Expected Return in Night and Day: Role of Trading Volume *

Abstract

We document a novel high-volume overnight premium (intraday discount). Specifically, stocks with high trading volume exhibit remarkable future outperformance (underperformance) during the overnight (intraday) period. This phenomenon is also prevalent in global equity markets, robust to alternative measures of trading volume, and persistent for at least three years after portfolio formation. In a tractable equilibrium model, we argue such heterogeneous night/day effect is the result of the cycling pattern of belief dispersion being high/low at market open/close, due to the overconfidence of late-informed investors. To justify this idea, we use high-frequency returns and document the decreasing cross-sectional dispersion of market betas over the intraday period is especially strong for high-volume stocks. We also propose several testable model predictions and empirically shows that our main results are more pronounced when uninformed investors are more optimistic, or more dominant in overall trading activity.

JEL Classification: G11, G12

Keywords: overnight and intraday returns, trading volume

1 Introduction

The trading volume constitutes the bedrock of asset pricing theory. A fundamental question is its joint behavior with asset returns. However, the empirical relation between expected return and trading volume is rather mixed. In the seminal work of Wang (1994), the return-volume relation is suggested to have rich implications on investor heterogeneity. On the other hand, as mentioned by Lou et al. (2019), investor clientele can be distinguished by the different time points of trading, especially at market open and close. In other words, overnight (close-to-open) and intraday (open-to-close) components of expected returns can be used to capture investor heterogeneity in demand. Based on the above, we are motivated to consider whether the relation to the trading volume differs for the expected returns in overnight and intraday periods. This paper targest this question that is less examined in the literature.

Specifically, we document a novel finding that the relation to the trading volume is strikingly opposite for expected returns in overnight and intraday. Using 3-month average turnover as trading volume, we document stocks with higher trading volume significantly outperform (underperform) in the overnight (intraday) period of the subsequent month. A graphical representation of out main results is in Figure 1. Specifically, with a univariate sort on trading volume from July 1992 to December 2022, the overnight returns in the subsequent month strictly monotonically increase with trading volume, from -0.15% to 2.29%. On the other hand, the intraday returns in the subsequent month strictly monotonically decrease with trading volume, from 0.71% to -1.56%. The monthly overnight and intraday return of the high-minus-low (H-L) trading volume portfolio is thus 2.44% ($t = 6.44$) and -2.26% ($t = -6.18$). We label these as the *high-volume overnight premium* (*intraday discount*) as the profits are primarily from the leg of high-volume stocks.

We build a simple and tractable rational expectation equilibrium model to demonstrate the intuition behind our finding. Precisely, investors face asymmetric information at market open, and are overconfident when extracting signals from the price. The combination of these two features generates implied belief dispersion across investors and the market's overreaction to information, thereby increasing the price. As information spreads, the disagreement is resolved at market close, and the price is corrected. As a result, the expected market prices exhibit an ebbs-and-flows pattern within daily cycles of market opens and closes.

Our model considers a continuum of investors trading a risky asset in four periods, which represent market opens and closes in two days. Cash flow news arrives every day. The key frictions arise from information asymmetry and investors' overconfidence. At market open, early informed investors receive a noisy signal about today's news, while others are late informed after the open period. One may interpret the early-informed investors as institutional investors, and the late-informed as retail investors. These

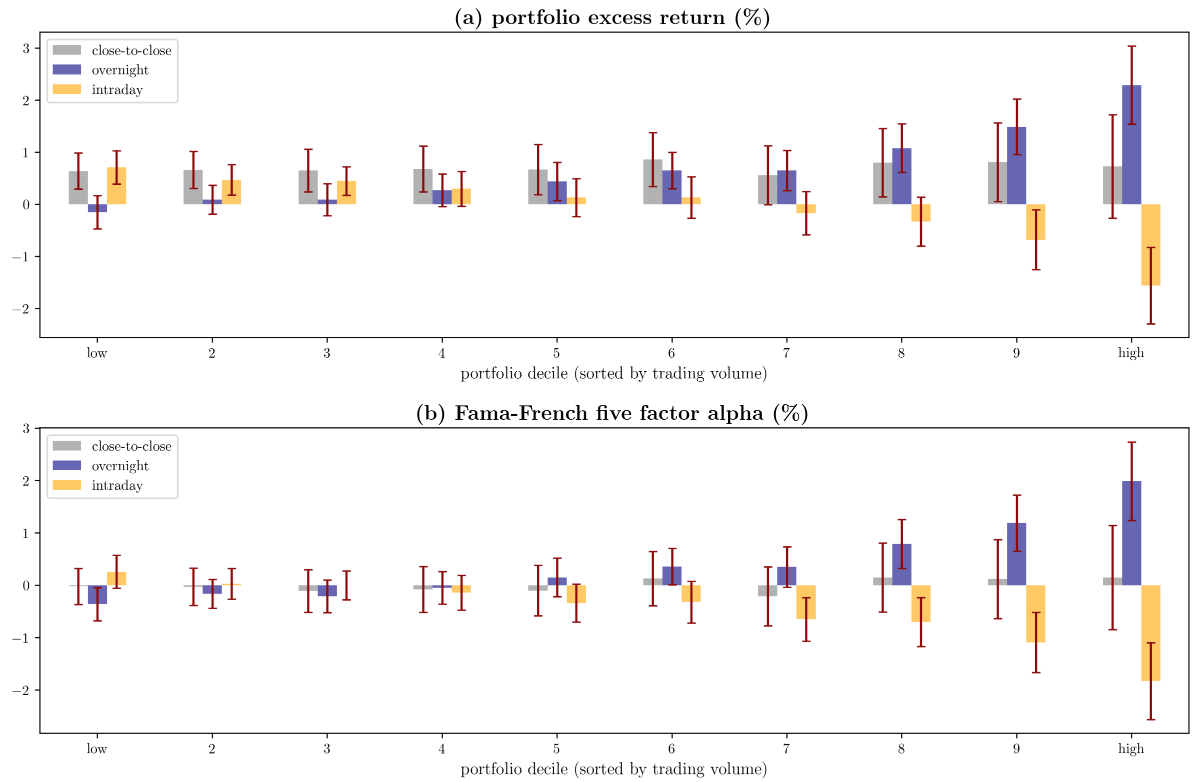


Figure 1 Value-weighted portfolios sorted by trading volume: night and day

This figure plots average monthly returns (close-to-close, overnight and intraday) of portfolios sorted into deciles of trading volume, defined as 3-month average turnover (ratio of traded shares to shares outstanding). Panel (b) plots the results adjusted by Fama-French 5 factors. Portfolios are value-weighted and rebalanced each month. The sample period covers from June 1992 to December 2022, excluding stocks with month-end prices lower than \$5 and market equity lower than 20th NYSE breakpoints at portfolio formation. 95% confidence intervals are displayed.

late-informed investors, having no private signals at market open, extract signals from the price, but mistakenly believe the precision of their extracted signal is greater than it actually is. At market close, all the investors have a noisy signal and trade in a rational manner. The news is then realized after market close.

The central model implication is that trading volume positively (negatively) predicts close-to-open (open-to-close) price changes, i.e., overnight (intraday) returns, while is to close-to-close returns. To see why trading volume can be a strong predictor, we highlight that the underlying mechanism falls in overconfidence and disagreement. They are not driven by specific information flows, but reflect collective clientele characteristics in the market (e.g., the fraction of late-informed traders). This shares a similar insight with the "tug-of-war mechanism" in Lou et al. (2019), in which the overnight/intraday patterns are due to different clients in different trading periods. The clientele characteristics change slowly in practice, and thus are persistently predictable. On the other hand, trading volume naturally relates to clients' disagreement, and therefore helps uncover these collective characteristics. In this sense, we further draw testable evidence and economic interpretations. First, high-volume stocks should exhibit a larger disagreement at market open and subsequently a steeper intraday decline in disagreement. Second, the overnight/intraday return-volume relationship should be amplified by a larger share of retail investors, and more pronounced in a less certain situation.

High-volume overnight premium (intraday discount) is substantial in predictability. The statistical and economic significance is barely changed after being adjusted by a wide range of asset pricing factor models. See Panel (b) of Figure 1 for the adjusted returns with respect to Fama and French (2015) five factors (FF5). Hendershott et al. (2020) shows that risk in overnight and intraday periods are very different, which implies factors realized in overnight/intraday periods largely improve the performance of pricing models on expected returns in respective periods. Importantly, our high-volume overnight premium and intraday discount are not susceptible to such concerns, with economically significant returns remained after related risk-based adjustment.

The predictability is also not driven by industry effects, considering trading volume quite differs across industries (high-tech v.s. manufacturing). High-volume overnight premium (intraday discount) is very stably profitable, yielding considerable returns even in recession periods. The implied annualized Sharpe ratio of overnight premium (discount) in the sample is 1.16 (-1.12), more than two times the magnitude of the market Sharpe (0.54). The profitability of the strategy survives moderate trading costs (e.g. 5bps daily) for rebalancing positions to earn overnight and intraday returns. Finally, the strategy is very persistent, especially for the overnight premium. The FF5-adjusted H-L portfolio returns in overnight and intraday remain significant for at least three years after portfolio formation, due to the persistence on high-volume leg.

High-volume overnight premium (intraday discount) is a highly robust asset pricing result. We perform a wide range of robustness tests to jusfity its widespread predictive power: (1) the results is robust to 5 alternative trading volume measures (e.g. transaction-based volume, dollar-based volume, attention-related abnormal volume); (2) economic significance of long-short strategy is largely retained after controlled with leading return predictors; (3) robust in two havles of sample period, across stocks listed on different exchanges, and whether economy is in recession or expansion; (4) robust using alternative portfolio breakpoints or sample filters. The majority of results are very close to the baseline results in the statistical significance (absolute t-stat larger than 5) and economic magnitude (averge monthly return of 2% of long-short strategy).

