Information Choice, Uncertainty, and Expected Returns

David Gempesaw*
Pennsylvania State University

December 31, 2018

Abstract

In this paper, I test the rational expectations equilibrium model of information choice and investment choice developed by Van Nieuwerburgh and Veldkamp (2010). By estimating a variable from the model called the learning index, I introduce a new approach to empirically measure investors' information choices and assess the effects of these choices. I find negative cross-sectional relationships between the learning index and expected stock returns and volatility. In addition, the learning index is associated with proxies for information demand and the amount of information in prices. Taken together, my findings provide evidence in support of the model's predictions.

*Smeal College of Business, Pennsylvania State University, University Park, PA 16802. E-mail address: dgempesaw@psu.edu. I am grateful to my dissertation committee members Charles Cao, Kai Du, Jeremiah Green, Stephen Lenkey, and Timothy Simin. I also wish to thank Mehmet Canayaz, Jess Cornaggia, Kimberly Cornaggia, Matthew Gustafson, David Haushalter, Peter Iliev, William Kracaw, Anh Le, Giang Nguyen, Joel Vanden, Mitch Warachka, Alexey Zhdanov, and seminar participants at Miami University, Pennsylvania State University, Rowan University, and the 2018 Financial Management Association Annual Meeting for helpful comments.

1 Introduction

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 investor performance by analyzing observable outcomes such as portfolio holdings and investment returns.[2] These outcomes are functions of an investor's ability and decision to acquire and use information. In contrast, less is known about the impact of information choices on outcomes for the underlying assets.

In this paper, I investigate the collective role of investors' learning decisions in determining the cross section of expected risk and 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 general equilibrium model in which investors are able to reduce uncertainty about the future payoffs of certain risky assets before making investment decisions. The model generates predictions regarding the relationships between aggregate learning, uncertainty, and expected returns. 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 given asset, and is increasing in prior expected returns, expected holdings of the asset, and expected pricing errors. In equilibrium, higher values of the learning index correspond to a greater degree of learning.

[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).

The objective of this paper is to test these predictions by empirically estimating the learning index. The conclusions from these tests have direct implications for empirical asset pricing. Most 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. 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 reveals which assets are most valuable to learn about. In addition, unlike other theories that rely on untestable assumptions about investors' unobservable information sets, the theory of Van Nieuwerburgh and Veldkamp (2010) can be tested with observable variables — the learning index is estimated based on past 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. Using this measure, I first test the predicted relationship between learning and expected returns. Univariate portfolio analyses indicate a negative cross-sectional relationship 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 -0.44% per month or -5.4% per year. After adjusting portfolio returns for exposure to the market, size, value, profitability, investment, and momentum factors of Fama and French (2018), the difference in risk-adjusted return between the extreme quintiles is -0.45% per month or -5.5% per year. As an alternative approach, I use Fama and MacBeth (1973) 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

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

indicate that the difference between the stock with the highest and lowest learning index in an average cross section is approximately −0.4% 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. Using value-weighted quintile portfolios sorted on LI, I find that the average change in return volatility in the following month is 3.87% lower for high LI stocks compared to low LI stocks (the unconditional sample average change in return volatility is 1.31%). I decompose changes in return volatility into systematic and idiosyncratic components, and find that information choices predict cross-sectional differences in both components of risk. This result suggests 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 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 relationship between LI and expected return derives from investors' decisions to reduce risk through learning.

After demonstrating the explanatory power of LI for the cross section of stock returns and volatilities, I perform three sets of analyses to better understand the information content of LI. First, I form value-weighted quintile portfolios based on LI and track the differences in expected returns and changes in risk between the extreme quintiles over a long-term horizon. If investors learn fundamental information and trade based on that information, prices should move towards their intrinsic values and not revert in the future. The difference in risk-adjusted monthly return between extreme LI quintiles is negative and significant for up to five 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 differences in volatility changes for several months after portfolio formation. The impact of learning is more persistent for the idiosyncratic component of volatility than for the systematic component, consistent with the idea that LI reflects learning about firm-specific information.

Second, to test the notion that the learning index is representative of learning choices, I investigate the contemporaneous relationship between LI and a number of variables reflecting investor attention or information demand. Using a bivariate portfolio sorting approach to control for the effects of firm size, I find that higher values of LI are positively associated with abnormal trading activity, analyst coverage, forecast revisions, forecast accuracy, SEC EDGAR filing downloads, and news reading activity on Bloomberg terminals. These relationships provide support for the use of LI as a proxy for the learning decisions of investors.

Third, I examine the relation between LI and the information environment surrounding quarterly earnings announcements. Learning about a firm prior to an earnings announcement should reduce the amount of new information revealed in the announcement. After controlling for size, I find that stocks with a higher learning index tend to have smaller market reactions to earnings announcements and a smaller post-earnings announcement drift. High LI stocks also experience a higher degree of abnormal trading activity in the month prior to an earnings announcement and in the three-day window surrounding the announcement. These findings are consistent with more information being acquired for high LI stocks and incorporated into prices prior to earnings announcements. In sum, the results reinforce the idea that the learning index is representative of the information choices of investors.

This paper contributes to a line of research featuring empirical applications of noisy rational expectations equilibrium models focused on the information content of prices.[4] The theoretical models underlying these papers typically assume 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 upon a model that treats information choice as endogenous and that generates predictions

[4]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 indicating 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 Admati (1985) and investigate its relationship with expected returns.

regarding the effects of this choice. 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, these authors concentrate on identifying patterns in mutual fund investment and performance that vary with the business cycle, whereas I focus on directly estimating the learning index at the asset level and using it in cross-sectional analyses.

In addition, my paper adds to a literature investigating the empirical relationship between information, expected returns, and risk. For example, 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 theoretically-motivated prediction of investors' endogenous learning decisions.

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 remainder of this paper is organized as follows. Section 2 discusses the model of Van Nieuwerburgh and Veldkamp (2010), describes the learning index, and outlines the model's relevant predictions. Section 3 describes the procedure to empirically estimate the learning index for individual assets. Section 4 summarizes the data. Section 5 presents empirical results on the cross-sectional relationships between the learning index and expected returns and volatilities. Section 6 examines the long-term predictability of the learning index as well as its relationship to measures of information demand and market activity surrounding earnings announcements. Section 7 provides robustness checks. Section 8 concludes.

2 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). These 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 the model with mean-variance preferences and entropy-based learning.[5]

The model contains multiple risky assets and multiple investors with mean-variance preferences. 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,

[5] 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). Specialized learning is more consistent with the empirical observation that concentrated portfolios outperform diversified ones, implying that investors with informational advantages choose to specialize in their information and portfolio choices (e.g., Kacperczyk, Sialm, and Zheng (2005) or Ivković, Sialm, and Weisbenner (2008)). When combined with the entropy technology, the assumption that investors exhibit constant absolute risk aversion (CARA) preferences leads to indifference between any allocation of learning capacity. On the other hand, an investor with mean-variance preferences chooses specialization in learning.

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: $q^* = \frac{1}{\rho}\hat{\Sigma}^{-1}(\hat{\mu} - pr)$, 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.[6]

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

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

The first term of (1) is the prior squared Sharpe ratio of PC i. Alternatively, this term can be viewed as the product of two terms. $\Gamma_i'(\mu - pr)$ is the prior expected return of PC i, while $\Gamma_i'(\mu - pr)\Lambda_i^{-1}$ is equivalent to ρ times the expected investment in PC i for an investor who has zero learning capacity: $\rho\Gamma_i' E[q]$. These two terms indicate that the value of learning about a PC is increasing in expected excess return and in expected holdings within the investor's

[6] 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.

portfolio. 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. Thus, the value of learning about a given asset is increasing in the asset's prior expected excess return, expected holdings, and expected noise in its price.

In general equilibrium, ex-ante identical investors continue to specialize in learning about a single factor, but will choose 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 asset, the expected return on that asset is reduced, which reduces the value of learning about that asset. 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 will employ a mixed strategy and randomize 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.[7] In contrast, the unconditional CAPM beta is measured 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. This discrepancy between the empirically estimated unconditional risk exposure and the unobserved conditional risk exposure leads to factor model pricing errors.

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 analyzing the cross-sectional explanatory power of LI for future stock returns and risk-adjusted returns. Second, I test the hypothesized relationship between learning and risk by investigating the predictive power of LI for the cross section of return volatility, systematic volatility, and idiosyncratic volatility. Third, I test the notion that LI captures information choices by examining the long-term predictive power of LI as well as its relationship with a number of variables or outcomes that are likely associated with investor learning. Finally, I provide a number of robustness checks using alternative asset pricing models, alternative measures of risk, subperiod analyses, and an alternative set of test assets.

3 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

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

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. 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 variance-covariance matrix Σ of standardized payoffs from Step 1.[8] Decompose Σ into a diagonal eigenvalue matrix Λ and an eigenvector matrix Γ: $\Sigma = \Gamma\Lambda\Gamma'$. Construct principal component prices ($\Gamma' p$), payoffs ($\Gamma' f$), and returns ($\Gamma'(f - pr)$).

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 separate regressions for each principal component of its price on a constant and its payoff.[9] The payoff coefficient Λ_B and the regression R^2 are used to compute

[8]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.

[9]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 second and third term.[10]

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 is a weighted sum of PC learning indexes where the weights are based on the contribution of the stock to each PC.[11] The stock learning index is rank-transformed to [0,1] to facilitate interpretation and comparability across cross sections, although the main results are qualitatively similar without this transformation.

