FISHER COLLEGE OF BUSINESS

Fisher College of Business Working Paper Series

Charles A. Dice Center for Research in Financial Economics

Have we solved the idiosyncratic volatility puzzle?

Kewei Hou, The Ohio State University

Roger K. Loh, Singapore Management University

Dice Center WP 2012-28
Fisher College of Business WP 2012-03-028

Revision: December 2012
Original: June 2011

Have we solved the idiosyncratic volatility puzzle?*

Kewei Hou
Ohio State University

Roger K. Loh
Singapore Management University

First Draft: June 2011
This Draft: December 2012

Abstract

We propose a simple methodology to evaluate a large number of potential explanations for the negative relation between idiosyncratic volatility and subsequent stock returns (the idiosyncratic volatility puzzle). We find that surprisingly many existing explanations explain less than 10% of the puzzle. On the other hand, explanations based on investors' lottery preferences, short-term return reversal, and earnings shocks show greater promise in explaining the puzzle. Together they account for 60-80% of the negative idiosyncratic volatility-return relation. Our methodology can be applied to evaluate competing explanations for a broad range of topics in asset pricing and corporate finance.

Keywords: Idiosyncratic volatility; Cross-section of stock returns; Lottery preferences; Market frictions
JEL Classification Codes: G12; G14

---

* We thank Jack Bao, Hank Bessembinder, Steve Dimmock, Fangjian Fu, Dong Hong, Chuan-Yang Hwang, Jung-Min Kim, William Leon, Angie Low, René Stulz, Mitch Warachka, and seminar participants at Nanyang Technological University, National University of Singapore, Ohio State University, Seoul National University, Singapore Management University, SUNY Buffalo, University of Central Florida, University of Delaware, and University of Exeter for their comments and suggestions. We also thank Karl Diether, Fangjian Fu, and Keith Vorkink for sharing data.
Address Correspondence: Kewei Hou, Ohio State University, Department of Finance, Fisher College of Business, 820 Fisher Hall, 2100 Neil Avenue, Columbus, OH 43210, USA. Tel: (614) 292-0552, hou.28@osu.edu. Roger Loh, Singapore Management University, Lee Kong Chian School of Business. 50 Stamford Rd, #04-01, Singapore 178899, Singapore. Tel: (65) 6828-0243, rogerloh@smu.edu.sg.

Have we solved the idiosyncratic volatility puzzle?

1. Introduction

Ang, Hodrick, Xing, and Zhang (2006), in a highly influential paper, document a negative relation between idiosyncratic volatility and subsequent stock returns.[1] This result is puzzling because traditional asset pricing theories either predict no relation between idiosyncratic volatility and expected returns under the assumptions that markets are complete and frictionless and investors are well-diversified, or predict a positive relation under the assumptions that markets are incomplete and investors face sizeable frictions and hold poorly-diversified portfolios (see, e.g., Merton (1987) and Hirshleifer (1988)). Consequently, many papers have been written trying to explain the puzzle, with each paper proposing a different economic mechanism linking idiosyncratic volatility to subsequent stock returns.[2] However, to date there has been no comprehensive examination about which explanations best explain the puzzle. Further complicating this matter is the fact that existing studies typically differ in terms of empirical methodology and sample construction, thus making direct comparisons of their results difficult.

Motivated by these concerns, this paper provides a simple unified framework to evaluate a large number of candidate explanations of the puzzle. Most studies in this literature typically promote a new explanation of the puzzle while controlling for a limited number of existing explanations. We believe that our paper provides the most comprehensive examination of

[1] Ang, Hodrick, Xing, and Zhang (2009) show that the relation also exists in international markets.

[2] The long list of candidate explanations includes those based on expected idiosyncratic skewness (Boyer, Mitton, and Vorkink (2010)), coskewness (Chabi-Yo and Yang (2009)), maximum daily return (Bali, Cakici, and Whitelaw (2011)), retail trading proportion (Han and Kumar (2012)), one-month return reversal (Fu (2009) and Huang, Liu, Rhee, and Zhang (2009)), illiquidity (Bali and Cakici (2008) and Han and Lesmond (2011)), uncertainty (Johnson (2004)), short-sale constraints (Boehme, Danielsen, Kumar, and Sorescu (2009) and George and Hwang (2011)), financial distress (Avramov, Chordia, Jostova, and Philipov (2012)), investor attention (George and Hwang (2011)), growth options (Cao, Simin, and Zhao (2008) and Chen and Petkova (2012)), average variance beta (Chen and Petkova (2012)), and earnings shocks (Jiang, Xu, and Yao (2009) and Wong (2011)).

existing explanations to date. More importantly, our methodology allows us to quantify the fraction of the puzzle that is explained by each candidate explanation, either by itself or after controlling for other competing explanations.

To summarize our approach, we start from Fama and Macbeth (1973) cross-sectional regressions of individual month t stock returns on month t–1 idiosyncratic volatility. We find, as many papers do, that the estimated regression coefficient, which we denote as γ_t, is on average negative and highly statistically significant. Next, we decompose the γ_t coefficient into one or more components each related to a candidate explanation of the puzzle (e.g., skewness) and a residual component. The ratio of the component related to a particular candidate explanation to the original γ_t coefficient then measures the fraction of the idiosyncratic volatility puzzle that is captured by that explanation, and the ratio of the residual component to γ_t measures the fraction of the puzzle left unexplained by all candidate explanations considered. Our approach ensures that the components related to the candidate explanations and the residual component add up to γ_t. This makes for intuitive interpretation and easy comparisons when we pit existing explanations against one another.[3]

To guide our analysis, we break up existing explanations into three groups. The first group of explanations attributes the idiosyncratic puzzle to lottery preferences of investors (they propose different proxies for the lottery feature of a stock, e.g., skewness, coskewness, realized and expected idiosyncratic skewness, maximum daily return, and retail trading proportion). The second group of explanations appeals to various forms of market frictions (e.g., one-month return reversal, illiquidity, short-sale constraints, and investor attention) to try to explain the puzzle. Explanations that do not fall naturally into the first two groups (e.g., uncertainty, leverage,

[3] Our approach can be easily used to compare competing explanations of a broad range of topics in empirical asset pricing and corporate finance.

financial distress, growth options, expected idiosyncratic volatility, and pre- and post-formation earnings shocks) are then included in the third group.

Using the sample of CRSP common stocks from 1963-2009, we find that surprisingly many existing explanations, when evaluated alone, explain less than 10% of the idiosyncratic volatility puzzle. This holds true for the explanations based on skewness, coskewness, realized idiosyncratic skewness, illiquidity, short-sale constraints, attention, uncertainty, leverage, financial distress, growth options, average variance beta, and expected idiosyncratic volatility. For example, skewness and financial distress can only explain 6.9% and 5.4%, respectively, of the puzzle. Or consider illiquidity. Despite being highly correlated with idiosyncratic volatility, it fails to capture any fraction of the puzzle.

On the other hand, explanations based on expected idiosyncratic skewness, maximum daily return, retail trading proportion, one-month return reversal, pre- and post-formation earnings shocks show promise in explaining the puzzle. In particular, post-formation earnings shocks alone can explain 38.5% of the puzzle, followed by expected idiosyncratic skewness at 35.0%, retail trading proportion at 27.1%, one-month return reversal at 24.5%, and pre-formation earnings shocks at 12.9%. For the maximum daily return variable proposed by Bali et al. (2011), it turns out that it can explain the entire puzzle. The problem, however, is that this variable is close to being perfectly collinear with idiosyncratic volatility (correlation of about 90%). When we break this mechanical relation using alternative measures of maximum return, the explained fraction drops to 23-54%, which is still impressive.

Finally, we include all explanations that on their own can explain more than 10% of the puzzle (and are also statistically significant) in a multivariate framework, so that we can evaluate the marginal contribution of each explanation. We are also interested in the fraction of the puzzle they can collectively explain. We find that after controlling for competing explanations, maximum daily return (using less collinear versions) explains 28-54% of the puzzle and expected idiosyncratic skewness explains 13-20% of the puzzle, depending on the specification. Together, these two lottery preference-based explanations capture a good 48-67% of the puzzle. On the other hand, the one-month return reversal, a market friction-based explanation, explains 3-11% and earnings shocks (pre- and post-formation) explain 11-14% of the puzzle in the multivariate analysis. Collectively, the above explanations account for roughly 60-80% of the puzzle. We therefore conclude that a significant portion (20-40%) of the idiosyncratic volatility puzzle remains unexplained by existing explanations.

The rest of the paper is organized as follows. Section 2 describes the data and methodology and gives an overview of the various explanations that have been proposed for the idiosyncratic volatility puzzle. Section 3 evaluates the explanations one at a time, Section 4 considers multiple explanations at the same time as well as robustness tests, and Section 5 concludes.

2. Data and methodology

2.1. Stock return and idiosyncratic volatility data

We start our sample from the standard CRSP common stock (share codes of 10 or 11) universe from July 1963 to December 2009. Monthly returns are adjusted for delisting following Shumway (1997). To be included in the analysis, we require a firm to have non-missing values for size and book-to-market equity (*B/M*), where *Size* is the most recent June-end market cap of the firm and *B/M* is computed according to Fama and French (2006). We apply a price screen of one dollar to remove microcap stocks although we also relax or tighten this screen in robustness tests (Section 4.2).

We compute idiosyncratic volatility (*IVOL*) following Ang et al. (2006) as the standard deviation of the residuals from a regression of daily stock returns in month t–1 on the Fama and French (1993) factors. The estimates for *IVOL* start in July 1963 because this is the first available month of the daily Fama-French factor returns downloaded from Ken French's website. Month t–1 estimates of *IVOL* are matched to month t returns from August 1963 to December 2009.

2.2. Candidate variables related to lottery preferences of investors

A battery of candidate variables is constructed as potential explanations of the idiosyncratic volatility puzzle. The first group of explanations concerns lottery preferences of investors. Barberis and Huang (2008) argue that under cumulative prospect theory, investors overweigh small chances of large gains (hence the lottery preferences). As a result, they prefer positively-skewed stocks, causing them to be overpriced, which would then earn low subsequent returns. Several papers attribute the idiosyncratic volatility puzzle to idiosyncratic volatility being a proxy for skewness. We measure skewness using the daily returns in month t–1 (denoted *Skew*). In addition to the raw skewness measure, we also compute several alternative measures of skewness. Chabi-Yo and Yang (2009) develop a model showing that the effect of idiosyncratic volatility on stock returns is related to a stock's coskewness with the market portfolio. We measure coskewness (*Coskew_CY*) as the regression coefficient of squared daily individual stock returns on market returns. We also consider Harvey and Siddique (2000)'s measures of coskewness (*Coskew_HS*) and idiosyncratic skewness (*Idioskew_HS*) which are calculated by regressing daily individual stock returns on market returns and squared market returns.

Coskew_HS is the regression coefficient on squared market returns and *Idioskew_HS* is the skewness of the residuals.

Boyer et al. (2010) argue that realized idiosyncratic skewness is a poor proxy for expected idiosyncratic skewness. Instead, they use the forecasts from a regression model to proxy for expected idiosyncratic skewness and show that it helps explain the idiosyncratic volatility puzzle.[4] We obtain the estimates of expected idiosyncratic skewness (*E(Idioskew)*) from the authors and the data cover the period from 1988 to 2005.

We also consider the maximum daily return (*Maxret*) and the retail trading proportion (*RTP*) of a stock, which are proposed by Bali et al. (2011) and Han and Kumar (2012), respectively, as indicators for stocks that are preferred by lottery-seeking retail investors. *Maxret* is measured using daily returns in month *t*–1. *RTP* is measured as the fraction of the dollar trading volume in month *t*–1 that comes from trades less than $5000. It is computed using the Trades and Quotes (TaQ) database from 1993 to 2000.[5]

2.3. Candidate variables related to market frictions

The second group of explanations attributes the idiosyncratic volatility puzzle to market frictions. Fu (2009) and Huang et al. (2009) argue that once we control for the one-month return reversal effect, which is likely driven by the bid-ask bounce and other microstructure biases, the

[4] Their expected idiosyncratic skewness measure is estimated by regressing idiosyncratic skewness on lagged idiosyncratic skewness, idiosyncratic volatility, momentum, turnover, dummy variables for small firms and medium sized firms, two-digit SIC dummies, and a Nasdaq dummy. Boyer et al. (2010) measure idiosyncratic skewness using the residuals from a regression of the past five years of daily returns on the Fama-French factors. See Boyer et al. (2010) for more details.

[5] Following the literature, we exclude the post-decimalization years because it is more difficult in those years to identify retail trades using dollar screens due to greater incidence of order-splitting by institutions.

negative idiosyncratic volatility-return relation is no longer significant.[6] We measure the one-month reversal effect using the month t–1 return (*Lagret*).

Illiquidity can also affect the idiosyncratic volatility-return relation. We examine several measures of illiquidity. The Amihud (2002) measure (*Amihud*) is computed as the month t–1 average daily absolute return divided by the daily dollar trading volume (in millions). We also include the reciprocal of the average price (*Invprc*) and the fraction of trading days in month t–1 with a zero return (*Zeroret*) to control for illiquidity effects related to the price level of a stock (Bali and Cakici (2008)) and the zero-return proportion (Han and Lesmond (2011)), respectively.

Papers have also appealed to short-sale constraints (Boehme et al. (2009) and George and Hwang (2011)) and investor attention (George and Hwang (2011)) as possible explanations of the puzzle. We use short interest in month t–1 (*ShortInt*) obtained from Karl Diether to proxy for shorting demand and the most recently reported quarterly institutional ownership from Thomson 13F filings (*InstOwn*) for shorting supply. These two variables are related to short-sale constraints (Cohen, Diether, and Malloy (2007)) as high demand or low supply represents constrained shorting ability. We measure investor attention using the number of I/B/E/S analysts issuing earnings forecasts for a stock in month t–1 (*Nanalyst*).

Finally, we consider the price delay variable (*Delay*) of Hou and Moskowitz (2005). It measures the severity of market frictions affecting a stock using the delay with which its price responds to information. It is estimated using a regression of a stock's weekly returns on contemporaneous and past four weeks' returns on the market portfolio. The data on price delay are provided by the authors for the period 1965 to 2002.

