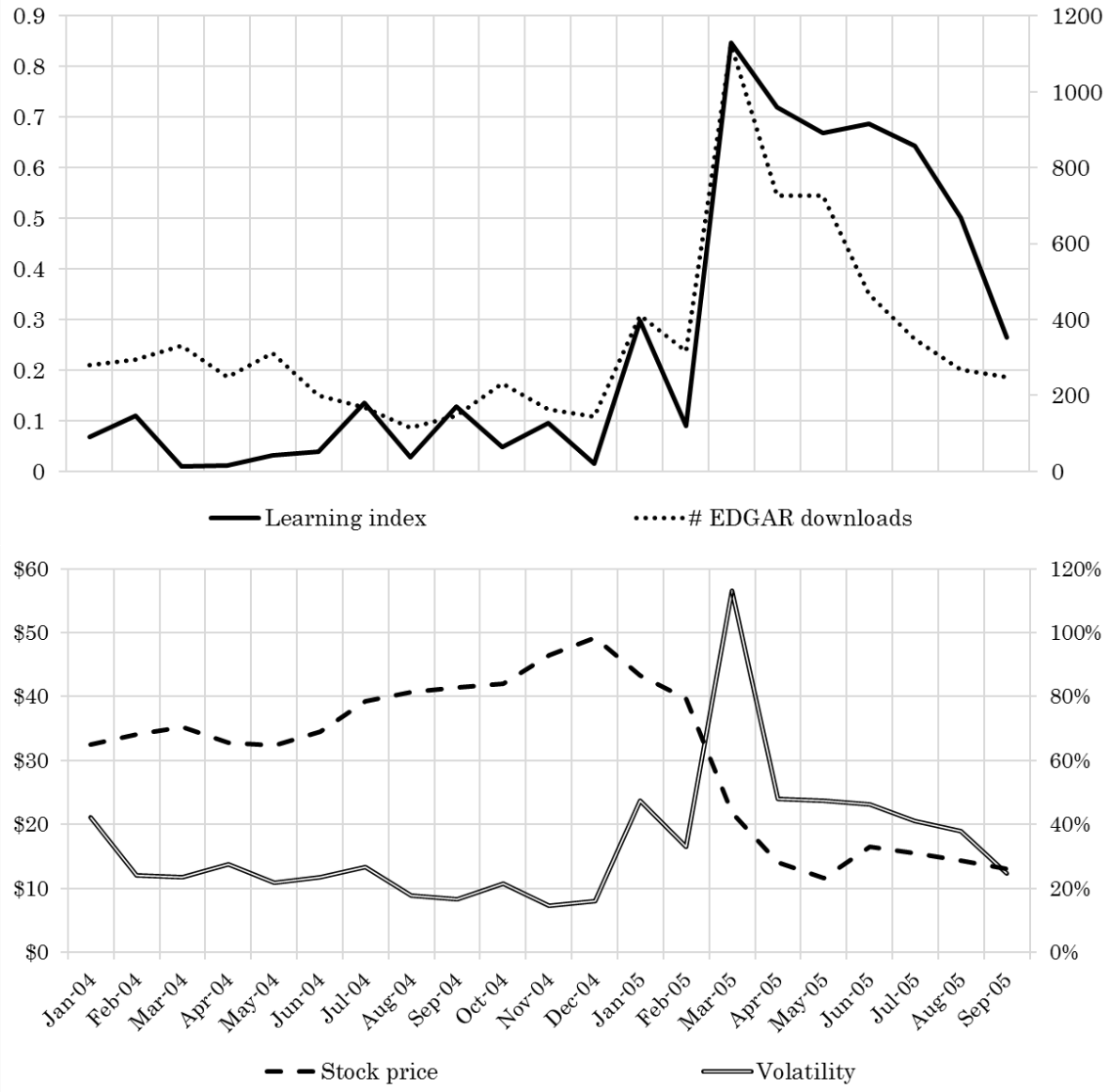
Xu, Y. (2007). Extracting factors with maximum explanatory power. *Working paper, University of Texas at Dallas*.

Zhang, Y. (2008). Analyst responsiveness and the post-earnings-announcement drift. *Journal of Accounting and Economics 46*, 201–215.

Zhao, X. (2017). Does information intensity matter for stock returns? Evidence from Form 8-K filings. *Management Science 63*, 1382–1404.

Figures and Tables

Figure 1: Illustrative example: Doral Financial Corporation



This figure plots monthly values of the learning index, number of downloads of SEC filings on the EDGAR database, stock price, and return volatility for Doral Financial Corporation from January 2004 through September 2005. Notable dates: On January 18, 2005, Doral Financial announced an impairment charge to its mortgage loan portfolio in an earnings press release for the fourth quarter of 2004, resulting in a 12% decline in its stock price on the following day. On March 15, 2005, Doral Financial publicly disclosed its improper use of the spot rate methodology for valuation purposes in its Form 10-K, resulting in a 44% decline in its stock price over the next three days.

Table 1: Cross-sectional summary statistics

			Percentiles		
	Mean	SD	25^{th}	50^{th}	75^{th}
LI	0.50	0.29	0.25	0.50	0.75
β^{MKT}	1.06	0.57	0.66	0.99	1.39
$SIZE$	6.50	1.24	5.50	6.24	7.27
BM	0.71	0.48	0.38	0.63	0.92
$PROF$	0.82	0.85	0.39	0.67	1.06
INV	0.17	0.30	0.03	0.10	0.20
MOM	20.70	47.05	−4.28	12.79	34.42
$ILLIQ$	0.22	1.07	0.02	0.06	0.18
STR	1.76	10.00	−3.85	1.07	6.46
LTR	1.11	2.07	0.17	0.64	1.36
$RVOL$	34.20	18.00	22.40	30.47	41.89
$IVOL$	24.46	14.13	15.34	21.38	30.02
$SVOL$	22.74	12.71	14.15	20.15	28.42
ROE	3.38	5.17	1.88	3.37	4.97
$ROEVOL$	4.33	14.11	0.90	1.64	3.24
AGE	23.62	17.84	10.21	17.88	32.89
$DIVD$	0.71	0.42	0.42	1.00	1.00
LEV	2.19	4.03	0.32	0.75	1.67
$INVPRC$	4.11	2.48	2.44	3.48	5.06
R	1.10	9.57	−4.25	0.73	5.98
# of stocks per month	1,627	346	1,600	1,654	1,757

This table reports time-series averages of monthly cross-sectional means, standard deviations, and quartiles of key variables in the paper. The sample includes all NYSE, AMEX, and NASDAQ domestic common stocks with stock price greater than \$5 and market capitalization greater than the 20^{th} percentile of NYSE stocks at the end of each month. The table summarizes the following characteristics: Learning index (LI), market beta (β^{MKT}), firm size ($SIZE$), book-to-market ratio (BM), profitability ($PROF$), investment (INV), momentum (MOM), illiquidity ($ILLIQ$), short-term reversal (STR), long-term reversal (LTR), return volatility ($RVOL$), idiosyncratic volatility ($IVOL$), systematic volatility ($SVOL$), return on equity (ROE), volatility of return on equity ($ROEVOL$), firm age (AGE), dividend dummy ($DIVD$), leverage (LEV), inverse stock price ($INVPRC$), and monthly return (R). See Table A1 for complete variable definitions. The last row in the table reports time-series summary statistics for the number of stocks in the sample per month. The sample period is July 1964 through December 2016.

Table 2: Cross-sectional correlations

	LI	β^{MKT}	$SIZE$	BM	$PROF$	INV	MOM	$ILLIQ$	STR	LTR	$IVOL$
LI	1.00										
β^{MKT}	−0.12	1.00									
$SIZE$	−0.09	−0.02	1.00								
BM	0.07	−0.13	−0.13	1.00							
$PROF$	−0.07	0.11	−0.02	−0.29	1.00						
INV	−0.06	0.20	−0.05	−0.20	0.18	1.00					
MOM	−0.28	0.08	−0.02	−0.09	0.05	0.01	1.00				
$ILLIQ$	0.02	−0.09	−0.19	0.03	−0.01	−0.01	−0.02	1.00			
STR	−0.02	0.01	−0.01	0.02	0.01	−0.01	0.02	0.03	1.00		
LTR	−0.11	0.16	0.02	−0.28	0.17	0.32	−0.03	−0.01	−0.02	1.00	
$IVOL$	0.06	0.35	−0.28	−0.07	0.07	0.15	0.06	0.05	0.18	0.09	1.00

This table reports time-series averages of monthly cross-sectional correlations between variables used as return predictors: Learning index (LI), market beta (β^{MKT}), firm size ($SIZE$), book-to-market ratio (BM), profitability ($PROF$), investment (INV), momentum (MOM), illiquidity ($ILLIQ$), short-term reversal (STR), long-term reversal (LTR), and idiosyncratic volatility ($IVOL$). See Table A1 for complete variable definitions. The sample period is July 1964 through December 2016.

Table 3: Transition probabilities for portfolios sorted by learning index

	$LI1_{t+1}$	$LI2_{t+1}$	$LI3_{t+1}$	$LI4_{t+1}$	$LI5_{t+1}$
Panel A: 1-month transition matrix					
$LI1_t$	71.3	23.4	4.6	0.7	0.1
$LI2_t$	23.6	45.5	24.4	5.8	0.8
$LI3_t$	4.5	24.7	41.8	24.1	4.8
$LI4_t$	0.6	5.7	24.4	45.3	24.0
$LI5_t$	0.1	0.7	4.8	24.1	70.3
	$LI1_{t+6}$	$LI2_{t+6}$	$LI3_{t+6}$	$LI4_{t+6}$	$LI5_{t+6}$
Panel B: 6-month transition matrix					
$LI1_t$	49.5	25.7	13.8	7.6	3.4
$LI2_t$	25.3	27.4	22.0	15.9	9.5
$LI3_t$	13.4	21.8	24.5	22.7	17.6
$LI4_t$	7.0	15.1	22.2	27.0	28.6
$LI5_t$	3.6	9.4	17.3	27.3	42.3
	$LI1_{t+12}$	$LI2_{t+12}$	$LI3_{t+12}$	$LI4_{t+12}$	$LI5_{t+12}$
Panel C: 12-month transition matrix					
$LI1_t$	32.9	22.9	18.1	14.7	11.3
$LI2_t$	23.0	22.0	20.3	18.4	16.4
$LI3_t$	17.7	20.2	21.0	20.8	20.3
$LI4_t$	13.5	18.2	20.9	23.0	24.4
$LI5_t$	9.3	15.2	20.1	24.8	30.6
	$LI1_{t+24}$	$LI2_{t+24}$	$LI3_{t+24}$	$LI4_{t+24}$	$LI5_{t+24}$
Panel D: 24-month transition matrix					
$LI1_t$	21.3	20.6	20.0	19.6	18.6
$LI2_t$	19.2	19.8	20.2	20.5	20.3
$LI3_t$	17.9	19.3	20.3	21.0	21.4
$LI4_t$	17.1	18.9	20.4	21.3	22.2
$LI5_t$	16.1	18.6	20.4	21.8	23.2
	$LI1_{t+36}$	$LI2_{t+36}$	$LI3_{t+36}$	$LI4_{t+36}$	$LI5_{t+36}$
Panel E: 36-month transition matrix					
$LI1_t$	20.2	19.9	19.8	20.1	20.0
$LI2_t$	19.3	19.7	20.2	20.5	20.2
$LI3_t$	18.5	19.7	20.6	20.8	20.4
$LI4_t$	17.9	19.8	20.5	20.8	21.0
$LI5_t$	17.2	19.2	20.2	21.2	22.2

At the end of each month, stocks are sorted into quintiles based on values of the learning index (LI). For each LI quintile in month t, the table reports the time-series average of the percentage of stocks that fall in each LI quintile in month $t+1$ (Panel A), $t+6$ (Panel B), $t+12$ (Panel C), $t+24$ (Panel D), and $t+36$ (Panel E). Within each panel, percentages are calculated using only the stocks that exist in both the initial month and the final month.

Table 4: Explaining the cross-section of expected returns: Portfolios of stocks sorted by learning index

Panel A: Value-weighted portfolios				
Quintile	Excess return	FF3 α	FF5 α	FF6 α
1 (Low LI)	1.138	0.241	0.315	0.246
2	0.976	0.060	0.033	0.025
3	0.861	−0.051	−0.118	−0.044
4	0.758	−0.154	−0.221	−0.116
5 (High LI)	0.638	−0.254	−0.357	−0.271
5−1	-0.500^{***}	-0.495^{***}	-0.672^{***}	-0.517^{***}
t-stat	(−3.50)	(−4.29)	(−5.75)	(−3.50)

Panel B: Equal-weighted portfolios				
Quintile	Excess return	FF3 α	FF5 α	FF6 α
1 (Low LI)	1.396	0.284	0.314	0.297
2	1.249	0.135	0.123	0.151
3	1.050	−0.045	−0.078	−0.026
4	0.991	−0.098	−0.162	−0.090
5 (High LI)	0.825	−0.227	−0.305	−0.255
5−1	-0.571^{***}	-0.511^{***}	-0.618^{***}	-0.551^{***}
t-stat	(−4.75)	(−5.08)	(−6.40)	(−4.45)

At the end of each month, stocks are sorted into quintiles based on values of the learning index (LI). The table reports the next month value-weighted (Panel A) and equal-weighted (Panel B) quintile average monthly excess return and risk-adjusted excess return (alpha or α). FF3 α is computed with respect to the Fama and French (1993) three-factor model which includes market, size, and value factors. FF5 α is computed with respect to the Fama and French (2015) five-factor model which adds profitability and investment factors to the three aforementioned factors. FF6 α is computed with respect to the Fama and French (2018) six-factor model which adds a momentum factor to the five aforementioned factors. The row labeled "5 – 1" presents the difference in monthly return and alpha between the highest and lowest quintile portfolios. Newey and West (1987) t-statistics and 10%(*), 5%(**), and 1%(***) significance levels for two-sided tests are given for the 5 – 1 portfolio. The sample period is July 1964 to December 2016.

Table 5: Explaining the cross-section of expected returns: Cross-sectional regressions

	Panel A: Equal-weighted coefficient average			Panel B: Precision-weighted coefficient average		
	(1)	(2)	(3)	(4)	(5)	(6)
LI	-0.691^{***}		-0.416^{***}	-0.633^{***}		-0.405^{***}
	(-4.38)		(-4.60)	(-4.71)		(-4.91)
β^{MKT}		0.061	0.042		-0.063	-0.077
		(0.44)	(0.31)		(-0.51)	(-0.63)
$SIZE$		-0.121^{***}	-0.129^{***}		-0.099^{***}	-0.108^{***}
		(-3.40)	(-3.67)		(-2.99)	(-3.30)
BM		0.085	0.085		0.165^{**}	0.164^{**}
		(0.90)	(0.90)		(2.23)	(2.20)
$PROF$		0.096^{*}	0.093^{*}		0.095^{***}	0.094^{***}
		(1.93)	(1.91)		(2.71)	(2.71)
INV		-0.501^{***}	-0.503^{***}		-0.470^{***}	-0.473^{***}
		(-5.55)	(-5.66)		(-6.03)	(-6.13)
MOM		0.005^{**}	0.004^{**}		0.004^{***}	0.003^{***}
		(2.58)	(2.37)		(3.37)	(2.83)
$ILLIQ$		0.591	0.556		-0.056^{**}	-0.057^{**}
		(0.59)	(0.57)		(-2.34)	(-2.36)
STR		-0.033^{***}	-0.034^{***}		-0.032^{***}	-0.033^{***}
		(-6.47)	(-6.62)		(-6.64)	(-6.79)
LTR		-0.053^{***}	-0.056^{***}		-0.034^{***}	-0.040^{***}
		(-2.82)	(-3.20)		(-2.84)	(-3.41)
$IVOL$		-0.011^{***}	-0.009^{***}		-0.011^{***}	-0.009^{***}
		(-4.24)	(-3.67)		(-4.42)	(-3.85)
Adj R^2	0.010	0.090	0.092	0.010	0.090	0.092

This table presents results from two-stage cross-sectional regressions. At the end of each month, I estimate a cross-sectional regression of next month excess stock return on a set of explanatory variables. Each column reports results for a different regression specification. Explanatory variables include an intercept term, the learning index (LI), market beta (β^{MKT}), firm size ($SIZE$), book-to-market ratio (BM), profitability ($PROF$), investment (INV), momentum (MOM), illiquidity ($ILLIQ$), short-term reversal (STR), long-term reversal (LTR), and idiosyncratic volatility ($IVOL$). See Table A1 for complete variable definitions. Panel A reports equal-weighted average slope coefficients and Panel B reports Litzenberger and Ramaswamy (1979) precision-weighted average slope coefficients. The average adjusted R^2 is reported in the last row. The intercept term is not reported for brevity. Newey and West (1987) t-statistics are given in parentheses. 10%(*), 5%(**), and 1%(***) significance levels for two-sided tests are denoted. This regression analysis is based on 814,089 stock-month observations from July 1966 to December 2016 with no missing values for all variables.