4 Data

I obtain daily and monthly data on 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, stocks are required to have a price greater than \$5 and market capitalization above the 20$^{\text{th}}$ NYSE percentile in order to be included in the sample. Data for market, size, value, profitability, investment, and momentum risk factors are obtained from Kenneth French's website.[12] 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, and Bloomberg. The learning index is estimated over the period July 1964 to December 2016, but

[10] 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 variance-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, $\Lambda_i^{-1}\Lambda_{Ci}^2\sigma_x^2 = \frac{1-R^2}{R^2}\Lambda_{Bi}^2$.

[11] 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 can be interpreted as 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.

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

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 (β^{MKT}) 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 of 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 estimated from a regression of excess stock returns on the Fama and French (2018) six-factor model, which includes market, size, value, profitability, investment, and momentum risk factors.[13] I also compute total return volatility ($RVOL$) as the standard deviation of daily 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 these 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

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

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.

Table 1 presents time series averages of monthly cross-sectional summary statistics for the aforementioned stock characteristics. In the average month, the average stock in the sample has a market beta of 1.06, market capitalization of $3.64 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 1,615 (1,649).

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 firm size, 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, and 24-month periods. The extreme LI quintiles exhibit a relatively high level of persistence from one month to the next. About 73.2% (71.8%) of the stocks in the lowest (highest) LI quintile remain in the same quintile in the next quarter. This result is likely mechanical given the high degree of overlap in the data used to calculate LI_t and LI_{t+1}. As the length of time between the initial month and the final month increases (and the degree of overlap 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 setting, 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. This result is consistent with the predictions of the theoretical model in that the expected benefits of learning about a particular stock tend to decline as more investors learn about the stock.

5 Empirical results

5.1 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 Fama and MacBeth (1973) cross-sectional regressions.

5.1.1 Portfolios of stocks sorted by LI

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 ($R_{p,t} - R_{f,t}$) 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 ($R_{M,t} - R_{f,t}$), 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:

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

Table 4 presents average excess returns and risk-adjusted excess returns for value-weighted (Panel A) and equal-weighted (Panel B) portfolios. I also 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 0.685% in the month following portfolio formation, while the lowest LI quintile has an average excess monthly return of 1.126%. The difference in excess returns between these quintiles is −0.441% per month (−5.4% 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 average excess 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 −0.441% with a t-statistic of −3.78. 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 −0.623% (−0.451%) per month or −7.7% (−5.5%) 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 0.840% and quintile 1 has an average excess return of 1.372% per month. The average monthly return of the $5-1$ portfolio is −0.532%. The average differences in three-factor, five-factor, and six-factor alphas between the extreme quintiles are −0.479%, −0.609%, and −0.541% per month (−5.9%, −7.6%, and −6.7% 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 a reduction in expected return and risk-adjusted return. 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 my 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 the 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 returns 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 whether she receives a good or bad signal about that asset's future payoff. 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 higher expected portfolio returns compared to an uninformed investor.

5.1.2 Fama-MacBeth cross-sectional regressions

In this section, I use two-stage Fama and MacBeth (1973) regressions to examine the cross-sectional relation between the learning index and expected returns while controlling for other determinants of returns. 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.[14] The full cross-sectional

[14] 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

model estimated at the end of each month is

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

In the second stage, I calculate the time series averages of the cross-sectional regression coefficients. As an alternative 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 -0.655 with a t-statistic of -4.14. Since values of LI range from 0 to 1, the reported 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, holding all other variables constant. 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 results based on a similar set of three regressions but reports precision-weighted average slope coefficients. In this setting, I continue to find a negative and significant relationship between LI and subsequent returns. In the univariate regression, the precision-weighted average coefficient on LI is -0.593 and is significant at the 1% level. Using the full multivariate specification in Column 6, the coefficient of interest is -0.401 with a t-statistic of -5.10. These results reinforce the conclusion that learning is associated with a decrease in

past month, share turnover, institutional ownership, number of institutional owners, number of analysts, the call-put option implied volatility spread, and the Stambaugh, Yu, and Yuan (2015) mispricing measure.

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 relationship between illiquidity and expected returns. While theory suggests a positive relationship between these two variables, Bali, Engle, and Murray (2016) show that the empirical relationship 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 summary statistics in Table 1 to get a sense of the relative economic importance 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 decrease in next month return of $9.98 \times 0.037 \approx 0.37\%$ on average, all else equal. The explanatory power of the learning index for next month returns is comparable to that of idiosyncratic volatility, firm size, and investment. Increases of one standard deviation in $IVOL$, $SIZE$, INV, and LI are associated with average changes in expected monthly return of -0.14%, -0.14%, -0.14%, and -0.12% respectively, holding all other variables constant.

5.2 Explaining the cross section of return 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.

5.2.1 Portfolios of stocks sorted by LI

I first conduct a univariate portfolio sorting analysis using quintiles sorted on LI. Because return volatility is serially correlated, I use the percentage change in volatility as the dependent variable in the sorting analysis. This measure can be viewed as an estimate of a stock's posterior variance relative to its prior variance. My objective is to determine whether there is a difference in the average volatility change between the extreme LI quintiles. The expectation is that volatility changes of high LI stocks should be lower on average compared to volatility changes of low LI stocks.[15]

I measure the change in return volatility ($\Delta RVOL$) 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 measure percentage changes in systematic volatility and idiosyncratic volatility. $\Delta SVOL$ is monthly systematic volatility divided by average monthly systematic volatility over the previous 12 months, minus one and multiplied by 100. Similarly, $\Delta IVOL$ 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 volatility changes across quintiles. Table 6 presents value-weighted (Panel A) and equal-weighted (Panel B) portfolio average changes in volatility. In Panel A, the change in return volatility is 3.866% 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 variation in

[15] As is typically the case in portfolio analyses, I am not directly interested in the level of the dependent variable (average volatility change) for any particular quintile over the sample period. The decision to learn about certain stocks does not imply that I should empirically observe decreases in 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 which are learned about to experience a decrease in volatility. In reality, stock volatility may change over time for reasons unrelated to learning. Thus, the theory does not directly provide a time series prediction about whether volatility is increasing or decreasing on average for any given quintile. It only leads to a cross-sectional prediction regarding the comparison of the average volatility change for stocks subject to a high degree of learning relative to that of stocks subject to a lower degree of learning.

both systematic and idiosyncratic volatility changes. On average, the change in systematic (idiosyncratic) volatility in the month following portfolio formation is 4.386% (2.851%) lower for high LI stocks compared to low LI stocks, with a t-statistic of -5.56 (-5.38). I arrive at similar conclusions if volatility changes are weighted equally within each portfolio. On average, the differences in $\Delta RVOL$, $\Delta SVOL$, and $\Delta IVOL$ between extreme equal-weighted portfolios is -3.268%, -3.746%, and -2.767%, 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 reduction in both the firm-specific and systematic components of risk. These findings do not necessary imply that the choice to learn about a stock involves the discovery of market-wide or macroeconomic information. Rather, my findings provide support for the notion that learning news about a firm can lower firm-specific uncertainty as well as uncertainty arising from co-movement with the market or other common risk factors.[16]

5.2.2 Fama-MacBeth cross-sectional regressions

Next, I use Fama and MacBeth (1973) 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

[16] For robustness, I consider defining volatility changes 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) in volatility compared to the prior 1, 3, 6, or 12 months. I also measure systematic and idiosyncratic volatility relative to alternative factor model specifications. Finally, instead of examining future volatility changes, I use a bivariate portfolio sorting approach to examine the relationship between LI and the future volatility level, conditional on the past 12-month average historical volatility level. The conclusions are qualitatively similar under each of these robustness checks.

coefficients. The full cross-sectional model estimated at the end of each month is

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

where the dependent variable 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 relationship 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 regress 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 relationship between LI and volatility. The coefficient on LI is -1.333 and is significant at the 1% level. This finding supports the prediction that learning leads to a reduction in uncertainty. 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 relationship 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 relationship between the learning index and volatility. The coefficient on LI in the full specification is -1.331 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 1.331 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 -1.160 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 as reported in Table 1. For the average stock in an average cross-section, an increase in the value of LI from 0 to 1 (all else being equal) is associated with a change in return volatility of $\frac{-1.331}{34.21} \approx -3.89\%$ based on the equal-weighted average LI coefficient, or $\frac{-1.160}{34.21} \approx -3.39\%$ based on the precision-weighted average LI coefficient. These changes are comparable to the predicted changes in return volatility from Table 6.

Based on precision-weighted average coefficient estimates in Column 6, variation in the current month return, next 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 STR, R, and $INVPRC$ are associated with average changes of -0.86, 0.85, 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 decreases in return volatility in the following month of 0.34, 0.34, 0.30, and 0.22 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 relationship between learning and systematic risk in the following month. The coefficient on LI is -1.038 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.

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 have 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 -0.791 (t-statistic$= -5.58$). 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 0 to 1 holding all other variables constant is associated with a change in $SVOL$ of $\frac{-0.791}{22.75} \approx -3.47\%$. 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 -0.686 (t-statistic= -3.77). 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 relationships 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 0.929 percentage points lower on average than the stock with the lowest value of LI, all else equal. For the average stock in an average cross-section, this change corresponds to a decrease of $\frac{0.929}{24.45} \approx 3.80\%$ in relative terms. 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 reduction in return volatility. When combined with the findings in Section 5.1, the results suggest that this reduction in risk corresponds to a reduction in risk premium or expected return.[17]

[17] In untabulated analyses, I investigate the predictive power of the three individual components of the general equilibrium learning index. While the model suggests that investors choose to learn about the asset with the highest sum of the three terms in LI, each individual term is expected to be positively associated with the expected benefits of learning. Using portfolio sorting, I find negative and significant relationships between the individual component of the learning index and the various measures of risk and return. Because of high multicollinearity between the three LI components, it is difficult to assess the independent effects of each component on expected returns and risk in a multivariate setting. Nevertheless, the results from portfolio sorting indicate that the explanatory power of each individual component is comparable to that of the sum of the components.