[6] A recent paper by Chen, Jiang, Xu, and Yao (2012) challenges the robustness of this result.

2.4. Candidate variables related to other explanations

The third group of explanations consists of those that do not fall naturally into the lottery preference or market friction category. First, idiosyncratic volatility could proxy for the fundamental uncertainty surrounding a stock. Johnson (2004) argues that uncertainty is negatively related to future stock returns because stock is a call option on a levered firm's underlying assets. We measure uncertainty using analyst dispersion (*Dispersion*) which is the standard deviation of analysts' FY1 forecasts scaled by the absolute value of the mean consensus forecast. We also include leverage (*Leverage*, Compustat long-term debt (LTDEBT) over total assets (AT)) which is important for the pricing of uncertainty in Johnson's framework.

A recent paper by Avramov et al. (2012) shows that many cross-sectional return anomalies, including the idiosyncratic volatility puzzle, exist only among financially distressed firms. We follow their paper and measure financial distress as the negative credit rating change of a firm in month t–1 (*NegRatingChg*). Credit rating is the S&P long-term credit rating reported in Compustat (SPLTICRM) and is converted into ascending numerical scores (1 represents an AAA rating and 22 represents a D rating, etc.) so that positive changes to the numerical ratings indicate rating downgrades.

We obtain Fu (2009) EGARCH estimates of monthly expected idiosyncratic volatility (*Egarch*). Fu (2009) finds that the variable is positively related to stock returns. We also include a measure of R&D intensity (*R&D*), defined as the Compustat R&D expenditure (XRD) divided by total assets (AT), as a proxy for growth options (Cao et al. (2008) and Chen and Petkova (2012)). In addition, Chen and Petkova (2012) argue that a stock's exposure to the average variance component of the market variance explains the idiosyncratic volatility puzzle. We replicate their measure of average variance beta (*AvgVarβ*) and include it in the analysis.[7]

We also examine *SUE* (the most recent quarter's standardized unexpected earnings as of month *t*–1) and *NextSUE* (next quarter's standardized unexpected earnings). Wong (2011) shows that high idiosyncratic volatility stocks suffer negative earnings shocks both before and after portfolio formation, which could explain the poor return performance of those stocks.[8] Standardized unexpected earnings are measured as the Compustat quarterly earnings before extraordinary items (item IBQ) minus the earnings four quarter ago, divided by the standard deviation of the difference over the last eight quarters.

2.5. Decomposition methodology

Our decomposition methodology is based on Fama-Macbeth cross-sectional regressions. For each month *t*, we regress the cross-section of individual stock returns on their month *t*–1 $IVOL$ as follows:

$$R_{it} = \alpha_t + \gamma_t IVOL_{it-1} + \varepsilon_{it}. \tag{1}$$

The negative idiosyncratic volatility-return relation (the idiosyncratic volatility puzzle) shows up as a statistically significant time-series average γ_t coefficient. For our baseline sample, the average γ_t coefficient (×100 and reported in basis points (bps)) equals –17.128 bps with a *t*-statistic of –3.69 (see Table 2). For simplicity, we use raw returns in the regressions. Our results are robust to using characteristic-adjusted returns to control for other known determinants of

---

[7] We follow Chen and Petkova (2012) and estimate the average variance beta for the sample period of 1968 to 2009. Each month *t*, a stock's past 60 months (24-month minimum) of returns are regressed on changes in the average variance (AV) of the market portfolio, changes in the average correlation (AC) of the market portfolio, and the Fama-French factors. AV is the average of the individual stock daily return variances. AC is the average pairwise correlation between stocks.

[8] See also Jiang et al. (2009).

average returns, such as size, B/M, and momentum. Those results are reported in Sections 3.2 and 4.2.

Next, we regress $IVOL_{it-1}$ on a candidate explanatory variable ($Candidate_{it-1}$):

$$IVOL_{it-1} = a_{t-1} + \delta_{t-1} Candidate_{it-1} + \mu_{it-1}. \tag{2}$$

This regression allows us to assess the relation between idiosyncratic volatility and the candidate variable as any candidate variable that can potentially explain the puzzle must be correlated with idiosyncratic volatility (although a high correlation in and of itself does not guarantee that the candidate variable will explain a large fraction of the puzzle). We then use the regression coefficient estimates to decompose $IVOL_{it-1}$ into two orthogonal components: $\delta_{t-1} Candidate_{it-1}$ is the component of $IVOL_{it-1}$ that is related to the candidate variable and $(a_{t-1} + \mu_{it-1})$ is the residual component that is unrelated to the candidate variable.

The final step is to use linearity of covariances to decompose the estimated γ_t coefficient from Equation 1:

$$\begin{aligned}
\gamma_t &= \frac{Cov[R_{it}, IVOL_{it-1}]}{Var[IVOL_{it-1}]} \\
&= \frac{Cov[R_{it}, (\delta_{t-1} Candidate_{it-1} + a_{t-1} + \mu_{it-1})]}{Var[IVOL_{it-1}]} \\
&= \frac{Cov[R_{it},\ \delta_{t-1} Candidate_{it-1}]}{Var[IVOL_{it-1}]} + \frac{Cov[R_{it}, (a_{t-1} + \mu_{it-1})]}{Var[IVOL_{it-1}]} \\
&= \gamma_t^C + \gamma_t^R.
\end{aligned} \tag{3}$$

The time-series average of γ_t^C divided by the time-series average of γ_t then measures the fraction of the idiosyncratic volatility puzzle explained by the candidate variable, and the average γ_t^R divided by the average γ_t measures the fraction of the puzzle left unexplained by the candidate variable. Using the time-series standard errors of γ_t^C and γ_t^R, we can also determine whether the candidate and residual components are statistically significant.

Our decomposition approach is different from the conventional approach that past studies use to evaluate a candidate variable, which usually involves including the candidate variable as a control in the regression of returns on idiosyncratic volatility:

$$R_{it} = \tilde{\alpha}_t + \tilde{\gamma}_t^R IVOL_{it-1} + \tilde{\gamma}_t^C Candidate_{it-1} + \tilde{\varepsilon}_{it}. \tag{4}$$

In this regression, if the average coefficient on *IVOL* is insignificant, researchers typically conclude that the candidate variable explains the idiosyncratic volatility puzzle. However, if the average coefficient on *IVOL* remains significant, this conventional approach does not allow one to easily quantify the fraction of the puzzle that is explained by a candidate variable, especially when it is evaluated against other competing variables. The important advantage of our decomposition approach is that by requiring that γ_t^C and γ_t^R add up to the original γ_t coefficient, we can make a direct statement about the fraction of the idiosyncratic volatility puzzle that is explained by the candidate variable. In addition, our approach can easily accommodate multiple candidate variables at the same time so we can objectively quantify the marginal contribution of each variable in a horse race.

It is important to point out that a candidate variable that is highly correlated with idiosyncratic volatility may not necessarily explain a large fraction of the puzzle according to our approach. This is because the part of idiosyncratic volatility that is related to the candidate variable may not be the part that is responsible for the negative relation between idiosyncratic volatility and returns. In the appendix, we show that γ_t^C from our decomposition approach (Equation 3) is related to the coefficients from the conventional approach (Equation 4) in the following way: $\gamma_t^C = \left(\frac{\tilde{\gamma}_t^C}{\delta_{t-1}} + \tilde{\gamma}_t^R\right) \times \frac{Var[\delta_{t-1} Candidate_{it-1}]}{Var[IVOL_{it-1}]}$ for each month *t*.[9] This suggests that γ_t^C not only depends on the fraction of the variation of idiosyncratic volatility explained by the candidate variable $\left(\frac{Var[\delta_{t-1} Candidate_{it-1}]}{Var[IVOL_{it-1}]}\right)$, but also on the component of the candidate variable that is uncorrelated with idiosyncratic volatility but correlated with future returns (as captured by $\tilde{\gamma}_t^C$ from Equation 4). Consequently, a candidate variable that is highly correlated with idiosyncratic volatility could actually have a small or even negative contribution to the puzzle if the component of the candidate variable that is uncorrelated with idiosyncratic volatility predicts returns positively. Empirically, we show in Section 3 that this is indeed the case for a number of candidate variables we investigate. The bottom line is that our decomposition approach is not simply picking up candidate variables based solely on their correlations with idiosyncratic volatility. Rather, we attribute a high explanatory power to a variable for capturing a significant fraction of the negative relation between idiosyncratic volatility and returns.

3. Evaluating candidate explanations one at a time

3.1. Sample descriptive statistics

Panel A of Table 1 reports the descriptive statistics of our sample. There are more than two million firm-month observations in our baseline sample. The average return is 1.1% per month with a standard deviation of 15.9%. The average *IVOL* estimated using daily returns is 2.7%. The average market beta, size, B/M ratio, and momentum (buy-and-hold return from month *t*–12 to *t*–2) are 1.135, $1.3 billion, 0.887, and 16.4%, respectively.

[Insert Table 1 here]

[9] We also consider the general case of multiple candidate variables in the appendix.

The rest of Panel A reports the three groups (lottery preference, market friction, and others) of candidate variables. Among the lottery preference variables, the average skewness is 0.264, suggesting that stock returns are on average positively skewed. The average maximum daily return is 7.2%. The average retail trading proportion is 16%, indicating that retail investors typically do not account for a large fraction of the trading of a stock. Among the market friction variables, last month's return has an average value of 1.6%. The average value of zero return proportion, an illiquidity proxy, is 21.9%, which indicates that on average about one-fifth of the trading days in a month have zero returns. The average values of the two short-sale constraint proxies—short interest (demand) and institutional ownership (supply)—are 1.5% and 35.4%, respectively. Finally, among the other candidate variables, analyst dispersion has an average value of 19.5%. For the negative rating change variable, the majority of the monthly observations are equal to zero since long-term credit ratings only change very infrequently. As a result, the average value of the variable is minuscule (0.012). The average monthly EGARCH expected idiosyncratic volatility is 12.1%, and the average pre- and post-formation earnings shocks are 17.0% and 16.2%, respectively.

Panel B of Table 1 reports the correlations between the variables. The average correlation between month t–1 *IVOL* and month t returns is –3.3%, which is consistent with the negative idiosyncratic volatility-return relation documented in the literature. The second column of Panel B shows that *IVOL* is positively correlated with skewness, coskewness (*Coskew_CY*), idiosyncratic skewness, expected idiosyncratic skewness, maximum daily return, retail trading proportion, last month's return, Amihud's illiquidity measure, inverse of price, price delay, analyst dispersion, negative rating change, EGARCH expected idiosyncratic volatility, R&D intensity, and average variance beta, and negatively correlated with coskewness (*Coskew_HS*), short interest, institutional ownership, analyst coverage, leverage, and pre- and post-formation earnings shocks. These correlations are generally consistent with the various explanations that have been proposed for the idiosyncratic volatility puzzle. For example, the average correlation between *IVOL* and skewness is 19.3%, which is consistent with the lottery preference explanation that *IVOL* proxies for skewness. Or consider pre- and post-formation earnings shocks. The average correlations between them and *IVOL* are –10.2% and –9.4%, respectively. These correlations are in line with Wong's (2011) conclusion that the weak earnings performance of high idiosyncratic stocks is responsible for their poor return performance. Among all the candidate variables, the one that has the highest correlation with *IVOL* is maximum daily return (average correlation of 88.3%), suggesting that it is well placed to explain the idiosyncratic volatility puzzle. However, collinearity could be a concern for this variable.

3.2. The idiosyncratic volatility puzzle

To set the stage, Table 2 reports the results of monthly Fama-MacBeth cross-sectional regressions of month *t* individual stock returns on month *t*–1 *IVOL* and different candidate variables.[10]

[Insert Table 2 here]

Model 1 of Panel A regresses raw returns on *IVOL* alone. The sample period is August 1963 to December 2009 with an average of 3,654 stocks per month. The average coefficient on *IVOL* is –17.128 bps (*t*=–3.69) and its magnitude and statistical significance are in line with the

[10] Asparouhova, Bessembinder, and Kalcheva (2012) show that microstructure noise introduces an upward bias to stock returns, which could potentially bias the inferences from Fama-MacBeth regressions. We follow their paper by using (one plus) month *t*–1 return as the weight in the Fama-MacBeth regressions and find that our results are robust to this noise-adjustment procedure.

findings in the literature.[11] Models 2-8 of Panel A add the lottery preference-based candidate variables one at a time to Model 1. For each model, the number of observations and sample period may be different from that of Model 1 due to data availability of the candidate variable of interest. The results from Models 2-8 show that in all but one case, the coefficient on *IVOL* remains negative and statistically significant. Only when maximum daily return (*Maxret*) is included in the regression does the coefficient on *IVOL* become positive and marginally significant, which is consistent with the findings of Bali et al. (2011).

Models 9-16 of Panel A regress characteristics-adjusted returns instead of raw returns on *IVOL* and the lottery preference variables. We form characteristic-based benchmark portfolios following Daniel, Grinblatt, Titman, and Wermers (1997) (DGTW) and subtract their returns from the raw returns of individual stocks to control for the size, B/M, and momentum effects in average returns.[12] This is equivalent to including those characteristics as control variables in the cross-sectional regressions. Model 9 shows that controlling for the premiums associated with size, B/M, and momentum strengthens the statistical significance of the idiosyncratic volatility-return relation (t=–7.80 compared to –3.69 from Model 1) while leaving the economic magnitude of the effect almost unchanged (–16.300 vs. –17.128 bps). Models 10-16 confirm the results from Models 2-8 that except for maximum daily return, controlling for the other lottery preference variables does not remove the negative and significant relation between idiosyncratic volatility and returns.

[11] The average coefficients from all of our cross-sectional regressions are multiplied by 100 and reported in basis points.

[12] DGTW benchmark portfolios are constructed as follows. At the end of June of each year, firms are first sorted into quintiles based on their market capitalization using NYSE breakpoints. Then, within each size quintile, firms are sorted into quintiles according to their B/M ratios. In the last step, firms within each double-sorted size-B/M portfolio are further sorted into momentum quintiles every month, based on their returns over the prior twelve months skipping the most recent month. Equally-weighted monthly returns are computed for each characteristic benchmark portfolio.

