A Research Proposal on a Cross-Sectional Factor from the Intraday-Overnight Return Differential

Yilong Li Wei Liu

September 28, 2025

Abstract

This document outlines a research proposal to investigate the cross-sectional properties of the well-documented intraday-overnight return phenomenon. We leverage a long-history dataset of daily open, high, low, and close (OHLC) prices to construct a new pricing factor based on a stock's historical intraday-overnight return differential. The central hypothesis is that this factor's premium is driven by modern market frictions, particularly those related to retail investor order flow and institutional trading constraints. This research aims to introduce a new, economically motivated pricing factor and provide evidence that its performance is linked to limits to arbitrage.

Contents

1 Research Proposal: A Cross-Sectional Factor from the Intraday-Overnight Return Differential

1.1 Motivation and Contribution

A persistent stylized fact in equity markets is that the long-run market risk premium is earned almost entirely during the overnight (close-to-open) session, while the average intraday (open-to-close) return is near zero. While this pattern is known, the cross-sectional heterogeneity of this phenomenon is less explored. This project proposes to construct a new pricing factor based on a stock's historical intraday-overnight return differential. The core innovation is to test the hypothesis that this factor's premium is driven by modern market frictions, particularly those related to retail investor order flow and institutional trading constraints.

1.2 Detailed Empirical Methodology

1.2.1 Step 1: Constructing the Intraday Reversal (IVR) Factor

Sorting Variable Calculation: At the beginning of each month t, for each stock i, we will calculate a characteristic based on its historical return patterns. First, for each day s in the lookback period, let $O_{i,s}$ and $C_{i,s}$ be the open and close prices for stock i on day s, and let $C_{i,s-1}$ be the close price on the prior day. We define the daily intraday return as $R_{i,s}^{O \to C} = (C_{i,s} - O_{i,s})/O_{i,s}$ and the daily overnight return as $R_{i,s}^{C \to O} = (O_{i,s} - C_{i,s-1})/C_{i,s-1}$. We then compute the sorting variable, $IVR_{i,t}$, using the arithmetic average of these daily returns over the past 12 months (from month $t-12$ to $t-1$):

$$IVR_{i,t} = \text{Avg}(R_{i,s}^{C \to O}) - \text{Avg}(R_{i,s}^{O \to C}) \quad \text{for } s \in [t-12, t-1] \tag{1}$$

This variable captures a stock's tendency to experience negative intraday price pressure that subsequently reverses overnight.

Methodological Refinements for Sorting Variable: To ensure the robustness of our sorting variable, we will implement and test two methodological refinements.

1. **Robustness to Outliers (t-statistic):** To mitigate the influence of outliers inherent in simple return averages, we will construct an alternative sorting variable, $IVR_{i,t}^{\text{t-stat}}$, based on the t-statistic of the daily return difference. Let $\Delta R_{i,s} = R_{i,s}^{C \to O} - R_{i,s}^{O \to C}$. The t-statistic is calculated over the 12-month lookback window as:

$$IVR_{i,t}^{\text{t-stat}} = \frac{\text{Avg}(\Delta R_{i,s})}{\text{StdDev}(\Delta R_{i,s})/\sqrt{D_{12M}}} \tag{2}$$

   where D_{12M} is the number of trading days in the 12-month window. This creates a signal-to-noise ratio that gives more weight to consistent return patterns.

2. **Adjustment for Non-Trading Days:** To handle overnight returns that span multiple calendar days (e.g., weekends and holidays), we will create an adjusted overnight return series. Let $N_{days,s}$ be the number of calendar days between the close on day $s-1$ and the open on day s. The adjusted return, $\tilde{R}_{i,s}^{C \to O}$, is:

$$\tilde{R}_{i,s}^{C \to O} = \frac{R_{i,s}^{C \to O}}{N_{days,s}} \tag{3}$$

The main sorting variable in Equation 1 will then be re-calculated using these adjusted returns to ensure comparability across all observations.

Factor Portfolio Construction: The IVR factor is a zero-investment (long-short) portfolio formed monthly.

1. **Stock Ranking:** At the end of each month t, all stocks in our investment universe are ranked in descending order based on their $IVR_{i,t}$ characteristic.

2. **Quintile Formation:** The ranked stocks are sorted into five quintile portfolios. Quintile 5 (Q5) contains stocks with the highest IVR (the long leg), and Quintile 1 (Q1) contains stocks with the lowest IVR (the short leg).

3. **Portfolio Weighting and Return Calculation:** Within each quintile, stocks are value-weighted. Let $ME_{i,t}$ be the market capitalization (stock price times shares outstanding) of stock i at the end of month t, and let $R_{i,t+1}$ be the return of stock i in the subsequent month $t+1$. The portfolio return for each leg for month $t+1$ is the value-weighted average of the individual stock returns within that quintile. For example, the return of the long leg (Q5) is calculated as:
$$R_{Long,t+1} = \sum_{i \in Q5} \frac{ME_{i,t}}{\sum_{j \in Q5} ME_{j,t}} R_{i,t+1}$$
The final factor return, denoted $R_{IVR,t+1}$, is the difference between the return of the long leg (Q5) and the short leg (Q1).