6 Supplementary analyses

6.1 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 LI reflects investors learning fundamental information and incorporating this information into prices, I expect that prices move towards fundamental value and do not reverse in the long run. Alternatively, if the explanatory power of LI 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 value-weighted average returns between the highest LI quintile and the lowest LI quintile $(5-1)$ in each of the 36 months after portfolio formation. Figure 1 presents average monthly returns and Fama and French (2018) six-factor risk-adjusted returns (alpha). The average return of the $5-1$ portfolio is most negative in the month immediately following portfolio formation and subsequently moves towards zero. By month $t+5$, the negative average return spread is no longer significant at the 10% level. On a risk-adjusted basis, the return spread between the highest and lowest LI quintiles is negative and significant until month $t+6$. Beyond this point, all risk-adjusted returns are not statistically different from zero. The results indicate that the explanatory power of LI continues in a declining manner over a period of several months. After adjusting for co-movement with systematic risk factors, I find that the monthly return differences between extreme LI quintiles are not reversed over the subsequent three years. This finding supports the notion that the cross-sectional explanatory power of LI for returns reflects the effects of investors learning and trading on fundamental information.

Next, I repeat the portfolio sorting analysis and track the difference in value-weighted average volatility changes between the extreme LI quintiles over the subsequent 36 months. In Figure 2, the average spread in $\Delta RVOL$ is negative and significant for seven months after portfolio formation. Beyond this point, all values of $\Delta RVOL$ are not statistically different from zero. This result suggests that the cross-sectional relationship between learning and risk is not attributable to temporary decreases in volatility. When I decompose volatility changes into systematic and idiosyncratic components, I find that the changes in each type of volatility exhibit different patterns over the long run. For the $5-1$ portfolio, average values of $\Delta IVOL$ are negative and significant until month $t+12$, and are not statistically different from zero for the next 24 months. This finding suggests that learning results in a permanent reduction in idiosyncratic risk. On the other hand, the changes in systematic volatility predicted by LI tend to reverse over the long run. The spread in average $\Delta SVOL$ is negative and significant until month $t+5$ and then turns positive and significant beginning in month $t+12$.

Thus, while learning appears to reduce return co-movement with systematic risk factors over the short run, this effect is not as permanent as the effect of learning on the idiosyncratic component of risk. 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 1 is associated with a reversal in systematic risk. There is no evidence of reversal in idiosyncratic volatility and risk-adjusted returns. In untabulated analyses, I arrive at similar conclusions using alternative factor model specifications for risk adjustment and equal-weighted rather than value-weighted portfolio averages. In aggregate, the results in this section demonstrate that the effects of learning on idiosyncratic risk and risk-adjusted returns are generally long-lasting.

6.2 Relationship with measures of investor attention or information demand

To reinforce the notion that the learning index represents collective information choices, I examine the contemporaneous cross-sectional relationship 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, EDGAR filing download activity, and Bloomberg news reading activity. In practice, information choices 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 measured by firm size. Stocks are sorted at the end of each month into quintiles based on size. Then, within each size quintile, stocks are sorted 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.[18]

Table 10 reports portfolio average values of six different proxies for investor attention 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 measure trading activity as share turnover, or the total number of shares traded within a month divided by shares outstanding. I then estimate the change in monthly share turnover ($\Delta TURN$) 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 is available from CRSP for the full sample period (July 1964 to December 2016). On average, high (low) LI stocks experience a 9.29% (3.71%) increase in monthly share turnover relative to average monthly turnover during the past year. The difference in abnormal turnover between the extreme quintiles is 5.59% with a t-statistic of 6.68. This difference suggests an increased level of trading activity among stocks subject to a greater degree of investor learning.

The next three proxies relate to analyst coverage, forecast revisions, and forecast accuracy. Greater analyst coverage can lead to 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. Harford, Jiang, Wang,

[18] In untabulated analyses, I use a similar bivariate sorting approach to examine patterns in returns and volatility changes across LI quintiles. The results from these analyses are qualitatively similar to those in Table 4 and 6.

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 ($nANALYST$) 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 change in forecast accuracy. First, I measure the error in the mean forecast for the nearest fiscal quarter as the average EPS forecast divided by the actual EPS, minus one. Then, I compute the monthly percentage change in the absolute value of the forecast error (ΔAFE) as the current month absolute error in mean forecast divided by the prior month absolute error in mean forecast, minus one and multiplied by 100. This measure is computed by firm within a given forecast period so that forecast errors are not compared across different fiscal quarters.

The evidence indicates a positive association between the expected benefits of gathering information (as measured by 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 8.60 (7.52) analysts. The difference in coverage is approximately one analyst with a t-statistic of 6.52. On average, 2.52 (2.12) analysts covering a high (low) LI stock revise their forecasts from the prior month. The average difference in the number of forecast revisions is 0.40 and is significant at the 1% level. I also find a significant relationship between changes in forecast precision and the learning index. For all quintiles, the average monthly percentage change in forecast precision is negative. This pattern implies that on average, analysts' estimates become more precise (relative to the actual realized value) as the fiscal quarter end approaches. Stocks in the highest (lowest) LI quintile have an average reduction in absolute error of the mean forecast of 12.46% (10.90%). The difference between the extreme LI quintiles is -1.56% on average (t-statistic = -4.96). Therefore, while the EPS forecasts for all stocks in the sample become closer on average over time to the actual realized EPS, the increase in precision is greater for those stocks subject to a greater degree of learning.[19]

The fifth proxy is based on download activity from the SEC EDGAR system. The SEC provides data beginning in January 2003 containing a record of downloads of filings from EDGAR. Using this data, Crane, Crotty, and Umar (2018) provide evidence on the value of this information by showing that hedge funds' use of publicly-available SEC filings predicts fund performance. Using 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.[20] After controlling for the size of the firm, I find that the filings of firms with the highest (lowest) values of LI are downloaded 830.89 (715.92) times within a month on average. The difference in average EDGAR downloads between these quintiles is 114.97 with a t-statistic of 3.58. 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, Da, and Israelsen (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. This variable is available beginning in February 2010. 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 3.04 (2.77) days within a month on average. The difference in abnormal attention days between high and low LI stocks is 0.27 with a t-statistic of 4.35. Overall, the patterns documented in this section serve as additional evidence of a positive relationship between the learning index and information demand.

[19] In untabulated analysis, I find qualitatively similar results when I compute forecast errors using the median forecast rather than the mean forecast.

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

6.3 Information environment surrounding earnings announcements

In this section, I 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, then the information acquired may be incorporated into prices beforehand. If this is true, then the average market reactions to earnings announcements of high LI stocks should be smaller in magnitude than those of low LI stocks.

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 (as of June) and book-to-market ratio (as of the prior December). I measure the absolute value of the cumulative abnormal return on the day of the announcement ($ACAR_d$), during a three-day window around the announcement ($ACAR_{d-1,d+1}$), and during the period from two days after the announcement through one day after the following quarterly announcement ($ACAR_{NextQtr}$).

I also examine abnormal trading activity using two measures. The first measure $\Delta TURN$ is share turnover during the month prior to the earnings announcement divided by average monthly turnover over the prior 12 months, minus one and multiplied by 100. Similar to Lerman, Livnat, and Mendenhall (2008), the second measure $\Delta TURN_{d-1,d+1}$ is average daily turnover during the three-day period around the announcement date d divided by average daily turnover from days $d-63$ through $d-8$, minus one and multiplied by 100.

As in the previous section, I use the 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. Since I am interested in examining the contemporaneous relationship between LI and abnormal trading activity, the sorting procedure is performed using contemporaneous values of size and LI when analyzing $\Delta TURN_{d-1,d+1}$. Due to data availability, the sample period for this analysis is October 1971 to December 2016.

Table 11 reports average values and associated t-statistics of the abnormal return and trading activity measures for each portfolio. After controlling for firm size, stocks with high values of LI tend to have smaller market reactions to quarterly earnings announcements. On average, the $ACAR$ on the event date is 0.141% smaller for high LI stocks compared to low LI stocks. Over a three-day window, the difference between the extreme LI quintiles is -0.215%. These estimates are significant at the 1% and 5% 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 high LI stocks. The average spread in $ACAR_{NextQtr}$ between the high and low LI quintiles is -0.494% with a t-statistic of -2.40.

The last two columns in Table 11 indicate a higher degree of abnormal trading activity prior to and surrounding earnings announcements for high LI stocks relative to low LI stocks. On average, high (low) LI stocks experience a 2.581% (-0.432%) 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 3.014% with a t-statistic of 3.18. During the three-day window around the announcement, high (low) LI stocks experience a 67.817% (53.477%) increase in average daily turnover relative to average daily turnover during the preceding non-event period. The difference between extreme quintiles is 14.340% with a t-statistic of 8.88. In sum, the results in this section suggest that stocks with higher values of LI are subject to a greater level of trading activity around quarterly earnings announcements, and tend to have smaller abnormal market reactions to these events. These findings reinforce the interpretation of the LI as a proxy for investors' information choices.

7 Robustness checks

7.1 Alternative asset pricing models

In this section, I investigate the robustness of the relationship between the learning index and expected returns by repeating the portfolio sorting analyses from Section 5.1.1 using four alternative factor model specifications for risk adjustment.[21] 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.