Panels B and C of Table 2 investigate the candidate variables related to market frictions and other explanations, respectively. The results show that the coefficient on *IVOL* is always negative and statistically significant, irrespective of the candidate variables considered or returns used in the regressions (raw vs. characteristic-adjusted).

The main takeaway from Table 2 is that the negative idiosyncratic volatility-return relation remains significant after controlling for almost all of the candidate explanatory variables (except for the maximum daily return variable). But the question remains: Even if these candidate variables cannot completely explain away the idiosyncratic volatility puzzle, can they at least explain part of it? If so, what fraction of the puzzle can the candidate variables capture? We investigate this next using the decomposition methodology described in Section 2.5.

3.3. Candidate variables related to lottery preferences of investors

We first examine the candidate variables related to lottery preferences of investors. We start off with a detailed account of the decomposition analysis using skewness (*Skew*) in Panel A of Table 3. Stage 1 reports monthly Fama-MacBeth cross-sectional regressions of month t returns on month t–1 *IVOL* using firm-month observations with non-missing *IVOL* and *Skew*. This is to ensure that the sample is kept constant when we later add *Skew* to the analysis. The average coefficient on *IVOL* is –17.540 bps with a t-statistic of –3.78.

In Stage 2, we add *Skew* to the cross-sectional regressions. This is identical to Model 2 in Table 2. The average coefficient on *Skew* is –0.085 bps with a t-statistic of –4.18, which is consistent with Barberis and Huang (2008)'s argument that investors overprice positively-skewed stocks and as a result the future returns of those stocks are low. Controlling for *Skew*, we see that the average coefficient on *IVOL* is still significantly negative (–16.357 bps, t=–3.45).

In Stage 3, we regress *IVOL* each month on *Skew* to study the relation between the two variables. The average coefficient on *Skew* is 0.369 bps with a *t*-statistic of 32.17, suggesting that part of *IVOL* is indeed related to the skewness of a security. However, the adjusted R-squared shows that only 4.3% of the variation in *IVOL* can be explained by *Skew*. The Stage 3 estimated coefficients allow us to separate *IVOL* each month into two components: the first one ($\delta_{t-1}Skew_{it-1}$) is the component of *IVOL* that is related to *Skew* and the second ($a_{t-1} + \mu_{it-1}$) is the residual component that is unrelated to *Skew*.

In Stage 4, we follow Equation 3 and use the above two components of *IVOL* to decompose the Stage 1 *IVOL* coefficient (γ_t) into a component that is related to *Skew* (γ_t^{Skew}) and a residual component (γ_t^R). The time-series averages of γ_t^{Skew} and γ_t^R are –1.211 bps (*t*=–3.28) and–16.328 bps (*t*=–3.67), respectively. Since by construction the two coefficients sum up to the Stage 1 coefficient of –17.540 bps, we can readily calculate the fraction of the Stage 1 coefficient attributable to *Skew* as $\frac{-1.211}{-17.540}$ =6.9%, and the fraction attributable to the residual component is $\frac{-16.328}{-17.540}$ =93.1%. We therefore conclude that skewness can only explain a very small fraction of the idiosyncratic volatility puzzle.

[Insert Table 3 here]

We also examine the other skewness variables in Panel A of Table 3. The results show that most of them also fail to explain a significant fraction of the puzzle. The coskewness measures of Chabi-Yo and Yang (2009) (*Coskew_CY*) and Harvey and Siddique (2000) (*Coskew_HS*) and the idiosyncratic skewness measure of Harvey and Siddique (2000) (*Idioskew_HS*) explain only 4.2%, 0.7%, and 4.9% of the puzzle, respectively.

The last column of Panel A examines the expected idiosyncratic skewness measure (*E(Idioskew)*) of Boyer et al. (2010). The Stage 2 average coefficient of regressing returns on this variable is negative (–0.351 bps) and statistically significant (t=–2.06) (the average coefficient on *IVOL* is also significantly negative in Stage 2). In Stage 3, *E(Idioskew)* explains 21.4% of the variation in *IVOL*. Finally, the Stage 4 decomposition shows that *E(Idioskew)* captures a sizable 35.0% of the puzzle, which is consistent with the findings of Boyer et al. (2010).

Panel B of Table 3 first investigates the maximum daily return measure (*Maxret*) of Bali et al. (2011) and shows that the portion of the puzzle that is explained is 105.8%.[13] It thus appears that *Maxret* completely explains the idiosyncratic volatility puzzle. However, given the near collinearity between *Maxret* and *IVOL* (a correlation of 88.3%, Table 1 Panel B), one might be concerned that this finding is mechanical. To mitigate this concern, we examine a number of alternative measures of maximum daily return that are less collinear with *IVOL* but at the same time maintain the ability to capture the lottery feature of a stock. The first measure we consider is the maximum daily return for the three-month period ending in month t–1 (*Maxret(3mth)*), as it is possible that lottery-seeking investors will use daily returns beyond the past one month to determine the lottery feature of a stock. The correlation between *Maxret(3mth)* and *IVOL* is 0.70, which suggests that collinearity is less of an issue. We find that *Maxret(3mth)* explains 53.8% of the puzzle, which is still very impressive. We also examine *Maxret(6mth)* and *LagMaxret(3mth)*, which are the maximum daily returns for the six month period ending in month t–1 and for the three-month period ending in month t–2, respectively. These two variables are even less correlated with *IVOL* (correlations of 0.63 and 0.48, respectively), and they explain 35.7% and 23.2% of the puzzle, respectively.

The last column of Panel B shows that the retail trading proportion measure (*RTP*) of Han and Kumar (2012) explains 27.1% of the idiosyncratic volatility puzzle. However, the *RTP*

[13] The reason this fraction is above 100% is because the adding-up constraint in Stage 4 requires the *Maxret* component and the residual component must add up to the Stage 1 coefficient on *IVOL*.

portion of the puzzle (average γ_t^{RTP} in Stage 4 is –4.932 bps) is not statistically significant (t=–1.23).

Overall, the results from Table 3 suggest that among all the lottery preference-based variables we examine, the expected idiosyncratic skewness measure of Boyer et al. (2010) and the maximum daily return measure of Bali et al. (2011) are the most promising candidates in explaining the idiosyncratic volatility puzzle, whereas the rest of the variables fail to capture significant fractions of the puzzle.

3.4. Candidate variables related to market frictions

Table 4 examines the candidate variables related to market frictions. We first consider the one-month reversal effect studied in Fu (2009) and Huang et al. (2009), which is likely driven by the bid-ask bounce and other microstructure effects. Panel A of Table 4 shows that the month t–1 return (*Lagret*) explains about a quarter (24.5%) of the idiosyncratic volatility puzzle, with the residual component capturing the remaining 75.5% of the puzzle.

[Insert Table 4 here]

The rest of Panel A examines the impact of illiquidity on the puzzle using the Amihud (2002) measure (*Amihud*), the reciprocal of the average price (*Invprc*), and the zero-return proportion (*Zeroret*) of Han and Lesmond (2011). The results show that none of them can explain a significant fraction of the puzzle. In fact, the fractions explained are all negative at –8.8%, –12.4%, and –1.4% for *Amihud*, *Invprc*, and *Zeroret*, respectively. Intuitively, the reason that these illiquidity proxies have negative contributions to the idiosyncratic volatility puzzle is because they are all positively correlated with *IVOL* (especially for *Amihud* and *Invprc*) but their return predictability (after controlling for idiosyncratic volatility) is also positive, which is in the opposite direction of the idiosyncratic volatility puzzle. The low explanatory power of these illiquidity candidate variables despite their high correlations with idiosyncratic volatility shows that our decomposition approach does not necessarily attribute a high explanatory power to a candidate variable just because it has a high correlation with idiosyncratic volatility.

Panel B of Table 4 investigates the other market friction-based candidate variables including short interest (*ShortInt*), institutional ownership (*InstOwn*), analyst coverage (*Nanalyst*), and the price delay measure (*Delay*) of Hou and Moskowitz (2005). *ShortInt* and *InstOwn* are both related to short-sale constraints, *Nanalyst* proxies for investor attention, and *Delay* captures the delay with which a firm's stock price responds to market-wide information. The results show that none of these variables explains more than 10% of the puzzle. *InstOwn* captures 7.1% of the puzzle, followed by *Nanalyst* at 5.6%, and then *ShortInt* and *Delay* at –0.4% and –18.9%, respectively.

In sum, with the exception of the one-month reversal effect, most of the market friction variables we examine have very little success in explaining the idiosyncratic volatility puzzle.

3.5. Candidate variables related to other explanations

Table 5 examines candidate variables that cannot be grouped into the lottery preference or market friction category. We first look at analyst dispersion (*Dispersion*) and leverage (*Leverage*), which are important for the pricing of uncertainty in Johnson (2004). We find that *Dispersion* can only explain a small fraction (6.4%) of the idiosyncratic volatility puzzle, whereas the contribution of *Leverage* is negligible (–0.8%).

We also investigate the negative rating change measure (*NegRatingChg*) of Avramov et al. (2012). The coverage of this variable is very sparse (averaging 1,000 firms per year from 1986-2009) due to the limited availability of the S&P long-term credit rating in Compustat. The results show that *NegRatingChg* is able to explain 5.4% of the puzzle.

The next set of candidate variables investigated include the expected idiosyncratic volatility measure (*Egarch*) of Fu (2009), R&D intensity (*R&D*), and average variance beta (*AvgVarβ*) of Chen and Petkova (2012). None of these variables shows much promise in explaining the puzzle, with the fractions explained equal to –91.5%, 0.9%, and 0.9% for *Egarch*, *R&D*, and *AvgVarβ*, respectively.[14]

[Insert Table 5 here]

Finally, we examine whether pre- and post-formation earnings shocks (*SUE* and *NextSUE*, respectively) can explain the idiosyncratic volatility puzzle. We find that *SUE* captures 12.9% of the puzzle whereas *NextSUE* captures 38.5% of the puzzle, which is consistent with the findings in Wong (2011).

4. Evaluating multiple candidate explanations at the same time

4.1. The most promising candidate variables

After investigating each of the candidate variables in isolation, we now turn to multivariate analysis and evaluate against each other the variables that account for at least 10% of the puzzle and are also statistically significant.[15] We want to know the marginal contribution of each variable after controlling for competing variables. In addition, we are interested in the total fraction of the puzzle those candidate variables can collectively explain.

The candidate variables that can explain 10% or more of the puzzle include the expected idiosyncratic skewness measure (*E(Idioskew)*) of Boyer et al. (2010), different variants of the maximum daily return measure (*Maxret*) of Bali et al. (2011), past one month's return (*Lagret*),

[14] The large negative contribution of *Egarch* is due to the high correlation between *Egarch* and *IVOL* (0.451, Table 1 Panel B) combined with the strong positive return predictability of *Egarch* (average coefficient of 15.883 bps with a *t*-stat of 14.20 in Stage 2 regressions).

[15] Choosing 10% as the cutoff is admittedly arbitrary. That said, our inferences are robust to using higher or lower levels of cutoff or removing the cutoff (and thereby using the entire set of candidate variables).

and pre- and post-formation earnings shocks (*SUE* and *NextSUE*, respectively). We put these variables through our decomposition framework and the linear adding-up constraint ensures that their contributions plus that of the residual component add up to 100% of the puzzle.

[Insert Table 6 here]

Table 6 reports the results. In Model 1, we measure maximum daily return using *Maxret(3mth)* to avoid the near collinearity between *IVOL* and the original *Maxret* variable. We see that the five variables (*E(Idioskew)*, *Maxret(3mth)*, *Lagret*, *SUE*, and *NextSUE*) collectively explain 81.4% of the puzzle and the residual component accounts for the remaining 18.6% of the puzzle. The best performing variable is *Maxret(3mth)* which alone captures 56.3% of the puzzle, followed by *E(Idioskew)* at 15.6%, and then *NextSUE*, *Lagret*, and *SUE* at 6.1%, 1.9%, and 1.5%, respectively. We note that controlling for the two lottery preference-based variables (*Maxret(3mth)* and *E(Idioskew)*) takes away most of the explanatory power of the other candidate variables. For example, the market friction variable *Lagret* now accounts for a mere 1.9% of the puzzle, which is much lower than its contribution in the univariate analysis (24.5%).

Model 2 replaces *Maxret(3mth)* with *LagMaxret(3mth)*, which has a even lower correlation with *IVOL* (0.70 vs. 0.48). The total fraction of the puzzle explained by the five candidate variables declines to 76.0%, with *LagMaxret(3mth)*, *E(Idioskew)*, *Lagret*, *NextSUE*, and *SUE* each contributing 31.5%, 21.6%, 12.9%, 8.0%, and 2.0%, respectively. In Model 3, we drop *NextSUE* from Model 2 to focus on candidate variables that are strictly ex ante (measured prior to the returns being predicted). The total fraction explained now declines further to 65.4%, with *LagMaxret(3mth)*, *E(Idioskew)*, *Lagret*, and *SUE* contributing 29.7%, 21.8%, 11.4%, and 2.5%, respectively.

The main takeaway from Table 6 is that about 19-35% of the idiosyncratic volatility puzzle remains unexplained after we control for the candidate variables that have the most success in the univariate analysis. In this multivariate setting, the best performing candidate variables are the maximum daily return measure and the expected idiosyncratic skewness measure, both of which are related to lottery preferences of investors.

4.2. Robustness tests

Table 7 performs a number of robustness checks. First, Bali and Cakici (2008), George and Hwang (2011), and Chen et al. (2012) all emphasize the importance of price screens for the idiosyncratic volatility puzzle. For our main analysis, we exclude firms with a price below one dollar at the end of the previous month. Panel A of Table 7 repeats the tests in Table 6 by imposing a more stringent five dollar price screen.[16] The results show that the total fraction explained by the candidate variables is smaller relative to that in Table 6, and the residual component now captures 39-53% of the puzzle. This is consistent with the finding in George and Hwang (2011) and Chen et al. (2012) that imposing more stringent price screens actually strengthens the idiosyncratic volatility puzzle. On the other hand, when we remove all price screens in Panel B of Table 7, we see that a larger fraction of the puzzle is explained by the candidate variables and the residual component now only accounts for 7-22% of the puzzle.