Table 6: Explaining the cross-section of volatility: Portfolios of stocks sorted by learning index

	Panel A: Value-weighted portfolios			Panel B: Equal-weighted portfolios		
Quintile	$ARVOL$	$ASVOL$	$AIVOL$	$ARVOL$	$ASVOL$	$AIVOL$
1 (Low LI)	3.502	4.954	2.321	3.106	4.635	2.286
2	1.618	2.703	1.125	1.666	3.054	1.056
3	0.855	2.050	0.297	1.244	2.497	0.691
4	0.934	1.835	0.544	0.583	1.697	0.172
5 (High LI)	−0.472	0.315	−0.452	−0.488	0.512	−0.730
5−1	-3.975^{***}	-4.639^{***}	-2.773^{***}	-3.595^{***}	-4.123^{***}	-3.015^{***}
t-stat	(−6.01)	(−5.79)	(−5.13)	(−6.12)	(−6.16)	(−5.82)

At the end of each month, stocks are sorted into quintiles based on values of the learning index (LI). The table reports next month value-weighted (Panel A) and equal-weighted (Panel B) quintile average abnormal return volatility ($ARVOL$), systematic volatility ($ASVOL$), and idiosyncratic volatility ($AIVOL$) relative to the respective average volatility in the prior 12 months. See Table A1 for complete variable definitions. The row labeled "5 – 1" presents the difference in abnormal volatility between the highest and lowest quintile portfolios. Newey and West (1987) t-statistics and 10%(*), 5%(**), and 1%(***) significance levels for two-sided tests are given for the 5 – 1 portfolio. The sample period is July 1964 to December 2016.

Table 7: Explaining the cross-section of total volatility: Cross-sectional regressions

	Panel A: Equal-weighted coefficient average			Panel B: Precision-weighted coefficient average		
	(1)	(2)	(3)	(4)	(5)	(6)
LI	-1.369^{***}		-1.355^{***}	-1.112^{***}		-1.190^{***}
	(-5.23)		(-7.22)	(-5.90)		(-7.68)
ROE		-0.042^{***}	-0.045^{***}		-0.039^{***}	-0.041^{***}
		(-6.59)	(-6.85)		(-7.98)	(-8.45)
$ROEVOL$		0.045^{***}	0.045^{***}		0.005^{***}	0.005^{***}
		(3.27)	(3.24)		(2.71)	(2.64)
AGE		-0.008^{***}	-0.009^{***}		-0.008^{***}	-0.008^{***}
		(-5.48)	(-5.64)		(-5.96)	(-6.09)
$DIVD$		-0.777^{***}	-0.769^{***}		-0.712^{***}	-0.708^{***}
		(-7.81)	(-8.00)		(-8.40)	(-8.64)
LEV		0.011	0.000		-0.010	-0.016
		(0.32)	(-0.01)		(-0.59)	(-0.96)
$INVPRC$		0.284^{***}	0.287^{***}		0.272^{***}	0.275^{***}
		(10.10)	(10.22)		(10.21)	(10.47)
R		0.088^{***}	0.089^{***}		0.095^{***}	0.096^{***}
		(3.96)	(3.96)		(4.25)	(4.25)
$SIZE$		-0.001	-0.031		-0.043	-0.069
		(-0.01)	(-0.59)		(-0.85)	(-1.37)
BM		-0.880^{***}	-0.842^{***}		-0.790^{***}	-0.757^{***}
		(-6.99)	(-6.80)		(-7.16)	(-6.99)
MOM		0.000	-0.003		0.004	0.001
		(0.10)	(-0.94)		(1.59)	(0.33)
STR		-0.106^{***}	-0.102^{***}		-0.099^{***}	-0.096^{***}
		(-10.61)	(-10.49)		(-11.35)	(-11.26)
Lagged volatilities	Yes	Yes	Yes	Yes	Yes	Yes
Adj R^2	0.493	0.532	0.533	0.493	0.532	0.533

This table presents results from two-stage cross-sectional regressions. At the end of each month, I estimate a cross-sectional regression of next month return volatility ($RVOL$) on a set of explanatory variables. Each column reports results for a different regression specification. Explanatory variables include an intercept term, the learning index (LI), return on equity (ROE), volatility of return on equity ($ROEVOL$), firm age (AGE), a dividend dummy ($DIVD$), leverage (LEV), inverse of stock price ($INVPRC$), firm size ($SIZE$), book-to-market ratio (BM), momentum (MOM), short-term reversal (STR), next month return (R), and 12 lagged values of volatility. See Table A1 for complete variable definitions. Panel A reports equal-weighted average slope coefficients and Panel B reports Litzenberger and Ramaswamy (1979) precision-weighted average slope coefficients. The average adjusted R^2 is reported in the last row. The intercept term and coefficient estimates for lagged volatilities are not reported for brevity. Newey and West (1987) t-statistics are given in parentheses. 10%(*), 5%(**), and 1%(***) significance levels for two-sided tests are denoted. This regression analysis is based on 686,245 stock-month observations from December 1974 to December 2016 with no missing values for all variables.

Table 8: Explaining the cross-section of systematic volatility: Cross-sectional regressions

	Panel A: Equal-weighted coefficient average			Panel B: Precision-weighted coefficient average		
	(1)	(2)	(3)	(4)	(5)	(6)
LI	-1.052^{***}		-0.793^{***}	-0.808^{***}		-0.652^{***}
	(-5.27)		(-5.54)	(-5.88)		(-5.68)
ROE		-0.029^{***}	-0.031^{***}		-0.028^{***}	-0.029^{***}
		(-6.46)	(-6.58)		(-7.67)	(-8.02)
$ROEVOL$		0.034^{***}	0.034^{***}		0.003^{**}	0.003^{**}
		(3.17)	(3.15)		(2.26)	(2.22)
AGE		-0.006^{***}	-0.006^{***}		-0.005^{***}	-0.006^{***}
		(-4.18)	(-4.37)		(-4.99)	(-5.14)
$DIVD$		-0.592^{***}	-0.593^{***}		-0.529^{***}	-0.532^{***}
		(-7.29)	(-7.47)		(-7.57)	(-7.82)
LEV		0.032	0.025		0.006	0.002
		(1.20)	(0.94)		(0.46)	(0.17)
$INVPRC$		0.203^{***}	0.205^{***}		0.192^{***}	0.193^{***}
		(9.87)	(9.90)		(10.31)	(10.50)
R		0.054^{***}	0.054^{***}		0.060^{***}	0.060^{***}
		(3.85)	(3.85)		(4.39)	(4.39)
$SIZE$		0.127^{**}	0.108^{*}		0.079	0.063
		(2.31)	(1.96)		(1.51)	(1.20)
BM		-0.633^{***}	-0.621^{***}		-0.556^{***}	-0.547^{***}
		(-6.33)	(-6.29)		(-6.60)	(-6.54)
MOM		0.005^{*}	0.003		0.008^{***}	0.006^{***}
		(1.70)	(1.14)		(3.62)	(2.81)
STR		-0.061^{***}	-0.059^{***}		-0.054^{***}	-0.053^{***}
		(-7.69)	(-7.64)		(-8.30)	(-8.27)
Lagged volatilities	Yes	Yes	Yes	Yes	Yes	Yes
Adj R^2	0.439	0.478	0.479	0.439	0.478	0.479

This table presents results from two-stage cross-sectional regressions. At the end of each month, I estimate a cross-sectional regression of next month systematic volatility ($SVOL$) on a set of explanatory variables. Each column reports results for a different regression specification. Explanatory variables include an intercept term, the learning index (LI), return on equity (ROE), volatility of return on equity ($ROEVOL$), firm age (AGE), a dividend dummy ($DIVD$), leverage (LEV), inverse of stock price ($INVPRC$), firm size ($SIZE$), book-to-market ratio (BM), momentum (MOM), short-term reversal (STR), next month return (R), and 12 lagged values of $SVOL$. See Table A1 for complete variable definitions. Panel A reports equal-weighted average slope coefficients and Panel B reports Litzenberger and Ramaswamy (1979) precision-weighted average slope coefficients. The average adjusted R^2 is reported in the last row. The intercept term and coefficient estimates for lagged volatilities are not reported for brevity. Newey and West (1987) t-statistics are given in parentheses. 10%(*), 5%(**), and 1%(***) significance levels for two-sided tests are denoted. This regression analysis is based on 686,245 stock-month observations from December 1974 to December 2016 with no missing values for all variables.

Table 9: Explaining the cross-section of idiosyncratic volatility: Cross-sectional regressions

	Panel A: Equal-weighted coefficient average			Panel B: Precision-weighted coefficient average		
	(1)	(2)	(3)	(4)	(5)	(6)
LI	-0.702^{***}		-0.934^{***}	-0.578^{***}		-0.849^{***}
	(-3.83)		(-6.92)	(-4.11)		(-7.19)
ROE		-0.039^{***}	-0.041^{***}		-0.034^{***}	-0.035^{***}
		(-6.49)	(-6.56)		(-8.07)	(-8.36)
$ROEVOL$		0.040^{***}	0.040^{***}		0.006^{***}	0.006^{***}
		(3.44)	(3.39)		(3.41)	(3.31)
AGE		-0.008^{***}	-0.008^{***}		-0.008^{***}	-0.008^{***}
		(-6.99)	(-7.08)		(-7.13)	(-7.21)
$DIVD$		-0.744^{***}	-0.730^{***}		-0.700^{***}	-0.689^{***}
		(-9.04)	(-9.16)		(-10.48)	(-10.66)
LEV		-0.031	-0.039^{*}		-0.029^{**}	-0.033^{***}
		(-1.33)	(-1.66)		(-2.43)	(-2.77)
$INVPRC$		0.257^{***}	0.258^{***}		0.245^{***}	0.247^{***}
		(11.19)	(11.32)		(10.83)	(11.05)
R		0.071^{***}	0.071^{***}		0.074^{***}	0.074^{***}
		(4.07)	(4.08)		(4.21)	(4.21)
$SIZE$		-0.140^{***}	-0.161^{***}		-0.149^{***}	-0.167^{***}
		(-5.16)	(-5.94)		(-5.78)	(-6.45)
BM		-0.776^{***}	-0.745^{***}		-0.701^{***}	-0.672^{***}
		(-7.87)	(-7.64)		(-7.86)	(-7.65)
MOM		-0.003^{*}	-0.006^{***}		-0.001	-0.003^{*}
		(-1.80)	(-3.00)		(-0.51)	(-1.87)
STR		-0.077^{***}	-0.075^{***}		-0.075^{***}	-0.073^{***}
		(-11.01)	(-10.91)		(-11.47)	(-11.37)
Lagged volatilities	Yes	Yes	Yes	Yes	Yes	Yes
Adj R^2	0.420	0.454	0.455	0.420	0.454	0.455

This table presents results from two-stage cross-sectional regressions. At the end of each month, I estimate a cross-sectional regression of next month idiosyncratic volatility ($IVOL$) on a set of explanatory variables. Each column reports results for a different regression specification. Explanatory variables include an intercept term, the learning index (LI), return on equity (ROE), volatility of return on equity ($ROEVOL$), firm age (AGE), a dividend dummy ($DIVD$), leverage (LEV), inverse of stock price ($INVPRC$), firm size ($SIZE$), book-to-market ratio (BM), momentum (MOM), short-term reversal (STR), next month return (R), and 12 lagged values of $IVOL$. See Table A1 for complete variable definitions. Panel A reports equal-weighted average slope coefficients and Panel B reports Litzenberger and Ramaswamy (1979) precision-weighted average slope coefficients. The average adjusted R^2 is reported in the last row. The intercept term and coefficient estimates for lagged volatilities are not reported for brevity. Newey and West (1987) t-statistics are given in parentheses. 10%(*), 5%(**), and 1%(***) significance levels for two-sided tests are denoted. This regression analysis is based on 686,245 stock-month observations from December 1974 to December 2016 with no missing values for all variables.

Table 10: Contemporaneous price impact of learning: Portfolios of stocks sorted by changes in learning index

	Panel A: Value-weighted portfolios		Panel B: Equal-weighted portfolios	
Quintile	*MAXDRET*	*MAXWRET*	*MAXDRET*	*MAXWRET*
	Dependent variable calculated over 1-month horizon			
1 (Low $LI_t - LI_{t-1}$)	3.676	4.426	4.576	5.306
2	3.697	4.493	4.641	5.456
3	3.754	4.589	4.681	5.527
4	3.797	4.675	4.759	5.684
5 (High $LI_t - LI_{t-1}$)	3.943	4.915	4.953	6.004
5−1	0.267***	0.490***	0.376***	0.698***
t-stat	(7.90)	(10.24)	(11.95)	(14.07)
	Dependent variable calculated over 3-month horizon			
1 (Low $LI_t - LI_{t-3}$)	5.184	7.268	6.578	8.958
2	5.097	7.228	6.549	9.000
3	5.142	7.312	6.589	9.098
4	5.210	7.401	6.680	9.227
5 (High $LI_t - LI_{t-3}$)	5.448	7.671	6.976	9.603
5−1	0.264***	0.403***	0.398***	0.645***
t-stat	(3.65)	(4.17)	(5.27)	(5.98)

At the end of each month, stocks are sorted into quintiles based on the change in the learning index (LI) over either a 1-month or 3-month horizon. The 1-month (3-month) change in LI is computed as the value of LI at the end of the current month t minus the value of LI at the end of month $t-1$ ($t-3$). The table reports the next month value-weighted (Panel A) and equal-weighted (Panel B) quintile average maximum daily return (*MAXDRET*) and maximum weekly return (*MAXWRET*) over the respective horizon. See Table A1 for complete variable definitions. The row labeled "5 – 1" presents the difference in maximum return between the highest and lowest quintile portfolios. Newey and West (1987) t-statistics and 10%(*), 5%(**), and 1%(***) significance levels for two-sided tests are given for the 5 – 1 portfolio. The sample period is July 1964 to December 2016.