Table 12 reports risk-adjusted excess returns for value-weighted (Panel A) and equal-weighted (Panel B) quintile portfolios sorted by LI. Using value-weighted portfolio returns, the difference in alpha between extreme LI quintiles ranges from −0.444% per month (−5.5% per year) based on the Stambaugh and Yuan (2017) factor model to −0.543% per month (−6.7% 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 relationship between the LI and risk-adjusted returns is robust.

7.2 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. Implied volatility can be viewed as a measure of the market's expectation of an asset's volatility over the remaining life of the

[21] Liquidity factor data are obtained from Lubos Pastor's website: faculty.chicagobooth.edu/lubos.pastor/research. Short-term and long-term reversal factor data are obtained from Kenneth French's website: mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. Data for the Stambaugh and Yuan (2017) mispricing factors are obtained from Robert Stambaugh's website: finance.wharton.upenn.edu/~stambaug. Data for the Hou, Xue, and Zhang (2015) q-factor model are provided by Kewei Hou.

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 relationship between the learning index and changes in the implied volatilities of calls ($CVOL$) and puts ($PVOL$). $\Delta CVOL$ 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 changes in put-implied volatility ($\Delta PVOL$) in a similar manner.

Table 13 reports value-weighted (Panel A) and equal-weighted (Panel B) average percentage change in $CVOL$ and $PVOL$ for LI-sorted portfolios. I find a negative cross-sectional relationship between the learning index and the market's expectation of next month volatility. Higher values of LI are associated with lower value-weighted average and equal-weighted averages of $\Delta CVOL$ and $\Delta PVOL$. For robustness, I perform two additional untabulated analyses of the relationship between the learning index and implied volatility. First, I investigate the explanatory power of LI for next month (instead of current month) changes in implied volatility. Second, I use a bivariate portfolio sorting approach to examine the explanatory power of LI for the current (or next month) level of implied volatility while controlling for the past 12-month average level of implied volatility. The conclusions from these tests are qualitatively similar. In total, my findings suggest that the learning index carries explanatory power not only for future realized volatility, but also for the market's expectation of future volatility.

7.3 Learning and CAPM beta

According to Van Nieuwerburgh and Veldkamp (2010), assets which 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 12 support this idea. The model predicts that learning results in a reduction in the conditional covariance of assets with the market. In this section, I investigate this prediction by examining the cross-sectional relationship between the learning index and CAPM beta.

Using portfolios sorted on LI, I first estimate β_p^{MKT} at the portfolio level from a regression of quintile average next month excess returns on the next month excess return of the market. Similar to the prior portfolio sorting analyses of volatility, I also examine monthly percentage changes in individual stock betas. β_m^{MKT} is market beta measured at the stock level using a regression of excess daily returns on lagged, current, and lead excess daily market returns within a month. $\Delta\beta_m^{MKT}$ is the percentage change in β_m^{MKT} in the month following portfolio formation relative to average β_m^{MKT} in the prior 12 months.

Table 14 reports the next month value-weighted (Panel A) and equal-weighted (Panel B) quintile market beta and average percentage change in beta. Consistent with the theoretical prediction that learning reduces co-movement with the market, I find that the high LI quintile has a lower market beta compared to that of the low LI quintile. The difference in portfolio beta between extreme LI quintiles is -0.226% and is significant at the 1% level. I also find that higher values of LI are associated with lower changes in CAPM beta. On average, the percentage change in beta in the month following portfolio formation is -3.408% lower for high LI stocks compared to low LI stocks, with a t-statistic of -2.80. Results based on equally-weighted quintile portfolio betas and average percentage changes in beta are qualitatively similar.

To further investigate this relationship, I estimate multivariate cross-sectional regressions similar to those used in Section 5.2. Instead of volatility, I use next month beta (β_m^{MKT}) as the dependent variable and lagged values of beta as controls. Table 15 presents the results. After controlling for various determinants of risk, I continue to find a negative and significant cross-sectional relationship between LI and CAPM beta. Taken together, these findings suggest that learning about an asset results in a lower conditional covariance with the market. This relationship corresponds to the dispersion in systematic volatility and factor model pricing errors documented in the prior analyses.

7.4 Subperiod analysis

To examine how the results vary over time, I repeat the portfolio sorting analyses over the subperiods July 1964 to December 1989 and January 1990 to December 2016. Table 16 presents value-weighted (Panel A) and equal-weighted (Panel B) averages of next month excess return, Fama and French (2018) six-factor risk-adjusted excess return, and percentage change in 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 relationships between LI and the measures of risk and return are evident in both subperiods. In the earlier part of the sample period, LI strongly predicts cross-sectional variation in raw returns. A portion of this explanatory power appears to be attributable to exposure to systematic risk factors; the predictive ability of LI for risk-adjusted returns and idiosyncratic volatility changes is weaker during this time. Cross-sectional differences in volatility changes between extreme LI portfolios tend to be larger in the later part of the sample.

In Table 17, I estimate Fama-MacBeth cross-sectional regressions of excess return and volatility over two subperiods. For these analyses, I use the full set of control variables from equations 3 and 4, 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 1996 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 -0.421 (-0.408). In Panel B, the average coefficient in the first (second) period is -0.411 (-0.396). 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 17 report equal-weighted and precision-weighted average LI coefficients 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 January 1996 to December

2016.[22] For each measure of volatility, both the equal-weighted average LI coefficient and precision-weighted average LI coefficient 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 16. The equal-weighted and precision-weighted average coefficient on LI is negative and significant for each volatility measure and subperiod. 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.

7.5 Alternative test assets: Industry portfolios

In this section, I apply the learning index estimation procedure as described in Section 3 to the 49 Fama-French industry portfolios, which are formed based on four-digit Standard Industrial Classification (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).

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 18 reports value-weighted (Panel A) and equal-weighted (Panel B) averages of next month excess return, risk-adjusted return, and percentage change in total return volatility. The average difference in excess returns (alphas) between the extreme LI terciles is −0.265% (−0.387%) on a value-weighted basis. These estimates are significant at the 5% and 1% level, respectively. I also find a negative and significant relationship between LI and changes in volatility. On average, industries in the high LI tercile experience a change in 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 the choice to learn more

[22] Conclusions are similar if I split the sample at December 1989, although a lower number of observations prior to this date reduces statistical power.

about particular industries leads to a reduction in expected risk and return.

8 Conclusion

This study examines the importance of information choice in determining the cross section of expected risk and 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 optimal learning decisions and the resulting impact on risk and risk premiums. To test these predictions, I estimate the learning index from the model. I find that the learning index is negatively related to both future returns and future volatilities. The reductions in risk-adjusted returns and idiosyncratic risk are persistent and do not reverse in the long run. In addition, I find that the learning index is cross-sectionally related to measures of investor attention, information demand, and the amount of information in prices. Taken together, the evidence suggests that the learning index is representative of collective information choices by investors, and that these choices are important in the determination of investors' expectations about risk and return. My findings support the theoretical predictions of Van Nieuwerburgh and Veldkamp (2010) and illustrate a new approach that can be used to empirically measure information choices.

References

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

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.

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.

Fama, E. and J. MacBeth (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy 81*, 607–636.

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*.

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.

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.

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*.

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*.

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

Table 1: Cross-sectional summary statistics

	Mean	SD	Percentiles 25th	50th	75th
LI	0.50	0.29	0.25	0.50	0.75
β_{MKT}	1.06	0.57	0.66	0.99	1.39
$SIZE$	6.51	1.25	5.51	6.26	7.28
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	20.62	46.86	−4.29	12.78	34.39
$ILLIQ$	0.20	0.84	0.02	0.06	0.18
STR	1.76	9.98	−3.85	1.08	6.46
LTR	1.11	2.06	0.17	0.64	1.36
$RVOL$	34.21	17.94	22.43	30.49	41.89
$IVOL$	24.45	14.08	15.34	21.37	30.00
$SVOL$	22.75	12.68	14.18	20.18	28.44
ROE	3.36	5.20	1.86	3.35	4.96
$ROEVOL$	4.34	14.11	0.90	1.65	3.26
AGE	23.66	17.84	10.22	17.92	33.02
$DIVD$	0.71	0.42	0.41	1.00	1.00
LEV	2.19	4.04	0.32	0.75	1.67
$INVPRC$	4.13	2.53	2.44	3.49	5.08
R	1.09	9.57	−4.26	0.72	5.97
# of stocks	1,615	352	1,572	1,649	1,754

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 $5 and market capitalization greater than the 20$^{\text{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}	−0.13	1.00									
$SIZE$	−0.10	−0.02	1.00								
BM	0.07	−0.13	−0.13	1.00							
$PROF$	−0.07	0.11	−0.01	−0.29	1.00						
INV	−0.06	0.19	−0.05	−0.20	0.17	1.00					
MOM	−0.27	0.08	−0.02	−0.10	0.05	0.01	1.00				
$ILLIQ$	0.03	−0.08	−0.20	0.03	−0.01	−0.01	−0.02	1.00			
STR	−0.02	0.01	−0.01	0.02	0.01	−0.01	0.03	0.29	1.00		
LTR	−0.11	0.16	0.02	−0.28	0.16	0.32	−0.01	−0.04	−0.02	1.00	
$IVOL$	0.06	0.35	−0.28	−0.07	0.07	0.15	0.05	0.10	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 LI-sorted portfolios