[Insert Table 7 here]

In Panel C, we revert to the one-dollar price screen and use DGTW-adjusted returns to control for the size, B/M, and momentum effects in average returns. The results show that 10-26% of the puzzle remains unexplained by the candidate variables.

[16] To conserve space, we only report the Stage 4 decomposition results.

Panel D of Table 7 re-examines the findings in Table 6 using a three-month return window instead of the one-month window used in our main analysis. In other words, the cross-sectional regressions estimated every month now use the three-month buy-and-hold returns beginning in month *t* as the independent variable.[17] We find that the unexplained fraction of the puzzle goes up to 27-53%.

Panel E of Table 7 reports a similar analysis as in Panel D but uses DGTW-adjusted returns. Here, 11-39% of the puzzle cannot be explained. Finally, Panel F of Table 7 adds a five-dollar price screen to Panel E, and we find that 25-52% of the puzzle remains unaccounted for by the candidate variables.

[Insert Figure 1 here]

We summarize the results from the multivariate analysis by plotting in Figure 1 the average fraction of the puzzle explained across the seven specifications in Tables 6 and 7. The first pie chart (Model 1) shows that on average, the various candidate variables explain a total of 80.4% of the idiosyncratic volatility puzzle while the residual component captures the remaining 19.6%. Comparing across different candidate variables, we see that the two lottery preference-based variables (*Maxret(3mth)* and *E(Idioskew)*) combine to explain 67.3% of the puzzle (53.9% for *Maxret(3mth)* and 13.4% for *E(Idioskew)*), followed by the two earnings shock proxies (*SUE* and *NextSUE*) which together explain 10.5%, and the market friction-related *Lagret* which explains 2.8% of the puzzle.

The second and third pie charts show that, when we replace *Maxret(3mth)* with *LagMaxret(3mth)* (Model 2) and also drop the ex post variable *NextSUE* (Model 3), the average unexplained fraction rises to 25.8% and 40.0%, respectively. The two lottery preference-based variables (*LagMaxret(3mth)* and *E(Idioskew)*) continue to dominate the other candidate variables

[17] We use Newey-West standard errors with three lags to account for overlapping returns.

in the multivariate analysis. They combine to explain close to 50% of the puzzle in both models. The contributions of *Lagret* and *SUE* are 11.3% and 2.3%, respectively, in Model 2 and 9.6% and 2.8%, respectively, in Model 3. Finally, *NextSUE* explains 11.6% of the puzzle in Model 2.

5. Conclusion

In this paper, we propose a simple methodology to examine a large number of explanations that have been proposed in the literature for the negative relation between idiosyncratic volatility and subsequent stock returns (the idiosyncratic volatility puzzle). The main advantage of our approach is that it allows us to objectively quantify the contribution of each explanation either by itself or when evaluated against competing explanations.

We find that surprisingly many existing explanations explain less than 10% of the idiosyncratic volatility puzzle. On the other hand, the maximum daily return measure of Bali et al. (2011), the expected idiosyncratic skewness measure of Boyer et al. (2010), one-month return reversal, and pre- and post-formation earnings shocks show much greater success in explaining the puzzle, although together they still leave 20-40% of the puzzle unexplained. Our approach can be easily adapted to evaluate competing explanations of a broad range of topics in empirical asset pricing and corporate finance.

Appendix

In this appendix, we demonstrate the relation between our decomposition methodology (Equation 3) and the conventional approach of regressing returns on idiosyncratic volatility and a candidate variable (Equation 4). Specifically, for each month *t*, we can substitute Equation 2 into Equation 4 and obtain:

$$\begin{aligned} R_{it} &= \tilde{\alpha}_t + \tilde{\gamma}_t^R(a_{t-1} + \mu_{it-1} + \delta_{t-1}Candidate_{it-1}) + \tilde{\gamma}_t^C Candidate_{it-1} + \tilde{\epsilon}_{it} \\ &= \tilde{\alpha}_t + \tilde{\gamma}_t^R(a_{t-1} + \mu_{it-1}) + (\tilde{\gamma}_t^C + \delta_{t-1}\tilde{\gamma}_t^R)Candidate_{it-1} + \tilde{\epsilon}_{it} \\ &= \tilde{\alpha}_t + \tilde{\gamma}_t^R(a_{t-1} + \mu_{it-1}) + \bar{\gamma}_t^C Candidate_{it-1} + \tilde{\epsilon}_{it}, \end{aligned} \tag{5}$$

where $\bar{\gamma}_t^C$, which equals $\tilde{\gamma}_t^C + \delta_{t-1}\tilde{\gamma}_t^R$, is identical to the coefficient of regressing returns on the candidate variable alone because $(a_{t-1} + \mu_{it-1})$ and $Candidate_{it-1}$ are uncorrelated by construction. We can then rewrite γ_t^C from Equation 3 as follows:

$$\begin{aligned} \gamma_t^C &= \frac{Cov[R_{it}, \delta_{t-1}Candidate_{it-1}]}{Var[IVOL_{it-1}]} \\ &= \frac{Cov[R_{it}, \delta_{t-1}Candidate_{it-1}]}{Var[\delta_{t-1}Candidate_{it-1}]} \times \frac{Var[\delta_{t-1}Candidate_{it-1}]}{Var[IVOL_{it-1}]} \\ &= \frac{\bar{\gamma}_t^C}{\delta_{t-1}} \times \frac{Var[\delta_{t-1}Candidate_{it-1}]}{Var[IVOL_{it-1}]} \\ &= \left(\frac{\tilde{\gamma}_t^C}{\delta_{t-1}} + \tilde{\gamma}_t^R\right) \times \frac{Var[\delta_{t-1}Candidate_{it-1}]}{Var[IVOL_{it-1}]}. \end{aligned} \tag{6}$$

To show the relation in a more general case involving *k* candidate variables, we simplify the notation by denoting $IVOL_{it-1}$ as **V** and R_{it} as **R**, both are $n\times 1$ vectors where *n* is the number of firms in the month *t* cross-sectional regression. The time subscripts are suppressed for brevity. We also denote an $n\times 1$ vector of ones by $\boldsymbol{\iota}$. We can then rewrite Equation 1 as:

$$\mathbf{R} = \boldsymbol{\iota}\alpha + \mathbf{V}\gamma + \boldsymbol{\varepsilon}. \tag{7}$$

Next, we regress $\mathbf{V}$ on $\boldsymbol{\iota}$ and the $n\times k$ matrix of k candidate variables (measured contemporaneously with $\mathbf{V}$ in month t–1) denoted by $\mathbf{C} = (\mathbf{C_1} \quad \cdots \quad \mathbf{C_k})$, where $\mathbf{C_j}$ is an $n\times 1$ vector:

$$\mathbf{V} = \boldsymbol{\iota}a + \mathbf{C}\boldsymbol{\delta}^{\mathbf{C}} + \boldsymbol{\mu}. \tag{8}$$

$\boldsymbol{\delta}^{\mathbf{C}}$ is the $k\times 1$ vector of coefficients. In the last step, we decompose the idiosyncratic volatility-return relation γ into k components each related to a candidate variable and a residual component:

$$\begin{aligned}\gamma &= (\mathbf{v}'\mathbf{v})^{-1}\mathbf{V}'\mathbf{r} \\ &= (\mathbf{v}'\mathbf{v})^{-1}(\mathbf{C}\boldsymbol{\delta}^{\mathbf{C}} + \boldsymbol{\iota}a + \boldsymbol{\mu})'\mathbf{r} \\ &= (\mathbf{v}'\mathbf{v})^{-1}(\boldsymbol{\iota}a + \boldsymbol{\mu})'\mathbf{r} + (\mathbf{v}'\mathbf{v})^{-1}(\mathbf{C}\boldsymbol{\delta}^{\mathbf{C}})'\mathbf{r},\end{aligned} \tag{9}$$

where $\mathbf{v}$ and $\mathbf{r}$ (both $n\times 1$ vectors) are demeaned versions of $\mathbf{V}$ and $\mathbf{R}$, respectively. The first term in the last line of Equation 9 represents the unexplained component of the idiosyncratic volatility puzzle and the second term represents the combined contribution of all k candidate variables. The contribution of the jth candidate variable is then $\gamma_j^C = (\mathbf{v}'\mathbf{v})^{-1}\left(\mathbf{C_j}\delta_j^C\right)'\mathbf{r}$.

Now, take the conventional approach of regressing $\mathbf{R}$ on $\mathbf{V}$ and $\mathbf{C}$, namely,

$$\mathbf{R} = \boldsymbol{\iota}\tilde{\alpha} + \mathbf{V}\tilde{\gamma}^R + \mathbf{C}\tilde{\boldsymbol{\gamma}}^{\mathbf{C}} + \tilde{\boldsymbol{\varepsilon}}. \tag{10}$$

We can rewrite Equation 10 by substituting in Equation 8 as follows:

$$\begin{aligned}\mathbf{R} &= \boldsymbol{\iota}\tilde{\alpha} + (\boldsymbol{\iota}a + \mathbf{C}\boldsymbol{\delta}^{\mathbf{C}} + \boldsymbol{\mu})\tilde{\gamma}^R + \mathbf{C}\tilde{\boldsymbol{\gamma}}^{\mathbf{C}} + \tilde{\boldsymbol{\varepsilon}} \\ &= \boldsymbol{\iota}\tilde{\alpha} + (\boldsymbol{\iota}a + \boldsymbol{\mu})\tilde{\gamma}^R + \mathbf{C}(\tilde{\boldsymbol{\gamma}}^{\mathbf{C}} + \boldsymbol{\delta}^{\mathbf{C}}\tilde{\gamma}^R) + \tilde{\boldsymbol{\varepsilon}}.\end{aligned} \tag{11}$$

Because $\mathbf{C}$ and $(\boldsymbol{\iota}a + \boldsymbol{\mu})$ are uncorrelated by construction, the coefficient on the jth candidate variable $\left(\tilde{\gamma}_j^C + \delta_j^C\tilde{\gamma}^R\right)$ should be identical to the slope coefficient when $\mathbf{R}$ is regressed on the regression residual of $\mathbf{C_j}$ on the other $k-1$ candidate variables. Specifically, we define an $n\times(k+1)$ matrix $\mathbb{C} = (\boldsymbol{\iota} \quad \mathbf{C_1} \quad \cdots \quad \mathbf{C_k})$, an $(k+1)\times(k+1)$ matrix $\mathbf{J}$ which is an identify matrix except that the

(j+1)th diagonal term is set to zero, and $\boldsymbol{\theta}$ which is the (k+1)×1 vector of coefficients from regressing $\mathbf{C_j}$ on $\mathbb{C}\mathbf{J}$. Then we have:

$$\left(\tilde{\gamma}_j^C + \delta_j^C \tilde{\gamma}^R\right) = \left[\left(\mathbf{C_j} - \mathbb{C}\mathbf{J}\boldsymbol{\theta}\right)'\left(\mathbf{C_j} - \mathbb{C}\mathbf{J}\boldsymbol{\theta}\right)\right]^{-1}\left(\mathbf{C_j} - \mathbb{C}\mathbf{J}\boldsymbol{\theta}\right)'\mathbf{r}$$

$$\left(\tilde{\gamma}_j^C + \delta_j^C \tilde{\gamma}^R\right) = \left[\left(\mathbf{C_j} - \mathbb{C}\mathbf{J}\boldsymbol{\theta}\right)'\left(\mathbf{C_j} - \mathbb{C}\mathbf{J}\boldsymbol{\theta}\right)\right]^{-1}\left[\mathbf{C_j}'\mathbf{r} - (\mathbb{C}\mathbf{J}\boldsymbol{\theta})'\mathbf{r}\right]$$

$$\mathbf{C_j}'\mathbf{r} = \left[\left(\mathbf{C_j} - \mathbb{C}\mathbf{J}\boldsymbol{\theta}\right)'\left(\mathbf{C_j} - \mathbb{C}\mathbf{J}\boldsymbol{\theta}\right)\right]\left(\tilde{\gamma}_j^C + \delta_j^C \tilde{\gamma}^R\right) + (\mathbb{C}\mathbf{J}\boldsymbol{\theta})'\mathbf{r}. \tag{12}$$

We can then rewrite Equation 9 to give us the relation between γ_j^C (the contribution of the jth candidate variable to the idiosyncratic volatility puzzle) and $\tilde{\gamma}_j^C$ (the coefficient on the jth candidate variable in Equation 10):

$$\begin{aligned}
\gamma_j^C &= (\mathbf{v}'\mathbf{v})^{-1}\left(\mathbf{C_j}\delta_j^C\right)'\mathbf{r} \\
&= (\mathbf{v}'\mathbf{v})^{-1}\delta_j^C \mathbf{C_j}'\mathbf{r} \\
&= (\mathbf{v}'\mathbf{v})^{-1}\delta_j^C\left\{\left[\left(\mathbf{C_j} - \mathbb{C}\mathbf{J}\boldsymbol{\theta}\right)'\left(\mathbf{C_j} - \mathbb{C}\mathbf{J}\boldsymbol{\theta}\right)\right]\left(\tilde{\gamma}_j^C + \delta_j^C \tilde{\gamma}^R\right) + (\mathbb{C}\mathbf{J}\boldsymbol{\theta})'\mathbf{r}\right\}.
\end{aligned} \tag{13}$$

When k = 1, the above relation collapses to $(\mathbf{v}'\mathbf{v})^{-1}\delta_j^C\left(\mathbf{C_j}'\mathbf{C_j}\right)\left(\tilde{\gamma}_j^C + \delta_j^C \tilde{\gamma}^R\right)$, which is the matrix form of Equation 6.

References

Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2006, The cross-section of volatility and expected returns, *Journal of Finance* 61, 259-299.

Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2009, High idiosyncratic volatility and low returns: International and further U.S. evidence, *Journal of Financial Economics* 1-23.

Asparouhova, Elena, Hendrik Bessembinder, and Ivalina Kalcheva, 2012, Noisy prices and inference regarding returns, *Journal of Finance*, forthcoming.

Avramov, Doron, Tarun Chordia, Gergana Jostova, and Alexander Philipov, 2012, Anomalies and financial distress, *Journal of Financial Economics*, forthcoming.