1.2.2 Step 2: Testing the Factor's Performance

To test Hypothesis 1, we determine if the IVR factor earns abnormal returns after accounting for exposure to known risk factors. We will run a time-series regression of the IVR factor's returns on the Fama-French 5 factors plus the momentum (UMD) factor. Let $R_{IVR,t}$ be the return of the IVR factor in month t.

$$\begin{aligned} R_{IVR,t} = \alpha &+ \beta_{Mkt}(R_{m,t} - R_{f,t}) + \beta_{SMB} SMB_t + \beta_{HML} HML_t \\ &+ \beta_{RMW} RMW_t + \beta_{CMA} CMA_t + \beta_{UMD} UMD_t + \epsilon_t \end{aligned} \tag{4}$$

Here, $R_{m,t} - R_{f,t}$ is the market excess return, and SMB (Small Minus Big), HML (High Minus Low), RMW (Robust Minus Weak), CMA (Conservative Minus Aggressive), and UMD (Up Minus Down) are the standard factor returns. The intercept, α, represents the factor's abnormal return. A positive and statistically significant alpha would support H1.

1.2.3 Step 3: Testing the Link to Market Frictions

To test Hypothesis 2, we investigate whether the IVR premium is stronger for stocks with characteristics associated with market frictions. We use two complementary methods.

Portfolio Double Sorts: This non-parametric test provides a clear visualization of the results. At the end of each month t, we perform a two-way independent sort:

1. First, we sort stocks into terciles based on a friction proxy (e.g., retail ownership).

2. Second, within each friction tercile, we sort stocks into quintiles based on their $IVR_{i,t}$ characteristic.

3. We then calculate the IVR factor premium (Q5 - Q1) separately within the "High Friction" tercile and the "Low Friction" tercile.

Hypothesis 2 predicts that the IVR premium will be significantly larger and more statistically significant in the High Friction tercile.

Fama-MacBeth Regressions with Interaction Terms: This parametric test quantifies the marginal effect of frictions. For each month t, we run a cross-sectional regression of next month's excess returns on the IVR characteristic, a friction proxy, and an interaction term between the two. Let $R_{i,t+1}$ be the raw return of stock i in month $t+1$, let $R_{f,t+1}$ be the risk-free rate, and let $\text{Friction}_{i,t}$ be the value of the friction proxy for stock i at month t.

$$R_{i,t+1} - R_{f,t+1} = \gamma_{0,t} + \gamma_{1,t} IVR_{i,t} + \gamma_{2,t}\text{Friction}_{i,t} + \gamma_{3,t}(IVR_{i,t} \times \text{Friction}_{i,t}) + \text{Controls}'_{i,t}\mathbf{\Gamma}_t + \epsilon_{i,t+1} \quad (5)$$

The key coefficient is $\gamma_{3,t}$, which captures the incremental effect of the IVR characteristic for a one-unit change in the friction proxy. After running this regression for every month in the sample, we test the significance of the time-series average, $\bar{\gamma}_3$. A significant and positive $\bar{\gamma}_3$ would support H2. The friction proxies to be tested include:

- **Retail Ownership:** Proxied by low institutional ownership (from 13F filings).
- **Short-Sale Constraints:** Measured by low institutional ownership and high short interest.
- **Market Sentiment:** An aggregate time-series measure like the Baker-Wurgler index.

The $\text{Controls}_{i,t}$ vector will include standard firm characteristics such as size, book-to-market, and momentum to ensure the interaction effect is not driven by other known anomalies.

1.2.4 Step 4: Robustness Tests

To ensure our findings are not an artifact of specific methodological choices, we will conduct a series of robustness tests. The primary test involves replicating the main analyses using the alternative sorting variable defined in our methodological refinements.

1. **Replicating Portfolio Sorts:** We will repeat the factor construction in Step 1 and the performance tests in Step 2 using $IVR_{i,t}^{\text{t-stat}}$ (from Equation 2) as the ranking variable. Consistent results for the factor's alpha would demonstrate that the premium is not sensitive to outliers in daily returns.
2. **Replicating Regressions:** We will repeat the Fama-MacBeth regressions from Step 3, replacing the main $IVR_{i,t}$ characteristic with $IVR_{i,t}^{\text{t-stat}}$. Finding that the interaction term remains significant would confirm that the link to market frictions is a robust phenomenon.

1.3 Testable Hypotheses

- **H1:** The Intraday Reversal (IVR) factor will earn a significant positive alpha, as tested in Equation 4, suggesting it is a priced characteristic in the cross-section.
- **H2:** The IVR factor premium will be significantly stronger among stocks with high market frictions. This will be evidenced by the portfolio double sorts and by a positive and significant average coefficient on the interaction term ($\bar{\gamma}_3$) in the Fama-MacBeth regression specified in Equation 5.