Panel A: 1-month transition matrix

	$LI1_{t+1}$	$LI2_{t+1}$	$LI3_{t+1}$	$LI4_{t+1}$	$LI5_{t+1}$
$LI1_t$	73.2	22.5	3.8	0.5	0.0
$LI2_t$	22.7	47.7	24.1	5.0	0.5
$LI3_t$	3.7	24.3	43.9	24.0	4.1
$LI4_t$	0.4	5.0	24.0	47.1	23.5
$LI5_t$	0.0	0.5	4.2	23.5	71.8

Panel B: 6-month transition matrix

	$LI1_{t+6}$	$LI2_{t+6}$	$LI3_{t+6}$	$LI4_{t+6}$	$LI5_{t+6}$
$LI1_t$	51.0	25.7	13.5	6.9	2.9
$LI2_t$	25.3	27.8	22.4	15.7	8.8
$LI3_t$	13.0	22.1	24.6	23.2	17.1
$LI4_t$	6.6	14.9	22.4	27.8	28.3
$LI5_t$	3.1	8.8	16.9	27.1	44.1

Panel C: 12-month transition matrix

	$LI1_{t+12}$	$LI2_{t+12}$	$LI3_{t+12}$	$LI4_{t+12}$	$LI5_{t+12}$
$LI1_t$	33.3	23.1	18.0	14.6	10.9
$LI2_t$	23.3	22.2	20.2	18.3	16.1
$LI3_t$	17.6	20.2	21.0	21.1	20.1
$LI4_t$	13.3	18.3	21.0	23.2	24.3
$LI5_t$	9.0	15.0	20.0	24.6	31.4

Panel D: 24-month transition matrix

	$LI1_{t+24}$	$LI2_{t+24}$	$LI3_{t+24}$	$LI4_{t+24}$	$LI5_{t+24}$
$LI1_t$	21.5	20.6	20.0	19.7	18.2
$LI2_t$	19.5	19.8	20.2	20.3	20.2
$LI3_t$	17.8	19.3	20.4	21.1	21.4
$LI4_t$	17.2	19.2	20.3	21.3	22.0
$LI5_t$	16.0	18.5	20.3	21.9	23.3

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), and $t+24$ (Panel D). 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 (LI)

Panel A: Value-weighted portfolio returns

Quintile	Excess return	FF3 α	FF5 α	FF6 α
1 (Low LI)	1.126	0.248	0.330	0.241
2	0.966	0.059	0.022	0.040
3	0.804	−0.103	−0.162	−0.098
4	0.727	−0.179	−0.258	−0.154
5 (High LI)	0.685	−0.193	−0.293	−0.210
5−1	-0.441^{***}	-0.441^{***}	-0.623^{***}	-0.451^{***}
t-stat	(−3.14)	(−3.78)	(−5.47)	(−3.12)

Panel B: Equal-weighted portfolio returns

Quintile	Excess return	FF3 α	FF5 α	FF6 α
1 (Low LI)	1.372	0.277	0.318	0.295
2	1.248	0.147	0.124	0.158
3	1.041	−0.044	−0.079	−0.023
4	0.943	−0.131	−0.188	−0.112
5 (High LI)	0.840	−0.202	−0.291	−0.246
5−1	-0.532^{***}	-0.479^{***}	-0.609^{***}	-0.541^{***}
t-stat	(−4.31)	(−4.71)	(−6.18)	(−4.29)

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: Fama-MacBeth cross-sectional regressions

	Panel A: Equal-weighted coefficient average			Panel B: Precision-weighted coefficient average		
	(1)	(2)	(3)	(4)	(5)	(6)
LI	-0.655^{***}		-0.416^{***}	-0.593^{***}		-0.401^{***}
	(-4.14)		(-4.90)	(-4.41)		(-5.10)
β_{MKT}		0.052	0.028		-0.069	-0.087
		(0.37)	(0.21)		(-0.55)	(-0.70)
$SIZE$		-0.120^{***}	-0.129^{***}		-0.098^{***}	-0.108^{***}
		(-3.38)	(-3.68)		(-2.96)	(-3.29)
BM		0.071	0.075		0.153^{**}	0.155^{**}
		(0.75)	(0.79)		(2.04)	(2.06)
$PROF$		0.100^{**}	0.098^{**}		0.096^{***}	0.096^{***}
		(1.98)	(1.98)		(2.72)	(2.73)
INV		-0.498^{***}	-0.496^{***}		-0.464^{***}	-0.466^{***}
		(-5.38)	(-5.47)		(-5.83)	(-5.92)
MOM		0.005^{**}	0.004^{**}		0.004^{***}	0.004^{***}
		(2.53)	(2.30)		(3.28)	(2.81)
$ILLIQ$		0.539	0.521		-0.092^{***}	-0.094^{***}
		(0.56)	(0.55)		(-2.78)	(-2.86)
STR		-0.040^{***}	-0.040^{***}		-0.037^{***}	-0.037^{***}
		(-7.78)	(-7.70)		(-7.89)	(-7.82)
LTR		-0.051^{***}	-0.057^{***}		-0.034^{***}	-0.041^{***}
		(-2.70)	(-3.12)		(-2.78)	(-3.36)
$IVOL$		-0.011^{***}	-0.010^{***}		-0.011^{***}	-0.010^{***}
		(-4.42)	(-3.83)		(-4.51)	(-3.96)
Adj R^2	0.009	0.090	0.092	0.009	0.090	0.092

This table presents results from Fama and MacBeth (1973) cross-sectional regressions. At the end of each month, I estimate a cross-sectional regression of the next month excess stock return on a set of explanatory variables. Panel A reports equal-weighted average slope coefficients, and Panel B reports Litzenberger and Ramaswamy (1979) precision-weighted average slope coefficients. Each column presents results for a different regression specification. Explanatory variables include an intercept term, the learning index (LI), 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 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 807,566 stock-month observations from July 1966 to December 2016 with no missing values for all variables.

Table 6: Explaining the cross section of return volatility: Portfolios of stocks sorted by learning index (LI)

	Panel A: Value-weighted average			Panel B: Equal-weighted average		
Quintile	$\Delta RVOL$	$\Delta SVOL$	$\Delta IVOL$	$\Delta RVOL$	$\Delta SVOL$	$\Delta IVOL$
1 (Low LI)	3.353	4.749	2.244	2.965	4.473	2.179
2	1.689	2.880	1.012	1.723	3.149	1.038
3	1.356	2.628	0.700	1.201	2.454	0.656
4	0.767	1.676	0.506	0.686	1.821	0.283
5 (High LI)	−0.513	0.363	−0.607	−0.303	0.726	−0.588
5−1	-3.866^{***}	-4.386^{***}	-2.851^{***}	-3.268^{***}	-3.746^{***}	-2.767^{***}
t-stat	(−5.94)	(−5.56)	(−5.38)	(−5.42)	(−5.46)	(−5.17)

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 percentage change in return volatility ($\Delta RVOL$), systematic volatility ($\Delta SVOL$), and idiosyncratic volatility ($\Delta IVOL$) 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 monthly change in 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 return volatility: Fama-MacBeth cross-sectional regressions

	Panel A: Equal-weighted coefficient average			Panel B: Precision-weighted coefficient average		
	(1)	(2)	(3)	(4)	(5)	(6)
LI	−1.333***		−1.331***	−1.075***		−1.160***
	(−5.05)		(−7.22)	(−5.78)		(−7.64)
ROE		−0.043***	−0.047***		−0.040***	−0.042***
		(−7.08)	(−7.27)		(−8.08)	(−8.53)
$ROEVOL$		0.043***	0.043***		0.005***	0.005**
		(3.21)	(3.18)		(2.60)	(2.50)
AGE		−0.009***	−0.009***		−0.008***	−0.008***
		(−5.62)	(−5.77)		(−6.07)	(−6.23)
$DIVD$		−0.784***	−0.777***		−0.716***	−0.713***
		(−7.96)	(−8.13)		(−8.47)	(−8.71)
LEV		0.011	0.000		−0.010	−0.016
		(0.31)	(−0.01)		(−0.60)	(−0.97)
$INVPRC$		0.283***	0.286***		0.270***	0.273***
		(10.05)	(10.20)		(10.17)	(10.42)
R		0.087***	0.088***		0.094***	0.094***
		(3.97)	(3.97)		(4.25)	(4.25)
$SIZE$		−0.001	−0.033		−0.045	−0.072
		(−0.02)	(−0.63)		(−0.87)	(−1.42)
BM		−0.881***	−0.854***		−0.790***	−0.765***
		(−7.03)	(−6.94)		(−7.21)	(−7.10)
MOM		0.000	−0.003		0.004	0.001
		(0.12)	(−0.90)		(1.63)	(0.39)
STR		−0.106***	−0.102***		−0.099***	−0.096***
		(−10.61)	(−10.45)		(−11.41)	(−11.27)
Lagged Volatilities	✓	✓	✓	✓	✓	✓
Adj R^2	0.493	0.532	0.533	0.493	0.532	0.533

This table presents results from Fama and MacBeth (1973) 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. Panel A reports equal-weighted average slope coefficients, and Panel B reports Litzenberger and Ramaswamy (1979) precision-weighted average slope coefficients. Each column presents 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. 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 680,543 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: Fama-MacBeth cross-sectional regressions