Beyond the U.S. market, high-volume overnight premium (intraday discount) widely exists in equity markets worldwide, for both developed and emerging markets. 24 out of 30 countries from the MSCI index are revealed with qualitatively same pattern as in the U.S. market. The premium (discount) is more profitable for emerging markets, among which South Korea, Thailand, and Taiwan are documented with a magnitude 3 times of the U.S. counterpart (more than 6% per month). These findings are also robust to the adjustment of risk-based factor models. Interestingly, the Chinese equity market is found to have the exactly opposite pattern, i.e. high-volume overnight discount (intraday premium), aligning with the robust fact of negative overnight-intraday return gap of stocks in China.

We control for additional firm characteristics in Fama-MacBeth regressions, including market beta, size, book-to-market, momentum, illiquidity, volatilities and lottery features. Trading volume remains to be a very weak predictor of close-to-close returns. Consistent to our main finding, the coefficient of one-month ahead overnight (intraday) returns on trading volume is very significantly positive (negative). The regression results suggest a similar statistical significance as in portfolio-level analysis. With multiple control variables included, around 60% of the magnitude in the coefficient on trading volume is retained, compared to the analogue in the univariate regressions. This is in accordance with the results in robustness checks that main results remain strong amid several predictors substantially correlated with trading volume (e.g. about 0.3 with market beta). Additionally, the results above are robust by considering the fixed effects of 48 industries in Fama and French (1997).

Next, we propose several empirical results to support the mechanism explaining the baseline finding, i.e., high-volume overnight premium and intraday discount. We first reveal the intraday dynamics of investor disagreement, heterogeneous across stocks with different trading volumes, from a high-frequency perspective. Specifically, we use high-frequency data to estimate systematic risks at different points of time in the trading day, which can be understood as continuous-time process of market betas. We show that for all stocks, the dispersion of such systematic risks across stocks is decreasing over the trading day, i.e. shrinking intraday beta dispersion. Aligned with our prediction, such intraday beta dispersion is decreasing in a much faster speed for stocks with high trading volume. By attributing beta dispersion as investor disagreement, the overnight-intraday return gap is due to the large disagreement at market open, which is gradually reduced over the trading day. This phenomenon is more pronounced in high-volume stocks, naturally leading to the emergence of high-volume overnight premium and intraday discount.

Our simple model also implies that the strength of the return-volume relation is related to parameters on how investors digest information in the economy. Concretely, the magnitude of the premium (discount) will be larger when (1) the fundamental-related shock is more uncertain; (2) late-informed investors are more confident in their own signals. We test these predictions by attributing indexes related to *investor sentiment, economic policy uncertainty* to capture the time variation in uncertainty in fundamental & public news, and overconfidence of uninformed investors. The model predictions are supported in our data. High-volume overnight premium (intraday discount) is significantly larger in periods when: (1) individual investors are more bullish on the market; (2) major media (newspapers) convey more certainty on economy policy. The results above survive the FF5 factors and are mostly due to the time-variation of returns in the group of high-volume stocks.

Furthermore, the model also predicts that premium (discount) will be stronger when uninformed investors are more prevalent in the economy. We focus on retail investors to proxy for uninformed investors, as they are usually inferior in information, more subject to behavioral bias, and disagreed. Using the algorithm pioneered by Boehmer et al. (2021), we are able to identify the (marketable) orders of retail trading. This helps us to decompose the trading volume into its retail and non-retail components. From the results of Fama-MacBeth regressions, we find that the future overnight and intraday returns significantly load on the retail volume, in the same direction as the total trading volume. This empirically supports the model prediction that a higher mass of uninformed investors enlarges the overnight premium (intraday discount). Appealingly, the coefficient of future overnight returns on non-retail volume is also significantly negative, as opposed to that of retail trading. This actually points to the 'tug-of-war' story, i.e. different investor clientele exert opposite price pressure in different time periods. However, for intraday returns, the non-retail impact is rather muted.

Finally, we consider several leading explanations on overnight/intraday returns as alternative choices to our main story. Based on heterogeneous investor clientele argument in Lou et al. (2019), persistence of overnight/intraday returns, incured by 'tug-of-war' between demand of these clienteles, plays an important role in explaining their dynamic variation. Using a double-sort portfolio exercise, we later confirm that the positive/negative relation of trading volume with future overnight and intraday returns are independent to their contemporaneous analogs. Therefore, trading volume seems to capture a novel dimensions of expected overnight/intraday returns, compared to the persistent component of night/day returns implied by 'tug-of-war' story.

To characterize the cross-section of expected overnight and intraday returns, we show that the power of trading volume is not subsumed by two previous explanations in literature: systematic risk (market beta) (Hendershott et al., 2020) and mispricing (Bogousslavsky, 2021). In these studies, the overnight (intraday) returns in the future are documented to be higher (lower) for stocks that are more sensitive to the market (high beta) or are more underpriced. First of all, we confirm that the high-volume overnight premium (intraday discount) is still in presence conditional on all levels of market beta and mispricing measures (Stambaugh et al., 2015). Meanwhile, it turns out that trading volume has additional explanatory power. As for the market beta, the relation with overnight returns is strongly positive for stocks with higher volume and weak for stocks with lower volume. Briefly, *trading volume strengthens risk-return tradeoff* (*in overnight*).

On the other hand, the mispricing, i.e. spread between underpriced and overpriced stocks, mainly materializes in intraday and is corrected partially in overnight. This is especially robust for stocks with higher trading volume, while the mispricing is economically negligible for low-volume stocks. In short, *trading volume amplifies mispricing* (*in intraday*), which further sharpens the volume amplification effect on mispricing in Han et al. (2022). On top of these, we also confirm that our main finding is not dominated by certain days of the week, which is shown to have a significant impact on asset pricing anomalies (Birru, 2018; Bogousslavsky, 2021).

Related Studies. From our perspective, Hendershott et al. (2020) and Bogousslavsky (2021) are probably the most relevant to our work, in the manner of explaining the cross-section variation of expected overnight and intraday returns. We distinguish from these studies by showing that the importance of their results is actually concentrated in stocks with high trading volume. Their findings are generally related to the different risks in the overnight period, while we attribute our results to the role of belief dispersion manifested in volume, under an economy with information asymmetry. Another highly related study would be Han et al. (2022), where they dissect the weak expected return-volume relation via the cross-sectional variation in mispricing. We take up the same challenge by exploring the time-variation (overnight, intraday) in expected returns for the full cross-section, and further confirm their finding is mainly from the intraday period.

Our study contributes to the vast literature on asset pricing implications of trading volume. Early works focus on how trading volume affects the dynamics of return autocorrelations (Campbell et al., 1993; Chordia and Swaminathan, 2000; Llorente et al., 2002). This paper fits more into the strand on the return predictability of trading volume. For the overall trading volume, Conrad et al. (1994) and Lee and Swaminathan (2000) disentangle its weak predictability by emphasizing the cross-sectional heterogeneity of past performance in short/medium horizons. Among other volume-related predictors, *abnormal trading volume*, capturing visibility to investors, is reported to generate robust premium in both U.S. (Gervais et al., 2001) and international markets (Kaniel et al., 2012), with substantial real effects on both firm-level (Israeli et al., 2022) and macroeconomy (Wang, 2021). Meanwhile, Pan et al. (2016) document that speculation-based abnormal trading actually predicts the opposite direction. Furthermore, other features reflected in trading volume also significantly predict expected return, e.g. favor of retail investors (Hvidkjaer, 2008) and salient thinking on high volume (Sun et al., 2023). Our study contributes to this literature by resolving the weak return-volume relation based on the time variation of investor demand.

This paper is closely connected to the literature of overnight returns. A series of studies focus on the persistent gap of returns realized in the overnight and intraday periods. In addition to risk and mispricing mentioned above, leading explanations include but are not limited to price pressure of opposing investor clientele (Lou et al., 2019), information asymmetry at market open (Lu et al., 2023), attention-induced retail activity (Berkman et al., 2012) and biased expectations (Jones et al., 2022; Saadon and Schreiber, 2023). We add trading volume, a simple measure superior in performance, to the explanations on the night-day return gap and emphasize the role of belief dispersion in overpricing induced at market open.

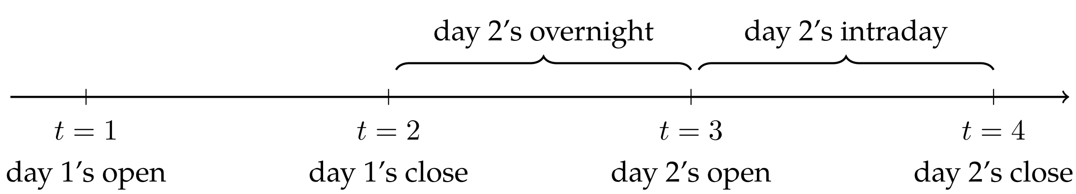
We proceed with this paper in the following order: Section 2 sketches an equilibrium model of heteregeneous investors to explain return-volume relation in night and day. Section 3 elaborates on the main empirical results. Section 4 illustrates the main mechanism from a high-frequency perspective. .Section 5 investigates several economic interpretations that consistent to our model predictions. Section 6 provides robustness checks for the main empirical finding. Section 7 concludes.

2 The Model

This section introduces a simple and tractable model to draw the main predictions and intuitions.