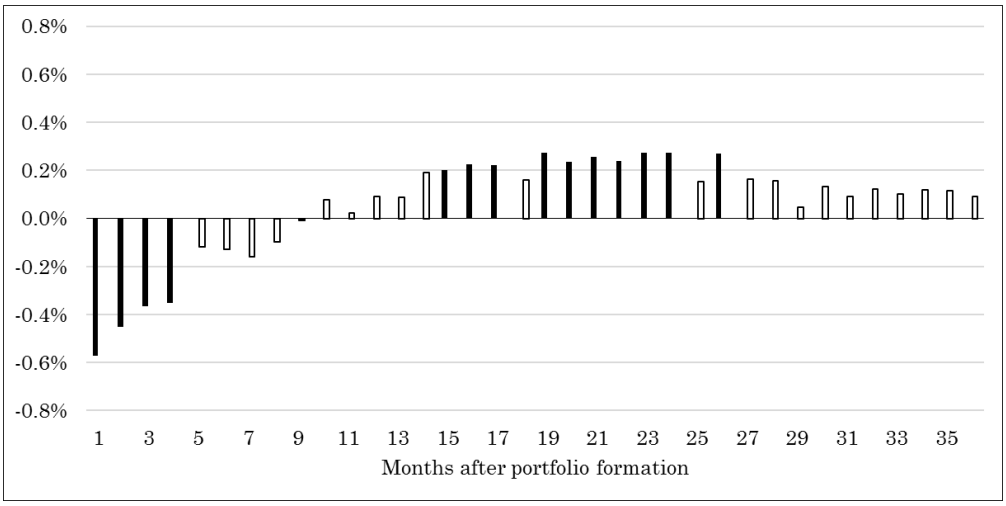
Table 11: Contemporaneous volatility impact of learning: Portfolios of stocks sorted by changes in learning index

	Panel A: Value-weighted portfolios		Panel B: Equal-weighted portfolios	
Quintile	$MAX\|DRET\|$	$MAX\|WRET\|$	$MAX\|DRET\|$	$MAX\|WRET\|$
	Dependent variable calculated over 1-month horizon			
1 (Low $LI_t - LI_{t-1}$)	4.155	5.581	5.196	6.812
2	4.185	5.672	5.284	7.002
3	4.251	5.793	5.355	7.135
4	4.355	6.007	5.487	7.404
5 (High $LI_t - LI_{t-1}$)	4.600	6.419	5.849	7.993
5−1	0.446^{***}	0.838^{***}	0.654^{***}	1.181^{***}
t-stat	(10.35)	(13.88)	(13.05)	(17.51)
	Dependent variable calculated over 3-month horizon			
1 (Low $LI_t - LI_{t-3}$)	5.730	8.115	7.296	10.092
2	5.640	8.039	7.276	10.116
3	5.724	8.193	7.357	10.266
4	5.871	8.413	7.536	10.533
5 (High $LI_t - LI_{t-3}$)	6.299	8.944	8.131	11.277
5−1	0.569^{***}	0.829^{***}	0.836^{***}	1.185^{***}
t-stat	(6.26)	(6.67)	(7.93)	(8.77)

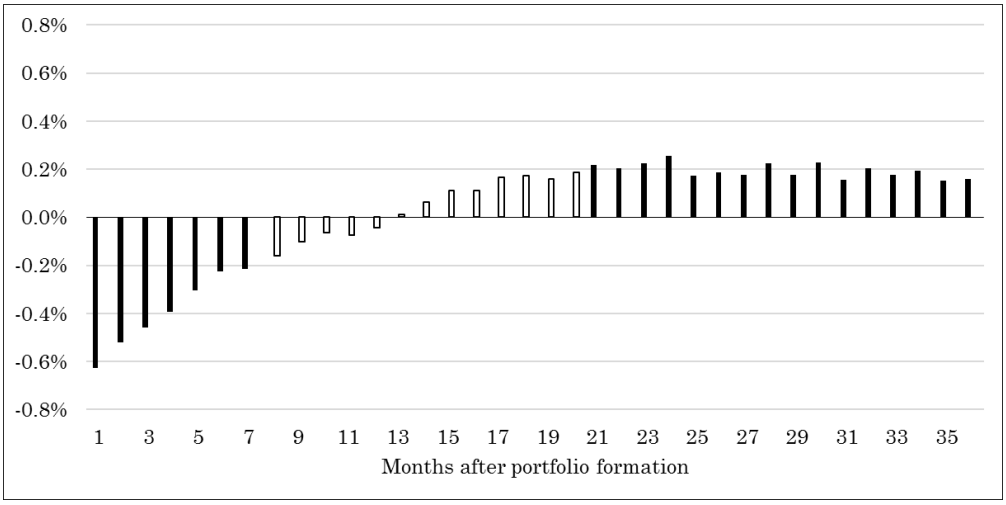
At the end of each month, stocks are sorted into quintiles based on the change in the learning index (LI) over either a 1-month or 3-month horizon. The 1-month (3-month) change in LI is computed as the value of LI at the end of the current month t minus the value of LI at the end of month $t-1$ ($t-3$). The table reports the next month value-weighted (Panel A) and equal-weighted (Panel B) quintile average maximum absolute daily return ($MAX|DRET|$) and maximum absolute weekly return ($MAX|WRET|$) over the respective horizon. See Table A1 for complete variable definitions. The row labeled "5 – 1" presents the difference in maximum return between the highest and lowest quintile portfolios. Newey and West (1987) t-statistics and 10%(*), 5%(**), and 1%(***) significance levels for two-sided tests are given for the 5 – 1 portfolio. The sample period is July 1964 to December 2016.

Figure 2: Long-term return predictability: $LI5 - LI1$ portfolio average return

Value-weighted $LI5 - LI1$ portfolio return



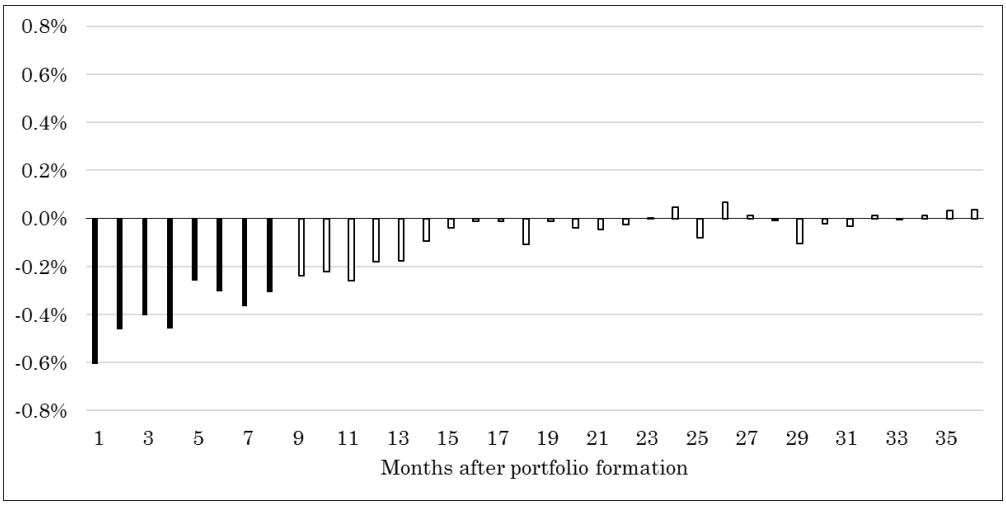
Equal-weighted $LI5 - LI1$ portfolio return



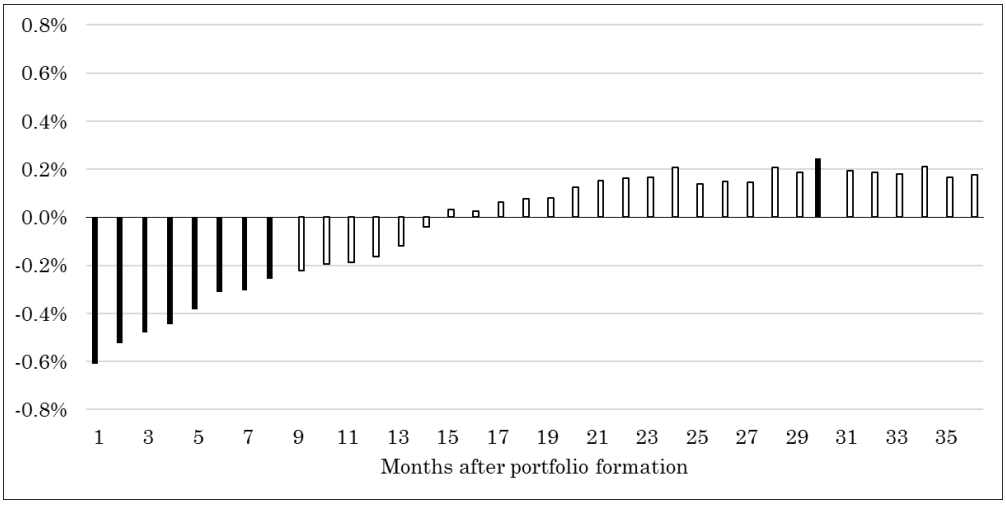
At the end of each month, I sort stocks into quintiles based on values of the learning index (LI) and track the difference in returns between the highest LI quintile and the lowest LI quintile ($5 - 1$ portfolio) in each of the 36 months following portfolio formation. Stocks are required to have non-missing return observations for all 36 months to be included in the sample. The figure presents value-weighted and equal-weighted monthly excess returns for the $5 - 1$ portfolio. Black bars indicate statistical significance at the 10% level. The sample period is July 1964 to December 2016.

Figure 3: Long-term return predictability: $LI5 - LI1$ portfolio risk-adjusted average return

Value-weighted $LI5 - LI1$ portfolio alpha



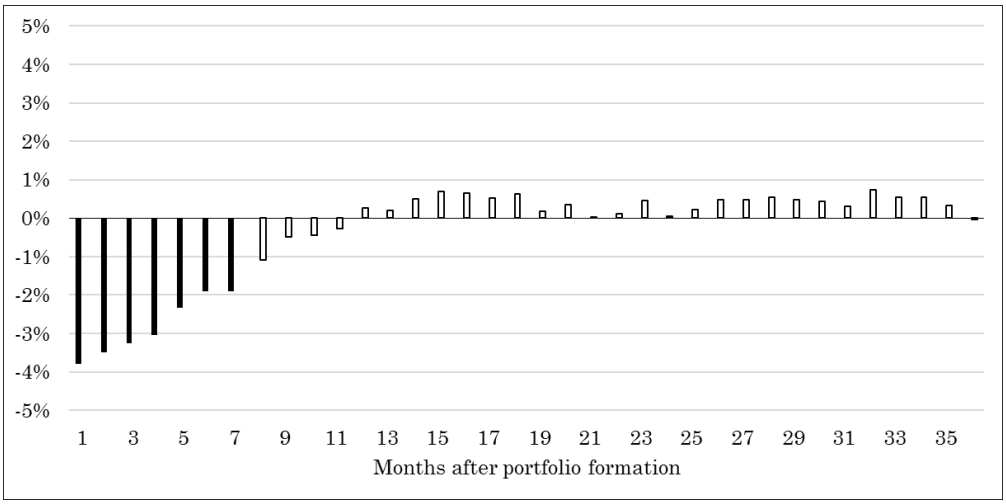
Equal-weighted $LI5 - LI1$ portfolio alpha



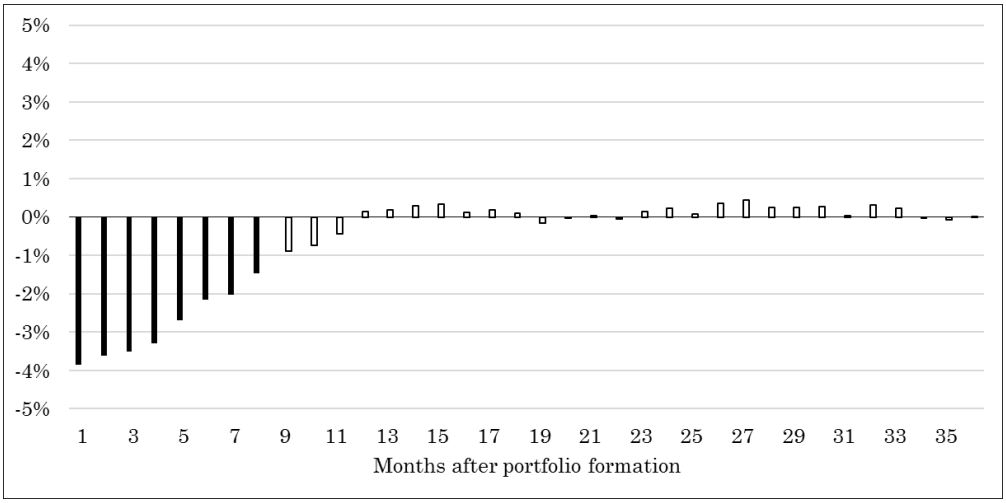
At the end of each month, I sort stocks into quintiles based on values of the learning index (LI) and track the difference in risk-adjusted returns between the highest LI quintile and the lowest LI quintile ($5 - 1$ portfolio) in each of the 36 months following portfolio formation. Stocks are required to have non-missing return observations for all 36 months to be included in the sample. The figure presents value-weighted and equal-weighted monthly risk-adjusted excess returns (alpha) for the $5 - 1$ portfolio. Black bars indicate statistical significance at the 10% level. Returns are risk-adjusted using the Fama and French (2018) six-factor model. The sample period is July 1964 to December 2016.

Figure 4: Long-term volatility predictability: $LI5 - LI1$ portfolio average abnormal return volatility

Value-weighted $LI5 - LI1$ portfolio average $ARVOL$



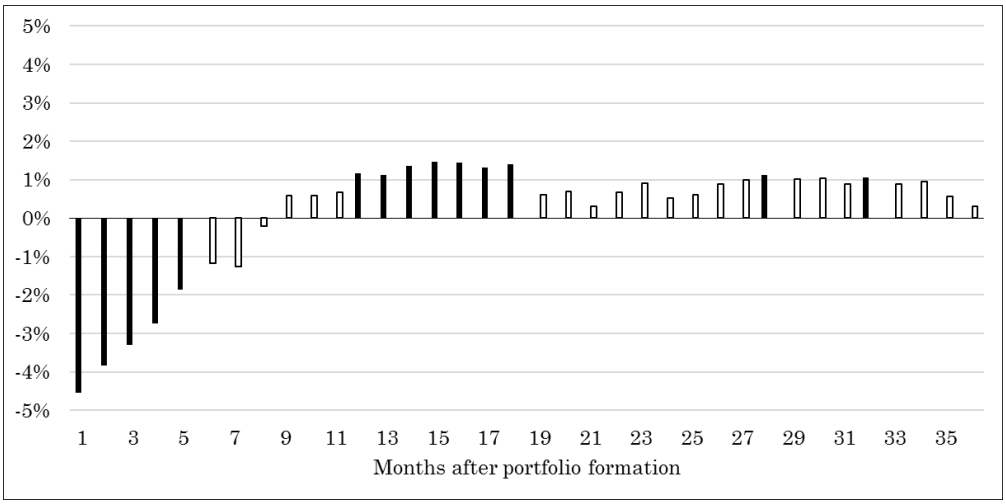
Equal-weighted $LI5 - LI1$ portfolio average $ARVOL$



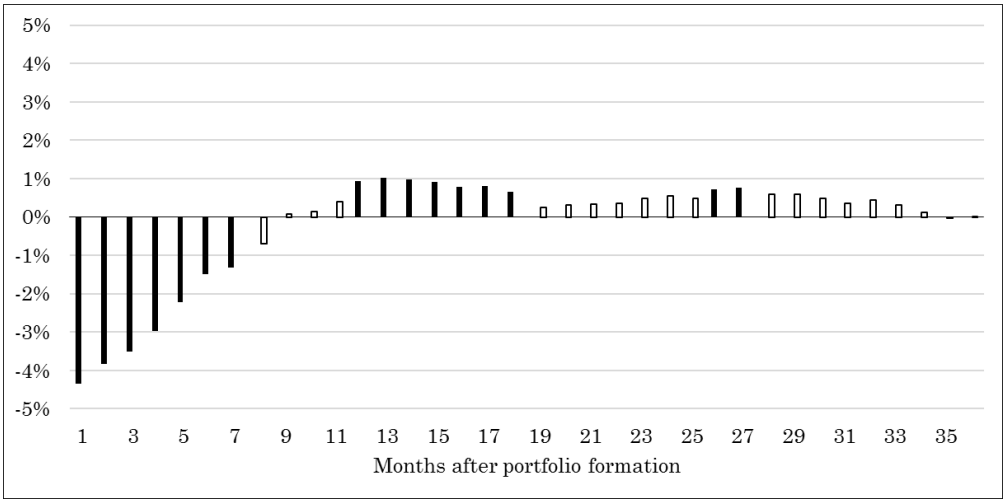
At the end of each month, I sort stocks into quintiles based on values of the learning index (LI) and track the difference in abnormal return volatility ($ARVOL$) between the highest LI quintile and the lowest LI quintile ($5 - 1$ portfolio) in each of the 36 months following portfolio formation. Stocks are required to have non-missing volatility observations for all 36 months to be included in the sample. The figure presents value-weighted and equal-weighted average abnormal return volatility for the $5 - 1$ portfolio. Black bars indicate statistical significance at the 10% level. See Table A1 for complete variable definitions. The sample period is July 1964 to December 2016.