	Panel A: Equal-weighted coefficient average			Panel B: Precision-weighted coefficient average		
	(1)	(2)	(3)	(4)	(5)	(6)
LI	-1.038^{***}		-0.791^{***}	-0.793^{***}		-0.647^{***}
	(−5.05)		(−5.58)	(−5.77)		(−5.74)
ROE		-0.030^{***}	-0.032^{***}		-0.028^{***}	-0.030^{***}
		(−6.99)	(−7.09)		(−7.72)	(−8.09)
$ROEVOL$		0.033^{***}	0.034^{***}		0.003^{**}	0.003^{**}
		(3.07)	(3.05)		(2.10)	(2.05)
AGE		-0.006^{***}	-0.006^{***}		-0.005^{***}	-0.006^{***}
		(−4.28)	(−4.48)		(−5.06)	(−5.26)
$DIVD$		-0.603^{***}	-0.604^{***}		-0.535^{***}	-0.537^{***}
		(−7.40)	(−7.54)		(−7.59)	(−7.82)
LEV		0.032	0.025		0.006	0.002
		(1.20)	(0.94)		(0.46)	(0.16)
$INVPRC$		0.202^{***}	0.204^{***}		0.190^{***}	0.192^{***}
		(9.76)	(9.80)		(10.20)	(10.35)
R		0.054^{***}	0.054^{***}		0.059^{***}	0.059^{***}
		(3.86)	(3.86)		(4.38)	(4.38)
$SIZE$		0.128^{**}	0.108^{**}		0.079	0.063
		(2.32)	(1.97)		(1.49)	(1.19)
BM		-0.634^{***}	-0.626^{***}		-0.557^{***}	-0.550^{***}
		(−6.37)	(−6.40)		(−6.66)	(−6.64)
MOM		0.005^{*}	0.003		0.008^{***}	0.006^{***}
		(1.72)	(1.14)		(3.67)	(2.85)
STR		-0.061^{***}	-0.059^{***}		-0.054^{***}	-0.053^{***}
		(−7.60)	(−7.53)		(−8.26)	(−8.20)
Lagged Volatilities	✓	✓	✓	✓	✓	✓
Adj R^2	0.439	0.478	0.479	0.439	0.478	0.479

This table presents results from Fama and MacBeth (1973) 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. Panel A reports equal-weighted average slope coefficients, and Panel B reports Litzenberger and Ramaswamy (1979) precision-weighted average slope coefficients. Each column presents 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. 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 680,543 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: Fama-MacBeth cross-sectional regressions

	Panel A: Equal-weighted coefficient average			Panel B: Precision-weighted coefficient average		
	(1)	(2)	(3)	(4)	(5)	(6)
LI	-0.686^{***}		-0.929^{***}	-0.563^{***}		-0.841^{***}
	(-3.77)		(-7.04)	(-4.12)		(-7.29)
ROE		-0.040^{***}	-0.042^{***}		-0.034^{***}	-0.036^{***}
		(-6.75)	(-6.78)		(-8.20)	(-8.45)
$ROEVOL$		0.039^{***}	0.038^{***}		0.006^{***}	0.005^{***}
		(3.41)	(3.36)		(3.41)	(3.29)
AGE		-0.008^{***}	-0.009^{***}		-0.008^{***}	-0.008^{***}
		(-7.14)	(-7.21)		(-7.25)	(-7.35)
$DIVD$		-0.744^{***}	-0.730^{***}		-0.700^{***}	-0.690^{***}
		(-9.18)	(-9.28)		(-10.57)	(-10.74)
LEV		-0.031	-0.039^{*}		-0.029^{**}	-0.033^{***}
		(-1.33)	(-1.67)		(-2.44)	(-2.79)
$INVPRC$		0.255^{***}	0.258^{***}		0.243^{***}	0.246^{***}
		(11.20)	(11.37)		(10.82)	(11.07)
R		0.070^{***}	0.070^{***}		0.072^{***}	0.073^{***}
		(4.08)	(4.09)		(4.21)	(4.22)
$SIZE$		-0.142^{***}	-0.164^{***}		-0.151^{***}	-0.170^{***}
		(-5.21)	(-6.05)		(-5.83)	(-6.58)
BM		-0.777^{***}	-0.753^{***}		-0.701^{***}	-0.677^{***}
		(-7.90)	(-7.76)		(-7.88)	(-7.74)
MOM		-0.003^{*}	-0.006^{***}		-0.001	-0.003^{*}
		(-1.79)	(-2.97)		(-0.49)	(-1.83)
STR		-0.078^{***}	-0.075^{***}		-0.075^{***}	-0.073^{***}
		(-11.22)	(-11.05)		(-11.72)	(-11.56)
Lagged Volatilities	✓	✓	✓	✓	✓	✓
Adj R^2	0.420	0.454	0.455	0.420	0.454	0.455

This table presents results from Fama and MacBeth (1973) 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. Panel A reports equal-weighted average slope coefficients, and Panel B reports Litzenberger and Ramaswamy (1979) precision-weighted average slope coefficients. Each column presents 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. 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 680,543 stock-month observations from December 1974 to December 2016 with no missing values for all variables.

Table 10: Relationship with measures of investor attention: Portfolios of stocks sorted by learning index (LI) controlling for size

Sample Begins:	Jul 1964	Jul 1984	Jul 1984	Jul 1984	Jan 2003	Feb 2010
Quintile	$\Delta TURN$	$nANALYST$	$nREV$	ΔAFE	$EDGAR$	BBG
1 (Low LI)	3.71	7.52	2.12	−10.90	715.92	2.77
2	6.34	7.82	2.23	−11.44	733.64	2.85
3	7.94	8.11	2.35	−11.86	770.09	2.93
4	8.87	8.38	2.45	−12.51	788.43	2.97
5 (High LI)	9.29	8.60	2.52	−12.46	830.89	3.04
5−1	5.59^{***}	1.07^{***}	0.40^{***}	-1.56^{***}	114.97^{***}	0.27^{***}
t-stat	(6.68)	(6.52)	(5.21)	(−4.96)	(3.59)	(4.35)

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 for six proxies of investor attention or information demand: change in monthly share turnover ($\Delta TURN$), number of analyst forecasts ($nANALYST$), number of analyst forecast revisions ($nREV$), change in absolute forecast error (ΔAFE), 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 11: Information environment surrounding quarterly earnings announcements: Portfolios of stocks sorted by learning index (LI) controlling for size

Quintile	$ACAR_d$	$ACAR_{d-1,d+1}$	$ACAR_{NextQtr}$	$\Delta TURN$	$\Delta TURN_{d-1,d+1}$
1 (Low LI)	2.547	4.433	12.169	−0.432	53.477
2	2.480	4.366	11.858	0.668	54.919
3	2.401	4.271	11.797	1.742	60.873
4	2.451	4.291	11.715	2.214	61.838
5 (High LI)	2.407	4.218	11.675	2.581	67.817
5−1	-0.141^{***}	-0.215^{**}	-0.494^{**}	3.014^{***}	14.340^{***}
t-stat	(−2.63)	(−2.56)	(−2.40)	(3.18)	(8.88)

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 for three proxies of market reaction and two proxies of abnormal trading activity. Market reaction proxies include the absolute value of the cumulative abnormal return on the earnings announcement date d ($ACAR_d$), over the three-day period around the announcement date ($ACAR_{d-1,d+1}$), and during the period from two days after the earnings announcement date through one day after the firm's next quarterly earnings announcement date ($ACAR_{NextQtr}$). Abnormal trading activity proxies include change in monthly share turnover during the month prior to the announcement ($\Delta TURN$), and change in daily turnover over the three-day period around the announcement date ($\Delta TURN_{d-1,d+1}$). See Table A1 for complete variable definitions. For $\Delta TURN_{d-1,d+1}$, the sorting procedure is performed using contemporaneous values of size and LI. 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 12: Alternative asset pricing models: Portfolios of stocks sorted by learning index (LI)

Panel A: Value-weighted portfolio returns				
Quintile	7-factor α	9-factor α	SY (2017) α	HXZ (2015) α
1 (Low LI)	0.235	0.277	0.253	0.663
2	0.049	0.042	0.062	0.446
3	−0.093	−0.110	−0.067	0.265
4	−0.167	−0.197	−0.111	0.172
5 (High LI)	−0.230	−0.261	−0.191	0.120
5−1	−0.465***	−0.539***	−0.444***	−0.543***
t-stat	(−3.08)	(−3.10)	(−3.31)	(−3.66)

Panel B: Equal-weighted portfolio returns				
Quintile	7-factor α	9-factor α	SY (2017) α	HXZ (2015) α
1 (Low LI)	0.302	0.322	0.429	0.739
2	0.170	0.160	0.228	0.600
3	−0.025	−0.045	0.024	0.372
4	−0.113	−0.150	−0.075	0.271
5 (High LI)	−0.255	−0.284	−0.232	0.129
5−1	−0.558***	−0.606***	−0.661***	−0.610***
t-stat	(−4.25)	(−3.91)	(−5.06)	(−4.79)

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 13: Explaining the cross section of implied volatility: Portfolios of stocks sorted by learning index (LI)

	Panel A: Value-weighted average		Panel B: Equal-weighted average	
Quintile	$\Delta CVOL$	$\Delta PVOL$	$\Delta CVOL$	$\Delta PVOL$
1 (Low LI)	1.222	1.227	0.673	0.587
2	0.405	0.405	0.187	0.230
3	0.091	0.095	−0.187	−0.123
4	−0.149	−0.222	−0.582	−0.493
5 (High LI)	−1.361	−1.331	−1.748	−1.650
5−1	-2.584^{***}	-2.558^{***}	-2.421^{***}	-2.237^{***}
t-stat	(−3.55)	(−3.59)	(−3.44)	(−3.23)

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 percentage change in call-implied volatility ($\Delta CVOL$) and put-implied volatility ($\Delta PVOL$) 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 monthly change in 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 14: Learning and CAPM beta: Portfolios of stocks sorted by learning index (LI)