2.1 Setup

Asset payoffs and timeline. There are two assets, a single risky asset in unit net supply, and a riskless asset in perfectly elastic supply at interest rate of zero. There are four trading periods, which can be treated as two days, whereas each day is divided into two periods, referring to market open and close, respectively. The timeline is plotted below. $t = 1, 2$ $(3, 4)$ represent "day 1's open and close" ("day 2's open and close"), respectively. Therefore, the time interval between $t = 2, 3$ $(3, 4)$ is interpreted as day 2's overnight (intraday) period. The two-day setting allows us to analyze predictability. As shown later, each day is set similarly on the information structure, ensuring the generality of our implications.



Standing at day t $(t = 1, 2)$, investors anticipate the value of the risky asset equals

$$v_t = v_{t-1} + \theta_t + \xi_t, \tag{2.1}$$

and is paid as a terminal divided. v_{t-1} is the common prior before day t's open. Assume $v_0 > 0$ is commonly known. The innovations θ_t and the fundamental uncertainty ξ_t are mutually uncorrelated and i.i.d. over time with zero means and variances τ_θ^{-1} and τ_ξ^{-1}, respectively.[1] After day t's close, i.e.,

[1]Eq. (2.1) implicitly assumes that investors treat future innovations as part of the fundamental uncertainty, whereas the

after trading in period $2t$, the exact value of θ_t is realized and becomes part of the common prior. After trading in the last period $t = 4$, the fundamental uncertainty is resolved, and the risky asset pays the terminal divided.

Investors, information, and belief. There is a continuum $[0, 1]$ of investors. The innovation θ_1 (θ_2) can be viewed as yesterday's (today's) cash flow news. At the market open period $2t - 1$ ($t = 1, 2$), a mass $(1-s)$ of investors are early informed by a noisy private signal $s_t = \theta_t + \epsilon_t$, where $s \in (0, 1)$, $\epsilon_t \sim \mathcal{N}(0, \tau_\epsilon^{-1})$. ϵ_1 and ϵ_2 are independent. The early-informed investors can be interpreted as institutional investors in contrast to retail investors. Following the common approach in noisy rational expectations models (e.g., Grossman and Stiglitz, 1980; Hirshleifer et al., 1994), we assume the early informed investors receive the identical signal and behave competitively. They hold the noisy signal and have unbiased belief about τ_ϵ. At the market close of each day, i.e., period $2t$ ($t = 1, 2$), the rest s mass of "late-informed" investors receive the private signal s_t.

The late-informed investors can be viewed as retail investors with limited capacity to access private information in practice. They extract signals from the price when not receiving signals. The key friction is that these late-informed investors are overconfident: they mistakenly believe that the early informed investors' signal is more precise than it actually is. As a result, they overestimate the precision of the signal extracted from the price. We assume they believe the extracted signal has a noise with variance $(m\tau_\epsilon)^{-1}$, $m > 1$.[2]

The overconfidence here is subtly distinguished from several similar forms of overconfidence in the literature. First, Luo et al. (2021) consider that the late-informed investors exhibit a form of skepticism, i.e., believe that early-informed investors' signal is less precise than it actually is. This can also be interpreted as "underestimating an opponent" (Johnson and Fowler, 2011). This form of overconfidence results in underreaction to private information. We alternatively consider late-informed investors' overconfidence in the other way to better align with our empirical findings, where the overnight premium is more likely to reflect the market's overreaction to private information. Second, vast literature consider the form of overconfidence in overestimating the precision of private signals (e.g., Odean, 1998; Daniel et al., 1998). We inherit the core insights on the resulting overreaction, while consider the case when investors without private information are overconfident. This relates to the realistic when distinguishing early and

variance of uncertainty does not change in any single day. This ensures the same information structure for each day, and aligns with investors' myopic optimization, as introduced later.

[2] As introduced below, supply shocks prevent the prices from being fully revealed, lowering the precision of the extracted signal. However, once the overconfidence is introduced, the investors fail to correctly infer the actual precision of the extracted signal, even if they notice the effect of supply shocks. Here we actually make a reduced-form assumption for tractability: the overconfidence dominates the effect of supply shocks, resulting in that the late-informed investors believe the extracted signal has a greater precision than the true precision of the original private signal.

late informed investors, as the early (late) informed are more institutional (retail) and more (less) likely to exhibit expert judgments. Our empirical evidence also confirms that the retail trading dominates the effect.

Investors' problem and market clearing. Each investor $i \in [0,1]$ is myopic with constant absolute risk aversion (CARA) utility over terminal wealth w_T.[3] She can trade and decide the allocation of risky asset $x_{i,t}$ in any periods $t \in \{1,2,3,4\}$ based on her information set $\mathcal{F}_{i,t}$, i.e.,

$$\max_{x_{i,t}} \mathrm{E}[-e^{-\frac{1}{\gamma}w_{i,T}}|\mathcal{F}_{i,t}], \tag{2.2}$$

where γ represents risk tolerance and is identical across all investors for simplicity.

In each period $t \in \{1,2,3,4\}$, the aggregate supply z_t is stochastic due to noise trading. $z_t \sim \mathcal{N}(z, \tau_z^{-1})$, which are mutually independent of other uncertainties and are i.i.d. over time. Without loss of generality, each investor initially holds z_0 risky asset, which is randomly chosen from $\mathcal{N}(z, \tau_z^{-1})$. The supply shocks prevent the prices from being fully revealed. The market clearing condition sets the aggregate demand equal to the aggregate supply, i.e.,

$$\int_{i\in[0,1]} x_{i,t}\mathrm{d}i = z_t, \quad t \in \{1,2,3,4\}. \tag{2.3}$$

Note that the late-informed investors have a reduced-form belief about the precision of their extracted signals (see footnote 2), then it is unnecessary to assume a certain distribution of z_t. Importantly, we assume $z > 0$, representing that the investors as a whole usually hold a positive amount of risky assets. One may relate assumption to short selling constraints (SSC), which force the investors to hold non-negative allocations. The assumption here is considerably looser than the SSC, and is realistic as the total number of circulating stocks are always positive, whereas noise traders rarely dominate the holdings.

2.2 Equilibrium Price and Overnight / Intraday Returns

Following the standard approach in identifying rational equilibrium models, we start with the conjecture of the equilibrium price p_t in each period t. In particular, consider p_t as a linear function of the private signal of the day and the supply shock z_t. Under this conjecture, the price is normally distributed, and the investor $i \in [0,1]$'s optimization problem (2.2) solves

$$x^*_{i,t} = \gamma[\mathrm{Var}(v|\mathcal{F}_{i,t})]^{-1}[\mathrm{E}(v|\mathcal{F}_{i,t}) - p_t]. \tag{2.4}$$

[3]Myopic optimization is for simpler exposition and algebraic convenience (as also in, e.g., Nagel, 2012; Hirshleifer, 2020).

Combining to the market clearing condition (2.3), we solve the price and verify the conjecture. Proposition 2.1 verifies the linear conjecture and solves for the equilibrium price p_t of each period.

Proposition 2.1. (***Equilibrium Price***) *The equilibrium prices p_t $(t \in \{1, 2, 3, 4\})$ satisfy*

$$p_1 = v_0 + \left(1 + \frac{1}{A_1 + s(A_2 - A_1)}\right)(\theta_1 + \epsilon_1) - \frac{1}{A_1 + s(A_2 - A_1)} z_1, \tag{2.5}$$

$$p_2 = v_0 + \theta_1 + \epsilon_1 - \frac{1}{A_1} z_2, \tag{2.6}$$

$$p_3 = v_0 + \theta_1 + \left(1 + \frac{1}{A_1 + s(A_2 - A_1)}\right)(\theta_2 + \epsilon_2) - \frac{1}{A_1 + s(A_2 - A_1)} z_3, \tag{2.7}$$

$$p_4 = v_0 + \theta_1 + \theta_2 + \epsilon_2 - \frac{1}{A_1} z_4, \tag{2.8}$$

where A_1 and A_2 are constant,

$$A_1 = \frac{\gamma \tau_\xi (\tau_\theta + \tau_\epsilon)}{\tau_\xi + \tau_\theta + \tau_\epsilon} < \frac{\gamma \tau_\xi (\tau_\theta + m\tau_\epsilon)}{\tau_\xi + \tau_\theta + m\tau_\epsilon} = A_2. \tag{2.9}$$

Proposition 2.1 directly indicates that there are predictable overnight and intraday returns in terms of expectations, where the expectation can correspond to pooling a set of observations in practice.

Corollary 1. (***Overnight Premium and Intraday Discount***) *The expected overnight (intraday) return in day-2 $\mathrm{E}(r_N^{\text{day } 2})$ $(\mathrm{E}(r_D^{\text{day } 2}))$ is*

$$\textbf{\textit{overnight}}: \quad \mathrm{E}(r_N^{\text{day } 2}) := \mathrm{E}(P_3 - P_2) = \left(\frac{1}{A_1} - \frac{1}{A_1 + s(A_2 - A_1)}\right) z > 0, \tag{2.10}$$

$$\textbf{\textit{intraday}}: \quad \mathrm{E}(r_D^{\text{day } 2}) := \mathrm{E}(P_4 - P_3) = \left(\frac{1}{A_1 + s(A_2 - A_1)} - \frac{1}{A_1}\right) z < 0. \tag{2.11}$$

Overconfidence and implied belief dispersion. The key mechanism arises from the subjective return precision wedge between late- and early-informed investors, $A_2 > A_1$. Late-informed investors are overconfident on the informativeness of their extracted signal. As a result, they overreact to the private information in the market, trade more aggressively, jacking up the price at market open. This effect is crucially related to the supply shock. The logic is: the supply shock prevents prices from being fully revealed, generating investors' disperse beliefs about the expected value, $\mathrm{E}(v|\mathcal{F}_{i,\sim \text{market open}})$. This implied belief dispersion consequently enables the overreaction of a particular (the late-informed) group to generate an impact on the price. On the other hand, at market close, the investors are all informed so that the disagreement is resolved. The high open price due to overreaction is corrected. The above process is cyclical, leaving no predictable expected close-to-close returns.