Bali, Turan G. and Nusret Cakici, 2008, Idiosyncratic volatility and the cross section of expected returns, *Journal of Financial and Quantitative Analysis* 43, 29-58.

Bali, Turan G., Nusret Cakici, and Robert F. Whitelaw, 2011, Maxing out: Stocks as lotteries and the cross-section of expected returns, *Journal of Financial Economics* 99, 427-446.

Barberis, Nicholas and Ming Huang, 2008, Stocks as lotteries: The implications of probability weighting for security prices, *American Economic Review* 98, 2066-2100.

Boehme, Rodney D., Bartley R. Danielsen, Praveen Kumar, and Sorin M. Sorescu, 2009, Idiosyncratic risk and the cross-section of stock returns: Merton (1987) meets Miller (1977), *Journal of Financial Markets* 12, 438-468.

Boyer, Brian, Todd Mitton, and Keith Vorkink, 2010, Expected idiosyncratic skewness, *Review of Financial Studies* 23, 169-202.

Cao, Charles, Timothy Simin, and Jing Zhao, 2008, Can growth options explain the trend in idiosyncratic risk?, *Review of Financial Studies* 21, 2599-2633.

Chabi-Yo, Fousseni and Jun Yang, 2009, Default risk, idiosyncratic coskewness and equity returns, Working paper, Ohio State University.

Chen, Linda H., George J. Jiang, Danielle D. Xu, and Tong Yao, 2012, Dissecting the idiosyncratic volatility anomaly, Working paper, University of Arizona.

Chen, Zhanhui and Ralitsa Petkova, 2012, Does idiosyncratic volatility proxy for risk exposure?, *Review of Financial Studies* 25, 2745-2787.

Cohen, Lauren, Karl B. Diether, and Christopher Malloy, 2007, Supply and demand shifts in the shorting market, *Journal of Finance* 65, 2061-2096.

Daniel, Kent, Mark Grinblatt, Sheridan Titman, and Russ Wermers, 1997, Measuring mutual fund performance with characteristic-based benchmarks, *Journal of Finance* 52, 1035-1058.

Fama, Eugene F. and Kenneth R. French, 1993, Common risk factors in the returns of stocks and bonds, *Journal of Financial Economics* 33, 3-56.

Fama, Eugene F. and Kenneth R. French, 2006, The value premium and the CAPM, *Journal of Finance* 61, 2163-2185.

Fama, Eugene F. and James D. Macbeth, 1973, Risk, return and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607-636.

Fu, Fangjian, 2009, Idiosyncratic risk and the cross-section of expected stock returns, *Journal of Financial Economics* 91, 24-37.

George, Thomas J. and Chuan-Yang Hwang, 2011, Why do firms with high idiosyncratic volatility and high trading volume volatility have low returns?, Working paper, Nanyang Technological University.

Han, Bing and Alok Kumar, 2012, Speculative trading and asset prices, *Journal of Financial & Quantitative Analysis*, forthcoming.

Han, Yufeng and David Lesmond, 2011, Liquidity Biases and the Pricing of Cross-sectional Idiosyncratic Volatility, *Review of Financial Studies* 24, 1590-1629.

Harvey, Campbell R. and Akhtar Siddique, 2000, Conditional skewness in asset pricing tests, *Journal of Finance* 55, 1263-1295.

Hirshleifer, David, 1988, Residual risk, trading costs, and commodity futures risk premia, *Review of Financial Studies* 1, 173-193.

Hou, Kewei and Tobias J. Moskowitz, 2005, Market frictions, price delay, and the cross-section of expected returns, *Review of Financial Studies* 18, 981-1020.

Huang, Wei, Qianqiu Liu, S. Ghon Rhee, and Liang Zhang, 2009, Return reversals, idiosyncratic risk, and expected returns, *Review of Financial Studies* 23, 147-168.

Jiang, George J., Danielle Xu, and Tong Yao, 2009, The information content of idiosyncratic volatility, *Journal of Financial and Quantitative Analysis* 44, 1-28.

Johnson, Timothy C., 2004, Forecast dispersion and the cross-section of stock returns, *Journal of Finance* 59, 1957-1978.

Merton, Robert C., 1987, A simple model of capital market equilibrium with incomplete information, *Journal of Finance* 42, 483-510.

Shumway, 1997, The delisting bias in CRSP data, *Journal of Finance* 52, 327-340.

Wong, Peter, 2011, Earnings shocks and the idiosyncratic volatility discount in the cross-section of expected returns, Working paper, Ohio State University.

Table 1: Sample descriptive statistics

Sample statistics from 1963-2009 are reported. Panel A shows the distribution of firm characteristics and Panel B shows the average correlations. The sample consists of all CRSP common stocks with share prices greater than $1 at the end of the previous month. N is the total number of firm-month observations. *IVOL* is the stdev of residuals from a regression of daily stock returns in month *t*–1 on the Fama-French factors. *Beta* is the regression coefficient of the past three years of monthly returns on market returns. *Size* and *B/M* are measured and aligned as in Fama and French (2006), and *Momentum* is the buy-and-hold month *t*–12 to *t*–2 return. *Skew* is the month *t*–1 skewness of raw daily returns. *Coskew_CY* is the coskewness measure in Chabi-Yo and Yang (2009). *Coskew_HS* and *Idioskew_HS* are the coskewness and idiosyncratic skewness measures in Harvey and Siddique (2000). *Maxret* is the maximum daily return in month *t*–1. *E(Idioskew)* is the expected idiosyncratic skewness measure in Boyer et al. (2010). *RTP* is the retail trading proportion computed from TaQ. *Lagret* is the month *t*–1 return. *Amihud* is the illiquidity measure in Amihud (2002). *Zeroret* is the fraction of trading days in month *t*–1 with a zero return. *Invprc* is the reciprocal of the average price in month *t*–1. *ShortInt* is the short interest in month *t*–1. *InstOwn* is the most recent 13F reported institutional ownership. *Nanalyst* is the number of analysts covering the stock. *Delay* is the price delay measure in Hou and Moskowitz (2005). *Dispersion* is the dispersion in analysts' FY1 forecasts. *Leverage* is the Compustat long-term debt over total assets. *NegRatingChg* is the negative S&P rating change as in Avramov et al. (2012). *Egarch* is the EGARCH estimate of expected idiosyncratic volatility in Fu (2009). *R&D* is the Compustat R&D expense scaled by total assets. *AvgVarβ* is a stock's exposure to the average variance component of the market variance as in Chen and Petkova (2012). *SUE* and *NextSUE* are the standardized unexpected earnings from the previous quarter and the following quarter, respectively.

Panel A: Distribution of firm characteristics

Variable	Mean	Std Dev	N	1st Pctl	10th Pctl	25th Pctl	50th Pctl	75th Pctl	90th Pctl	99th Pctl
Return	0.011	0.159	2035201	-0.367	-0.145	-0.063	0.000	0.072	0.168	0.511
IVOL	0.027	0.022	2036817	0.003	0.009	0.013	0.021	0.034	0.052	0.108
Size ($m)	1287.3	8706.6	2036844	1.9	8.5	24.3	93.2	434.8	1702.4	21506.0
B/M	0.887	2.795	2036844	-0.214	0.178	0.363	0.671	1.142	1.807	4.587
Momentum	0.164	0.711	2036844	-0.753	-0.417	-0.183	0.064	0.346	0.747	2.587
Beta	1.135	0.937	2009258	-0.813	0.187	0.577	1.040	1.570	2.197	4.079
Lottery preference variables										
Skew	0.264	1.047	2026600	-2.871	-0.802	-0.251	0.221	0.763	1.434	3.344
Coskew_CY	0.006	0.953	2036827	-0.450	-0.064	-0.017	0.001	0.022	0.078	0.527
Coskew_HS	-7.342	777.0	2036827	-630.7	-145.2	-50.8	-2.8	38.9	126.2	595.6
Idioskew_HS	0.238	0.989	2026600	-2.640	-0.791	-0.267	0.201	0.723	1.365	3.102
E(Idioskew)	0.921	0.590	861130	-0.205	0.231	0.515	0.874	1.276	1.679	2.476
Maxret	0.072	0.078	2036827	0.000	0.020	0.031	0.053	0.088	0.143	0.345
RTP	0.160	0.208	528457	0.000	0.008	0.022	0.069	0.215	0.465	1.000
Market friction variables										
Lagret	0.016	0.170	2036747	-0.347	-0.141	-0.062	0.000	0.074	0.173	0.550
Amihud	6.560	74.066	1885419	0.000	0.002	0.016	0.184	1.694	9.501	108.89
Invprc	0.145	0.181	2036844	0.012	0.024	0.038	0.072	0.170	0.372	0.873
Zeroret	0.219	0.215	2036844	0.000	0.000	0.048	0.158	0.318	0.500	0.947
ShortInt	1.479	3.353	587840	0.000	0.003	0.030	0.204	1.292	4.246	16.620
InstOwn	0.354	0.281	1528983	0.001	0.028	0.105	0.295	0.562	0.779	1.000
Nanalyst	7.361	7.184	879515	1.000	1.000	2.000	5.000	10.000	18.000	32.000
Delay	0.085	0.140	1526870	-0.001	0.002	0.018	0.056	0.124	0.216	0.385
Other variables										
Dispersion	0.195	1.236	731414	0.000	0.009	0.019	0.042	0.109	0.300	2.556
Leverage	0.175	0.185	2032080	0.000	0.000	0.018	0.130	0.277	0.417	0.733
NegRatingChg	0.012	0.260	282152	0.000	0.000	0.000	0.000	0.000	0.000	1.000
Egarch	0.121	0.084	1895485	0.029	0.051	0.069	0.099	0.146	0.211	0.452
R&D	0.072	0.156	914645	0.000	0.000	0.006	0.029	0.085	0.174	0.620
AvgVarβ	0.293	7.771	1658203	-20.739	-6.324	-2.413	0.068	2.692	7.116	23.837
SUE	0.170	17.704	1617093	-7.719	-1.624	-0.471	0.191	1.071	2.383	6.443
NextSUE	0.162	17.804	1598064	-7.783	-1.631	-0.476	0.187	1.064	2.373	6.424

Table 1 (Cont'd)

Panel B: Average correlations between firm characteristics

Variable	Ret	IVOL	Beta	Size	BM	Mom	Skew	CoskwC	CoskwH	Iskw	Eiskw	Maxret	RTP	Lagret	Amihud	Invprc	Zeroret	Short	IO	Anlyst	Delay	Disp	Lev	NegRtg	Egarch	R&D	AVβs	SUE
IVOL	-0.033	1																										
Beta	-0.013	0.185	1																									
Size	-0.001	-0.138	-0.044	1																								
B/M	0.019	0.006	-0.093	-0.054	1																							
Momentum	0.032	-0.085	-0.017	0.004	0.013	1																						
Skew	-0.010	0.193	0.036	-0.022	0.010	-0.019	1																					
Coskew_CY	-0.007	0.057	0.017	-0.004	-0.001	-0.012	0.101	1																				
Coskew_HS	0.002	-0.042	-0.009	0.012	0.004	-0.017	0.046	0.131	1																			
Idioskew_HS	-0.008	0.175	0.034	-0.023	0.009	-0.013	0.870	0.022	-0.115	1																		
E(Idioskew)	-0.030	0.459	0.020	-0.171	0.086	-0.178	0.073	0.006	-0.019	0.064	1																	
Maxret	-0.036	0.883	0.157	-0.101	0.002	-0.065	0.474	0.110	-0.012	0.418	0.369	1																
RTP	-0.019	0.504	-0.079	-0.142	0.118	-0.186	0.049	0.003	-0.015	0.039	0.591	0.388	1															
Lagret	-0.039	0.190	-0.008	-0.003	0.017	0.001	0.349	0.082	-0.016	0.331	0.007	0.387	-0.045	1														
Amihud	0.003	0.313	-0.026	-0.051	0.105	-0.093	0.008	0.003	-0.004	0.007	0.214	0.225	0.369	-0.013	1													
Invprc	-0.005	0.543	0.072	-0.139	0.084	-0.209	0.051	0.020	-0.013	0.044	0.589	0.416	0.753	0.006	0.434	1												
Zeroret	0.007	0.014	-0.148	-0.157	0.137	-0.136	-0.006	-0.007	0.002	-0.005	0.399	0.007	0.426	-0.049	0.218	0.432	1											
ShortInt	-0.023	-0.014	0.202	0.046	-0.127	0.037	0.003	0.004	0.001	0.001	-0.180	0.001	-0.185	-0.023	-0.085	-0.159	-0.238	1										
InstOwn	0.010	-0.287	0.081	0.171	-0.064	0.032	-0.044	-0.003	0.019	-0.040	-0.510	-0.217	-0.524	-0.013	-0.154	-0.416	-0.390	0.290	1									
Nanalyst	0.008	-0.287	-0.018	0.496	-0.081	0.017	-0.041	-0.003	0.024	-0.037	-0.375	-0.214	-0.375	0.004	-0.143	-0.341	-0.349	0.220	0.471	1								
Delay	0.005	0.368	-0.086	-0.185	0.190	-0.005	0.039	0.013	-0.010	0.036	0.642	0.292	0.697	0.033	0.388	0.612	0.505	-0.251	-0.539	-0.547	1							
Dispersion	-0.011	0.128	0.063	-0.036	0.039	-0.086	0.014	0.011	-0.002	0.012	0.104	0.101	0.149	-0.020	0.055	0.174	0.068	0.021	-0.056	-0.056	0.096	1						
Leverage	-0.003	-0.007	-0.013	-0.010	-0.012	-0.006	-0.004	0.001	0.000	-0.003	-0.043	-0.008	-0.033	-0.002	-0.023	0.009	-0.015	0.029	0.044	0.035	-0.056	0.036	1					
NegRating	-0.014	0.111	0.003	-0.008	0.020	-0.084	-0.010	-0.004	0.002	-0.011	0.031	0.075	0.065	-0.057	0.023	0.083	0.025	0.015	-0.024	-0.009	0.024	0.023	0.003	1				
Egarch	0.046	0.451	0.219	-0.142	-0.036	0.014	0.046	0.021	-0.020	0.041	0.368	0.362	0.313	0.006	0.161	0.420	0.119	0.079	-0.228	-0.247	0.332	0.108	-0.026	0.046	1			
R&D	-0.003	0.133	0.144	-0.036	-0.189	0.019	0.024	0.009	-0.010	0.021	0.118	0.117	0.069	0.008	-0.032	0.080	-0.031	0.060	-0.119	-0.073	0.017	0.038	-0.166	-0.005	0.165	1		
AvgVarβ	-0.005	0.040	0.099	-0.003	-0.003	0.004	0.007	0.007	0.002	0.004	-0.016	0.034	0.006	0.000	0.011	0.036	-0.008	0.025	-0.003	0.004	0.015	0.008	-0.005	0.000	0.042	0.036	1	
SUE	0.029	-0.102	-0.017	0.047	-0.059	0.206	0.007	-0.004	-0.004	0.009	-0.083	-0.065	-0.075	0.058	-0.047	-0.119	-0.085	-0.008	0.053	0.050	-0.081	-0.072	-0.019	-0.048	-0.082	0.012	-0.001	1
NextSUE	0.080	-0.094	-0.019	0.043	-0.050	0.165	0.021	0.001	-0.006	0.023	-0.071	-0.051	-0.060	0.080	-0.039	-0.100	-0.070	-0.017	0.044	0.049	-0.070	-0.053	-0.018	-0.044	-0.080	0.011	-0.001	0.310