	Panel A: Value-weighted average		Panel B: Equal-weighted average	
Quintile	β_p^{MKT}	$\Delta\beta_m^{MKT}$	β_p^{MKT}	$\Delta\beta_m^{MKT}$
1 (Low LI)	1.098	2.328	1.207	1.939
2	1.026	0.793	1.147	0.661
3	0.948	1.204	1.102	0.062
4	0.925	−0.120	1.076	−0.902
5 (High LI)	0.871	−1.080	0.998	−2.134
5−1	−0.226***	−3.408***	−0.209***	−4.073***
t-stat	(−3.25)	(−2.80)	(−3.35)	(−5.78)

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 market beta and average percentage change in beta. β_p^{MKT} is market beta measured at the quintile portfolio level using a regression of quintile average next month excess returns on the next month excess return of the market. β_m^{MKT} is market beta measured at the stock level using a regression of excess daily returns on lagged, current, and lead excess daily market returns within a month. $\Delta\beta_m^{MKT}$ is the percentage change in β_m^{MKT} in the month following portfolio formation relative to average β_m^{MKT} 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 15: Learning and CAPM beta: Fama-MacBeth cross-sectional regressions

	Panel A: Equal-weighted coefficient average			Panel B: Precision-weighted coefficient average		
	(1)	(2)	(3)	(4)	(5)	(6)
LI	-0.105^{***}		-0.040^{***}	-0.108^{***}		-0.046^{***}
	(-6.87)		(-4.00)	(-7.16)		(-4.80)
ROE		0.000	0.000		-0.001^{*}	-0.001^{*}
		(0.56)	(0.41)		(-1.70)	(-1.90)
$ROEVOL$		0.001	0.001		0.000	0.000
		(1.42)	(1.41)		(1.62)	(1.58)
AGE		-0.001^{***}	-0.001^{***}		-0.001^{***}	-0.001^{***}
		(-5.04)	(-5.22)		(-4.49)	(-4.71)
$DIVD$		-0.070^{***}	-0.070^{***}		-0.062^{***}	-0.063^{***}
		(-6.46)	(-6.48)		(-6.80)	(-6.88)
LEV		0.004^{**}	0.004^{*}		0.003^{*}	0.002^{*}
		(2.18)	(1.96)		(1.87)	(1.70)
$INVPRC$		0.011^{***}	0.011^{***}		0.010^{***}	0.010^{***}
		(7.96)	(8.14)		(8.11)	(7.99)
R		0.002	0.002		0.001	0.001
		(1.59)	(1.60)		(1.29)	(1.30)
$SIZE$		0.039^{***}	0.038^{***}		0.029^{***}	0.028^{***}
		(4.51)	(4.38)		(3.71)	(3.55)
BM		-0.058^{***}	-0.058^{***}		-0.049^{***}	-0.049^{***}
		(-5.54)	(-5.55)		(-4.71)	(-4.77)
MOM		0.001^{***}	0.001^{***}		0.001^{***}	0.001^{***}
		(3.56)	(3.30)		(4.59)	(4.23)
STR		-0.004^{***}	-0.004^{***}		-0.004^{***}	-0.003^{***}
		(-6.28)	(-6.18)		(-6.94)	(-6.86)
Lagged Betas	✓	✓	✓	✓	✓	✓
Adj R^2	0.156	0.201	0.202	0.156	0.201	0.202

This table presents results from Fama and MacBeth (1973) 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. Panel A reports equal-weighted average slope coefficients, and Panel B reports Litzenberger and Ramaswamy (1979) precision-weighted average slope coefficients. Each column presents 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. 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 16: Subperiod analysis: Portfolios of stocks sorted by learning index (LI)

	Panel A: Value-weighted average					Panel B: Equal-weighted average				
Quintile	Return	FF6 α	$\Delta RVOL$	$\Delta SVOL$	$\Delta IVOL$	Return	FF6 α	$\Delta RVOL$	$\Delta SVOL$	$\Delta IVOL$
					Sample Period: July 1964 – December 1989					
1 (Low LI)	1.220	0.184	3.118	4.862	1.844	1.416	0.240	2.572	4.073	1.942
2	0.929	−0.004	2.347	3.901	1.473	1.255	0.146	1.877	3.378	1.345
3	0.784	−0.006	2.044	3.637	1.282	1.052	0.079	1.715	2.974	1.403
4	0.731	−0.044	1.554	2.743	1.293	0.957	−0.005	1.443	2.632	1.215
5 (High LI)	0.717	−0.056	0.882	2.076	0.712	0.872	−0.083	0.896	1.917	0.861
5−1	−0.502***	−0.240	−2.236***	−2.786***	−1.132*	−0.544***	−0.323*	−1.676*	−2.156**	−1.081
t-stat	(−3.04)	(−1.19)	(−2.85)	(−3.00)	(−1.85)	(−3.72)	(−1.94)	(−1.92)	(−2.12)	(−1.46)
					Sample Period: January 1990 – December 2016					
1 (Low LI)	1.039	0.148	3.490	4.606	2.465	1.331	0.314	3.331	4.845	2.401
2	1.000	0.089	1.100	2.011	0.519	1.242	0.176	1.579	2.936	0.752
3	0.822	−0.087	0.754	1.802	0.085	1.031	−0.053	0.723	1.969	−0.040
4	0.724	−0.174	0.148	0.878	−0.205	0.931	−0.143	−0.020	1.065	−0.585
5 (High LI)	0.655	−0.247	−1.696	−1.047	−1.803	0.810	−0.299	−1.420	−0.384	−1.938
5−1	−0.384*	−0.394**	−5.187***	−5.654***	−4.268***	−0.520***	−0.613***	−4.751***	−5.228***	−4.339***
t-stat	(−1.72)	(−2.18)	(−5.66)	(−4.84)	(−6.11)	(−2.66)	(−3.70)	(−6.67)	(−6.38)	(−7.03)

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), and percentage change in return volatility ($\Delta RVOL$), systematic volatility ($\Delta SVOL$), and idiosyncratic volatility ($\Delta IVOL$) 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 17: Subperiod analysis:
Learning index (LI) coefficient from Fama-MacBeth 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	-0.421^{***}	-0.408^{***}	-0.411^{***}	-0.396^{***}
	(-3.30)	(-3.36)	(-3.63)	(-3.84)
	Dec 1974 – Dec 1995	Jan 1996 – Dec 2016	Dec 1974 – Dec 1995	Jan 1996 – Dec 2016
$RVOL$	-0.877^{***}	-1.790^{***}	-0.770^{***}	-1.568^{***}
	(-4.97)	(-6.19)	(-4.76)	(-6.81)
$SVOL$	-0.455^{***}	-1.131^{***}	-0.359^{***}	-0.938^{***}
	(-3.22)	(-5.12)	(-2.91)	(-5.55)
$IVOL$	-0.569^{***}	-1.291^{***}	-0.525^{***}	-1.174^{***}
	(-4.37)	(-6.52)	(-4.09)	(-7.13)

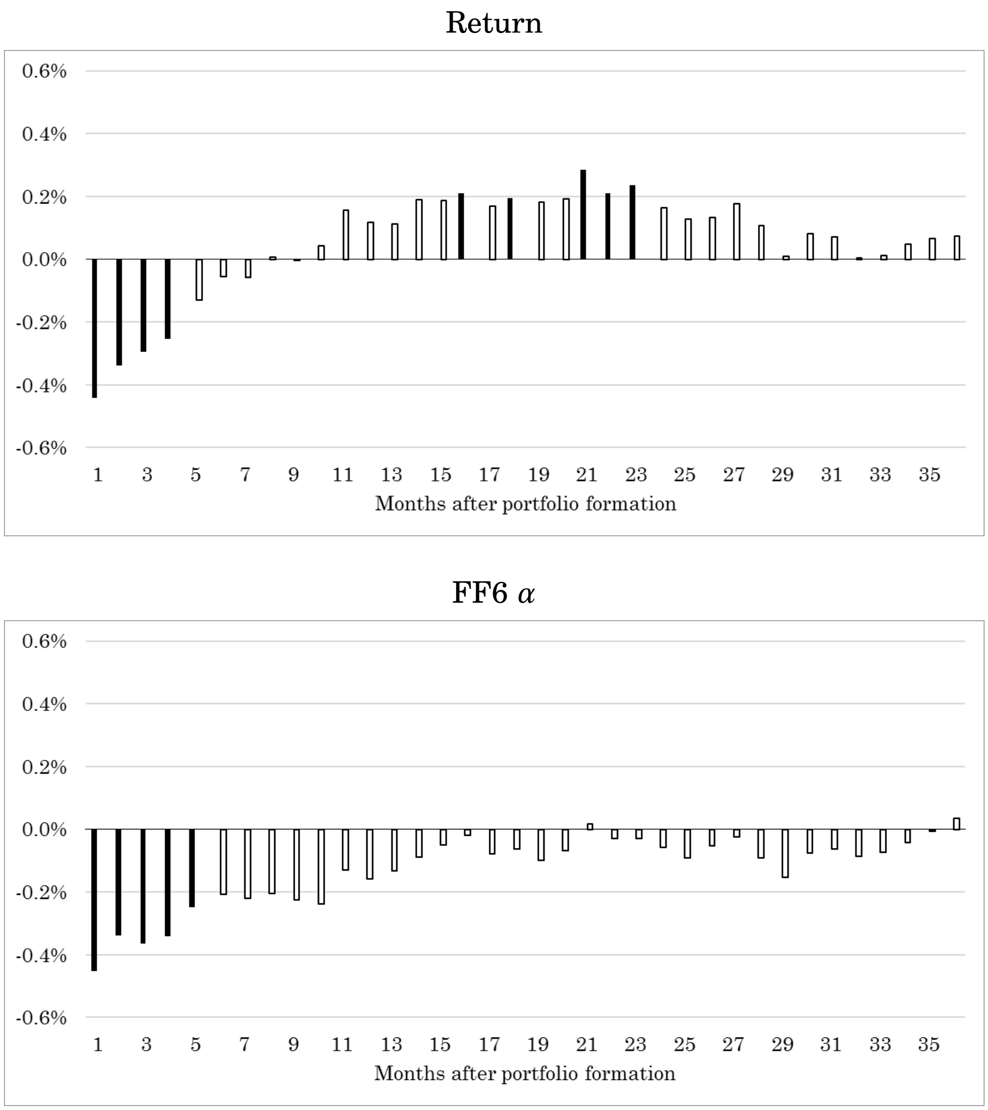
This table presents results from Fama and MacBeth (1973) cross-sectional regressions. At the end of each month, I estimate cross-sectional regressions for each of the dependent variables listed in the first column on the learning index (LI) and a set of control variables following the respective full specification described in the text (equations 3 and 4). 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 18: Alternative test assets:
49 industry portfolios