Information. Corollary 1 also suggests that the results (i.e., the expected overnight premium and the intraday return) are not driven by any private information with specific directions or magnitudes. In addition, although not included in our model, public information is less likely to be the driving force of the results, since it is observed by all the investors and does not create disagreement in a rational expectation setting.

Clientele characteristics and persistence. Furthermore, note that the formulas in Corollary 1 also applies to other dates, e.g., $\mathrm{E}(r_D^{\text{day 1}}) = \mathrm{E}(P_2 - P_1) = \mathrm{E}(P_4 - P_3) = \mathrm{E}(r_D^{\text{day 2}})$. Similarly, it applies to any single day in an extended N-day setting, as long as the clientele-characteristic parameters do not change significantly (e.g., the uninformed share s and overconfidence m). This aligns with Lou et al. (2019) that the overnight-intraday expected returns potentially reflect the specific clientele patterns, and helps to understand the persistence of the results.

2.3 The Role of Trading Volume

Although the clientele-characteristic parameters are relatively opaque in practice, they change slowly that appear auto-correlation, which is captured by constant parameters for simply focusing on the mechanism of predictability. Historical statistics may naturally contain information about the clientele and thus exhibit predictive power for overnight and intraday returns. We hypothesize the aggregate trading volume to be a strong predictor in this sense, since client heterogeneity and disagreement naturally creates trades.

We consider the aggregate expected trading volume $\mathrm{E}(V^{\text{day 1}})$ as the indicator of trading volume in day 1, which measures each investor's trading volume as the absolute change in its expected shares of the risky asset over periods,

$$\mathrm{E}(V^{\text{day 1}}) = \frac{1}{2}\left[\begin{aligned}&\int_{i\in[0,1]} (|\overline{x_{i,1}} - \overline{z_0}| + |\overline{x_{i,2}} - \overline{x_{i,1}}|)\mathrm{d}i \\ &\qquad + |\overline{(1-z_1)} - \overline{(1-z_0)}| + |\overline{(1-z_2)} - \overline{(1-z_1)}|\end{aligned}\right], \tag{2.12}$$

where $\overline{X}$ denotes the unconditional expectation of X.[4]

Proposition 2.2. (***Trading Volume and Its Predictability***) *The aggregate expected trading volume in day 1 reads*

$$\mathrm{E}(V^{\textit{day 1}}) = \frac{2s(1-s)z(A_2 - A_1)}{A_1 + s(A_2 - A_1)}. \tag{2.13}$$

[4] The measure of volume here is analytically convenient for understanding directional characteristics, yet is with the loss of counting for the subsequent trading volume due to allocation uncertainties. Compared to measuring the expected aggregate trading volume, $\mathrm{E}(V^{\text{day 1}})$ guarantees ordinal relations, but is weakly smaller.

The expected overnight and intraday returns in day 2 can be rearranged as,

$$\begin{aligned} &\mathrm{E}(r_N^{day\ 2}) = \beta \mathrm{E}(V^{day\ 1}), \quad \mathrm{E}(r_D^{day\ 2}) = -\beta \mathrm{E}(V^{day\ 1}), \\ \textit{where} \quad &\beta = \frac{1}{2(1-s)A_1} = \frac{\tau_\xi + \tau_\theta + \tau_\epsilon}{2(1-s)\gamma\tau_\xi(\tau_\theta + \tau_\epsilon)} > 0. \end{aligned} \tag{2.14}$$

Proposition 2.2 indicates that, on average, a higher volume predicts a larger overnight premium and intraday discount, implying an excess overnight return of the high-minus-low portfolio sorted by trading volume and vice versa intraday. In addition, volume generally shows weak predictability for close-to-close returns, which aligns with the common knowledge.

We further draw several observations from (2.14), as concluded by Corollary 2.

Corollary 2. (***Testable Evidence and Economic Interpretations***)

(i) *High-volume stocks exhibit a higher investor disagreement at market open and a steeper intraday decline of disagreement;*

(ii) *The volume-return relation β increases in the share of late-informed investors s, and the variances of the fundamental uncertainty and signal, i.e., τ_ξ^{-1} and τ_θ^{-1}.*

First, β does not include the term related to overreaction, i.e., A_2, implying that the overconfidence and belief dispersion mechanism is also presented in the trading volume. As a result, a higher volume corresponds to a greater disagreement at market open. Since the disagreement is resolved at market close, the higher volume implies a sharper intraday decrease.[5] Second, by decoding β, we find several elements that can affect the economic significance of the role of trading volume. A greater s typically means there is a greater share of retail investors in practice. They generate more overreaction and aggressive trading, amplifying the return-volume relation. In contrast, in a "less-uncertain" situation, investors generate less implied belief dispersion, weakening the return-volume relation.

3 Empirical Analysis

In this section, we employ portfolio analysis and Fama-MacBeth regressions to examine the power and direction of trading volume in predicting future returns realized in overnight and intraday periods. Robustness tests and international evidence are also provided to corroborate the main findings.

[5] In practice, the disagreement may not be fully resolved, e.g., there are always uninformed investors or behavioral investors. They are more likely to increase the disagreement in the whole day rather than change it within the day, thereby distinguishing them from our mechanism.

3.1 Data and Summary Statistics

Key variables. We start with the forming returns realized in overnight (close-to-open) and intraday (open-to-close) periods, following the convention in Lou et al. (2019). Specifically, let $P_{i,t}^{close}$ and $P_{i,t}^{open}$ be the close and open prices for stock i on day t. The overnight and intraday returns at daily level are defined as:

$$\text{intraday: } r_{i,t}^{day} = \frac{P_{i,t}^{close}}{P_{i,t}^{open}} - 1 \qquad \text{overnight: } r_{i,t}^{night} = \frac{1 + r_{i,t}}{1 + r_{i,t}^{day}} - 1 \tag{3.1}$$

where $r_{i,t}$ is daily close-to-close (CTC) return, adjusted for dividend distribution and stock splits in overnight periods. CTC returns are obtained from CRSP using the variable *holding period return* (`RET`) for the U.S. market. For the international markets, CTC returns are obtained from Datastream using the percentage difference in daily *return index* (`RI`). For our main analysis, we make use of monthly level overnight and intraday returns by compounding their daily analogs within each month m.

$$r_{i,m}^{day} = \prod_{t \in m}(1 + r_{i,t}^{day}) - 1 \qquad r_{i,m}^{night} = \prod_{t \in m}(1 + r_{i,t}^{night}) - 1 \tag{3.2}$$

Another key variable is trading volume, measured by the turnover averaged in the past three months (Han et al., 2022). For each month, turnover is measured by the amount of traded shares as a fraction of the total number of outstanding shares at the end of the period. For NASDAQ-listed stocks, which are subject to the double counting issue, we follow Gao and Ritter (2010) to adjust trading volume in shares.[6] Without specific indication, we will refer to trading volume as the 3-month average turnover.

Data Source. We obtain equity data in the US market from the daily stock file of the Center for Research in Security Prices (CRSP). The sample runs from June 1992 to December 2022, during which open prices are available. We obtain all common stocks (CRSP share codes 10 and 11) listed on the NYSE, AMEX, and NASDAQ exchanges. To avoid microstructure noise effects, our sample excludes penny stocks (prices lower than $5 at portfolio formation) and microcaps (market equity lower than the 20th percentile NYSE size breakpoint). For control variables in regression nalysis, we consider a set of firm charactersitics, including size, book-to-market ratio, momentum, illiquidity, idiosyncratic volatility, maximum return and skewness. Detailed definition of the controls are contained in Table A.I. Firm-month level data can be obtained from Open Source Asset Pricing[7] (OSAP in short), contributed by Chen and Zimmermann

[6]For observations before January 2001, share volume is scaled by 2. For observations in the remaining months of 2001, share volume is scaled by 1.8. For observations in 2002 and 2003, share volume is scaled by 1.6. No adjustment in share volume is required for observations after 2004.

[7]Website: https://www.openassetpricing.com

(2022). Other data sources for variable construction include Compustat, I/B/E/S, and Trades and Quotes (TAQ) data. We defer to elaborate the details when necessary.

Summary Statistics. Trading volume displays sufficient dispersions in cross-section over time. Figure 2 plots the three levels of trading volume percentiles (10th, 50th, and 90th) in cross section of all stocks included in the US sample. Generally, trading volume experienced a steady upward trend in the first half of the sample. It peaked around the 2007-2008 financial crisis when high-volume stocks had a 3-month average turnover of at least 50%. In the following 15 years, trading volume gradually reverted to a relatively stable level. Another observation is that trading volume is constantly right-skewed over time, manifested by the significantly larger gap between the 90th percentile and median.

[Insert Figure 2 here]

Figure 3 provides correlation-based statistics of trading volume and other control variables. In Panel A, we report monthly, quarterly, and yearly autocorrelations[8] of main variables, labeled as AR(1), AR(3), and AR(12), respectively. Trading volume is persistent in the short run (with monthly and quarterly correlations of 0.84 and 0.41, respectively) while relatively transitory in the long term (yearly autocorrelation of 0.08). This pattern is quite different from that of other persistent return predictors (e.g., size, beta, illiquidity), which exhibit longer memory in time series (slower decaying pattern in autocorrelogram).