Table 2: The negative relation between idiosyncratic volatility and returns

Fama-Macbeth cross-sectional regressions are estimated each month from August 1963 to December 2009. Stocks with prices less than $1 at the end of the previous month are excluded. For each month *t*, returns are regressed on independent variables and the time-series averages of the coefficients (×100) are reported. Idiosyncratic volatility (*IVOL*) is the standard deviation of residuals from a regression of daily stock returns in month *t*–1 on the Fama-French factors. For the lottery preference-based candidate variables in Panel A, *Skew* is the month *t*–1 skewness of raw daily returns. *Coskew_CY* is the coskewness measure in Chabi-Yo and Yang (2009). *Coskew_HS* and *Idioskew_HS* are the coskewness and idiosyncratic skewness measures in Harvey and Siddique (2000). *Maxret* is the maximum daily return in month *t*–1. *E(Idioskew)* is the expected idiosyncratic skewness measure in Boyer et al. (2010). *RTP* is the retail trading proportion computed from TaQ. For the market friction-based candidate variables in Panel B, *Lagret* is the month *t*–1 return. *Amihud* is the illiquidity measure in Amihud (2002). *Zeroret* is the fraction of trading days in month *t*–1 with a zero return. *Invprc* is the reciprocal of the average price in month *t*–1. *ShortInt* is the short interest in month *t*–1. *InstOwn* is the most recent 13F reported institutional ownership. *Nanalyst* is the number of analysts covering the stock. *Delay* is the price delay measure in Hou and Moskowitz (2005). For the other candidate variables in Panel C, *Dispersion* is the dispersion in analysts' FY1 forecasts. *Leverage* is the Compustat long-term debt over total assets. *NegRatingChg* is the negative S&P rating change as in Avramov et al. (2012). *Egarch* is the EGARCH estimate of expected idiosyncratic volatility in Fu (2009). *R&D* is the Compustat R&D expense scaled by total assets. *AvgVarβ* is a stock's exposure to the average variance component of the market variance as in Chen and Petkova (2012). *SUE* and *NextSUE* are the standardized unexpected earnings from the previous quarter and the following quarter, respectively. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels respectively with *t*-statistics in parentheses.

Panel A: IVOL and lottery preference variables

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12	Model 13	Model 14	Model 15	Model 16
	Raw return as dependent variable								DGTW-adjusted return as dependent variable							
Intercept	1.558***	1.575***	1.543***	1.554***	1.572***	2.085***	1.478***	1.705***	0.334***	0.352***	0.328***	0.331***	0.350***	0.710***	0.260***	0.290
	(8.36)	(8.39)	(8.40)	(8.35)	(8.37)	(6.89)	(8.04)	(4.28)	(5.73)	(6.06)	(5.55)	(5.68)	(6.05)	(6.04)	(4.30)	(1.25)
IVOL	-17.128***	-16.357***	-16.906***	-17.030***	-16.669***	-19.733***	11.971*	-17.544*	-16.300***	-15.364***	-16.326***	-16.207***	-15.589***	-16.488***	13.243***	-12.722**
	(-3.69)	(-3.45)	(-3.57)	(-3.69)	(-3.53)	(-3.91)	(1.68)	(-1.86)	(-7.80)	(-7.24)	(-7.71)	(-7.82)	(-7.36)	(-5.71)	(3.39)	(-2.33)
Skew		-0.085***								-0.108***						
		(-4.18)								(-5.78)						
Coskew_CY			-0.515***								-0.396***					
			(-2.95)								(-2.59)					
Coskew_HS				-0.000								-0.000				
				(-0.19)								(-0.44)				
Idioskew_HS					-0.066***								-0.099***			
					(-3.34)								(-5.43)			
E(Idioskew)						-0.351**								-0.409***		
						(-2.06)								(-5.27)		
Maxret							-9.579***								-9.941***	
							(-8.02)								(-10.33)	
RTP								0.239								0.057
								(0.27)								(0.13)
Avg Adj R^2	0.017	0.018	0.019	0.018	0.018	0.019	0.020	0.024	0.004	0.005	0.006	0.005	0.005	0.005	0.006	0.006
#firms/mth	3653.8	3635.6	3653.8	3653.8	3635.6	4005.3	3653.8	5503.1	3653.8	3635.6	3653.8	3653.8	3635.6	4005.3	3653.8	5503.1
Startdate	196308	196308	196308	196308	196308	198801	196308	199302	196308	196308	196308	196308	196308	198801	196308	199302
Enddate	200912	200912	200912	200912	200912	200512	200912	200101	200912	200912	200912	200912	200912	200512	200912	200101

Table 2 (Cont'd)

Panel B: IVOL and market friction variables

Variable	Model 17	Model 18	Model 19	Model 20	Model 21	Model 22	Model 23	Model 24	Model 25	Model 26	Model 27	Model 28	Model 29	Model 30	Model 31	Model 32
			Raw return as dependent variable								DGTW-adjusted return as dependent variable					
Intercept	1.476***	1.578***	1.553***	1.508***	2.057***	1.777***	1.655***	1.553***	0.243***	0.397***	0.327***	0.335***	0.479***	0.421***	0.290**	0.309***
	(8.14)	(8.59)	(8.38)	(6.92)	(7.74)	(8.85)	(6.56)	(7.00)	(4.08)	(6.82)	(5.51)	(6.15)	(3.11)	(3.84)	(2.45)	(4.69)
IVOL	-11.255**	-20.439***	-27.091***	-17.552***	-17.737***	-24.173***	-23.879***	-25.329***	-10.139***	-20.702***	-26.495***	-16.391***	-14.097***	-19.426***	-14.776***	-20.212***
	(-2.33)	(-4.32)	(-6.81)	(-3.83)	(-2.96)	(-5.08)	(-3.61)	(-5.03)	(-4.52)	(-9.03)	(-11.12)	(-7.93)	(-4.21)	(-7.22)	(-3.94)	(-7.48)
Lagret	-4.333***								-4.571***							
	(-10.35)								(-13.67)							
Amihud		0.042***								0.037***						
		(3.41)								(4.52)						
Invprc			2.680***								2.781***					
			(3.56)								(5.53)					
Zeroret				0.407								0.089				
				(1.02)								(0.34)				
ShortInt					-0.098***								-0.055***			
					(-2.98)								(-2.72)			
InstOwn						0.073								-0.083		
						(0.23)								(-0.47)		
Nanalyst							-0.009								-0.001	
							(-1.09)								(-0.17)	
Delay								3.492***								1.523***
								(3.96)								(3.62)
Avg Adj R^2	0.026	0.023	0.027	0.024	0.019	0.019	0.024	0.028	0.011	0.007	0.008	0.007	0.006	0.005	0.007	0.007
#firms/mth	3653.6	3382.6	3653.8	3653.8	2983.2	4108.7	2640.7	3435.6	3653.6	3382.6	3653.8	3653.8	2983.2	4108.7	2640.7	3435.6
Startdate	196308	196308	196308	196308	198807	197901	198204	196507	196308	196308	196308	196308	198807	197901	198204	196507
Enddate	200912	200912	200912	200912	200501	200912	200912	200206	200912	200912	200912	200912	200501	200912	200912	200206

Table 2 (Cont'd)

Panel C: IVOL and other variables

Variable	Model 33	Model 34	Model 35	Model 36	Model 37	Model 38	Model 39	Model 40	Model 41	Model 42	Model 43	Model 44	Model 45	Model 46	Model 47	Model 48
			Raw return as dependent variable								DGTW-adjusted return as dependent variable					
Intercept	1.497***	1.621***	1.416***	0.645***	1.596***	1.563***	1.492***	1.234***	0.194**	0.390***	0.390***	-0.658***	0.407***	0.351***	0.283***	0.047
	(6.29)	(8.76)	(5.47)	(3.92)	(7.67)	(7.81)	(6.97)	(5.75)	(2.00)	(6.76)	(4.00)	(-6.74)	(6.48)	(5.72)	(4.39)	(0.70)
IVOL	-21.258***	-17.332***	-30.374***	-51.235***	-19.365***	-21.215***	-16.992***	-10.199**	-10.622**	-16.374***	-26.684***	-49.692***	-18.359***	-16.980***	-14.789***	-8.211***
	(-2.79)	(-3.76)	(-3.64)	(-16.71)	(-3.90)	(-4.61)	(-3.76)	(-2.17)	(-2.54)	(-7.88)	(-5.68)	(-30.85)	(-6.63)	(-8.05)	(-6.60)	(-3.44)
Dispersion	-0.112								-0.122*							
	(-1.48)								(-1.79)							
Leverage		-0.339**								-0.308***						
		(-2.43)								(-2.72)						
NegRatingChg			-0.393**								-0.271*					
			(-2.21)								(-1.69)					
Egarch				15.883***								15.494***				
				(14.20)								(21.49)				
R&D					1.812**								2.210***			
					(2.11)								(3.50)			
AvgVarβ						-0.007								-0.004		
						(-1.16)								(-0.81)		
SUE							0.131***								0.111***	
							(11.91)								(16.54)	
NextSUE								0.373***								0.340***
								(27.22)								(33.19)
Avg Adj R^2	0.023	0.018	0.027	0.034	0.022	0.018	0.018	0.025	0.008	0.005	0.012	0.015	0.008	0.005	0.005	0.011
#firms/mth	2196.1	3645.3	1000.2	3561.1	1640.8	3266.1	3520.9	3479.7	2196.1	3645.3	1000.2	3561.1	1640.8	3266.1	3520.9	3479.7
Startdate	198204	196308	198602	196308	196308	196808	197110	197107	198204	196308	198602	196308	196308	196808	197110	197107
Enddate	200912	200912	200907	200712	200912	200912	200912	200909	200912	200912	200907	200712	200912	200912	200912	200909

Table 3: Decomposing the idiosyncratic volatility puzzle: Lottery preference explanations

Using Fama-Macbeth cross-sectional regressions, the negative relation between idiosyncratic volatility and returns is decomposed into a component that is related to a lottery preference-based candidate variable and a residual component. Stage 1 regresses month t returns on month t–1 $IVOL$ ($R_{it} = \alpha_t + \gamma_t IVOL_{it-1} + \varepsilon_{it}$). Stage 2 adds a candidate variable ($Candidate_{it-1}$) to the regression. Stage 3 regresses $IVOL$ on the candidate variable ($IVOL_{it-1} = a_{t-1} + \delta_{t-1} Candidate_{it-1} + \mu_{it-1}$) to decompose $IVOL_{it-1}$ into two orthogonal components: $\delta_{t-1} Candidate_{it-1}$ and $(a_{t-1} + \mu_{it-1})$. In Stage 4, the γ_t coefficient from Stage 1 is decomposed as follows:

$$\gamma_t = \frac{\mathrm{Cov}[R_{it}, IVOL_{it-1}]}{\mathrm{Var}[IVOL_{it-1}]} = \frac{\mathrm{Cov}[R_{it}, \delta_{t-1} Candidate_{it-1}]}{\mathrm{Var}[IVOL_{it-1}]} + \frac{\mathrm{Cov}[R_{it}, (a_{t-1} + \mu_{it-1})]}{\mathrm{Var}[IVOL_{it-1}]} = \gamma_t^C + \gamma_t^R .$$

The time-series average of γ_t^C divided by the time-series average of γ_t then measures the fraction of the negative idiosyncratic volatility-return relation explained by the candidate variable. The average γ_t^R divided by the average γ_t measures the fraction of the relation left unexplained by the candidate variable. Stocks with prices less than $1 at the end of the previous month are excluded from the analysis. *IVOL* is the standard deviation of residuals from a regression of daily stock returns in month t–1 on the Fama-French factors. *Skew* is the month t–1 skewness of raw daily returns. *Coskew_CY* is the coskewness measure in Chabi-Yo and Yang (2009). *Coskew_HS* and *Idioskew_HS* are the coskewness and idiosyncratic skewness measures in Harvey and Siddique (2000). *E(Idioskew)* is the expected idiosyncratic skewness measure in Boyer et al. (2010). *RTP* is the retail trading proportion computed from TaQ. *Maxret* is the maximum daily return in month t–1. *Maxret(3mth)* (*Maxret(6mth)*) is the maximum daily return for the three-month (six-month) period ending in month t–1. *LagMaxret(3mth)* is the maximum daily return for the three-month period ending in month t–2. Time-series averages of estimated coefficients (×100) are reported and *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels respectively with t-statistics in parentheses.