Figure 5: Long-term volatility predictability: $LI5-LI1$ portfolio average abnormal systematic volatility

Value-weighted $LI5-LI1$ portfolio average $ASVOL$



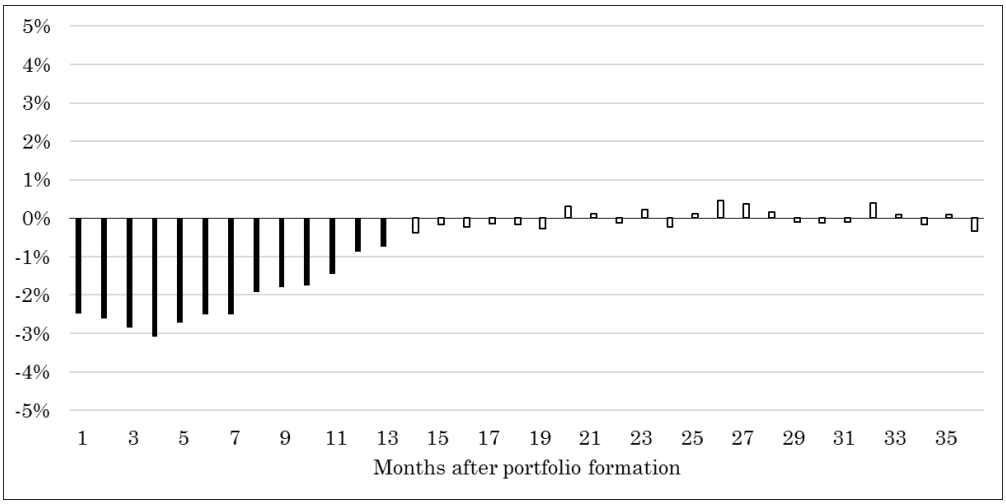
Equal-weighted $LI5-LI1$ portfolio average $ASVOL$



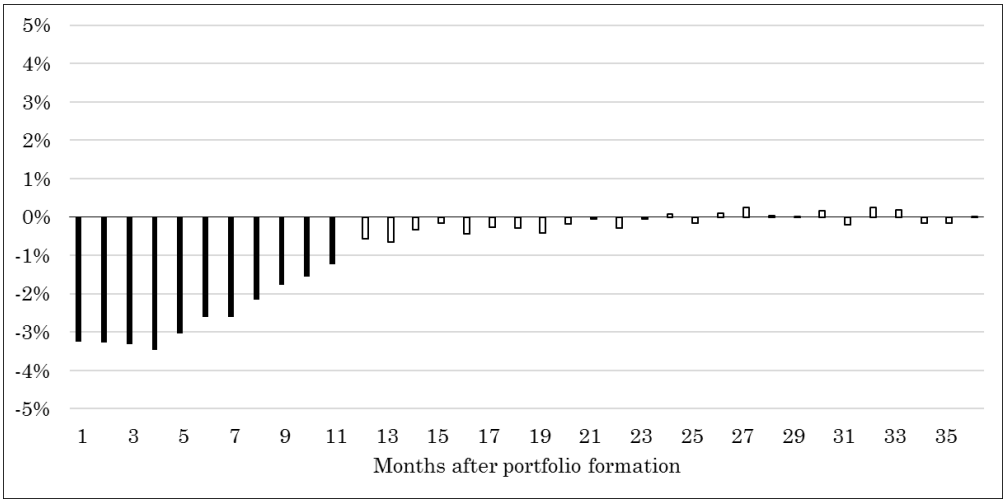
At the end of each month, I sort stocks into quintiles based on values of the learning index (LI) and track the difference in abnormal systematic volatility ($ASVOL$) between the highest LI quintile and the lowest LI quintile ($5-1$ portfolio) in each of the 36 months following portfolio formation. Stocks are required to have non-missing volatility observations for all 36 months to be included in the sample. The figure presents value-weighted and equal-weighted average abnormal systematic volatility for the $5-1$ portfolio. Black bars indicate statistical significance at the 10% level. Systematic and idiosyncratic components of volatility are measured using the Fama and French (2018) six-factor model. See Table A1 for complete variable definitions. The sample period is July 1964 to December 2016.

Figure 6: Long-term volatility predictability: $LI5-LI1$ portfolio average abnormal idiosyncratic volatility

Value-weighted $LI5-LI1$ portfolio average $AIVOL$



Equal-weighted $LI5-LI1$ portfolio average $AIVOL$



At the end of each month, I sort stocks into quintiles based on values of the learning index (LI) and track the difference in abnormal idiosyncratic volatility ($AIVOL$) between the highest LI quintile and the lowest LI quintile ($5-1$ portfolio) in each of the 36 months following portfolio formation. Stocks are required to have non-missing volatility observations for all 36 months to be included in the sample. The figure presents value-weighted and equal-weighted average abnormal idiosyncratic volatility for the $5-1$ portfolio. Black bars indicate statistical significance at the 10% level. Systematic and idiosyncratic components of volatility are measured using the Fama and French (2018) six-factor model. See Table A1 for complete variable definitions. The sample period is July 1964 to December 2016.

Table 12: Relationship with measures of information flow: Portfolios of stocks sorted by learning index controlling for firm size

Sample period begins:	Jul 1964	Jul 1984	Jul 1984	Jul 1984	Mar 2003	Mar 2010
Quintile	$ATURN$	$nFCST$	$nREV$	ΔFA	$EDGAR$	BBG
1 (Low LI)	3.676	7.483	2.076	2.294	742.025	2.805
2	6.705	7.767	2.206	2.843	774.872	2.933
3	8.189	8.047	2.299	3.428	815.069	3.010
4	9.166	8.361	2.456	4.102	838.857	3.047
5 (High LI)	8.547	8.621	2.505	4.319	871.655	3.166
5−1	4.871^{***}	1.139^{***}	0.428^{***}	2.025^{***}	129.630^{***}	0.361^{***}
t-stat	(5.78)	(6.76)	(5.42)	(4.63)	(3.54)	(5.63)

At the end of each month, stocks are sorted into quintiles based on market capitalization. Within each size quintile, stocks are sorted based on values of the learning index (LI). Each LI subquintile is combined across size quintiles into a single quintile. This approach creates portfolios of stocks with differences in LI but similar distributions of size. The table reports the time-series means of quintile averages of six proxies of investor attention or information demand: abnormal monthly share turnover ($ATURN$), number of analyst forecasts ($nFCST$), number of analyst forecast revisions ($nREV$), change in forecast accuracy (ΔFA), number of SEC filing downloads from EDGAR ($EDGAR$), and number of days with abnormal news reading activity on Bloomberg (BBG). See Table A1 for complete variable definitions. The row labeled "5 – 1" presents the difference in the respective dependent variable between the highest and lowest quintile portfolios. Newey and West (1987) t-statistics and 10%(*), 5%(**), and 1%(***) significance levels for two-sided tests are given for the 5 – 1 portfolio. The first row of the table header indicates the first month that data are available for the respective dependent variable. All sample periods end in December 2016.

Table 13: Learning index and analyst coverage: Cross-sectional regressions

Dependent variable: $\ln(1+nFCST)$								
	Panel A: Equal-weighted coefficient average				Panel B: Precision-weighted coefficient average			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LI	-0.036^{*}	0.175^{***}	0.119^{***}	0.094^{***}	-0.036^{*}	0.176^{***}	0.117^{***}	0.092^{***}
	(-1.84)	(8.23)	(10.84)	(9.56)	(-1.87)	(8.24)	(10.72)	(9.22)
$SIZE$		0.295^{***}	0.312^{***}	0.313^{***}		0.296^{***}	0.313^{***}	0.313^{***}
		(81.45)	(83.01)	(81.14)		(79.71)	(85.27)	(82.74)
$NASDAQ$			0.115^{***}	0.082^{***}			0.120^{***}	0.087^{***}
			(8.91)	(6.44)			(9.64)	(7.10)
BM			-0.118^{***}	-0.073^{***}			-0.109^{***}	-0.068^{***}
			(-13.34)	(-10.05)			(-13.04)	(-9.92)
β^{MKT}			0.063^{***}	0.035^{***}			0.056^{***}	0.030^{***}
			(5.67)	(3.35)			(5.30)	(3.03)
$INVPRC$			0.008^{***}	0.008^{***}			0.008^{***}	0.009^{***}
			(7.39)	(8.72)			(8.83)	(9.94)
$RVOL$/100			0.072^{***}	0.028^{**}			0.076^{***}	0.038^{***}
			(5.60)	(2.51)			(5.99)	(3.20)
MOM			-0.001^{***}	-0.002^{***}			-0.001^{***}	-0.001^{***}
			(-15.80)	(-17.04)			(-13.89)	(-14.03)
$\overline{TURN}$			0.002^{***}	0.002^{***}			0.002^{***}	0.002^{***}
			(22.33)	(22.26)			(19.08)	(18.65)
$NASDAQ * \overline{TURN}$			-0.001^{***}	-0.001^{***}			-0.001^{***}	-0.001^{***}
			(-7.70)	(-6.25)			(-7.34)	(-6.12)
Industry dummies	No	No	No	Yes	No	No	No	Yes
Adj R^2	0.005	0.374	0.485	0.501	0.005	0.374	0.485	0.501

This table presents results from two-stage cross-sectional regressions. At the end of each month, I estimate a cross-sectional regression of the natural logarithm of one plus the number of analyst forecasts for the nearest fiscal quarter ($\ln(1+nFCST)$) on a set of explanatory variables. Each column reports results for a different regression specification. Explanatory variables include an intercept term, the learning index (LI), firm size ($SIZE$), a dummy variable equal to one for stocks traded on the NASDAQ Stock Exchange ($NASDAQ$), book-to-market ratio (BM), market beta (β^{MKT}), inverse of stock price ($INVPRC$), return volatility ($RVOL$), momentum (MOM), average monthly turnover over the past 12 months ($\overline{TURN}$), and an interaction term between $NASDAQ$ and $\overline{TURN}$. Columns 4 and 8 include industry dummy variables based on two-digit SIC code. See Table A1 for complete variable definitions. Panel A reports equal-weighted average slope coefficients and Panel B reports Litzenberger and Ramaswamy (1979) precision-weighted average slope coefficients. The average adjusted R^2 is reported in the last row. Newey and West (1987) t-statistics are given in parentheses. 10%(*), 5%(**), and 1%(***) significance levels for two-sided tests are denoted. This regression analysis is based on 596,209 stock-month observations from July 1984 to December 2016 with no missing values for all variables.

Table 14: Learning index and EDGAR downloads: Cross-sectional regressions

Dependent variable: $\ln(1+EDGAR)$						
	Panel A: Equal-weighted coefficient average			Panel B: Precision-weighted coefficient average		
	(1)	(2)	(3)	(4)	(5)	(6)
LI	-0.105^{**}	0.181^{***}	0.137^{***}	-0.102^{**}	0.171^{***}	0.130^{***}
	(-2.32)	(3.56)	(4.90)	(-2.12)	(3.27)	(4.72)
$SIZE$		0.414^{***}	0.402^{***}		0.413^{***}	0.401^{***}
		(101.76)	(75.91)		(105.06)	(74.08)
$\ln(1+nFCST)$			0.062^{***}			0.062^{***}
			(5.99)			(5.79)
$\overline{TURN}$			0.001^{***}			0.001^{***}
			(17.00)			(14.88)
$IVOL$			0.011^{***}			0.011^{***}
			(20.75)			(20.24)
$MOM/100$			-0.055^{***}			-0.037^{*}
			(-2.69)			(-1.89)
BM			0.140^{***}			0.138^{***}
			(11.51)			(10.17)
IO			0.001			0.001^{***}
			(0.89)			(5.49)
$SP500$			0.144^{***}			0.146^{***}
			(16.95)			(16.60)
EAM			0.081^{***}			0.046^{***}
			(10.88)			(7.14)
Adj R^2	0.007	0.410	0.493	0.007	0.410	0.493

This table presents results from two-stage cross-sectional regressions. At the end of each month, I estimate a cross-sectional regression of the natural logarithm of one plus the number of EDGAR downloads during the month ($\ln(1+EDGAR)$) on a set of explanatory variables. Each column reports results for a different regression specification. Explanatory variables include an intercept term, the learning index (LI), firm size ($SIZE$), the natural logarithm of one plus the number of analyst forecasts ($\ln(1+nFCST)$), average monthly turnover over the past 12 months ($\overline{TURN}$), idiosyncratic volatility ($IVOL$), momentum (MOM), book-to-market ratio (BM), institutional ownership (IO), a dummy variable equal to one for stocks included in the S&P 500 index ($SP500$), and a dummy variable equal to one if the firm announces quarterly earnings during the month (EAM). See Table A1 for complete variable definitions. Panel A reports equal-weighted average slope coefficients and Panel B reports Litzenberger and Ramaswamy (1979) precision-weighted average slope coefficients. The average adjusted R^2 is reported in the last row. Newey and West (1987) t-statistics are given in parentheses. 10%(*), 5%(**), and 1%(***) significance levels for two-sided tests are denoted. This regression analysis is based on 212,797 stock-month observations from March 2003 to December 2016 with no missing values for all variables.

Table 15: Learning prior to earnings announcement month: Portfolios of stocks sorted by lagged learning index controlling for firm size

Quintile	$\lvert CAR\rvert_d$	$\lvert CAR\rvert_{d-1,d+1}$	$\lvert CAR\rvert_{q+1}$	$ATURN_{t-1}$
1 (Low LI_{t-1})	2.534	4.425	12.095	−0.409
2	2.477	4.325	11.876	0.889
3	2.416	4.304	11.812	2.236
4	2.450	4.327	11.724	2.464
5 (High LI_{t-1})	2.415	4.211	11.757	1.838
5−1	-0.119^{**}	-0.214^{***}	-0.338^{*}	2.247^{**}
t-stat	(−2.49)	(−2.65)	(−1.65)	(2.45)

At the end of each month, all stocks with a quarterly earnings announcement during the month are sorted into quintiles based on lagged market capitalization. Within each size quintile, stocks are sorted based on lagged values of the learning index (LI). Each LI subquintile is combined across size quintiles into a single quintile. This approach creates portfolios of stocks with differences in LI but similar distributions of size. The table reports time-series means of quintile averages of three measures of market reaction and a measure of abnormal trading activity. Market reaction proxies include the absolute value of the cumulative abnormal return on the earnings announcement date d $\left(\lvert CAR\rvert_d\right)$, over the three-day period around the announcement date $\left(\lvert CAR\rvert_{d-1,d+1}\right)$, and during the period from two days after the earnings announcement date through one day after the firm's next quarterly earnings announcement date $\left(\lvert CAR\rvert_{q+1}\right)$. The proxy for abnormal trading activity is abnormal share turnover in the month prior to the earnings announcement ($ATURN_{t-1}$). See Table A1 for complete variable definitions. The row labeled "5 – 1" presents the difference in the respective dependent variable between the highest and lowest quintile portfolios. Newey and West (1987) t-statistics and 10%(*), 5%(**), and 1%(***) significance levels for two-sided tests are given for the 5 – 1 portfolio. The sample period is October 1971 to December 2016.

Table 16: Variation in earnings announcement activity: Portfolios of stocks sorted by lagged learning index controlling for firm size

Quintile	$\lvert CAR\rvert_d$	$\lvert CAR\rvert_{d-1,d+1}$	$\lvert CAR\rvert_{q+1}$	$ATURN_{t-1}$
Panel A: Months with number of earnings announcements above yearly median				
1 (Low LI_{t-1})	2.360	4.156	11.399	2.143
2	2.302	4.013	11.150	3.803
3	2.316	4.049	11.348	4.856
4	2.302	4.067	11.346	4.827
5 (High LI_{t-1})	2.290	4.022	11.330	4.260
5−1	−0.070	−0.134	−0.069	2.117**
t-stat	(−1.12)	(−1.15)	(−0.25)	(2.25)
Panel B: Months with number of earnings announcements below yearly median				
1 (Low LI_{t-1})	2.710	4.694	12.793	−2.971
2	2.653	4.639	12.605	−2.036
3	2.517	4.561	12.278	−0.394
4	2.599	4.588	12.103	0.093
5 (High LI_{t-1})	2.541	4.400	12.186	−0.594
5−1	−0.169**	−0.295***	−0.608**	2.377
t-stat	(−2.32)	(−2.83)	(−2.19)	(1.57)

At the end of each month, all stocks with a quarterly earnings announcement during the month are sorted into quintiles based on lagged market capitalization. Within each size quintile, stocks are sorted based on lagged values of the learning index (LI). Each LI subquintile is combined across size quintiles into a single quintile. This approach creates portfolios of stocks with differences in LI but similar distributions of size. The sample period is split into two groups: months with high earnings announcement activity (above the yearly median) and months with low earnings announcement activity (below the yearly median). For each of these groups, the table reports time-series means of quintile averages of three measures of market reaction and a measure of abnormal trading activity. Market reaction proxies include the absolute value of the cumulative abnormal return on the earnings announcement date d $\left(\lvert CAR\rvert_d\right)$, over the three-day period around the announcement date $\left(\lvert CAR\rvert_{d-1,d+1}\right)$, and during the period from two days after the earnings announcement date through one day after the firm's next quarterly earnings announcement date $\left(\lvert CAR\rvert_{q+1}\right)$. The proxy for abnormal trading activity is abnormal share turnover in the month prior to the earnings announcement ($ATURN_{t-1}$). See Table A1 for complete variable definitions. The row labeled "5 − 1" presents the difference in the respective dependent variable between the highest and lowest quintile portfolios. Newey and West (1987) t-statistics and 10%(*), 5%(**), and 1%(***) significance levels for two-sided tests are given for the 5 − 1 portfolio. The sample period for analysis of market reaction and trading activity is October 1971 to December 2016.