Panel B of Figure 3 shows the pairwise correlations across variables. Highlighted in black rectangles, correlations of trading volume are generally moderate (between -0.2 and 0.3) with other characteristics. Market beta and illiquidity exhibit stronger (rank-based) comovement with trading volume, which is unsurprising as trading volume is a composite measure to reflect these factors. We later show in the multivariate analysis that the effect of trading volume on overnight/intraday returns is (at best) partially explained by other characteristics, even superseding the power of highly correlated variables. Therefore, trading volume may capture a different dimension in explaining the dynamics of overnight/intraday returns compared to the knowledge in previous studies.

[Insert Figure 3 here]

3.2 Baseline Result

In this subsection, we demonstrate a robust empirical pattern in the US market: *in cross-section, stocks with higher trading volume experience significantly higher (lower) overnight (intraday) returns in the subsequent month*. At the beginning of each month, we sort the stocks in our sample of the U.S. equity market into

[8] For each firm, we compute the autocorrelations at month t using a 60-month rolling window, requiring at least 24 available observations. Reported results are average of autocorrelations pooled at the firm-month level. Using pooled median yields basically same numbers.

ten groups based on trading volume computed at the end of last month. In each decile portfolio, stocks are value-weighted and held for one month. High-volume stocks on average have higher risk, better past performance, and exhibit more lottery-type features[9].

Panel A of Table I, a numeric presentation of Figure 1, provides the excess returns of each decile portfolio, measured in close-to-close (CTC) and its overnight and intraday components. We refer to the high-minus-low (H-L) portfolio as the spread of longing the portfolio with the highest decile of trading volume and shorting the one with the lowest decile. Regarding CTC returns, trading volume is not a sharp cross-sectional predictor, i.e., there is no monotonicity in returns across portfolios. The average monthly spread of the H-L portfolio is just 0.09% ($t = 0.20$).

[Insert Table I here]

The weak CTC return-volume relation is the net effect of two strongly opposite forces: overnight (intraday) expected returns are positively (negatively) related to trading volume. As for the H-L portfolio, the monthly overnight spread averages 2.44% (t-stat: 6.44) per month, and the intraday analog is -2.26% (t-stat: -6.18). Adjusted returns with respect to Fama-French five factors are also very similar in magnitude. In Figure OA1 of Online Appendix, we show the results using equal-weighted portfolios are nearly identical. The magnitudes of overnight and intraday returns are more pronounced for high-volume stocks while relatively muted for low-volume groups. Hence, we call these significant overnight (intraday) spreads of the H-L portfolio as *high-volume overnight premium* (*intraday discount*). The annualized overnight and intraday Sharpe ratios of high-volume overnight premium and intraday discount are 1.16 and -1.12, significantly larger (in magnitude) than the market Sharpe ratio of 0.54.

Trading volume significantly varies across industries. For example, stocks in high-tech industries are more actively traded than those in the mining industry. We aim to show the main result is not a pure industry fixed effect. In particular, firms are classified into Fama and French (1997) 48 categories (FF48) based on SIC codes. For each firm, volume is demeaned by its FF48 industry average. We then perform portfolio sorts based on industry-adjusted trading volume in Panel B of Table I. The main results still hold after considering the industry effect. Across portfolios from the lowest to highest decile of industry-adjusted trading volume, the average overnight returns increase from 0.34% to 2.14%, and the average intraday returns decrease from 0.21% to -1.38%. High-volume overnight premium and intraday discount remain highly comparable to original results, on average 1.80% (t-stat: 7.48) and -1.59% (t-stat: -5.53) per month, robust to risk-based adjustment.

Risk-adjusted performance. High-volume overnight premium and intraday discount are not fully accounted for by leading asset pricing factor models. Besides FF5-adjusted returns presented in Table I,

[9] See details in Section A.1 of Online Appendix.

we also consider eight more models (e.g., CAPM, Hou-Xue-Zhang, mispricing, etc.) to adjust for the overnight and intraday performance of the H-L portfolio. We report the results in Table OA.II in Online Appendix. Under all model specifications, adjusted returns (alphas) are highly significant and very close to the magnitude of average overnight and intraday returns of the H-L portfolio. Using equal-weighted portfolios yields similar conclusions.

We also show that overnight (intraday) returns of H-L portfolio also survive the adjustment of asset pricing factors realized in overnight (intraday) period. This exercise is motivated by the concern that whether standard factors measured in CTC returns are able to capture the variation in overnight/intraday returns. Hendershott et al. (2020) show the cross-section of stock overnight & intraday returns are related to systematic risks in different directions. We decompose asset pricing factors into their overnight and intraday analogs using high-frequency factor data from Aleti (2022)[10]. Table II provides overnight and intraday returns adjusted by their corresponding CAPM and FF5 factor returns components. Specifically, we run the following regressions and report α_N and α_D in Panel A and B, respectively:

$$\text{overnight: } r_{p,t}^N = \alpha^N + (\beta^N)'\mathbf{f}_t^N + \varepsilon_{p,t}^N \tag{3.3}$$

$$\text{intraday: } r_{p,t}^D = \alpha^D + (\beta^D)'\mathbf{f}_t^D + \varepsilon_{p,t}^D \tag{3.4}$$

where $r_{p,t}^N, r_{p,t}^D$ are overnight and intraday returns of volume-sorted decile portfolios. $\mathbf{f}_t^N, \mathbf{f}_t^D$ are vectors of monthly-level overnight/intraday analogs of asset pricing factors formed as above. We report results using the CAPM and FF5 models. The monotonicity across portfolios and significance of H-L spread are still highly robust here, even though FF5 captures a relatively large variation in H-L spread. Therefore, the high-volume overnight premium and intraday discount are not fully due to the heterogeneity in systematic risk during trading/non-trading sessions.

[Insert Table II here]

Profitability and persistence. High-volume overnight premium implies a steady, profitable strategy for investors. Figure 4 shows the cumulative performance, in terms of close-to-close, overnight, and intraday returns, with \$1 invested in July 1992 on H-L spread portfolio. It turns out the *high-volume overnight premium* (dash line in navy blue) earns a striking \$4755.26 after 30 years (annualized return: 32%). High-volume intraday discount (dashed dot line orange) also implies a raw cumulative profit of \$1620.17 (annualized return: 27%) when we use reversed strategy (buy low volume, sell high volume). To realize overnight/intraday returns, we are also faced with non-trivial trading costs incurred by opening

[10] Aleti (2022) provides overnight (labeled as return at 9:30) and 1-minute level returns at a daily level for 272 asset pricing factors. We first compound the 1-minute level factor returns to compute that in intraday factors at daily level. Then we compound overnight and intraday returns at monthly level. The sample is available from January 1996 to December 2020.

and closing positions on a daily frequency. In fact, the real-world performance of our strategy is also robust to this concern. By imposing a daily trading cost of 5 bps, the high-volume overnight premium and intraday discount still have an annualized return of 16.4% and 12%, respectively. Meanwhile, as shown in Figure 4, both strategies are immune to significant drawdowns, even in recession periods. The strategy profits are growing in a steady log-linear pattern, except for being boosted during the dot-com bubble, especially for overnight returns. Finally, there is no visible trend in close-to-close returns of H-L spread, except for a short-term surge around the turn of the millennium.

[Insert Figure 4 here]

Note that we rebalance the portfolios sorted by trading volume every month. It is of interest to examine whether both overnight premium and intraday discount have a persistent pattern over time. In Figure 5, we plot the average overnight and intraday returns of H-L spread in h months after portfolio formation. Horizon h varies from 3 to 36 months. The figure also provides analogous results for the portfolios of the extreme deciles. Panel (a) reveals that high-volume overnight premium is very persistent and significant up to at least 36 months, mainly from the leg of high volume. We can argue similarly for high-volume intraday discount in panel (b). In fact, overnight returns are still significant up to 5 years.[11] Unreported results show nearly identical results with adjustment of FF5 factors. To some extent, the evidence of long-term persistence echoes with the 'tug-of-war' story in Lou et al. (2019), which argues persistent price pressure is due to persistent demand from (night or day) investors to compete with their opposing clientele.

[Insert Figure 5 here]

One may also question whether the persistent performance of the long-short strategy is due to the persistent nature of trading volume. In other words, stocks in the highest volume decile will always remain in this group. In the online appendix, we compute the transition matrix of portfolio rebalancing. It turns out that the cross-sectional distribution of trading volume is not static. Portfolios without extreme deciles (decile 2-9) are rebalanced with a proportion of at least 60% after 3 months. Extreme deciles are also substantially updated with 30-50% of new component stocks (about half) after 12 months.

In all, we show in this subsection that high-volume stocks earn higher (lower) future returns in the overnight (intraday) period, which leads to the weak relation of CTC returns with trading volume. The average returns of high-volume overnight premium (intraday discount) are significant over a 30-year sample and robust to various asset pricing factor models. The overnight premium and intraday discount mainly originate from high-volume stocks and last for at least 3 years. Trading strategy based on this

[11] This is quite unusual in terms of long-run performance for market anomalies, as the monthly return of our strategy in distant future remains statistically and economically significant. This is in contrast with other strategies that exhibit purely increasing cumulative profits without reversal, while future monthly performance are insiginificant.

earns significant economic profits, even after considering moderate trading costs.