Panel A: Various skewness variables

Stage	Description	Variable	Skew		Coskew_CY		Coskew_HS		Idioskew_HS		E(Idioskew)	
1	Regress returns on IVOL	Intercept	1.574***	(8.37)	1.558***	(8.36)	1.558***	(8.36)	1.574***	(8.37)	1.831***	(7.33)
		IVOL	-17.540***	(-3.78)	-17.128***	(-3.69)	-17.128***	(-3.69)	-17.540***	(-3.78)	-23.618***	(-4.18)
2	Add candidate variable	Intercept	1.575***	(8.39)	1.543***	(8.40)	1.554***	(8.35)	1.572***	(8.37)	2.085***	(6.89)
		IVOL	-16.357***	(-3.45)	-16.906***	(-3.57)	-17.030***	(-3.69)	-16.669***	(-3.53)	-19.733***	(-3.91)
		Candidate	-0.085***	(-4.18)	-0.515***	(-2.95)	-0.000	(-0.19)	-0.066***	(-3.34)	-0.351**	(-2.06)
3	IVOL on candidate variable	Intercept	2.439***	(86.71)	2.518***	(83.75)	2.517***	(84.25)	2.452***	(86.02)	1.072***	(33.13)
		Candidate	0.369***	(32.17)	0.682***	(10.50)	-0.001**	(-2.26)	0.347***	(35.17)	1.992***	(42.81)
		Adj R^2	4.3%		3.8%		1.2%		3.5%		21.4%	
4	Decompose stage 1 IVOL coefficient	Candidate	-1.211***	(-3.28)	-0.728**	(-2.25)	-0.126	(-0.55)	-0.862***	(-2.59)	-8.259***	(-3.86)
			6.9%		4.2%		0.7%		4.9%		35.0%	
		Residual	-16.328***	(-3.67)	-16.400***	(-3.60)	-17.002***	(-3.73)	-16.678***	(-3.73)	-15.358***	(-3.81)
			93.1%		95.8%		99.3%		95.1%		65.0%	
		Total	-17.540***	(-3.78)	-17.128***	(-3.69)	-17.128***	(-3.69)	-17.540***	(-3.78)	-23.618***	(-4.18)
			100.0%		100.0%		100.0%		100.0%		100.0%	
Sample period			1963 to 2009		1963 to 2009		1963 to 2009		1963 to 2009		1988 to 2005	
Avg # firms/mth			3635.6		3653.8		3653.8		3635.6		4005.3	

Table 3 (Cont'd)

Panel B: Other lottery preference variables

Stage	Description	Variable	Maxret		Maxret(3mth)		Maxret(6mth)		LagMaxret(3mth)		RTP	
1	Regress returns on IVOL	Intercept	1.558***	(8.36)	1.491***	(8.04)	1.530***	(8.18)	1.516***	(8.09)	1.746***	(4.52)
		IVOL	-17.128***	(-3.69)	-17.419***	(-3.70)	-17.113***	(-3.65)	-17.098***	(-3.63)	-18.226*	(-1.77)
2	Add candidate variable	Intercept	1.478***	(8.04)	1.529***	(8.28)	1.547***	(8.41)	1.534***	(8.48)	1.705***	(4.28)
		IVOL	11.971*	(1.68)	-13.768***	(-3.58)	-17.127***	(-4.81)	-17.009***	(-4.72)	-17.544*	(-1.86)
		Candidate	-9.579***	(-8.02)	-1.397**	(-2.45)	-0.128	(-0.24)	-0.105	(-0.16)	0.239	(0.27)
3	IVOL on candidate variable	Intercept	0.784***	(78.04)	1.016***	(60.17)	1.126***	(57.52)	1.467***	(67.62)	2.382***	(68.19)
		Candidate	25.971***	(236.89)	15.603***	(108.43)	11.916***	(86.48)	11.071***	(79.59)	6.470***	(76.35)
		Adj R^2	78.0%		49.5%		40.4%		24.3%		25.6%	
4	Decompose stage 1 IVOL coefficient	Candidate	-18.116***	(-5.37)	-9.372***	(-3.04)	-6.115**	(-2.12)	-3.961*	(-1.75)	-4.932	(-1.23)
			105.8%		53.8%		35.7%		23.2%		27.1%	
		Residual	0.988	(0.64)	-8.047***	(-4.23)	-10.998***	(-5.30)	-13.137***	(-4.87)	-13.294*	(-1.79)
			-5.8%		46.2%		64.3%		76.8%		72.9%	
		Total	-17.128***	(-3.69)	-17.419***	(-3.70)	-17.113***	(-3.65)	-17.098***	(-3.63)	-18.226*	(-1.77)
			100.0%		100.0%		100.0%		100.0%		100.0%	
Sample period			1963 to 2009		1963 to 2009		1963 to 2009		1963 to 2009		1993 to 2001	
Avg # firms/mth			3653.8		3653.0		3652.8		3626.3		5503.1	

Table 4: Decomposing the idiosyncratic volatility puzzle: Market friction explanations

Using Fama-Macbeth cross-sectional regressions, the negative relation between idiosyncratic volatility and returns is decomposed into a component that is related to a market friction-based candidate variable and a residual component. Stage 1 regresses month *t* returns on month *t*–1 *IVOL*. Stage 2 adds a candidate variable to the regression. Stage 3 regresses *IVOL* on the candidate variable to decompose *IVOL* into a component that is related to the candidate variable and a residual component. In Stage 4, the Stage 1 coefficient of regressing returns on *IVOL* is decomposed into a component that is related to the candidate variable and a residual component. Stocks with prices less than $1 at the end of the previous month are excluded from the analysis. *IVOL* is the standard deviation of residuals from a regression of daily stock returns in month *t*–1 on the Fama-French factors. *Lagret* is the month *t*–1 return. *Amihud* is the illiquidity measure in Amihud (2002). *Invprc* is the reciprocal of the average price in month *t*–1. *Zeroret* is the fraction of trading days in month *t*–1 with a zero return. *ShortInt* is the short interest in month *t*–1. *InstOwn* is the most recent 13F reported institutional ownership. *Nanalyst* is the number of analysts covering the stock. *Delay* is the price delay measure in Hou and Moskowitz (2005). Time-series averages of estimated coefficients (×100) are reported and *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels respectively with *t*-statistics in parentheses.

Panel A

Stage	Description	Variable	Lagret		Amihud		Invprc		Zeroret	
1	Regress returns on IVOL	Intercept	1.558***	(8.36)	1.550***	(8.26)	1.558***	(8.36)	1.558***	(8.36)
		IVOL	-17.125***	(-3.69)	-17.330***	(-3.63)	-17.128***	(-3.69)	-17.128***	(-3.69)
2	Add candidate variable	Intercept	1.476***	(8.14)	1.578***	(8.59)	1.553***	(8.38)	1.508***	(6.92)
		IVOL	-11.255**	(-2.33)	-20.439***	(-4.32)	-27.091***	(-6.81)	-17.552***	(-3.83)
		Candidate	-4.333***	(-10.35)	0.042***	(3.41)	2.680***	(3.56)	0.407	(1.02)
3	IVOL on candidate variable	Intercept	2.392***	(81.42)	2.416***	(82.04)	1.681***	(89.51)	2.485***	(86.24)
		Candidate	2.295***	(20.72)	0.041***	(18.99)	6.736***	(62.66)	0.561***	(8.32)
		Adj R^2	8.2%		11.6%		30.3%		0.9%	
4	Decompose stage 1 IVOL coefficient	Candidate	-4.197***	(-4.00)	1.528	(1.28)	2.129	(0.76)	0.236	(0.62)
			24.5%		-8.8%		-12.4%		-1.4%	
		Residual	-12.929***	(-2.96)	-18.858***	(-4.54)	-19.256***	(-6.92)	-17.364***	(-3.83)
			75.5%		108.8%		112.4%		101.4%	
		Total	-17.125***	(-3.69)	-17.330***	(-3.63)	-17.128***	(-3.69)	-17.128***	(-3.69)
			100.0%		100.0%		100.0%		100.0%	
Sample period			1963 to 2009		1963 to 2009		1963 to 2009		1963 to 2009	
Avg # firms/mth			3653.6		3382.6		3653.8		3653.8	

Panel B

Stage	Description	Variable	ShortInt		InstOwn		Nanalyst		Delay	
1	Regress returns on IVOL	Intercept	1.911***	(6.65)	1.722***	(7.53)	1.568***	(6.41)	1.656***	(7.89)
		IVOL	-17.605***	(-2.89)	-23.599***	(-4.88)	-22.964***	(-3.38)	-17.878***	(-3.24)
2	Add candidate variable	Intercept	2.057***	(7.74)	1.777***	(8.85)	1.655***	(6.56)	1.553***	(7.00)
		IVOL	-17.737***	(-2.96)	-24.173***	(-5.08)	-23.879***	(-3.61)	-25.329***	(-5.03)
		Candidate	-0.098***	(-2.98)	0.073	(0.23)	-0.009	(-1.09)	3.492***	(3.96)
3	IVOL on candidate variable	Intercept	3.744***	(72.64)	3.544***	(63.04)	2.902***	(75.39)	1.897***	(92.49)
		Candidate	-0.017***	(-4.74)	-2.451***	(-33.38)	-0.063***	(-51.75)	8.619***	(43.59)
		Adj R^2	0.3%		9.2%		8.8%		15.0%	
4	Decompose stage 1 IVOL coefficient	Candidate	0.068	(0.30)	-1.680*	(-1.86)	-1.279	(-1.11)	3.370*	(1.77)
			-0.4%		7.1%		5.6%		-18.9%	
		Residual	-17.673***	(-2.96)	-21.919***	(-4.91)	-21.685***	(-3.48)	-21.248***	(-4.85)
			100.4%		92.9%		94.4%		118.9%	
		Total	-17.605***	(-2.89)	-23.599***	(-4.88)	-22.964***	(-3.38)	-17.878***	(-3.24)
			100.0%		100.0%		100.0%		100.0%	
Sample period			1988 to 2005		1979 to 2009		1982 to 2009		1965 to 2002	
Avg # firms/mth			2983.2		4108.7		2640.7		3435.6	

Table 5: Decomposing the idiosyncratic volatility puzzle: All other explanations

Using Fama-Macbeth cross-sectional regressions, the negative relation between idiosyncratic volatility and returns is decomposed into a component that is related to a candidate variable not belonging to the lottery preference or market friction category, and a residual component. Stage 1 regresses month *t* returns on month *t*–1 *IVOL*. Stage 2 adds a candidate variable. Stage 3 regresses *IVOL* on the candidate variable to decompose *IVOL* into a component that is related to the candidate variable and a residual component. In Stage 4, the Stage 1 *IVOL* coefficient is decomposed into a component that is related to the candidate variable and a residual component. Stocks with prices less than $1 at the end of the previous month are excluded. *IVOL* is the standard deviation of residuals from a regression of daily stock returns in month *t*–1 on the Fama-French factors. *Dispersion* is the dispersion in analysts' FY1 forecasts. *Leverage* is the Compustat long-term debt over total assets. *NegRatingChg* is the negative S&P rating change as in Avramov et al. (2012). *Egarch* is the EGARCH estimate of expected idiosyncratic volatility in Fu (2009). *R&D* is the Compustat R&D expense scaled by total assets. *AvgVarβ* is a stock's exposure to the average variance component of the market variance as in Chen and Petkova (2012). *SUE* and *NextSUE* are the standardized unexpected earnings from the previous quarter and the following quarter, respectively. Time-series averages of estimated coefficients (×100) are reported and *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels respectively with *t*-statistics in parentheses.

Panel A

Stage	Description	Variable	Dispersion		Leverage		NegRatingChg		Egarch	
1	Regress returns on IVOL	Intercept	1.498***	(6.27)	1.559***	(8.35)	1.436***	(5.53)	1.695***	(9.38)
		IVOL	-22.106***	(-2.83)	-17.150***	(-3.70)	-31.535***	(-3.77)	-21.328***	(-4.51)
2	Add candidate variable	Intercept	1.497***	(6.29)	1.621***	(8.76)	1.416***	(5.47)	0.645***	(3.92)
		IVOL	-21.258***	(-2.79)	-17.332***	(-3.76)	-30.374***	(-3.64)	-51.235***	(-16.71)
		Candidate	-0.112	(-1.48)	-0.339**	(-2.43)	-0.393**	(-2.21)	15.883***	(14.20)
3	IVOL on candidate variable	Intercept	2.232***	(67.71)	2.543***	(79.20)	1.966***	(61.09)	1.248***	(57.40)
		Candidate	0.230***	(16.03)	-0.087***	(-3.27)	0.681***	(18.06)	10.992***	(100.11)
		Adj R^2	2.0%		0.4%		2.1%		20.7%	
4	Decompose stage 1 IVOL coefficient	Candidate	-1.423	(-1.58)	0.138	(1.46)	-1.704***	(-2.61)	19.510***	(7.31)
			6.4%		-0.8%		5.4%		-91.5%	
		Residual	-20.683***	(-2.76)	-17.288***	(-3.77)	-29.831***	(-3.68)	-40.839***	(-17.28)
			93.6%		100.8%		94.6%		191.5%	
		Total	-22.106***	(-2.83)	-17.150***	(-3.70)	-31.535***	(-3.77)	-21.328***	(-4.51)
			100.0%		100.0%		100.0%		100.0%	
Sample period			1982 to 2009		1963 to 2009		1986 to 2009		1963 to 2007	
Avg # firms/mth			2196.1		3645.3		1000.2		3561.1	

Panel B

Stage	Description	Variable	R&D		AvgVarβ		SUE		NextSUE	
1	Regress returns on IVOL	Intercept	1.654***	(7.74)	1.569***	(7.84)	1.596***	(7.42)	1.499***	(6.97)
		IVOL	-19.263***	(-3.79)	-21.313***	(-4.61)	-19.323***	(-4.22)	-16.388***	(-3.44)
2	Add candidate variable	Intercept	1.596***	(7.67)	1.563***	(7.81)	1.492***	(6.97)	1.234***	(5.75)
		IVOL	-19.365***	(-3.90)	-21.215***	(-4.61)	-16.992***	(-3.76)	-10.199**	(-2.17)
		Candidate	1.812**	(2.11)	-0.007	(-1.16)	0.131***	(11.91)	0.373***	(27.22)
3	IVOL on candidate variable	Intercept	2.545***	(72.38)	2.531***	(87.70)	2.659***	(85.12)	2.648***	(85.14)
		Candidate	2.518***	(33.45)	0.010***	(6.26)	-0.067***	(-33.46)	-0.059***	(-35.23)
		Adj R^2	2.2%		0.6%		1.2%		1.0%	
4	Decompose stage 1 IVOL coefficient	Candidate	-0.179	(-0.32)	-0.202	(-1.48)	-2.487***	(-3.79)	-6.302***	(-18.56)
			0.9%		0.9%		12.9%		38.5%	
		Residual	-19.084***	(-3.92)	-21.111***	(-4.62)	-16.837***	(-8.37)	-10.086**	(-2.17)
			99.1%		99.1%		87.1%		61.5%	
		Total	-19.263***	(-3.79)	-21.313***	(-4.61)	-19.323***	(-4.22)	-16.388***	(-3.44)
			100.0%		100.0%		100.0%		100.0%	
Sample period			1963 to 2009		1968 to 2009		1971 to 2009		1971 to 2009	
Avg # firms/mth			1640.8		3266.1		3520.9		3479.7	

Table 6: Decomposing the idiosyncratic volatility puzzle: Multivariate analysis

Using Fama-Macbeth cross-sectional regressions, the negative relation between idiosyncratic volatility and returns is decomposed into a number of components each related to a candidate variable and a residual component. The candidate variables considered are those that explain more than 10% of the puzzle in the univariate analysis and are statistically significant. Stocks with prices less than $1 at the end of the previous month are excluded from the analysis. *IVOL* is the standard deviation of residuals from a regression of daily stock returns in month *t*–1 on the Fama-French factors. *Maxret(3mth)* (*LagMaxret(3mth)*) is the maximum daily return for the three-month period ending in month *t*–1 (*t*–2). *E(Idioskew)* is the expected idiosyncratic skewness measure in Boyer et al. (2010). *Lagret* is the month *t*–1 return. *SUE* and *NextSUE* are the standardized unexpected earnings from the previous quarter and the following quarter, respectively. Time-series averages of estimated coefficients (×100) are reported and *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels respectively with *t*-statistics in parentheses.