Table 17: Learning during earnings announcement month: Portfolios of stocks sorted by contemporaneous learning index controlling for firm size

Quintile	$\lvert CAR\rvert_d$	$\lvert CAR\rvert_{d-1,d+1}$	$\lvert CAR\rvert_{q+1}$	$ATURN_{d-1,d+1}$
1 (Low LI_t)	2.397	4.114	12.187	52.331
2	2.442	4.275	11.844	56.569
3	2.444	4.323	11.723	59.634
4	2.475	4.412	11.726	63.144
5 (High LI_t)	2.536	4.463	11.805	66.748
5−1	0.139^{***}	0.349^{***}	-0.381^{*}	14.417^{***}
t-stat	(3.17)	(4.44)	(−1.68)	(8.95)

At the end of each month, all stocks with a quarterly earnings announcement during the month are sorted into quintiles based on market capitalization. Within each size quintile, stocks are sorted based on contemporaneous values of the learning index (LI). Each LI subquintile is combined across size quintiles into a single quintile. This approach creates portfolios of stocks with differences in LI but similar distributions of size. The table reports time-series means of quintile averages of three measures of market reaction and a measure of abnormal trading activity around the earnings announcement date. Market reaction proxies include the absolute value of the cumulative abnormal return on the earnings announcement date d $\left(\lvert CAR\rvert_d\right)$, over the three-day period around the announcement date $\left(\lvert CAR\rvert_{d-1,d+1}\right)$, and during the period from two days after the earnings announcement date through one day after the firm's next quarterly earnings announcement date $\left(\lvert CAR\rvert_{q+1}\right)$. The proxy for abnormal trading activity is abnormal daily turnover around the earnings announcement date ($ATURN_{d-1,d+1}$). See Table A1 for complete variable definitions. The row labeled "5 − 1" presents the difference in the respective dependent variable between the highest and lowest quintile portfolios. Newey and West (1987) t-statistics and 10%(*), 5%(**), and 1%(***) significance levels for two-sided tests are given for the $5-1$ portfolio. The sample period is October 1971 to December 2016.

Table 18: Variation in earnings announcement activity: Portfolios of stocks sorted by contemporaneous learning index controlling for firm size

Quintile	$\lvert CAR\rvert_d$	$\lvert CAR\rvert_{d-1,d+1}$	$\lvert CAR\rvert_{q+1}$	$ATURN_{d-1,d+1}$
Panel A: Months with number of earnings announcements above yearly median				
1 (Low LI_t)	2.249	3.873	11.436	46.192
2	2.289	3.999	11.153	48.871
3	2.316	4.045	11.263	51.457
4	2.349	4.183	11.303	54.754
5 (High LI_t)	2.365	4.203	11.416	58.427
5−1	0.116^{**}	0.330^{***}	−0.020	12.235^{***}
t-stat	(1.98)	(2.84)	(−0.07)	(6.89)
Panel B: Months with number of earnings announcements below yearly median				
1 (Low LI_t)	2.545	4.355	12.940	58.492
2	2.597	4.551	12.538	64.296
3	2.573	4.601	12.184	67.841
4	2.600	4.642	12.151	71.566
5 (High LI_t)	2.708	4.724	12.195	75.099
5−1	0.163^{**}	0.369^{***}	-0.745^{**}	16.607^{***}
t-stat	(2.54)	(3.48)	(−2.41)	(6.34)

At the end of each month, all stocks with a quarterly earnings announcement during the month are sorted into quintiles based on market capitalization. Within each size quintile, stocks are sorted based on contemporaneous values of the learning index (LI). Each LI subquintile is combined across size quintiles into a single quintile. This approach creates portfolios of stocks with differences in LI but similar distributions of size. The sample period is split into two groups: months with high earnings announcement activity (above the yearly median) and months with low earnings announcement activity (below the yearly median). For each of these groups, the table reports time-series means of quintile averages of three measures of market reaction and a measure of abnormal trading activity around the earnings announcement date. Market reaction proxies include the absolute value of the cumulative abnormal return on the earnings announcement date d $\left(\lvert CAR\rvert_d\right)$, over the three-day period around the announcement date $\left(\lvert CAR\rvert_{d-1,d+1}\right)$, and during the period from two days after the earnings announcement date through one day after the firm's next quarterly earnings announcement date $\left(\lvert CAR\rvert_{q+1}\right)$. The proxy for abnormal trading activity is abnormal daily turnover around the earnings announcement date ($ATURN_{d-1,d+1}$). See Table A1 for complete variable definitions. The row labeled "5 – 1" presents the difference in the respective dependent variable between the highest and lowest quintile portfolios. Newey and West (1987) t-statistics and 10%(*), 5%(**), and 1%(***) significance levels for two-sided tests are given for the 5 – 1 portfolio. The sample period is October 1971 to December 2016.

Table 19: Learning during earnings announcement month: Portfolios of stocks sorted by changes in learning index controlling for firm size

Quintile	$\lvert CAR\rvert_d$	$\lvert CAR\rvert_{d-1,d+1}$	$\lvert CAR\rvert_{q+1}$	$ATURN_{d-1,d+1}$
1 (Low $LI_t - LI_{t-1}$)	2.267	3.966	11.782	52.841
2	2.384	4.095	11.809	53.768
3	2.435	4.254	11.953	57.461
4	2.504	4.429	11.900	61.328
5 (High $LI_t - LI_{t-1}$)	2.695	4.852	11.796	73.636
5−1	0.428^{***}	0.885^{***}	0.014	20.795^{***}
t-stat	(10.55)	(12.64)	(0.11)	(12.04)

At the end of each month, all stocks with a quarterly earnings announcement during the month are sorted into quintiles based on market capitalization. Within each size quintile, stocks are sorted based on the change in the learning index (LI), defined as the value of LI at the end of the current month t minus the value of LI at the end of month $t-1$. Each $LI_t - LI_{t-1}$ subquintile is combined across size quintiles into a single quintile. This approach creates portfolios of stocks with differences in $LI_t - LI_{t-1}$ but similar distributions of size. The table reports time-series means of quintile averages of three measures of market reaction and a measure of abnormal trading activity around the earnings announcement date. Market reaction proxies include the absolute value of the cumulative abnormal return on the earnings announcement date d $\left(\lvert CAR\rvert_d\right)$, over the three-day period around the announcement date $\left(\lvert CAR\rvert_{d-1,d+1}\right)$, and during the period from two days after the earnings announcement date through one day after the firm's next quarterly earnings announcement date $\left(\lvert CAR\rvert_{q+1}\right)$. The proxy for abnormal trading activity is abnormal daily turnover around the earnings announcement date ($ATURN_{d-1,d+1}$). See Table A1 for complete variable definitions. The row labeled "5 – 1" presents the difference in the respective dependent variable between the highest and lowest quintile portfolios. Newey and West (1987) t-statistics and 10%(*), 5%(**), and 1%(***) significance levels for two-sided tests are given for the 5 – 1 portfolio. The sample period is October 1971 to December 2016.

Table 20: Variation in earnings announcement activity: Portfolios of stocks sorted by changes in learning index controlling for firm size

Quintile	$\lvert CAR\rvert_d$	$\lvert CAR\rvert_{d-1,d+1}$	$\lvert CAR\rvert_{q+1}$	$ATURN_{d-1,d+1}$
Panel A: Months with number of earnings announcements above yearly median				
1 (Low $LI_t - LI_{t-1}$)	2.168	3.754	11.259	44.721
2	2.250	3.867	11.353	47.387
3	2.280	4.015	11.332	50.464
4	2.361	4.146	11.352	52.579
5 (High $LI_t - LI_{t-1}$)	2.508	4.522	11.267	64.623
5−1	0.341***	0.768***	0.008	19.902***
t-stat	(9.71)	(10.47)	(0.08)	(12.65)
Panel B: Months with number of earnings announcements below yearly median				
1 (Low $LI_t - LI_{t-1}$)	2.367	4.179	12.307	60.991
2	2.519	4.324	12.266	60.173
3	2.590	4.494	12.577	64.483
4	2.648	4.714	12.450	70.109
5 (High $LI_t - LI_{t-1}$)	2.883	5.183	12.327	82.683
5−1	0.516***	1.003***	0.020	21.692***
t-stat	(7.13)	(9.16)	(0.09)	(7.65)

At the end of each month, all stocks with a quarterly earnings announcement during the month are sorted into quintiles based on market capitalization. Within each size quintile, stocks are sorted based on the change in the learning index (LI), defined as the value of LI at the end of the current month t minus the value of LI at the end of month $t-1$. Each $LI_t - LI_{t-1}$ subquintile is combined across size quintiles into a single quintile. This approach creates portfolios of stocks with differences in $LI_t - LI_{t-1}$ but similar distributions of size. The sample period is split into two groups: months with high earnings announcement activity (above the yearly median) and months with low earnings announcement activity (below the yearly median). For each of these groups, the table reports time-series means of quintile averages of three measures of market reaction and a measure of abnormal trading activity around the earnings announcement date. Market reaction proxies include the absolute value of the cumulative abnormal return on the earnings announcement date d $\left(\lvert CAR\rvert_d\right)$, over the three-day period around the announcement date $\left(\lvert CAR\rvert_{d-1,d+1}\right)$, and during the period from two days after the earnings announcement date through one day after the firm's next quarterly earnings announcement date $\left(\lvert CAR\rvert_{q+1}\right)$. The proxy for abnormal trading activity is abnormal daily turnover around the earnings announcement date ($ATURN_{d-1,d+1}$). See Table A1 for complete variable definitions. The row labeled "5 – 1" presents the difference in the respective dependent variable between the highest and lowest quintile portfolios. Newey and West (1987) t-statistics and 10%(*), 5%(**), and 1%(***) significance levels for two-sided tests are given for the 5 – 1 portfolio. The sample period is October 1971 to December 2016.

Table 21: Learning costs and return predictability: Portfolios of stocks sorted by number of business segments and learning index controlling for firm size

Panel A: Value-weighted portfolios						
	Excess return			FF6 α		
	Low $nSEG$		High $nSEG$	Low $nSEG$		High $nSEG$
	1	2	3	1	2	3
1 (Low LI)	1.324	1.265	1.200	0.484	0.382	−0.047
2	1.170	1.203	1.129	0.222	0.159	−0.091
3	1.000	0.953	1.012	0.067	−0.044	−0.165
4	0.856	0.675	0.858	−0.133	−0.278	−0.228
5 (High LI)	0.519	0.767	0.691	−0.595	−0.188	−0.363
5−1	-0.805^{***}	-0.498^{*}	-0.508^{***}	-1.079^{***}	-0.570^{**}	-0.316^{*}
t-stat	(−3.47)	(−1.96)	(−3.28)	(−4.40)	(−2.58)	(−1.80)

Panel B: Equal-weighted portfolios						
	Excess return			FF6 α		
	Low $nSEG$		High $nSEG$	Low $nSEG$		High $nSEG$
	1	2	3	1	2	3
1 (Low LI)	1.537	1.382	1.469	0.398	0.215	0.079
2	1.326	1.312	1.424	0.221	0.071	0.047
3	1.064	1.012	1.154	−0.019	−0.215	−0.233
4	1.050	0.893	1.111	−0.098	−0.349	−0.252
5 (High LI)	0.695	0.854	0.903	−0.527	−0.314	−0.445
5−1	-0.842^{***}	-0.527^{***}	-0.567^{***}	-0.925^{***}	-0.528^{***}	-0.524^{***}
t-stat	(−5.11)	(−2.98)	(−4.13)	(−4.71)	(−3.35)	(−3.79)

This table presents results from trivariate portfolio sorting based on market capitalization, number of industry segments within the firm ($nSEG$), and the learning index (LI). At the end of each month, stocks are sorted into quintiles based on size. Within each size quintile, stocks are sorted into terciles based on $nSEG$. Each $nSEG$ tercile is combined across size quintiles into a single tercile. This approach creates three portfolios of stocks with differences in $nSEG$ but similar distributions of size. Within each of these three portfolios, stocks are sorted into quintiles based on LI. For each of the resulting 15 portfolios, the table reports the next month value-weighted (Panel A) and equal-weighted (Panel B) average excess return and risk-adjusted average excess return (alpha or α). Returns are risk-adjusted using the Fama and French (2018) six-factor model. See Table A1 for complete variable definitions. The row labeled "5 – 1" presents the difference in the respective dependent variable between the highest and lowest LI quintile portfolios within each $nSEG$ tercile. Newey and West (1987) t-statistics and 10%(*), 5%(**), and 1%(***) significance levels for two-sided tests are given for the 5 – 1 portfolio. The sample period is July 1975 to December 2016.

Table 22: Learning costs and volatility predictability: Portfolios of stocks sorted by number of business segments and learning index controlling for firm size

Panel A: Value-weighted portfolios						
	ARVOL			*RVOL*		
	Low *nSEG*		High *nSEG*	Low *nSEG*		High *nSEG*
	1	2	3	1	2	3
1 (Low *LI*)	2.944	3.314	3.275	32.173	30.192	27.697
2	0.472	0.797	1.325	30.518	28.850	27.292
3	−0.004	0.725	−0.216	30.739	28.705	26.611
4	−0.037	0.784	0.153	29.947	28.173	26.422
5 (High *LI*)	−1.084	−1.917	−0.931	29.045	26.870	26.320
5−1	-4.028^{***}	-5.231^{***}	-4.206^{***}	-3.128^{***}	-3.322^{***}	-1.376^{***}
t-stat	(−4.10)	(−6.17)	(−5.36)	(−2.94)	(−4.80)	(−2.65)

Panel B: Equal-weighted portfolios						
	ARVOL			*RVOL*		
	Low *nSEG*		High *nSEG*	Low *nSEG*		High *nSEG*
	1	2	3	1	2	3
1 (Low *LI*)	2.773	2.604	2.850	37.647	34.150	32.806
2	0.579	1.077	1.499	36.764	33.483	32.585
3	0.076	0.272	0.755	36.576	33.081	32.508
4	−0.545	−0.351	−0.185	36.411	32.747	32.096
5 (High *LI*)	−1.514	−1.659	−1.022	35.797	32.103	31.920
5−1	-4.286^{***}	-4.263^{***}	-3.871^{***}	-1.850^{***}	-2.047^{***}	-0.885^{**}
t-stat	(−5.94)	(−7.34)	(−5.88)	(−3.75)	(−3.57)	(−2.04)

This table presents results from trivariate portfolio sorting based on market capitalization, number of industry segments within the firm (*nSEG*), and the learning index (*LI*). At the end of each month, stocks are sorted into quintiles based on size. Within each size quintile, stocks are sorted into terciles based on *nSEG*. Each *nSEG* tercile is combined across size quintiles into a single tercile. This approach creates three portfolios of stocks with differences in *nSEG* but similar distributions of size. Within each of these three portfolios, stocks are sorted into quintiles based on *LI*. For each of the resulting 15 portfolios, the table reports the next month value-weighted (Panel A) and equal-weighted (Panel B) average abnormal volatility (*ARVOL*) and level of volatility (*RVOL*). When *RVOL* is the dependent variable, a bivariate dependent sorting approach is used to control for the past level of volatility across *LI* quintiles. See Table A1 for complete variable definitions. The row labeled "5 – 1" presents the difference in the respective dependent variable between the highest and lowest *LI* quintile portfolios within each *nSEG* tercile. Newey and West (1987) t-statistics and 10%(*), 5%(**), and 1%(***) significance levels for two-sided tests are given for the 5 – 1 portfolio. The sample period is July 1975 to December 2016.

Table 23: Learning costs and return predictability: Portfolios of stocks sorted by firm complexity and learning index controlling for firm size

	Excess return			FF6 α		
	Low $COMPLEX$		High $COMPLEX$	Low $COMPLEX$		High $COMPLEX$
	1	2	3	1	2	3
Panel A: Value-weighted portfolios						
1 (Low LI)	1.324	1.192	1.221	0.470	0.241	0.003
2	1.184	1.118	1.150	0.227	0.035	−0.053
3	1.012	1.028	0.945	0.068	0.024	−0.258
4	0.842	0.777	0.803	−0.162	−0.158	−0.325
5 (High LI)	0.568	0.616	0.735	−0.550	−0.353	−0.315
5−1	−0.756***	−0.576***	−0.486***	−1.020***	−0.595***	−0.318*
t-stat	(−3.16)	(−2.92)	(−2.82)	(−4.03)	(−3.10)	(−1.85)
Panel B: Equal-weighted portfolios						
1 (Low LI)	1.549	1.386	1.426	0.410	0.134	0.054
2	1.345	1.315	1.394	0.238	0.070	0.004
3	1.084	1.029	1.156	−0.019	−0.229	−0.206
4	1.033	0.977	1.073	−0.130	−0.247	−0.294
5 (High LI)	0.701	0.827	0.878	−0.530	−0.376	−0.456
5−1	−0.848***	−0.559***	−0.548***	−0.940***	−0.511***	−0.510***
t-stat	(−4.99)	(−3.45)	(−3.75)	(−4.60)	(−3.51)	(−3.80)

This table presents results from trivariate portfolio sorting based on market capitalization, firm complexity based on sales concentration among industry segments within the firm ($COMPLEX$), and the learning index (LI). At the end of each month, stocks are sorted into quintiles based on size. Within each size quintile, stocks are sorted into terciles based on $COMPLEX$. Each $COMPLEX$ tercile is combined across size quintiles into a single tercile. This approach creates three portfolios of stocks with differences in $COMPLEX$ but similar distributions of size. Within each of these three portfolios, stocks are sorted into quintiles based on LI. For each of the resulting 15 portfolios, the table reports the next month value-weighted (Panel A) and equal-weighted (Panel B) average excess return and risk-adjusted average excess return (alpha or α). Returns are risk-adjusted using the Fama and French (2018) six-factor model. See Table A1 for complete variable definitions. The row labeled "5 – 1" presents the difference in the respective dependent variable between the highest and lowest LI quintile portfolios within each $COMPLEX$ tercile. Newey and West (1987) t-statistics and 10%(*), 5%(**), and 1%(***) significance levels for two-sided tests are given for the 5 – 1 portfolio. The sample period is July 1975 to December 2016.