3.3 Robustness Tests

This subsection conducts a series of tests to show that high-volume overnight premium and intraday discount are still pronounced using alternative measures of trading volume, not dominated by characterisitics correlated with trading volume, and robust in different subsamples or using different portfolio formation methods. The results are presented in Table III. Our discussion below focus on raw returns of spread of value-weighted high-minus-low (H-L) trading volume portfolio. Results of FF5-adjusted returns are also provided and yield identical conclusions.

[Insert Table III here]

Alternative measures of trading volumes. We propose five alternative measures below and show that the main results still hold using thes alternatives. Results are in Panel A. Three measures are weakly (magnitude less than 0.1) correlated with the main variable, i.e., 3-month average turnover.

1. *six-month average turnover*. We first construct average turnover in the past 6 months, following Lee and Swaminathan (2000), as an alternative horizon compared to our main variable. We report a monthly H-L spread of 2.37% (t-stat=6.55) in overnight and -2.15% (t-stat=-5.91) in intraday, which is slightly lower than the results in the baseline finding. The correlation between average turnover using the past 3 and 6 months is about 0.8, a mechanical result of long-memory property in times-series of trading volume (Fleming and Kirby, 2011).

2. *growth of transaction amount*. In the main analysis, trading volume is formed using the amount of shares. As an alternative, we consider a transaction-based volume measure (Conrad et al., 1994), defined as the monthly growth rate of the total number of trades [12]. This measure just has a correlation of 0.03 with main measure. It turns out that H-L spread is still robust in night and day, respectively 0.57% (t-stat:4.88) and -0.71% (t-stat:-3.39) on average per month, relatively weak than main results in terms of economic magnitude.

3. *dollar volume*. For the third measure, we consider total volume measured in total dollars volume[13] as a proxy of the level of liquidity. This measure is also weakly correlated (about 0.09) with primary measure. Results based on dollar volume confirm the same idea as our baseline finding. The H-L portfolio earn a positive spread (1.00%, t-stat: 5.32) in overnight period and a negative spread (-0.96%, t-stat: -4.05) in intraday period.

[12] We extract data from WRDS intraday indicators, based on Trades-and-Quotes (TAQ) data. For data before the end of 2002, we use variable `NumTrades_t` from intraday indicators file. For data after beginning of 2003, we use variable `total_trade` from millisecond intraday indicators file.

[13] Similar to the source of *total number of trades*, we obtain dollar volume using variable `SumValue_t` in intraday indicators before 2002, and variables `total_dv_LR` in millisecond intraday indicators after 2003.

4. *abnormal trading volume*. We consider shock of the trading volume as our next alternative, usually considered as a proxy for investor attention. To cater to our analysis based on monthly data, we make several adjustments to the measure proposed initially in Gervais et al. (2001). Specifically, for each firm-month observation, we compute the average weekly turnover τ_1^w, using five trading days closest to month's end. It is similar to compute the turnover in previous nine weeks $\tau_2^w, \ldots, \tau_{10}^w$ using daily data up to 10 weeks before the month's end[14]. By ranking $(\tau_1^w, \tau_2^w, \ldots, \tau_{10}^w)$ increasingly, we can obtain the rank of τ_1^w (from 1 to 10), which is the number of decile groups the stock belongs to. A higher rank indicates that the stock had higher abnormal trading volume in the recent week. The monthly spread in overnight is 0.20% (t-stat: 2.24) between highest and lowest abnormal volume. However, results of intraday returns are instead muted. Therefore, results of abnormal trading volume can only weakly and partially echo our baseline findings.
5. *three-month average turnover, excluding announcement days*. We revise our main variable (3-month average turnover) by controlling the effect of surged trading due to information events. The earning announcement dates (eDay) are obtained from Compustat. When computing average turnover, we exclude the daily observations that belong to a three-day window [-1,+1] around eDays. The overnight and intraday spreads of the H-L portfolio are respectively 2.46% (t-stat: 6.63) and -2.24% (t-stat: -6.32), basically the same as baseline results. These findings imply that the high-volume overnight premium and intraday discount are not purely driven by earning announcements.

Portfolio sort controlled by other predictors. In Subsection 3.1, we point out that trading volume is reasonably relevant to well-known predictors. It is necessary to check whether these correlated variables also explain the baseline results. We include all firm characteristics defined in Panel B of Table A.I as control variables, except for short-term reversal and return skewness that are nearly uncorrelated with trading volume. For each control variable, we perform bivariate portfolio sort by first assigning stocks into deciles based on the control. Then, within each decile, we further divide stocks into ten groups based on trading volume. Finally, we average overnight/intraday returns of each of the volume-sorted decile portfolios across the ten control groups and then compute the spread between the highest and lowest decile of volume-sorted portfolios, i.e., H-L portfolio.

Panel B of Table III reports the monthly average overnight/intraday spread of the H-L portfolio, controlled by correlated predictors. We find the high-volume overnight premium and intraday discount are still pronounced after controlling with the correlated predictors. In particular, the economic significance (magnitude of spread in the H-L portfolio) of baseline results is largely retained. Therefore, the main results are not (entirely) superseded by the power of established predictors.

[14]For example, τ_2^w is average turnover of 6th-10th trading days that are closest to month end, or the week before the data used for τ_1^w. Therefore, τ_{10}^w is computed based on information of 45th-50th trading days closest to month end (10th closest week).

Subsample analysis. We start with examining the main findings in two subperiods: 1992–2008 and 2009–2022. The reason we divide the sample in this way is the contrasting trend pattern of trading volume in these two periods (see previous discussions on Figure 2). Panel C of Table III shows the relevant results. The high-volume overnight premium is more pronounced before 2008, with a monthly average spread of 3.32% (t-stat: 5.80) of the H-L portfolio, which is around two to three folds of result after 2008 (1.40%, t-stat: 3.63). Similar conclusions can be made for intraday returns: monthly average of intraday spread in the H-L portfolio are -3.27% (t-stat: -5.88) and -1.10% (t-stat: -3.06) before and after 2008.

Different exchanges have varying levels of trading intensity. We split the stocks of our sample into two halves: the group listed on NYSE/AMEX and the group listed on NASDAQ. We present the portfolio results in Panel D of Table III. Both the overnight and intraday spreads of the H-L portfolio are much larger for stocks listed on NASDAQ, with a monthly average of 3.37% (t-stat: 6.78) and -3.10% (t-stat: -6.36), respectively. Results for NYSE/AMEX-listed stocks are around half in magnitude compared to stocks listed in NASDAQ.

We also compare the performance of high-volume overnight premium (intraday discount) in regimes of economic recessions and expansions, indicated by NBER. Relevant results are exhibited in Panel E of Table III, The overnight/intraday spread of the long-short strategy is significant with similar magnitude in both regimes. Differences in the H-L spread between recessions and expansions are insignificant, about -0.30% (t-stat: -0.26) and -0.66% (t-stat: -0.38) in overnight and intraday period per month.

Alternative portfolio formation and sample filters. Panel F of Table III elaborates on our final robustness tests of baseline findings. We start with the choice of portfoio breakpoints. In the main exercise, stocks are sorted into decile portfolios based on breakpoints constructed from all stocks. For a more balanced average market share across decile portfolio, we use NYSE breakpoints as an alternative, yielding results (overnight: 2.21%, intraday: -1.89%) quite close to baseline ones.

Next, we examine whether the results still hold for microcaps (market equity below 20th NYSE size breakpoints). Microcaps are excluded in the main analysis, as they usually face higher arbitrage costs that drive worsened mispricing. We empirically testify to this notion, where the spread of the H-L portfolio (based on microcaps) is much pronounced among microcaps, with a monthly overnight return of 3.92% (t-stat: 8.34) and intraday return of -3.24% (t-stat: -8.78). Interestingly, the decile portfolio with the lowest volume earns quite significant negative (positive) returns in the overnight (intraday) period, contrasting with the relatively muted average return of low-volume leg in baseline findings. This may also point to the speculative nature among microcaps, even for those traded less actively.

Finally, we repeat our main exercise by excluding financial and utility firms based on SIC codes. This practice of exclusion serves as a convention in empirical research to rule out firms with different business nature relative to the overall market. It turns out that the high-volume overnight premium and intraday discount are slightly stronger after this adjustment. Therefore, our main results are not just driven by particular industries.

3.4 International Evidence

In this subsection, we show that the high-volume overnight premium (intraday discount) is a pervasive phenomenon in international stock markets. Briefly, the majority of 30 markets examined here exhibit the same qualitative evidence as that in the U.S. market, except that China appears to show a significantly opposite pattern.

We collect firm-day level equity data (return, volume, firm size) of 49 international markets in the MSCI index from Refinitiv Datastream. The sample period runs from January 1990 to December 2022. Following Appendix B.1 and B.2 in Griffin et al. (2010), static filters are applied to exclude non-common equities, e.g., preferred stocks, ETFs, REITs, and ADRs. For each market, firms are kept in the sample if they are listed on the main domestic exchange[15] and use the local currency of the respective market. Then, we apply the dynamic filter in Hendershott et al. (2020) to exclude abnormal return observations.[16] Finally, we only keep firm-day observations with a trading volume of at least 100 US dollars.

We further implement the following filters to align with data preprocessing in baseline results. For all international markets, we exclude penny stocks (month-end prices lower than USD $1) and microcaps (market equity lower than NYSE 5th percentile). The key difference here is in the definition of trading volume, where we change to 1-month turnover, i.e., in the recent past month.[17] To alleviate the small-sample estimation bias, results are provided only for markets with sufficient sample coverage. We require the sample length to be available for at least 120 months, and for each month, there are at least 50 stocks with non-missing observations of trading volume. In the end, 18 developed markets and 12 emerging markets are selected for this analysis.

