A Research Proposal on Decomposing Idiosyncratic Volatility

Yilong Li Wei Liu

September 28, 2025

Abstract

This document outlines a research proposal to investigate the idiosyncratic volatility (IVOL) puzzle, a cornerstone of behavioral finance. We leverage a long-history dataset of daily open, high, low, and close (OHLC) prices to decompose IVOL into its intraday and overnight components. The central hypothesis is that the anomalously low returns associated with high IVOL are driven primarily by high *intraday* volatility—a proxy for speculative activity—while overnight volatility, which is more related to the processing of fundamental news, carries a different or insignificant risk premium. This research aims to refine the "lottery-like" stock explanation for the IVOL anomaly (Barberis & Huang, 2008) and provide a more nuanced understanding of what constitutes "good" versus "bad" volatility.

Contents

1 Research Proposal: Decomposing Idiosyncratic Volatility 2

1.1 Motivation and Contribution 2

1.2 Detailed Empirical Methodology 2

1.2.1 Step 1: Volatility Decomposition 2

1.2.2 Step 2: Portfolio Sorts and Factor Construction 3

1.2.3 Step 3: Fama-MacBeth (1973) Regressions 3

1.3 Testable Hypotheses 4

1 Research Proposal: Decomposing Idiosyncratic Volatility

1.1 Motivation and Contribution

The idiosyncratic volatility (IVOL) puzzle is a cornerstone of behavioral finance: stocks with high IVOL have historically exhibited anomalously low future returns. The leading explanation is investor demand for "lottery-like" stocks (Barberis & Huang, 2008). This project aims to refine this theory by decomposing IVOL into its intraday and overnight components. The central hypothesis is that it is not just high volatility, but specifically high *intraday* volatility—representing speculative activity during market hours—that drives the anomaly. Overnight volatility, in contrast, may be more related to the processing of fundamental news and carry a different risk premium.

1.2 Detailed Empirical Methodology

Our empirical approach follows the standard in asset pricing literature by employing two complementary methods to test our hypotheses. First, we use a non-parametric approach by forming factor portfolios based on our new volatility characteristics (portfolio sorts). Second, we use a parametric approach to test for a risk premium after controlling for other known effects (Fama-MacBeth regressions). Presenting both ensures our results are robust and not an artifact of a single testing methodology.

1.2.1 Step 1: Volatility Decomposition

For each stock i, the intraday and overnight volatility measures are calculated at the end of each month t using a rolling window of the daily data within that month. Let d be the index for the trading days within month t, and let $O_{i,d}, H_{i,d}, L_{i,d}, C_{i,d}$ be the open, high, low, and close prices for stock i on day d.

1. **Intraday Volatility ($IVOL_{intra}$):** The idiosyncratic component of a stock's intraday volatility. We will use two primary measures for the raw monthly volatility:

    - **Standard Deviation of Intraday Returns:** First, for each day d in month t, we calculate the open-to-close return, denoted as $R_{i,d}^{O\to C} = (C_{i,d} - O_{i,d})/O_{i,d}$. Let $\bar{R}_i^{O\to C}$ be the average of these daily returns within month t. Then, the monthly intraday volatility is the standard deviation of this daily return series, provided there are at least 15 daily observations in the month:

    $$\text{VOL}_{intra,i,t} = \sqrt{\frac{1}{D_t - 1}\sum_{d=1}^{D_t}(R_{i,d}^{O\to C} - \bar{R}_i^{O\to C})^2}$$

    where D_t is the number of trading days in month t.

    - **Average Daily High-Low Range:** For each day d, we calculate the logarithmic price range, $Range_{i,d} = \ln(H_{i,d}) - \ln(L_{i,d})$. The monthly measure is the average of these daily values.

2. **Overnight Volatility ($IVOL_{over}$):** The idiosyncratic component of a stock's overnight volatility. Let $C_{i,d-1}$ be the closing price on the prior day. The raw monthly volatility is calculated as the standard deviation of the daily close-to-open returns, $R_{i,d}^{C\to O} = (O_{i,d} - C_{i,d-1})/C_{i,d-1}$, within month t.

To obtain the idiosyncratic component for each measure (e.g., $IVOL_{intra}$), we take the residual from a regression that controls for systematic sources of volatility. Specifically, at the end of each month t, for each stock i, we run a time-series regression of its raw monthly volatility measure against the corresponding volatility measures of common factors over a rolling window of the past 36 months (from month $t-35$ to t):

$$\begin{aligned} \mathrm{VOL}_{intra,i,s} = \alpha_i + \beta_{Mkt}\mathrm{VOL}_{intra,Mkt,s} + \beta_{SMB}\mathrm{VOL}_{intra,SMB,s} \\ + \beta_{HML}\mathrm{VOL}_{intra,HML,s} + \eta_{i,s} \quad \text{for } s \in [t-35, t] \end{aligned} \tag{1}$$

Here, $\mathrm{VOL}_{intra,Mkt,s}$ is the intraday volatility of the market portfolio, and $\mathrm{VOL}_{intra,SMB,s}$ and $\mathrm{VOL}_{intra,HML,s}$ are the intraday volatilities of the size and value factor portfolios. The idiosyncratic intraday volatility for stock i for month t, denoted $IVOL_{intra,i,t}$, is defined as the residual for the last month of this regression window, $\eta_{i,t}$. This procedure yields a single, well-defined value for each stock for each month. A parallel regression is run for the overnight volatility measure to obtain $IVOL_{over,i,t}$.