Table 24: Learning costs and volatility predictability: Portfolios of stocks sorted by firm complexity and learning index controlling for firm size

	ARVOL			*RVOL*		
	Low *COMPLEX*		High *COMPLEX*	Low *COMPLEX*		High *COMPLEX*
	1	2	3	1	2	3
1 (Low *LI*)	2.915	3.602	3.061	32.266	29.314	27.899
2	0.478	1.096	1.212	30.645	28.479	27.313
3	−0.149	0.310	0.117	30.784	28.211	26.681
4	−0.019	0.800	0.156	30.026	28.086	26.209
5 (High *LI*)	−1.219	−1.981	−1.072	29.122	26.741	26.187
5−1	−4.134***	−5.583***	−4.133***	−3.143***	−2.574***	−1.712***
t-stat	(−4.33)	(−5.80)	(−5.50)	(−2.94)	(−3.51)	(−3.18)

Panel A: Value-weighted portfolios

Panel B: Equal-weighted portfolios

	ARVOL			*RVOL*		
	Low *COMPLEX*		High *COMPLEX*	Low *COMPLEX*		High *COMPLEX*
	1	2	3	1	2	3
1 (Low *LI*)	2.762	2.653	2.886	37.734	34.272	32.650
2	0.607	0.956	1.598	36.807	33.705	32.514
3	0.013	0.220	0.774	36.604	33.333	32.417
4	−0.489	−0.408	−0.196	36.497	33.252	31.903
5 (High *LI*)	−1.499	−1.488	−1.105	35.853	32.855	31.655
5−1	−4.261***	−4.141***	−3.992***	−1.880***	−1.417***	−0.995**
t-stat	(−5.88)	(−6.69)	(−6.33)	(−3.86)	(−3.25)	(−2.31)

This table presents results from trivariate portfolio sorting based on market capitalization, firm complexity based on sales concentration among industry segments within the firm (*COMPLEX*), and the learning index (*LI*). At the end of each month, stocks are sorted into quintiles based on size. Within each size quintile, stocks are sorted into terciles based on *COMPLEX*. Each *COMPLEX* tercile is combined across size quintiles into a single tercile. This approach creates three portfolios of stocks with differences in *COMPLEX* but similar distributions of size. Within each of these three portfolios, stocks are sorted into quintiles based on *LI*. For each of the resulting 15 portfolios, the table reports the next month value-weighted (Panel A) and equal-weighted (Panel B) average abnormal volatility (*ARVOL*) and level of volatility (*RVOL*). When *RVOL* is the dependent variable, a bivariate dependent sorting approach is used to control for the past level of volatility across *LI* quintiles. See Table A1 for complete variable definitions. The row labeled "5 – 1" presents the difference in the respective dependent variable between the highest and lowest *LI* quintile portfolios within each *COMPLEX* tercile. Newey and West (1987) t-statistics and 10%(*), 5%(**), and 1%(***) significance levels for two-sided tests are given for the 5 – 1 portfolio. The sample period is July 1975 to December 2016.

Table 25: Alternative asset pricing models: Portfolios of stocks sorted by learning index

Panel A: Value-weighted portfolios				
Quintile	7-factor α	9-factor α	SY (2017) α	HXZ (2015) α
1 (Low LI)	0.255	0.297	0.267	0.702
2	0.008	0.016	0.050	0.406
3	−0.046	−0.068	−0.024	0.299
4	−0.113	−0.139	−0.089	0.224
5 (High LI)	−0.282	−0.314	−0.239	0.067
5−1	-0.537^{***}	-0.611^{***}	-0.506^{***}	-0.635^{***}
t-stat	(−3.51)	(−3.56)	(−3.69)	(−4.01)

Panel B: Equal-weighted portfolios				
Quintile	7-factor α	9-factor α	SY (2017) α	HXZ (2015) α
1 (Low LI)	0.306	0.325	0.432	0.749
2	0.152	0.152	0.220	0.579
3	−0.027	−0.052	0.016	0.373
4	−0.092	−0.125	−0.046	0.294
5 (High LI)	−0.254	−0.291	−0.240	0.122
5−1	-0.560^{***}	-0.615^{***}	-0.672^{***}	-0.627^{***}
t-stat	(−4.32)	(−4.03)	(−5.23)	(−5.04)

At the end of each month, stocks are sorted into quintiles based on values of the learning index (LI). The table reports the next month value-weighted (Panel A) and equal-weighted (Panel B) risk-adjusted excess return (alpha or α) for each quintile. 7-factor α is computed with respect to a seven factor model that includes the market, size, value, profitability, investment, and momentum factors of Fama and French (2018) as well as the liquidity factor of Pastor and Stambaugh (2003). 9-factor α is computed with respect to a nine factor model that includes the seven aforementioned factors as well as a short-term reversal factor and a long-term reversal factor. SY (2017) α is computed with respect to the Stambaugh and Yuan (2017) factor model. HXZ (2015) α is computed with respect to the Hou et al. (2015) q-factor model. The row labeled "5 – 1" presents the difference in alpha between the highest and lowest quintile portfolios. Newey and West (1987) t-statistics and 10%(*), 5%(**), and 1%(***) significance levels for two-sided tests are given for the 5 – 1 portfolio. The sample period is July 1964 to December 2016.

Table 26: Explaining the cross-section of volatility: Portfolios of stocks sorted by learning index controlling for past volatility

	Panel A: Value-weighted portfolios			Panel B: Equal-weighted portfolios		
Quintile	$RVOL$	$SVOL$	$IVOL$	$RVOL$	$SVOL$	$IVOL$
			Bivariate Dependent Sorting			
1 (Low LI)	28.719	21.217	18.279	34.450	23.181	24.392
2	27.687	20.137	17.950	34.089	22.812	24.252
3	27.313	19.696	18.046	33.822	22.581	24.154
4	27.205	19.399	17.996	33.693	22.423	24.080
5 (High LI)	26.663	18.958	17.947	33.288	22.020	23.948
5−1	-2.057^{***}	-2.258^{***}	-0.332	-1.162^{***}	-1.161^{***}	-0.444^{**}
t-stat	(-4.14)	(-5.77)	(-1.05)	(-3.97)	(-5.02)	(-2.57)
			Bivariate Independent Sorting			
1 (Low LI)	33.868	22.868	23.574	34.543	23.219	24.468
2	33.116	22.224	23.256	34.103	22.789	24.300
3	32.901	22.025	23.237	33.956	22.634	24.247
4	32.752	21.843	23.124	33.727	22.409	24.165
5 (High LI)	32.318	21.486	22.962	33.363	22.036	24.009
5−1	-1.550^{***}	-1.383^{***}	-0.613^{***}	-1.179^{***}	-1.183^{***}	-0.459^{***}
t-stat	(-5.46)	(-6.55)	(-3.46)	(-4.06)	(-5.21)	(-2.67)

The table presents results regarding the cross-sectional relationship between the learning index (LI) and the level of volatility in the following month, controlling for the past level of volatility using two methods of bivariate portfolio sorting. Dependent sorting: At the end of each month, stocks are sorted into quintiles based on average volatility over the past 12 months. Within each volatility quintile, stocks are sorted based on values of LI. Each LI subquintile is combined across volatility quintiles into a single quintile. Independent sorting: At the end of each month, stocks are sorted independently into quintiles based on average volatility over the past 12 months and values of LI. Average values of the dependent variable are computed within each of the resulting 25 portfolios and then averaged across volatility quintiles within each LI quintile. The table reports the next month value-weighted (Panel A) and equal-weighted (Panel B) quintile average level of return volatility ($RVOL$), systematic volatility ($SVOL$), and idiosyncratic volatility ($IVOL$), controlling for the respective average level of volatility in the prior 12 months. See Table A1 for complete variable definitions. The row labeled "5 − 1" presents the difference in monthly volatility between the highest and lowest quintile portfolios. Newey and West (1987) t-statistics and 10%(*), 5%(**), and 1%(***) significance levels for two-sided tests are given for the 5 − 1 portfolio. The sample period is July 1964 to December 2016.

Table 27: Explaining the cross-section of implied volatility: Portfolios of stocks sorted by learning index

	Panel A: Value-weighted portfolios		Panel B: Equal-weighted portfolios	
Quintile	$ACVOL$	$APVOL$	$ACVOL$	$APVOL$
1 (Low LI)	1.101	1.141	0.579	0.567
2	0.348	0.359	0.174	0.159
3	0.324	0.315	−0.024	−0.015
4	−0.026	−0.199	−0.668	−0.514
5 (High LI)	−1.477	−1.328	−1.746	−1.672
5−1	-2.577^{***}	-2.469^{***}	-2.325^{***}	-2.239^{***}
t-stat	(−3.45)	(−3.46)	(−3.36)	(−3.33)

At the end of each month, stocks are sorted into quintiles based on values of the learning index (LI). The table reports the next month value-weighted (Panel A) and equal-weighted (Panel B) quintile average abnormal call-implied volatility ($ACVOL$) and put-implied volatility ($APVOL$) in the current month relative to the average respective implied volatility in the prior 12 months. See Table A1 for complete variable definitions. The row labeled "5 – 1" presents the difference in abnormal implied volatility between the highest and lowest quintile portfolios. Newey and West (1987) t-statistics and 10%(*), 5%(**), and 1%(***) significance levels for two-sided tests are given for the 5 – 1 portfolio. The sample period is January 1996 to December 2016.

Table 28: Explaining the cross-section of implied volatility: Portfolios of stocks sorted by learning index controlling for past implied volatility

	Panel A: Value-weighted portfolios		Panel B: Equal-weighted portfolios	
Quintile	$CVOL$	$PVOL$	$CVOL$	$PVOL$
		Bivariate Dependent Sorting		
1 (Low LI)	33.529	33.903	41.602	42.109
2	31.880	32.163	41.519	42.052
3	31.202	31.547	41.377	41.869
4	30.612	30.951	41.167	41.700
5 (High LI)	29.740	30.097	40.752	41.305
5−1	-3.789^{***}	-3.806^{***}	-0.850^{**}	-0.804^{**}
t-stat	(−5.52)	(−5.33)	(−2.24)	(−2.17)
		Bivariate Independent Sorting		
1 (Low LI)	40.701	41.238	41.535	42.062
2	40.408	40.930	41.480	41.988
3	40.240	40.744	41.418	41.901
4	39.970	40.481	41.145	41.706
5 (High LI)	39.523	40.079	40.763	41.299
5−1	-1.178^{***}	-1.159^{***}	-0.771^{**}	-0.762^{**}
t-stat	(−2.88)	(−2.84)	(−1.97)	(−1.99)

The table presents results regarding the cross-sectional relationship between the learning index (LI) and the level of option-implied volatility in the following month, controlling for the past level of implied volatility using two methods of bivariate portfolio sorting. Dependent sorting: At the end of each month, stocks are sorted into quintiles based on average implied volatility over the past 12 months. Within each implied volatility quintile, stocks are sorted based on values of LI. Each LI subquintile is combined across implied volatility quintiles into a single quintile. Independent sorting: At the end of each month, stocks are sorted independently into quintiles based on average implied volatility over the past 12 months and values of LI. Average values of the dependent variable are computed within each of the resulting 25 portfolios and then averaged across implied volatility quintiles within each LI quintile. The table reports the next month value-weighted (Panel A) and equal-weighted (Panel B) quintile average level of call-implied volatility ($CVOL$) and put-implied volatility ($PVOL$), controlling for the respective average level of implied volatility in the prior 12 months. See Table A1 for complete variable definitions. The row labeled "5 – 1" presents the difference in monthly volatility between the highest and lowest quintile portfolios. Newey and West (1987) t-statistics and 10%(*), 5%(**), and 1%(***) significance levels for two-sided tests are given for the 5 – 1 portfolio. The sample period is July 1996 to December 2016.

Table 29: Explaining the cross-section of market beta: Portfolios of stocks sorted by learning index

	Panel A: Value-weighted portfolios		Panel B: Equal-weighted portfolios	
Quintile	$A\beta^{MKT}$	$A\beta_m^{MKT}$	$A\beta^{MKT}$	$A\beta_m^{MKT}$
1 (Low LI)	4.009	9.471	6.546	13.616
2	2.678	9.618	5.373	12.487
3	1.219	7.384	3.581	11.353
4	−0.211	8.626	2.100	9.975
5 (High LI)	−2.313	8.257	−0.611	8.147
5−1	-6.322^{***}	−1.214	-7.157^{***}	-5.469^{***}
t-stat	(−4.81)	(−0.50)	(−7.43)	(−3.02)

At the end of each month, stocks are sorted into quintiles based on values of the learning index (LI). The table reports the next month value-weighted (Panel A) and equal-weighted (Panel B) average abnormal market beta using two measures of beta. β^{MKT} is estimated from a regression of excess stock returns on lagged, current, and lead excess market returns using daily data from the past year. β_m^{MKT} is estimated in a similar manner using only daily data within the month. $A\beta^{MKT}$ is abnormal β^{MKT} in the month following portfolio formation relative to average β^{MKT} in the prior 12 months. $A\beta_m^{MKT}$ is defined in a similar manner using β_m^{MKT}. See Table A1 for complete variable definitions. The row labeled "5 – 1" presents the difference in the respective dependent variable between the highest and lowest quintile portfolios. Newey and West (1987) t-statistics and 10%(*), 5%(**), and 1%(***) significance levels for two-sided tests are given for the 5 – 1 portfolio. The sample period is July 1964 to December 2016.

Table 30: Explaining the cross-section of market beta: Portfolios of stocks sorted by learning index controlling for past market beta

	Panel A: Value-weighted portfolios		Panel B: Equal-weighted portfolios	
Quintile	β^{MKT}	β_m^{MKT}	β^{MKT}	β_m^{MKT}
		Bivariate Dependent Sorting		
1 (Low LI)	1.086	1.095	1.111	1.104
2	1.034	1.041	1.081	1.057
3	1.003	0.996	1.060	1.031
4	0.969	0.964	1.042	1.010
5 (High LI)	0.932	0.912	1.003	0.957
5−1	-0.154^{***}	-0.183^{***}	-0.107^{***}	-0.147^{***}
t-stat	(−6.84)	(−6.89)	(−12.20)	(−9.79)
		Bivariate Independent Sorting		
1 (Low LI)	1.085	1.075	1.112	1.105
2	1.055	1.032	1.081	1.055
3	1.034	1.003	1.059	1.028
4	1.013	0.985	1.040	1.005
5 (High LI)	0.977	0.938	1.003	0.954
5−1	-0.108^{***}	-0.137^{***}	-0.109^{***}	-0.151^{***}
t-stat	(−9.94)	(−8.27)	(−12.33)	(−9.91)

The table presents results regarding the cross-sectional relationship between the learning index (LI) and market beta in the following month, controlling for the past level of market beta using two methods of bivariate portfolio sorting. Dependent sorting: At the end of each month, stocks are sorted into quintiles based on average beta over the past 12 months. Within each implied volatility quintile, stocks are sorted based on values of LI. Each LI subquintile is combined across beta quintiles into a single quintile. Independent sorting: At the end of each month, stocks are sorted independently into quintiles based on average beta over the past 12 months and values of LI. Average values of the dependent variable are computed within each of the resulting 25 portfolios and then averaged across beta quintiles within each LI quintile. The table reports the next month value-weighted (Panel A) and equal-weighted (Panel B) quintile average beta, controlling for the average beta in the prior 12 months. Two measures of beta are considered. β^{MKT} is estimated from a regression of excess stock returns on lagged, current, and lead excess market returns using daily data from the past year. β_m^{MKT} is estimated in a similar manner using only daily data within the month. See Table A1 for complete variable definitions. The row labeled "5 – 1" presents the difference in the respective dependent variable between the highest and lowest quintile portfolios. Newey and West (1987) t-statistics and 10%(*), 5%(**), and 1%(***) significance levels for two-sided tests are given for the 5 – 1 portfolio. The sample period is July 1964 to December 2016.