[Insert Figure 6 here]

We repeat the exercise of univariate portfolio sorts based on trading volume for each market. Figure 6 presents the results. In Panel (a), we plot the average monthly overnight and intraday returns of the value-weighted high-minus-low (H-L) trading volume portfolio. Among all thirty markets, twenty-four have corroborative evidence (statistically significant H-L spreads) to support the baseline results in the

[15] Several countries have multiple main exchanges: Canada (Toronto & TSX Ventures), China (Shanghai & Shenzhen), Germany (Deutsche Boerse & Xetra), India (NSE & BSE), Japan (Tokyo & Osaka), and South Korea (KRX & KOSDAQ)

[16] We exclude observations with return index (`RI`) lower than 0.01 or non-positive open prices. With respect to abnormal reversal on daily returns, both r_t and r_{t-1} will be set to missing if $\max\{r_t, r_{t-1}\} > 100\%$ and $(1+r_t)\times(1+r_{t-1})-1 < 20\%$. Furthermore, observations of overnight and intraday returns will be kept only if their magnitude is less than 200%.

[17] U.S. results are nearly identical using a 1-month turnover. However, the pattern disappears for most international markets if a 3-month average turnover is used, probably due to the low price informativeness of trading volume in distant months.

U.S. market. In particular, this phenomenon is present in 15 out of 18 developed markets and 9 out of 12 emerging markets. The magnitude of the monthly H-L spread is generally more prominent in emerging markets. Compared to the U.S. result, For several markets (e.g., South Korea, Thailand, and Taiwan), a high-volume overnight premium (intraday discount) is documented, with a magnitude of about 2-3 times that in U.S. results. We also check the risk-adjusted performance of the international markets by controlling the H-L portfolio with Fama-French 5 factors[18] in the respective market. As exhibited in Panel (b), the magnitude of FF5-adjusted spread in the H-L portfolio is almost unchanged relative to raw returns.

Despite the prevalent evidence worldwide, the Chinese stock market appears to be a salient exception. In particular, we document a significantly and exactly opposite pattern in China, i.e., *stocks with higher volume are more likely to have lower overnight returns and higher intraday returns in the future*. We argue this might align with a well-documented *negative* overnight-intraday return gap in the Chinese market, in contrast to the U.S. evidence. Explanations of such negative return gaps (or negative overnight returns) in China include investors' required discount due to the intraday constraints, or 'T+1' rule (Qiao and Dam, 2020), and the gambling preferences of retail investors (Gu et al., 2023). However, testing which mechanism is attributable to the return-volume relation is beyond the scope of this paper, and we leave it to future studies.

3.5 Fama-MacBeth Regressions

Previously, we used univariate portfolio sort to show high-volume overnight premium (intraday discount). Despite of its intuitiveness, we ought to control for the impact of other confounding variables that are not fully revealed at portfolio-level analysis. In this subsection, a series of Fama-MacBeth regressions are performed for 1-month ahead excess returns (CTC, overnight, and intraday) on trading volume, the key predictor, and other firm characteristics. Regression results confirm similar return predictability of trading volume as in portfolio sort, after accounting for a wide range of related variables.

Table IV reports the results. All the independent variables are first winsorized at 1st and 99th percentiles and then normalized to have zero-mean and unit standard deviation cross-sectionally. Regressions (1)(4)(7) are the univariate models of excess returns (CTC, overnight, intraday) on trading volume. Similarly, CTC returns are weakly related to volume, as a net result of high-volume overnight premium and intraday discount. In an economic interpretation, the coefficient for overnight returns (0.74, t-stat: 5.97) implies an annualized Sharpe ratio of $5.97/\sqrt{30.5} = 1.08$ (where 30.5 is sample length in years) if trading on high-volume overnight premium. For intraday returns, the coefficient (-0.77, t-stat: -6.82)

[18] Asset pricing factors of global markets are from Global Factor Data, thanks to the contribution of Jensen et al. (2023). Website: https://jkpfactors.com/

suggests an even stronger result of $-6.82/\sqrt{30.5} = -1.23$. These results are quite close to those suggested in the portfolio analysis (overnight: 1.16, intraday: -1.12).

[Insert Table IV here]

In columns (2)(5)(8), we involve the control variables listed in subsection 3.1 for multivariate analysis. For both results in night and day, the statistical significance of coefficients on the trading volume is even stronger after involving controls. However, the magnitudes of coefficients drop around 40% (overnight: 0.40, t-stat: 7.28; intraday: -0.45, t-stat: -7.01), possibly attributed to variables with high correlation with trading volume (higher than 0.2 for market beta, volatility, maximum returns). We also control for the industry effect as mentioned in portfolio results. Multivariate regressions are performed by including dummies assigned based on Fama-French 48 industry classification, labeled as "Industry FE" in Table IV. Results are exhibited in columns (3)(6)(9). The coefficients of trading volume are barely changed (overnight: 0.39, t-stat: 6.95; intraday: -0.48, t-stat: -7.63) in both sense of statistical and economic significance.

In addition, for other controls in regressions, the sign of coefficients on overnight and intraday returns are generally in opposite directions. Most of them are consistent to the sign as in literature when predicting overnight returns (e.g. market beta, momentum, realized volatility, maximum returns), while the size effect mainly arises from predictability on intraday returns. To sum up, the high-volume overnight premium and intraday discount are neither explained by well-documented return predictors nor just a simple industry phenomenon.

4 A High-Frequency Perspective

Our model shows that investor disagreement is stronger at market open and shows decreasing pattern over the intraday trading time. This effect is even stronger for high-volume stocks. In this section, we provide empirical evidence to support such argument by showing that the cross-sectional dispersion of high-frequency systematic risks is much larger for volumes with high trading volume.

As for the notation of high-frequency prices and returns, we adopt the following expression:

$$r_{i,t-1+\frac{\tau}{M}} = P_{i,t-1+\frac{\tau}{M}}/P_{i,t-1+\frac{\tau-1}{M}} - 1, \text{ where } \tau = 1, \ldots, M \tag{4.1}$$

where M is the intraday sampling frequency of prices. In this section, we focus on 5-minute prices and returns. Therefore, $M = 6.5 \times (60/5) = 78$, i.e. there are 78 observations of 5-minute level returns for each firm and each day. Therefore, $r_{i,t-1+\tau/M}$ is interpreted as the realized return of stock i, within the τ-th 5-minute interval on day t. For example, $\tau = 3$ corresponds to the return realized in the interval of 9:40-9:45. Realized returns, as shown in the formula, is computed as percentage difference of prices sampled at beginning & end of the 5-minute intervals. The intraday transaction prices of U.S. common stocks are obtained from First Rate Data[19], where the sample runs from January 2000 to December 2022.

We estimate local market beta, capturing the individual asset's exposure to market risk within a local window of trading time, i.e. high-frequency systematic risks. We use i and m to denote the generic individual asset and aggregate market. For each stock-day observations, we estimate local market betas using a 90-minute rolling window, with forward step of size of 30 minutes. So the first beta we estimated is indicated for 11:00, using high-frequency returns between 9:30-11:00. The next local beta is indicated at 11:30 and so on so forth.

Following Andersen et al. (2023), the local market beta is estimated as follows:

$$
\begin{aligned}
\hat{\beta}_{i,t-1+\frac{\tau}{M}} &= \left(\widehat{Var}_{t,\kappa}\right)^{-1} \widehat{Cov}^{(j)}_{t,\kappa} \\
&= \left(\frac{M}{\Delta_M} \sum_{\tau'=\tau-23}^{\tau} \left[r_{m,t-1+\frac{\tau'}{M}}\right]^2 \cdot \mathbf{1}[|r_{m,t-1+\frac{\tau'}{M}}| \leq \nu^{(m)}_{t,M}]\right)^{-1} \times \\
&\quad \left(\frac{M}{\Delta_M} \sum_{\tau'=\tau-23}^{\tau} r_{i,t-1+\frac{\tau'}{M}} r_{m,t-1+\frac{\tau'}{M}} \cdot \mathbf{1}[|r_{m,t-1+\frac{\tau'}{M}}| \leq \nu^{(m)}_{t,M}] \cdot \mathbf{1}[|r_{i,t-1+\frac{\tau'}{M}}| \leq \nu^{(i)}_{t,M}]\right)
\end{aligned}
\tag{4.2}
$$

where $M = 78$ is sampling frequency (5-mintute in our case) within each trading day, and $\Delta_M = 1/M$ is the block size. Intuitively, the market beta is estimated using 'continuous' returns (both individual and market) within the local 90-minute window. To reduce estimation error from 'jump' returns, for each day t, we construct a threshold $\nu^{(j)}_{t,n}$ to eliminate such 'jumps' as follows:

$$
\nu^{i}_{t,M} = 4\sqrt{BV^{(i)}_{t,M}}\Delta_M^{0.49}, \text{ where } BV^{j}_{t,M} = \frac{\pi}{2}\sum_{\tau=2}^{n}\left|r_{i,t-1+\frac{\tau}{M}} \cdot r_{i,t-1+\frac{\tau-1}{M}}\right| \tag{4.3}
$$

where BV is bipower variance, a realized variance estimator robust to jumps. Market threshold $\nu^{(m)}_{t,M}$ is constructed similarly for high-frequency market return $r^{(m)}_{t,i}$. As demonstrated in equation (4.3), returns with magnitude larger than the threshold, i.e. 'jumps', will be set to 0.