	Panel A: Value-weighted average			Panel B: Equal-weighted average		
Tercile	Return	FF6 α	$\Delta RVOL$	Return	FF6 α	$\Delta RVOL$
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)

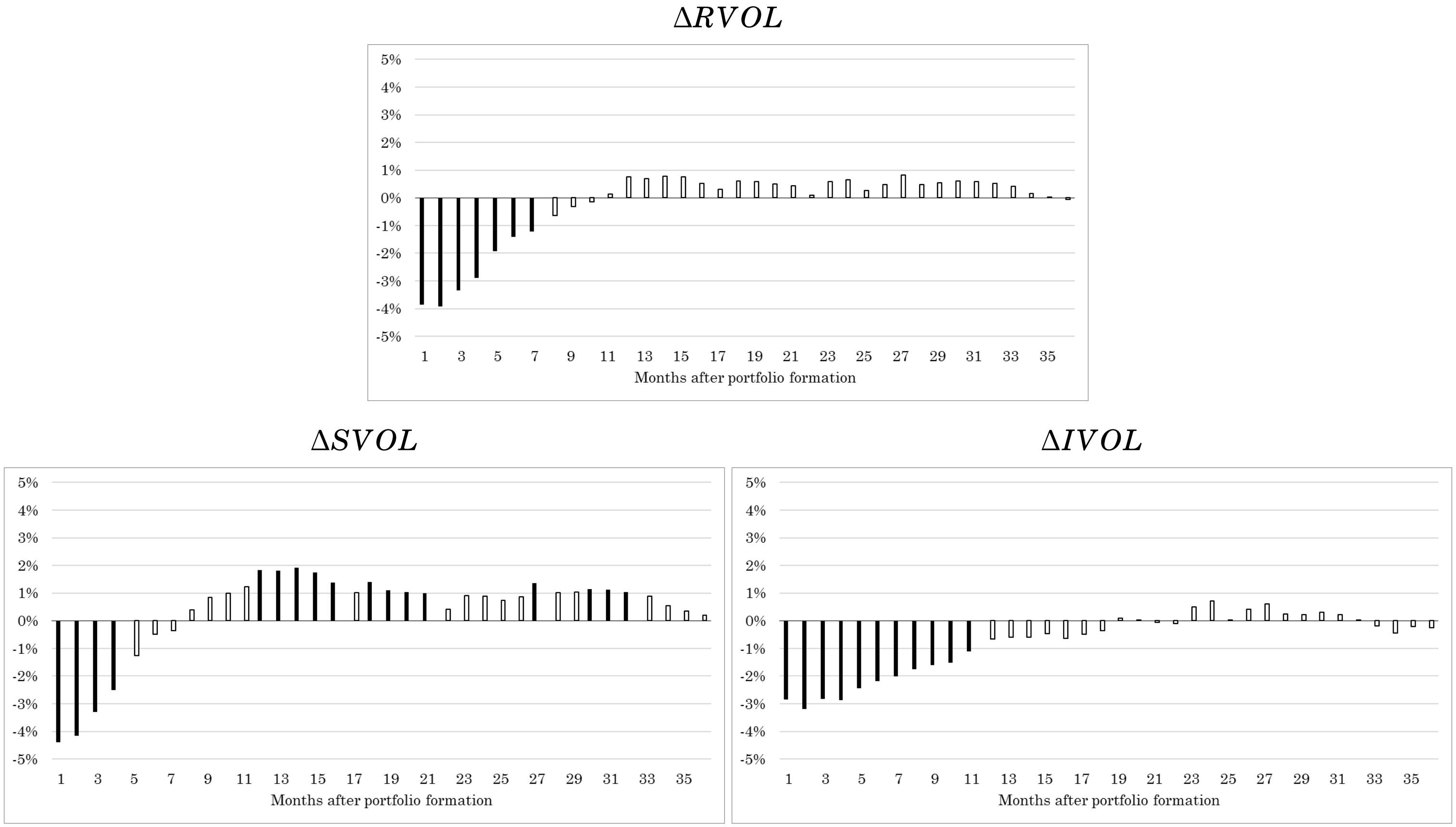
This table presents portfolio sorting results using 49 value-weighted industry portfolios as test assets. At the end of each month, industry portfolios are sorted into terciles based on values of the learning index (LI). The table reports the next month value-weighted (Panel A) and equal-weighted (Panel B) average of the following variables: excess return, risk-adjusted excess return (alpha or α), and percentage change in next month return volatility ($\Delta RVOL$) 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. 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 sample period is July 1964 to December 2016.

Figure 1: Long-term return predictability:
Monthly returns of $LI5 - LI1$ portfolio over next 36 months



At the end of each month, I sort stocks into quintiles based on values of the learning index (LI) and track the difference in value-weighted average returns between the highest LI quintile and the lowest LI quintile ($5-1$ portfolio) in each of the 36 months after portfolio formation. The figure presents average monthly excess returns and risk-adjusted excess returns (alpha or α) 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 2: Long-term volatility predictability:
Monthly volatility changes of $LI5 - LI1$ portfolio over next 36 months



At the end of each month, I sort stocks into quintiles based on values of the learning index (LI) and track the difference in value-weighted average percentage changes in return volatility ($\Delta RVOL$), systematic volatility ($\Delta SVOL$), and idiosyncratic volatility ($\Delta IVOL$) between the highest LI quintile and the lowest LI quintile ($5-1$ portfolio) in each of the 36 months after portfolio formation. The figure presents average volatility changes 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.

Appendix

Table A1: Variable definitions

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 empirical learning index reflects the value of learning about a given asset for the average investor. Higher values of the learning index correspond to a greater expected degree of learning. See Section 3 in the text for complete description of variable measurement.
β^{MKT}	Market beta, calculated from a regression of excess stock returns on lagged, current, and lead excess market returns using daily data from the past year.
$SIZE$	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 estimated from a regression of excess stock returns on the six-factor model of Fama and French (2018).
$SVOL$	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.
$\Delta RVOL$	Change in return volatility, defined as next month $RVOL$ divided by average monthly $RVOL$ over the previous 12 months, minus one and multiplied by 100.
$\Delta IVOL$	Change in the idiosyncratic component of volatility, defined as next month $IVOL$ divided by average monthly $IVOL$ over the previous 12 months, minus one and multiplied by 100.
$\Delta SVOL$	Change in the systematic component of volatility, defined as next month $SVOL$ divided by average monthly $SVOL$ over the previous 12 months, minus one and multiplied by 100.

Continued on next page

Variable	Definition
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.
$\Delta TURN$	Change in 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. $\Delta TURN_{d-1,d+1}$ is change in 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.
$nANALYST$	Number of analyst forecasts for the nearest fiscal quarter.
$nRev$	Number of analyst forecast revisions since the last month.
ΔAFE	Change in absolute value of the error in the mean forecast. The error in the mean forecast is measured for the nearest fiscal quarter as the average EPS forecast divided by the actual EPS, minus one. Monthly percentage change in the absolute value of the error in mean forecast is measured as the current month absolute error in mean forecast divided by the prior month absolute error in mean forecast, minus one and 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" variable is equal to 3 or 4 out of 4.
$ACAR$	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 (as of June) and book-to-market ratio (as of December). $ACAR_d$ is computed on the earnings announcement date d. $ACAR_{d-1,d+1}$ is computed over the three-day period around the announcement date. $ACAR_{NextQtr}$ 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.
$CVOL$	Call-implied volatility, measured based on an at-the-money call option with 30 days to maturity.
$PVOL$	Put-implied volatility, measured based on an at-the-money put option with 30 days to maturity.
$\Delta CVOL$	Change in call-implied volatility, defined as 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.

Continued on next page

Variable	Definition
$\Delta PVOL$	Change in put-implied volatility, defined as the difference between current month $PVOL$ and average $PVOL$ in the prior 12 months, scaled by average $PVOL$ in the prior 12 months and multiplied by 100.
β_p^{MKT}	Market beta measured at the quintile portfolio level using a regression of quintile average next month excess returns on the next month excess return of the market.
β_m^{MKT}	Market beta, calculated from a regression of excess stock returns on lagged, current, and lead excess market returns using daily data within a month.
$\Delta\beta_m^{MKT}$	Change in market beta, defined as monthly β_m^{MKT} divided by average monthly β_m^{MKT} over the previous 12 months, minus one and multiplied by 100.