Table 31: Explaining the cross-section of market beta: Cross-sectional regressions

	Panel A: Equal-weighted coefficient average			Panel B: Precision-weighted coefficient average		
	(1)	(2)	(3)	(4)	(5)	(6)
LI	-0.104^{***}		-0.039^{***}	-0.107^{***}		-0.045^{***}
	(-6.75)		(-3.67)	(-7.27)		(-4.51)
ROE		0.001	0.000		-0.001^{*}	-0.001^{*}
		(0.63)	(0.50)		(-1.67)	(-1.85)
$ROEVOL$		0.001	0.001		0.000	0.000
		(1.01)	(0.99)		(1.64)	(1.59)
AGE		-0.001^{***}	-0.001^{***}		-0.001^{***}	-0.001^{***}
		(-5.01)	(-5.18)		(-4.49)	(-4.72)
$DIVD$		-0.069^{***}	-0.069^{***}		-0.062^{***}	-0.063^{***}
		(-6.30)	(-6.33)		(-6.75)	(-6.85)
LEV		0.004^{**}	0.004^{*}		0.003^{*}	0.002^{*}
		(2.17)	(1.93)		(1.84)	(1.65)
$INVPRC$		0.011^{***}	0.011^{***}		0.010^{***}	0.010^{***}
		(7.97)	(8.24)		(8.16)	(8.08)
R		0.002	0.002		0.001	0.001
		(1.59)	(1.59)		(1.30)	(1.31)
$SIZE$		0.039^{***}	0.038^{***}		0.029^{***}	0.029^{***}
		(4.54)	(4.40)		(3.75)	(3.58)
BM		-0.057^{***}	-0.056^{***}		-0.048^{***}	-0.048^{***}
		(-5.42)	(-5.40)		(-4.62)	(-4.63)
MOM		0.001^{***}	0.001^{***}		0.001^{***}	0.001^{***}
		(3.51)	(3.25)		(4.58)	(4.22)
STR		-0.004^{***}	-0.004^{***}		-0.004^{***}	-0.003^{***}
		(-6.25)	(-6.18)		(-6.95)	(-6.90)
Lagged betas	Yes	Yes	Yes	Yes	Yes	Yes
Adj R^2	0.155	0.200	0.201	0.155	0.200	0.201

This table presents results from two-stage cross-sectional regressions. At the end of each month, I estimate a cross-sectional regression of next month market beta (β_m^{MKT}) on a set of explanatory variables. Each column reports results for a different regression specification. Explanatory variables include an intercept term, the learning index (LI), return on equity (ROE), volatility of return on equity ($ROEVOL$), firm age (AGE), a dividend dummy ($DIVD$), leverage (LEV), inverse of stock price ($INVPRC$), firm size ($SIZE$), book-to-market ratio (BM), momentum (MOM), short-term reversal (STR), next month return (R), and 12 lagged values of β_m^{MKT}. See Table A1 for complete variable definitions. Panel A reports equal-weighted average slope coefficients and Panel B reports Litzenberger and Ramaswamy (1979) precision-weighted average slope coefficients. The average adjusted R^2 is reported in the last row. The intercept term and coefficient estimates for lagged betas are not reported for brevity. Newey and West (1987) t-statistics are given in parentheses. 10%(*), 5%(**), and 1%(***) significance levels for two-sided tests are denoted. This regression analysis is based on 680,543 stock-month observations from December 1974 to December 2016 with no missing values for all variables.

Table 32: Additional control variables: Cross-sectional return regressions

Sample period begins:	Jul 1966				Jul 1984				Mar 2003	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
LI	-0.405^{***}	-0.402^{***}	-0.402^{***}	-0.422^{***}	-0.439^{***}	-0.420^{***}	-0.409^{***}	-0.349^{***}	-0.382^{***}	-0.426^{***}
	(-4.91)	(-4.90)	(-4.89)	(-5.16)	(-4.60)	(-4.50)	(-4.52)	(-3.41)	(-2.97)	(-3.21)
$MAXDRET$		0.035^{***}								
		(3.44)								
$MAX\|DRET\|$			0.012							
			(1.09)							
$ATURN$				0.004^{***}						
				(8.99)						
$nFCST$						0.011^{*}				
						(1.73)				
$nREV$							-0.003			
							(-0.38)			
ΔFA								-0.012^{***}		
								(-6.40)		
$EDGAR$/100										0.002
										(0.66)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj R^2	0.092	0.093	0.093	0.094	0.083	0.091	0.090	0.091	0.078	0.080

This table presents Litzenberger and Ramaswamy (1979) precision-weighted average slope coefficients from two-stage cross-sectional regressions. At the end of each month, I estimate a cross-sectional regression of next month excess stock return on a set of explanatory variables. Each column reports results for a different regression specification. Control variables of interest include the learning index (LI), maximum daily return during the month ($MAXDRET$), maximum absolute daily return during the month ($MAX|DRET|$), abnormal monthly share turnover ($ATURN$), number of analyst forecasts for the nearest fiscal quarter ($nFCST$), number of analyst forecast revisions ($nREV$), change in forecast accuracy (ΔFA), and number of EDGAR downloads ($EDGAR$). Untabulated control variables include an intercept term, the learning index (LI), market beta (β^{MKT}), firm size ($SIZE$), book-to-market ratio (BM), profitability ($PROF$), investment (INV), momentum (MOM), illiquidity ($ILLIQ$), short-term reversal (STR), long-term reversal (LTR), and idiosyncratic volatility ($IVOL$). See Table A1 for complete variable definitions. The average adjusted R^2 is reported in the last row. Newey and West (1987) t-statistics are given in parentheses. 10%(*), 5%(**), and 1%(***) significance levels for two-sided tests are denoted. The first row of the table header indicates the first month with non-missing data for all variables included in the respective specification. All sample periods end in December 2016.

Table 33: Additional control variables: Cross-sectional volatility regressions

Sample period begins:	Dec 1974				Jul 1984				Mar 2003	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
LI	-1.190^{***}	-1.156^{***}	-0.987^{***}	-1.173^{***}	-1.423^{***}	-1.500^{***}	-1.417^{***}	-1.497^{***}	-1.600^{***}	-1.584^{***}
	(-7.68)	(-7.44)	(-7.13)	(-7.64)	(-8.92)	(-8.79)	(-8.28)	(-8.17)	(-7.85)	(-7.41)
$MAXDRET$		-0.339^{***}								
		(-22.00)								
$MAX\|DRET\|$			-0.618^{***}							
			(-18.97)							
$ATURN$				-0.005^{***}						
				(-3.97)						
$nFCST$						0.088^{***}				
						(11.94)				
$nREV$							-0.003			
							(-0.26)			
ΔFA								-0.001		
								(-0.30)		
$EDGAR/100$										-0.006^{*}
										(-1.67)
Lagged volatilities	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj R^2	0.533	0.528	0.531	0.527	0.527	0.530	0.530	0.523	0.504	0.506

This table presents Litzenberger and Ramaswamy (1979) precision-weighted average slope coefficients from two-stage cross-sectional regressions. At the end of each month, I estimate a cross-sectional regression of next month return volatility ($RVOL$) on a set of explanatory variables. Each column reports results for a different regression specification. Control variables of interest include the learning index (LI), maximum daily return during the month ($MAXDRET$), maximum absolute daily return during the month ($MAX|DRET|$), abnormal monthly share turnover ($ATURN$), number of analyst forecasts for the nearest fiscal quarter ($nFCST$), number of analyst forecast revisions ($nREV$), change in forecast accuracy (ΔFA), and number of EDGAR downloads ($EDGAR$). Untabulated control variables include an intercept term, the learning index (LI), return on equity (ROE), volatility of return on equity ($ROEVOL$), firm age (AGE), a dividend dummy ($DIVD$), leverage (LEV), inverse of stock price ($INVPRC$), firm size ($SIZE$), book-to-market ratio (BM), momentum (MOM), short-term reversal (STR), next month return (R), and 12 lagged values of volatility. See Table A1 for complete variable definitions. The average adjusted R^2 is reported in the last row. Newey and West (1987) t-statistics are given in parentheses. 10%(*), 5%(**), and 1%(***) significance levels for two-sided tests are denoted. The first row of the table header indicates the first month with non-missing data for all variables included in the respective specification. All sample periods end in December 2016.

Table 34: Subperiod analysis: Portfolios of stocks sorted by learning index

	Panel A: Value-weighted portfolios					Panel B: Equal-weighted portfolios				
Quintile	Return	FF6 α	$ARVOL$	$ASVOL$	$AIVOL$	Return	FF6 α	$ARVOL$	$ASVOL$	$AIVOL$
					Sample period: July 1964 – December 1989					
1 (Low LI)	1.187	0.117	3.392	5.310	1.972	1.470	0.265	2.706	4.279	2.034
2	0.940	−0.044	1.841	3.121	1.397	1.262	0.127	1.655	3.033	1.251
3	0.889	0.057	1.570	2.847	1.125	1.056	0.034	1.698	2.982	1.358
4	0.841	0.095	1.775	2.958	1.297	1.031	0.063	1.489	2.624	1.277
5 (High LI)	0.642	−0.144	0.903	2.074	0.804	0.872	−0.093	0.614	1.569	0.633
5−1	−0.545***	−0.261	−2.489***	−3.236***	−1.168	−0.598***	−0.359**	−2.092**	−2.710***	−1.401*
t-stat	(−3.30)	(−1.35)	(−2.88)	(−3.12)	(−1.62)	(−3.95)	(−2.08)	(−2.35)	(−2.61)	(−1.84)
					Sample period: January 1990 – December 2016					
1 (Low LI)	1.091	0.215	3.525	4.606	2.470	1.326	0.299	3.484	4.970	2.522
2	1.011	0.104	1.419	2.349	0.838	1.237	0.184	1.677	3.074	0.872
3	0.835	−0.092	0.251	1.423	−0.491	1.043	−0.034	0.816	2.042	0.064
4	0.679	−0.202	0.243	0.969	−0.169	0.954	−0.154	−0.271	0.824	−0.868
5 (High LI)	0.633	−0.269	−1.673	−1.170	−1.630	0.781	−0.304	−1.525	−0.484	−2.013
5−1	−0.458**	−0.483**	−5.198***	−5.776***	−4.099***	−0.546***	−0.603***	−5.009***	−5.454***	−4.535***
t-stat	(−1.99)	(−2.59)	(−5.70)	(−4.99)	(−6.09)	(−2.95)	(−3.65)	(−7.64)	(−7.14)	(−8.26)

At the end of each month, stocks are sorted into quintiles based on values of the learning index (LI). The table reports the next month value-weighted (Panel A) and equal-weighted (Panel B) quintile average of the following variables: excess return, Fama and French (2018) six-factor risk-adjusted excess return (alpha or α), abnormal return volatility ($ARVOL$), abnormal systematic volatility ($ASVOL$), and abnormal idiosyncratic volatility ($AIVOL$) relative to the average return volatility in the prior 12 months. See Table A1 for complete variable definitions. The row labeled "5 – 1" presents the difference in the respective dependent variable between the highest and lowest quintile portfolios. Newey and West (1987) t-statistics and 10%(*), 5%(**), and 1%(***) significance levels for two-sided tests are given for the 5 – 1 portfolio. The first sample period is July 1964 to December 1989 and the second sample period is January 1990 to December 2016.

Table 35: Subperiod analysis: Learning index coefficient from cross-sectional regressions

	Panel A: Equal-weighted average LI coefficient		Panel B: Precision-weighted average LI coefficient	
Dependent Variable	Jul 1966 – Dec 1989	Jan 1990 – Dec 2016	Jul 1966 – Dec 1989	Jan 1990 – Dec 2016
Return	-0.422^{***}	-0.397^{***}	-0.411^{***}	-0.411^{***}
	(-2.90)	(-2.91)	(-3.67)	(-4.09)
	Dec 1974 – Dec 1995	Jan 1996 – Dec 2016	Dec 1974 – Dec 1995	Jan 1996 – Dec 2016
$RVOL$	-0.780^{***}	-1.935^{***}	-0.686^{***}	-1.713^{***}
	(-4.31)	(-7.14)	(-4.23)	(-7.98)
$SVOL$	-0.369^{***}	-1.220^{***}	-0.280^{**}	-1.026^{***}
	(-2.64)	(-5.83)	(-2.28)	(-6.45)
$IVOL$	-0.489^{***}	-1.382^{***}	-0.456^{***}	-1.262^{***}
	(-3.69)	(-7.45)	(-3.53)	(-8.20)

This table presents results from two-stage cross-sectional regressions. At the end of each month, I estimate cross-sectional regressions for each of the dependent variables listed in the first column on the learning index (LI) and a set of control variables following the respective full specification described in the text (equations 3.2 and 3.3). Specifically, in return regressions, I control for firm size ($SIZE$), book-to-market ratio (BM), profitability ($PROF$), investment (INV), momentum (MOM), illiquidity ($ILLIQ$), short-term reversal (STR), long-term reversal (LTR), and idiosyncratic volatility ($IVOL$). In volatility regressions, I control for return on equity (ROE), volatility of return on equity ($ROEVOL$), firm age (AGE), a dividend dummy ($DIVD$), leverage (LEV), inverse of stock price ($INVPRC$), firm size ($SIZE$), book-to-market ratio (BM), momentum (MOM), short-term reversal (STR), next month return (R), and 12 lagged values of volatility. See Table A1 for complete variable definitions. All regressions include an intercept term. The table reports only the average coefficient estimate for LI; coefficient estimates for control variables are not reported for brevity. Panel A reports the equal-weighted average coefficient on LI, and Panel B reports the Litzenberger and Ramaswamy (1979) precision-weighted average coefficient on LI. Newey and West (1987) t-statistics are given in parentheses. 10%(*), 5%(**), and 1%(***) significance levels for two-sided tests are denoted. For the return regressions, the first sample period is July 1966 to December 1989 and the second sample period is January 1990 to December 2016. For the volatility regressions, the first sample period is December 1974 to December 1995 and the second sample period is January 1996 to December 2016.

Table 36: Components of the learning index: Cross-sectional summary statistics and correlations

Panel A: Cross-sectional summary statistics					
			Percentiles		
	Mean	SD	25^{th}	50^{th}	75^{th}
LI_1	0.35	0.30	0.18	0.24	0.38
LI_2	0.10	0.13	0.02	0.05	0.12
LI_3	0.35	0.45	0.08	0.18	0.43
LI_u	0.79	0.84	0.29	0.48	0.95

Panel B: Cross-sectional correlations				
	LI_1	LI_2	LI_3	LI_u
LI_1		0.82	0.86	0.92
LI_2	0.74		0.98	0.96
LI_3	0.82	0.96		0.99
LI_u	0.91	0.94	0.98	

Panel A reports time-series averages of monthly cross-sectional means, standard deviations, and quartiles of the three non-rank-transformed components of the learning index (LI_1, LI_2, LI_3) and the sum of the components (LI_u). Panel B reports time-series averages of monthly cross-sectional Pearson (below the diagonal) and Spearman (above the diagonal) correlations between these four variables. The sample period is July 1964 through December 2016.