Justification of the Idiosyncratic Volatility Measure: It is crucial to clarify the economic interpretation of our $IVOL$ measure. While volatility itself must be non-negative, our measure, $\eta_{i,t}$, is a regression residual and can therefore be negative. This is by design. Our measure does not represent the *level* of idiosyncratic volatility, but rather the *abnormal component* of volatility, or the "volatility shock," for a given month. A negative value for $IVOL_{intra,i,t}$ is economically meaningful: it signifies that the stock was unexpectedly calm during month t relative to the volatility predicted by its systematic factor exposures. This approach is a valid method for isolating the component of volatility that is orthogonal to systematic volatility shocks, which is the precise characteristic we aim to test.

1.2.2 Step 2: Portfolio Sorts and Factor Construction

This non-parametric test examines if a simple trading strategy based on our volatility measures earns abnormal returns.

1. **Independent Sorts:** At the beginning of each month, we will independently sort stocks into quintiles based on their $IVOL_{intra}$ and $IVOL_{over}$ from the previous month.

2. **Factor Portfolios:** We will construct two long-short factor portfolios:

   - **Intraday Volatility Factor (IVF):** Long the lowest $IVOL_{intra}$ quintile, short the highest $IVOL_{intra}$ quintile.
   - **Overnight Volatility Factor (OVF):** Long the lowest $IVOL_{over}$ quintile, short the highest $IVOL_{over}$ quintile.

1.2.3 Step 3: Fama-MacBeth (1973) Regressions

As a complementary parametric test, we use Fama-MacBeth regressions to estimate the marginal risk premium associated with our volatility characteristics after controlling for other known effects.

First, we define the key variable for our test. The **excess return** of a stock i for month $t+1$, denoted $R_{i,t+1}$, is the stock's raw return minus the risk-free rate for that month.

Stage 1: Monthly Cross-Sectional Regressions For each month t in our sample period, we run a single cross-sectional regression of the *next month's* excess returns ($R_{i,t+1}$) on the firm characteristics known at the end of month t. The independent variables are our idiosyncratic volatility components, $IVOL_{intra,i,t}$ and $IVOL_{over,i,t}$. The regression for a given month t is:

$$\begin{aligned} R_{i,t+1} = \gamma_{0,t} + \gamma_{1,t} IVOL_{intra,i,t} + \gamma_{2,t} IVOL_{over,i,t} \\ + \text{Controls}'_{i,t}\mathbf{\Gamma}_t + \epsilon_{i,t+1} \quad \text{for } i = 1, \ldots, N_t \end{aligned} \tag{2}$$

Here, the regression is run across all firms $i = 1, \ldots, N_t$ available in that month. The term $\text{Controls}'_{i,t}\mathbf{\Gamma}_t$ represents the combined effect of a vector of standard firm characteristics, $\text{Controls}_{i,t}$, and their corresponding regression coefficients, $\mathbf{\Gamma}_t$. This is crucial to ensure that our volatility measures are not simply proxies for other known return predictors. The standard control variables to be included in the $\text{Controls}_{i,t}$ vector are:

- **Size:** The natural logarithm of the firm's market capitalization at the end of month t.
- **Value:** The book-to-market ratio for the firm.
- **Momentum:** The cumulative return of the stock from month $t-12$ to $t-2$.
- **Short-Term Reversal:** The stock's return in the previous month, t.

The output of this stage is a time-series of estimated coefficients for each month: $\{\hat{\gamma}_{1,t}\}_{t=1}^{T}$ and $\{\hat{\gamma}_{2,t}\}_{t=1}^{T}$.

Stage 2: Time-Series Analysis of Coefficients In the second stage, we analyze the time-series of coefficients generated in Stage 1. We compute the average coefficient for each characteristic over the entire sample period:

$$\bar{\gamma}_k = \frac{1}{T}\sum_{t=1}^{T}\hat{\gamma}_{k,t} \quad \text{for } k = 1, 2$$

The Fama-MacBeth t-statistic is then used to test if these average coefficients are statistically different from zero. The t-statistic for $\bar{\gamma}_k$ is calculated as:

$$t(\bar{\gamma}_k) = \frac{\bar{\gamma}_k}{\text{SE}(\hat{\gamma}_{k,t})/\sqrt{T}}$$

where $\text{SE}(\hat{\gamma}_{k,t})$ is the standard deviation of the time-series of the estimated coefficients. This approach provides a robust test of whether a characteristic is associated with a significant risk premium over time.

1.3 Testable Hypotheses

- **H1:** The Intraday Volatility Factor (IVF) will earn a significant positive premium. In the Fama-MacBeth regressions, the coefficient on the intraday volatility component ($\bar{\gamma}_1$) will be significantly negative, consistent with the lottery-stock anomaly.
- **H2:** The Overnight Volatility Factor (OVF) will earn an insignificant premium. The coefficient on the overnight volatility component ($\bar{\gamma}_2$) will be statistically insignificant or much smaller in magnitude than $\bar{\gamma}_1$, a result that would suggest that volatility related to overnight news processing is not associated with the same behavioral biases.