Stage	Description	Variable	Model 1		Model 2		Model 3	
1	Regress returns on IVOL	Intercept	1.742***	(6.85)	1.743***	(6.86)	1.823***	(7.19)
		IVOL	-19.537***	(-3.29)	-19.563***	(-3.29)	-22.241***	(-3.85)
2	Add candidate variables	Intercept	1.774***	(6.01)	1.770***	(6.03)	2.017***	(6.91)
		IVOL	-6.692	(-1.57)	-7.332*	(-1.70)	-11.855***	(-2.80)
		Maxret(3mth)	-1.247**	(-1.98)				
		LagMaxret(3mth)			-1.091	(-1.54)	-1.157*	(-1.75)
		Lagret	-3.914***	(-7.38)	-3.987***	(-7.79)	-3.212***	(-6.23)
		NextSUE	0.316***	(20.57)	0.317***	(20.51)		
		SUE	0.032***	(3.39)	0.032***	(3.37)	0.107***	(9.96)
		E(IdioSkew)	-0.245	(-1.56)	-0.251	(-1.65)	-0.326**	(-2.16)
3	IVOL on candidate variables	Intercept	0.600***	(22.15)	0.720***	(22.95)	0.726***	(23.58)
		Maxret(3mth)	14.435***	(50.97)				
		LagMaxret(3mth)			9.545***	(37.76)	9.054***	(38.25)
		Lagret	0.629***	(6.41)	1.982***	(16.03)	2.087***	(17.04)
		NextSUE	-0.017***	(-18.99)	-0.023***	(-19.92)		
		SUE	-0.014***	(-15.64)	-0.019***	(-15.48)	-0.026***	(-18.20)
		E(IdioSkew)	0.754***	(46.16)	1.155***	(37.56)	1.200***	(38.83)
		Adj R^2	57.6%		41.2%		39.9%	
4	Decompose stage 1 IVOL coefficient	Maxret(3mth)	-10.994***	(-3.54)				
			56.3%					
		LagMaxret(3mth)			-6.157***	(-3.18)	-6.613***	(-3.69)
					31.5%		29.7%	
		Lagret	-0.377	(-0.42)	-2.520**	(-2.07)	-2.538**	(-2.10)
			1.9%		12.9%		11.4%	
		NextSUE	-1.188***	(-12.34)	-1.568***	(-12.57)		
			6.1%		8.0%			
		SUE	-0.295***	(-5.73)	-0.393***	(-5.61)	-0.553***	(-6.74)
			1.5%		2.0%		2.5%	
		E(IdioSkew)	-3.043***	(-3.22)	-4.225***	(-2.95)	-4.841***	(-3.41)
			15.6%		21.6%		21.8%	
		Residual	-3.639*	(-1.83)	-4.700*	(-1.74)	-7.697***	(-2.85)
			18.6%		24.0%		34.6%	
		Total	-19.537***	(-3.29)	-19.563***	(-3.29)	-22.241***	(-3.85)
			100.0%		100.0%		100.0%	
		Sample		1988 to 2005		1988 to 2005		1988 to 2005
		Avg # firms		3654.8		3653.7		3749.7

Table 7: Robustness tests: Price screens, DGTW-adjustments, and different holding periods

Using Fama-Macbeth cross-sectional regressions, the negative relation between idiosyncratic volatility and returns is decomposed into a number of components each related to a candidate variable and a residual component. The candidate variables considered are those that explain more than 10% of the puzzle in the univariate analysis and are statistically significant. The main analysis of the paper uses raw returns over a one-month holding period and a $1 price screen. The robustness tests in this table change the price screen to $5 (or remove the price screen), use DGTW-adjusted returns, and increase the holding period to three months. *IVOL* is the standard deviation of residuals from a regression of daily stock returns in month t–1 on the Fama-French factors. *Maxret(3mth)* (*LagMaxret(3mth)*) is the maximum daily return for the three-month period ending in month t–1 (t–2). *E(Idioskew)* is the expected idiosyncratic skewness measure in Boyer et al. (2010). *Lagret* is the month t–1 return. *SUE* and *NextSUE* are the standardized unexpected earnings from the previous quarter and the following quarter, respectively. Time-series averages of estimated coefficients (×100) are reported and *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels respectively with t-statistics in parentheses. Standard errors are based on the Newey-West method with three lags for the three-month holding period.

Variable	Model 1			Model 2			Model 3		
	Coeff.	Frac.	*t*-stat	Coeff.	Frac.	*t*-stat	Coeff.	Frac.	*t*-stat
Panel A: $5 screen									
Maxret(3mth)	-9.078**	43.4%	(-2.20)						
LagMaxret(3mth)				-5.626**	26.9%	(-2.05)	-4.440*	22.7%	(-1.82)
Lagret	-0.623	3.0%	(-0.60)	-2.576*	12.3%	(-1.70)	-2.385	12.2%	(-1.56)
NextSUE	-1.971***	9.4%	(-11.03)	-2.650***	12.7%	(-11.93)			
SUE	-0.280***	1.3%	(-4.02)	-0.437***	2.1%	(-4.89)	-0.670***	3.4%	(-6.53)
E(Idioskew)	-0.918	4.4%	(-1.37)	-1.568	7.5%	(-1.53)	-1.715*	8.8%	(-1.69)
Residual	-8.052***	38.5%	(-2.83)	-8.063**	38.5%	(-2.10)	-10.356***	52.9%	(-2.62)
Total	-20.922***	100%	(-2.82)	-20.920***	100%	(-2.82)	-19.567***	100%	(-2.73)
Sample			1988 to 2005			1988 to 2005			1988 to 2005
Avg # firms			2886.5			2885.8			2947.9
Panel B: No price screen									
Maxret(3mth)	-8.602***	59.2%	(-2.63)						
LagMaxret(3mth)				-4.699**	32.4%	(-2.34)	-5.471***	30.5%	(-2.86)
Lagret	-0.187	1.3%	(-0.21)	-2.328**	16.0%	(-2.03)	-2.353**	13.1%	(-2.10)
NextSUE	-0.853***	5.9%	(-11.73)	-1.110***	7.6%	(-11.87)			
SUE	-0.224***	1.5%	(-5.47)	-0.288***	2.0%	(-4.78)	-0.389***	2.2%	(-5.57)
E(Idioskew)	-3.600***	24.8%	(-3.68)	-4.816***	33.2%	(-3.12)	-5.780***	32.2%	(-3.76)
Residual	-1.059	7.3%	(-0.57)	-1.279	8.8%	(-0.50)	-3.970	22.1%	(-1.53)
Total	-14.524**	100%	(-2.43)	-14.520**	100%	(-2.42)	-17.964***	100%	(-3.04)
Sample			1988 to 2005			1988 to 2005			1988 to 2005
Avg # firms			3786.3			3785.1			3904.1
Panel C: DGTW-adjusted return									
Maxret(3mth)	-10.421***	61.8%	(-7.44)						
LagMaxret(3mth)				-5.583***	33.0%	(-6.07)	-6.111***	31.4%	(-7.29)
Lagret	-0.717	4.3%	(-1.10)	-3.088***	18.3%	(-3.53)	-3.087***	15.8%	(-3.55)
NextSUE	-1.063***	6.3%	(-13.91)	-1.402***	8.3%	(-14.30)			
SUE	-0.236***	1.4%	(-8.09)	-0.329***	1.9%	(-8.61)	-0.448***	2.3%	(-9.98)
E(Idioskew)	-2.759***	16.4%	(-7.55)	-4.140***	24.5%	(-7.75)	-4.718***	24.2%	(-8.89)
Residual	-1.675	9.9%	(-1.36)	-2.356	13.9%	(-1.48)	-5.122***	26.3%	(-3.24)
Total	-16.871***	100%	(-5.92)	-16.899***	100%	(-5.93)	-19.487***	100%	(-7.05)
Sample			1988 to 2005			1988 to 2005			1988 to 2005
Avg # firms			3654.8			3653.7			3749.7

Table 7 (Cont'd)

Variable	Model 1			Model 2			Model 3		
	Coeff.	Frac.	*t*-stat	Coeff.	Frac.	*t*-stat	Coeff.	Frac.	*t*-stat
Panel D: 3-month return									
Maxret(3mth)	-13.040*	47.6%	(-1.92)						
LagMaxret(3mth)				-7.491*	27.3%	(-1.81)	-9.744**	27.5%	(-2.57)
Lagret	-0.361	1.3%	(-0.30)	-0.745	2.7%	(-0.36)	-0.534	1.5%	(-0.26)
NextSUE	-3.096***	11.3%	(-14.66)	-4.074***	14.8%	(-15.46)			
SUE	-0.683***	2.5%	(-6.87)	-0.909***	3.3%	(-7.06)	-1.265***	3.6%	(-8.25)
E(Idioskew)	-2.982	10.9%	(-1.63)	-3.324	12.1%	(-1.21)	-5.082*	14.3%	(-1.87)
Residual	-7.252*	26.5%	(-1.88)	-10.898**	39.7%	(-2.04)	-18.852***	53.1%	(-3.50)
Total	-27.413**	100%	(-2.28)	-27.441**	100%	(-2.28)	-35.478***	100%	(-3.04)
Sample			1988 to 2005			1988 to 2005			1988 to 2005
Avg # firms			3654.8			3653.7			3749.7
Panel E: 3-month DGTW-adjusted return									
Maxret(3mth)	-14.794***	58.2%	(-4.31)						
LagMaxret(3mth)				-7.942***	31.2%	(-3.83)	-10.055***	30.3%	(-5.40)
Lagret	-0.987	3.9%	(-1.39)	-2.569**	10.1%	(-2.15)	-2.332**	7.0%	(-1.97)
NextSUE	-2.798***	11.0%	(-15.44)	-3.665***	14.4%	(-16.38)			
SUE	-0.523***	2.1%	(-8.90)	-0.722***	2.8%	(-9.29)	-0.999***	3.0%	(-10.54)
E(Idioskew)	-3.560***	14.0%	(-5.47)	-5.239***	20.6%	(-5.23)	-6.946***	21.0%	(-7.13)
Residual	-2.774	10.9%	(-1.20)	-5.329*	20.9%	(-1.70)	-12.818***	38.7%	(-4.07)
Total	-25.437***	100%	(-4.36)	-25.467***	100%	(-4.36)	-33.150***	100%	(-5.89)
Sample			1988 to 2005			1988 to 2005			1988 to 2005
Avg # firms			3654.8			3653.7			3749.7
Panel F: 3-month DGTW-adjusted return, $5 screen									
Maxret(3mth)	-18.737***	50.7%	(-4.73)						
LagMaxret(3mth)				-10.278***	27.8%	(-3.77)	-8.981***	24.7%	(-3.80)
Lagret	-1.348	3.6%	(-1.54)	-2.590*	7.0%	(-1.76)	-2.360	6.5%	(-1.58)
NextSUE	-4.308***	11.7%	(-14.66)	-5.790***	15.7%	(-15.90)			
SUE	-0.495***	1.3%	(-5.60)	-0.735***	2.0%	(-6.55)	-1.060***	2.9%	(-8.01)
E(Idioskew)	-2.736***	7.4%	(-5.04)	-4.711***	12.7%	(-6.01)	-5.051***	13.9%	(-6.41)
Residual	-9.344***	25.3%	(-3.21)	-12.871***	34.8%	(-3.55)	-18.938***	52.0%	(-5.02)
Total	-36.968***	100%	(-5.86)	-36.975***	100%	(-5.86)	-36.390***	100%	(-5.90)
Sample			1988 to 2005			1988 to 2005			1988 to 2005
Avg # firms			2886.5			2885.8			2947.9

Figure 1: Average fractions of the idiosyncratic volatility puzzle captured by various explanations

The average fractions of the idiosyncratic volatility puzzle explained by various explanations across the seven specifications (one in Table 6 and six in Table 7) are reported in pie charts. *Maxret(3mth)* (*LagMaxret(3mth)*) is the maximum daily return for the three-month period ending in month *t*–1 (month *t*–2). *E(Idioskew)* is the expected idiosyncratic skewness measure in Boyer et al. (2010). *Lagret* is the month *t*–1 return. *SUE* and *NextSUE* are the standardized unexpected earnings from the previous quarter and the following quarter, respectively.