Table 37: Components of the learning index: Portfolio sorting

	Panel A: Value-weighted portfolios					Panel B: Equal-weighted portfolios				
Quintile	Return	FF6 α	$ARVOL$	$ASVOL$	$AIVOL$	Return	FF6 α	$ARVOL$	$ASVOL$	$AIVOL$
					Sorting Variable: 1st term of LI					
1 (Low LI_1)	1.130	0.257	2.754	4.186	1.657	1.370	0.273	2.196	3.727	1.355
2	1.021	0.061	1.699	2.919	1.049	1.236	0.133	1.414	2.783	0.758
3	0.853	−0.073	1.333	2.311	0.857	1.120	0.026	1.022	2.369	0.422
4	0.756	−0.085	1.008	2.116	0.531	0.969	−0.103	0.496	1.622	0.060
5 (High LI_1)	0.661	−0.245	−0.019	0.740	−0.071	0.817	−0.251	−0.143	0.880	−0.404
5−1	−0.470***	−0.501***	−2.772***	−3.446***	−1.728***	−0.553***	−0.524***	−2.339***	−2.847***	−1.759***
t-stat	(−3.30)	(−3.68)	(−4.60)	(−4.63)	(−3.30)	(−5.12)	(−4.78)	(−4.39)	(−4.74)	(−3.65)
					Sorting Variable: 2nd term of LI					
1 (Low LI_2)	1.118	0.215	3.638	5.007	2.612	1.386	0.285	3.216	4.803	2.361
2	0.951	0.021	1.553	2.796	0.746	1.209	0.115	1.557	2.875	0.953
3	0.850	−0.054	1.020	2.141	0.588	1.091	0.012	1.038	2.344	0.442
4	0.785	−0.110	0.688	1.551	0.457	0.959	−0.113	0.234	1.327	−0.165
5 (High LI_2)	0.641	−0.256	−0.704	0.173	−0.858	0.866	−0.222	−1.059	0.032	−1.401
5−1	−0.476***	−0.471***	−4.342***	−4.833***	−3.470***	−0.521***	−0.507***	−4.276***	−4.771***	−3.762***
t-stat	(−3.27)	(−3.24)	(−6.31)	(−5.84)	(−6.37)	(−4.41)	(−4.32)	(−7.15)	(−7.13)	(−7.02)
					Sorting Variable: 3rd term of LI					
1 (Low LI_3)	1.135	0.228	3.698	5.146	2.555	1.412	0.316	3.173	4.737	2.337
2	0.943	0.029	1.406	2.518	0.836	1.233	0.121	1.537	2.891	0.896
3	0.883	−0.034	1.065	2.225	0.574	1.052	−0.008	0.940	2.243	0.342
4	0.750	−0.143	0.800	1.688	0.519	0.972	−0.105	0.354	1.476	−0.066
5 (High LI_3)	0.639	−0.262	−0.729	0.133	−0.836	0.840	−0.247	−1.018	0.033	−1.318
5−1	−0.497***	−0.490***	−4.427***	−5.014***	−3.390***	−0.572***	−0.564***	−4.191***	−4.704***	−3.655***
t-stat	(−3.59)	(−3.32)	(−6.42)	(−5.96)	(−6.20)	(−4.80)	(−4.55)	(−6.81)	(−6.76)	(−6.72)

At the end of each month, stocks are sorted into quintiles based on one of the three terms in the learning index (LI) (see equation 1.1). The table reports the next month equal-weighted (Panel A) and value-weighted (Panel B) quintile average of the following variables: excess return, 11-factor risk-adjusted excess return (alpha or α), and percentage change in next month return volatility ($ARVOL$), systematic volatility ($ASVOL$), and idiosyncratic volatility ($AIVOL$) relative to the average return volatility in the prior 12 months. See Table A1 for complete variable definitions. The row labeled "5 – 1" presents the difference in the respective dependent variable between the highest and lowest quintile portfolios. Newey and West (1987) t-statistics and 10%(*), 5%(**), and 1%(***) significance levels for two-sided tests are given for the 5 – 1 portfolio. The sample period is July 1964 to December 2016.

Table 38: Components of the learning index: Cross-sectional return regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	VIF
LI_1	-0.446^{***}			-0.215^{***}			-0.073	4.097
	(-5.22)			(-3.90)			(-1.03)	
LI_2		-0.850^{***}			-0.493^{***}		-0.018	20.421
		(-4.29)			(-4.24)		(-0.06)	
LI_3			-0.270^{***}			-0.154^{***}	-0.102	27.838
			(-4.37)			(-4.19)	(-0.96)	
Control variables	No	No	No	Yes	Yes	Yes	Yes	
Adj R^2	0.004	0.004	0.005	0.091	0.090	0.091	0.091	

This table presents Litzenberger and Ramaswamy (1979) precision-weighted average slope coefficients from two-stage cross-sectional regressions. At the end of each month, I estimate a cross-sectional regression of next month excess stock return on a set of explanatory variables. The first seven columns report results for a different regression specification. The last column reports variance inflation factors from the regression in Column 7. Explanatory variables include an intercept term, the three (non-rank-transformed) components of the learning index (LI_1, LI_2, LI_3), market beta (β^{MKT}), firm size ($SIZE$), book-to-market ratio (BM), profitability ($PROF$), investment (INV), momentum (MOM), illiquidity ($ILLIQ$), short-term reversal (STR), long-term reversal (LTR), and idiosyncratic volatility ($IVOL$). See Table A1 for complete variable definitions. The average adjusted R^2 is reported in the last row. The intercept term and control variable coefficients are not reported for brevity. Newey and West (1987) t-statistics are given in parentheses. 10%(*), 5%(**), and 1%(***) significance levels for two-sided tests are denoted. This regression analysis is based on 814,089 stock-month observations from July 1966 to December 2016 with no missing values for all variables.

Table 39: Components of the learning index: Cross-sectional volatility regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	VIF
LI_1	-0.639^{***}			-0.609^{***}			-0.109	4.356
	(-4.86)			(-5.34)			(-1.01)	
LI_2		-1.551^{***}			-1.501^{***}		-0.002	21.604
		(-5.97)			(-6.96)		(-0.01)	
LI_3			-0.476^{***}			-0.469^{***}	-0.412^{***}	29.505
			(-5.69)			(-6.67)	(-3.50)	
Lagged volatilities	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Control variables	No	No	No	Yes	Yes	Yes	Yes	
Adj R^2	0.492	0.492	0.492	0.532	0.532	0.532	0.532	

This table presents Litzenberger and Ramaswamy (1979) precision-weighted average slope coefficients from two-stage cross-sectional regressions. At the end of each month, I estimate a cross-sectional regression of next month return volatility ($RVOL$) on a set of explanatory variables. The first seven columns report results for a different regression specification. The last column reports variance inflation factors from the regression in Column 7. Explanatory variables include an intercept term, the three (non-rank-transformed) components of the learning index (LI_1, LI_2, LI_3), return on equity (ROE), volatility of return on equity ($ROEVOL$), firm age (AGE), a dividend dummy ($DIVD$), leverage (LEV), inverse of stock price ($INVPRC$), firm size ($SIZE$), book-to-market ratio (BM), momentum (MOM), short-term reversal (STR), next month return (R), and 12 lagged values of volatility. See Table A1 for complete variable definitions. The average adjusted R^2 is reported in the last row. The intercept term and coefficient estimates for lagged volatilities and control variables are not reported for brevity. Newey and West (1987) t-statistics are given in parentheses. 10%(*), 5%(**), and 1%(***) significance levels for two-sided tests are denoted. This regression analysis is based on 686,245 stock-month observations from December 1974 to December 2016 with no missing values for all variables.

Table 40: Alternative test assets: Portfolios sorted by learning index

	Panel A: Value-weighted portfolios			Panel B: Equal-weighted portfolios		
Tercile	Return	α	$ARVOL$	Return	α	$ARVOL$
			Industry portfolios			
1 (Low LI)	1.098	0.071	3.002	1.144	0.024	2.937
2	1.001	0.017	2.343	1.005	−0.097	2.179
3 (High LI)	0.829	−0.188	1.774	0.863	−0.207	1.656
3−1	-0.268^{**}	-0.259^{**}	-1.227^{**}	-0.281^{***}	-0.231^{***}	-1.281^{***}
t-stat	(−2.22)	(−1.98)	(−2.26)	(−3.48)	(−2.76)	(−2.87)
			International equity indexes			
1 (Low LI)	1.511	0.142	3.576	1.133	−0.018	3.216
2	0.838	−0.335	2.565	1.032	−0.090	1.425
3 (High LI)	0.594	−0.485	0.201	1.027	−0.027	1.491
3−1	-0.917^{***}	-0.627^{**}	-3.376^{**}	−0.106	−0.009	-1.725^{*}
t-stat	(−2.88)	(−2.33)	(−1.97)	(−0.70)	(−0.06)	(−1.72)

This table presents portfolio sorting results using 49 value-weighted industry portfolios and 68 international equity indexes as test assets. At the end of each month, assets are sorted into terciles based on values of the learning index (LI). The table reports the next month value-weighted (Panel A) and equal-weighted (Panel B) average excess return, risk-adjusted excess return (alpha or α), and abnormal return volatility ($ARVOL$) relative to the average return volatility in the prior 12 months. See Table A1 for complete variable definitions. Industry portfolio returns are risk adjusted using the Fama and French (2018) six-factor model. International equity index returns are risk adjusted using the Asness et al. (2013) global three-factor model. The row labeled "3-1" presents the difference in the respective dependent variable between the highest and lowest tercile portfolios. Newey and West (1987) t-statistics and 10%(*), 5%(**), and 1%(***) significance levels for two-sided tests are given for the 3-1 portfolio. The industry portfolio sample period is July 1964 to December 2016 and the international equity index sample period is January 1985 to June 2018.

Appendix

Table A1: Variable definitions

Variables are listed in the order that they are introduced within the body of the paper.

Variable	Definition
LI	Learning index, based on the rational expectations general equilibrium model of information choice and investment choice developed by Van Nieuwerburgh and Veldkamp (2010). The learning index reflects the expected benefits of learning about an asset for a rational average investor. Higher values of the empirical learning index correspond to a greater expected degree of learning and information flow. See Section 2.1 in the text for complete description of variable measurement.
β^{MKT}	Market beta, estimated from a regression of excess stock returns on lagged, current, and lead excess market returns using daily data from the past year. β_m^{MKT} is estimated in a similar manner using only daily data within the month.
$SIZE$	Natural logarithm of market value of equity in millions of dollars.
BM	Book-to-market ratio, defined as book value of equity in the latest fiscal year ending in the prior calendar year divided by the market value of equity at the end of December of the prior calendar year.
$PROF$	Profitability, defined as annual revenues minus cost of goods sold, interest expense, and selling, general, and administrative expenses divided by book equity for the latest fiscal year ending in the prior calendar year.
INV	Investment, defined as the annual percentage change in total assets as a decimal.
MOM	Momentum, defined as the cumulative return in percent from month $t-11$ to month $t-1$.
$ILLIQ$	Illiquidity, defined as the absolute monthly return divided by the respective monthly trading volume in dollars, scaled by 10^5.
STR	Short-term reversal, defined as the monthly return in percent over the past month.
LTR	Long-term reversal, defined as the cumulative return as a decimal from month $t-59$ to month $t-12$.
$RVOL$	Return volatility, defined as the standard deviation of daily excess returns within a month.
$IVOL$	Idiosyncratic component of volatility, defined as the standard deviation of daily residuals within a month from estimation of the Fama and French (2018) six-factor model.
$SVOL$	Systematic component of volatility, defined as the square root of the difference between return variance ($RVOL^2$) and idiosyncratic variance ($IVOL^2$).
α	Risk-adjusted average excess return, defined as the intercept from a regression of excess returns on a set of risk factors.
$ARVOL$	Abnormal return volatility, defined as $RVOL$ divided by average $RVOL$ over the previous 12 months, minus one and multiplied by 100.
$AIVOL$	Abnormal idiosyncratic volatility, defined as $IVOL$ divided by average $IVOL$ over the previous 12 months, minus one and multiplied by 100.

Continued on next page

Variable	Definition
$ASVOL$	Abnormal systematic volatility, defined as $SVOL$ divided by average $SVOL$ over the previous 12 months, minus one and multiplied by 100.
ROE	Return on equity, defined as earnings before extraordinary items as of the most recent fiscal quarter end divided by common shareholders' equity as of the end of the previous quarter and multiplied by 100.
$ROEVOL$	Volatility of return on equity, defined as the standard deviation of return on equity over the prior 12 fiscal quarters.
AGE	Firm age, defined as the number of years the firm has existed on CRSP.
$DIVD$	Dummy variable equal to 1 if the firm paid dividends during the most recent fiscal quarter, and 0 otherwise.
LEV	Leverage, defined as total liabilities scaled by the market value of equity as of the most recent fiscal quarter end.
$INVPRC$	Inverse of the stock price, scaled by 100.
R	Monthly return in percent.
$MAXDRET$	Maximum daily return during a 1-month or 3-month horizon (specified in tables).
$MAXWRET$	Maximum weekly return during a 1-month or 3-month horizon (specified in tables).
$MAX\|DRET\|$	Maximum absolute daily return during a 1-month or 3-month horizon (specified in tables).
$MAX\|WRET\|$	Maximum absolute weekly return during a 1-month or 3-month horizon (specified in tables).
$ATURN$	Abnormal monthly share turnover, defined as monthly turnover (total number of shares traded within a month divided by shares outstanding) divided by average monthly turnover over the prior 12 months, minus one and multiplied by 100. $ATURN_{d-1,d+1}$ is abnormal daily share turnover, defined as average daily turnover over the three-day period around earnings announcement date d divided by average daily turnover over days $d-63$ through $d-8$, minus one and multiplied by 100.
$nFCST$	Number of analyst forecasts for the nearest fiscal quarter.
$nREV$	Number of analyst forecast revisions since the last month.
ΔFA	Change in forecast accuracy. The error in the mean forecast is defined for the nearest fiscal quarter as the absolute value of the difference between the mean EPS forecast and the actual EPS as a percentage of the actual EPS. Forecast accuracy is defined as one minus forecast error. The monthly change (in percent) in forecast accuracy is defined as the current month forecast accuracy minus the prior month forecast accuracy, multiplied by 100. This measure is computed by firm and forecast period.
$EDGAR$	Number of human downloads (according to the methodology of Ryans (2017)) of a company's SEC filings from EDGAR during the month.
BBG	Number of days within the month when Bloomberg's "News Heat - Daily Maximum Readership" measure is equal to 3 or 4 out of 4.
$NASDAQ$	Dummy variable equal to 1 for stocks traded on NASDAQ Stock Exchange.
$\overline{TURN}$	Average monthly share turnover (total number of shares traded within a month divided by shares outstanding) over the past 12 months.

Continued on next page

Variable	Definition
IO	Institutional ownership, defined as the percentage of shares outstanding held by 13F institutions as of the most recent quarter-end.
$SP500$	Dummy variable equal to 1 for stocks included in the S&P 500 index.
EAM	Dummy variable equal to 1 if the firm announces quarterly earnings during the month.
$\|CAR\|$	Absolute value of the cumulative abnormal return around a quarterly earnings announcement in percent. Abnormal returns are computed relative to the daily returns of a portfolio matched on size and book-to-market ratio. $\|CAR\|_d$ is computed on the earnings announcement date d. $\|CAR\|_{d-1,d+1}$ is computed over the three-day period around the announcement date. $\|CAR\|_{q+1}$ is computed over the period starting two days after the earnings announcement date through one day after the firm's next quarterly earnings announcement date.
$nSEG$	Number of business segments in different industries (defined based on 4-digit SIC code)
$COMPLEX$	Firm complexity, defined as $1-\sum_{j=1}^{J} s_j^2$, where s_j is the fraction of the firm's total sales generated by industry segment j.
$CVOL$	Call-implied volatility of an at-the-money call option with 30 days to maturity.
$PVOL$	Put-implied volatility of an at-the-money put option with 30 days to maturity.
$ACVOL$	Abnormal call-implied volatility, defined as $CVOL$ divided by average $CVOL$ over the previous 12 months, minus one and multiplied by 100.
$APVOL$	Abnormal put-implied volatility, defined as $PVOL$ divided by average $PVOL$ over the previous 12 months, minus one and multiplied by 100.
$A\beta^{MKT}$	Abnormal market beta, defined as β^{MKT} divided by average monthly β^{MKT} over the previous 12 months, minus one and multiplied by 100. $A\beta_m^{MKT}$ is defined in a similar manner using β_m^{MKT}.

VITA

David Gempesaw

EDUCATION

Pennsylvania State University, State College, PA 2014 – 2019
Doctor of Philosophy in Business Administration, Emphasis in Finance

Miami University, Oxford, OH 2013 – 2014
Master of Arts in Economics, Concentration in Financial Economics

University of Delaware, Newark, DE 2007 – 2010
Bachelor of Science with Honors, Summa Cum Laude
Majors: Finance and Accounting; Minors: Economics and Jazz Studies (Guitar)

RESEARCH INTERESTS

Empirical asset pricing, informed trading, institutional investors (e.g., mutual funds, hedge funds, exchange-traded funds)

SELECTED PAPERS

Information Choice, Uncertainty, and Expected Returns (Job Market Paper)

The Decline of Informed Trading in the Equity and Options Markets, with Charles Cao and Timothy Simin, Journal of Alternative Investments (2018), Volume 21 (2), 16–29

TEACHING EXPERIENCE

Instructor

FIN 410 - Derivatives Markets	SU17, FA17
FIN 100 - Introduction to Finance	SU15

Teaching Assistant

FIN 305 - Financial Management of the Business Enterprise	FA15, FA16, FA18
FIN 406 - Security Analysis and Portfolio Management	SP16, SP19

PROFESSIONAL EXPERIENCE

Ernst & Young, Philadelphia, PA 2009 – 2013
External Auditor, Financial Services Office, Assurance Services
Intern (06/2009 – 08/2009), Staff (09/2010 – 09/2012), Senior (10/2012 – 08/2013)

CERTIFICATION

Certified Public Accountant, State of Pennsylvania 2012