[Insert Figure 7 here]

We then sort all stocks into five groups, based on the trading volume in previous month, and compute the pooled-standard deviation of high frequency betas, grouped by the volume quintiles and half-hour intervals in the trading day. Figure 7 shows the relevant results. It can be seen that, for all volume quintiles, stocks have higher dispersions in betas at the market open, echoed with evidence of Andersen et al. (2023). Furthermore, stocks with high (low) volume stocks are on average more (less) dispersed in

[19] Website: https://firstratedata.com/

systematic risks over the trading period. The decreasing pattern of intraday beta dispersion is also much more faster among high volume stocks. The dispersion significantly goes down in the second half of trading session, when the arbitrageurs may detect the opportunities to correct the mispricing and make prices more informative.

Overall, this echos with our previous argument that investor disagreement is much larger at market open and is gradually resolved over trading period. This cycling pattern over the days naturally leads to, as suggested by our model, the gap between overnight and intraday returns. As the intraday beta dispersion is reduced more significantly for high-volume stocks, it is natural to expect a high-volume overnight permium and intraday discount.

5 Economic Interpretations

5.1 Time-Varying Performance across Market Regimes

The dynamic nature of market movements naturally leads to a time-varying return-volume relation, as trading volume is a first-order outcome of the stock market. As mentioned in the previous text, the heterogeneous return-volume relation in night and day is directly related to investor heterogeneity. More specifically, we care about uninformed/noise traders' deviation from full information and rationality.

This subsection considers two relevant sources of market-wide forces: investor sentiment and economic policy uncertainty. Investor sentiment captures the (over)confidence of less-informed investors, a proxy for aggregate expectation bias. Economic policy uncertainty brings up investor disagreement on the future of asset fundamentals, which may amplify the mispricing.

Empirically, we care whether there is a significant difference in return-volume relation across different market regimes, proxied by sentiment or uncertainty. Following Stambaugh et al. (2015), for each of the 10 volume-sorted portfolio, we perform a two-state time-series regression as below:

$$R_{p,t} = a_H d_{H,t} + a_L d_{L,t} + \gamma' \mathbf{FF5}_t + \varepsilon_{p,t} \tag{5.1}$$

where $R_{p,t}$ is the excess return (close-to-close, overnight, intraday) of portfolio p. $\mathbf{FF5}_t$ is the vector of Fama-French 5 factors realized at time t. $d_{H,t}$ and $d_{L,t}$ are high/low-regime indicators of certain market force (sentiment or uncertainty). $d_{H,t}$ is set to 1 when the certain index in month $t-1$ is higher than its sample median. Similar logic applies to $d_{L,t}$ as low-regime indicator. Therefore, we can interpret a_H and a_L as the FF5-adjusted return of portfolio p in the high and low regime of market force. We mainly care about the difference $a_H - a_L$, which demonstrates the differential return predictability of trading volume between periods of high and low market regimes. Table V shows the results.

[Insert Table V here]

Investor sentiment. Dumas et al. (2009) show that rational investors (with proper beliefs) are more likely to profit from overconfident investors who incur high sentiment (excess movement) in the market, which may amplify the mispricing. We construct a monthly index of investor sentiment using Survey of American Association of Individual Investors (AAII). AAII survey provides the percentage of investors' opinions (bull, neutral, bear) on the future performance of stock market every week. For each month, we average the weekly difference in the percentage difference between 'bull' and 'bear' investors. A higher value of index means investors are more inclined to a "bull" opinion, i.e., higher sentiment.

Panel A in Table V presents the results related to investor sentiment. For close-to-close returns, the predictability of trading volume is weak universally over time. In contrast, the high-volume overnight premium and intraday discount are much more pronounced following the high sentiment periods. Concretely, the monthly FF5-adjusted H-L spread in overnight is 3.12% (t-stat: 9.20) in high-sentiment periods, significantly larger than its equivalent of 1.58% (t-stat: 6.03) in low sentiment periods. The significant difference between the two periods is almost entirely from the portfolio with highest decile of volume (1.42%, t-stat: 3.23), while this difference remains insignificant for the portfolio with lowest decile. A similar conclusion can be made for high-volume intraday discount, where the FF5-adjusted H-L spread differs by -1.64% (t-stat: -3.46) between high and low sentiment periods, primarily from the leg of high-volume stocks (difference: -1.56%, t-stat: -3.37).

It is worthwhile to note that, although both high-volume overnight premium and intraday discount are both more pronounced after high sentiment periods, only results of intraday returns are (to some extent) aligned with sentiment-induced overpricing. To see this, Stambaugh et al. (2012) argues that the short-leg of an anomaly should be more profitable due to short-selling constraints that makes mispricing more difficult to be corrected. Based on our baseline findings, high-volume stocks indeed serve as the short leg for strategies on intraday returns, while appear to be the long leg for overnight returns. This may also relate to that "volume amplifies mispricing" (Han et al., 2022) is primarily for returns realized in intraday period[20].

Economic policy uncertainty. Intuitively, late-informed investors in low-uncertainty economy are more likely to be overconfident as they rely more on current information in hand (e.g., prices), which is the subset of information that early-informed investors have. Such stronger overconfidence may lead to worsened mispricing, especially overpricing in market open. To identify time-varying uncertainty, we employ Economic Policy Uncertainty (EPU) index (Baker et al., 2016), derived from the text in main-

[20]We elaborate relevant results later in the section of 'Alternative Explanations'.

stream U.S. published newspapers. A higher value of index indicates higher magnitude of uncertainty.

Panel B in Table V presents the results related to economic uncertainty. Consistent with our prediction, following low economic uncertainty periods, the high-volume overnight premium (intraday discount) is much stronger. For overnight returns, a significant monthly difference of 1.54% (t-stat: 3.73) is documented for FF5-adjusted spread of the H-L portfolio between periods following low uncertainty (3.09%, t-stat: 9.77) and high uncertainty (1.54%, t-stat: 5.40). Similar as results on sentiment, the primary source of such difference in overnight returns between two periods are from the high-volume leg, with a monthly FF5-adjusted difference of -1.24% (t-stat: -2.82) for the portfolio with highest decile of volume. We can argue similarly for time-variation in strategy of intraday returns. The FF5-adjusted H-L spread in intraday period is 0.97% (t-stat: 2.03) larger in magnitude, following periods of low uncertainty (-2.56%, t-stat: -7.36) compared to those with high uncertainty (-1.59%, t-stat: -4.51).

Overall, we show that high-volume overnight premium and intraday discount, or return-volume relations in night and day, are significantly stronger following the periods when investors are more bullish on the future market performance, or faced with less economic uncertainty.

5.2 Role of Retail Trading

In this subsection, we test the prediction that a larger mass of uninformed investors generates a larger overnight premium (intraday discount), due to the overestimation of their signal accuracy. Individual (retail) investors tend to underperform when they trade too much (Barber and Odean, 2000). A possible reason for the (excessively) high trading volume is overconfidence due to self-attribution bias (Gervais and Odean, 2001). The following empirical test will revisit the return predictability of trading volume by focusing on its component of retail investors, who are more likely to be uninformed and overconfident.

The total share volume of retail trades, identified by the algorithm of Boehmer et al. (2021), is obtained from from WRDS Intraday Indicators database[21], which is available from October 2006 to December 2021. Similar to the definition of trading volume, we define *retail volume* as the average monthly retail turnover (shares of retail trades divided by shares outstanding) in the past 3 months. Variable *non-retail volume* is just the difference between trading volume and retail volume. We perform Fama-MacBeth regressions on *retail volume* and *non-retail volume* to see how retail trading contributes to the high-volume overnight premium and intraday discount. In columns (1)(4)(7) of Table VI, we set the benchmark by regressing excess returns (CTC, overnight, and intraday) on trading volume, with control variables included. The results are highly consistent with what we find in baseline results in Table IV, with quite similar implied Sharpe ratios in overnight ($5.77/\sqrt{15} = 1.49$) and intraday ($-4.29/\sqrt{15} = -1.11$).

[21] (Millisecond) Intraday Indicators aggregate "(daily) monthly product" of TAQ data to the stock-day level for a wide range of trading-related variables.

[Insert Table VI here]

Our main focus is columns (2)(5)(8) of Table VI. For close-to-close returns, the relation between future returns and retail trading is rather weak. The role of retail trading is much sharpened when predicting overnight and intraday returns. For overnight returns, the positive coefficient of trading volume is driven by that of retail volume (0.44, t-stat: 6.06), while the predictability of non-retail trading is significantly negative (-0.10, t-stat: -2.23). The results point to the importance of activity of the retail investors in contributing high-volume overnight premium, echoing our model prediction that the magnitude of premium is increasing with the mass of uninformed traders, proxied with retail investors here. On the other hand, the significantly opposite impact of retail and non-retail trading somehow provides a way to distinguish the investor clientele that is considered to cause opposing price pressure as in the 'tug-of-war' story (Lou et al., 2019). For intraday returns, similarly, the negative predictability of trading volume also stems from retail volume (-0.36, t-stat: -3.33), i.e. a higher level of retail trading enlarges the high-volume intraday discount. However, the impact of non-retail trading on intraday returns is silenced. In columns (3)(6)(9) of Table VI, we consider the industry effects as in baseline Fama-MacBeth results. It turns out that results remain nearly unchanged. Therefore, the trading activity of retail investors is an important driver of high-volume overnight premium (intraday discount), which is not an industry-specific result.

6 Alternative Explanations

In this section, we consider several potential explanations on why trading volume are oppositely related to expected overnight and intraday returns. We show that these mechanisms are at best partial explanations and trading volume appears to be a predictor superior in predictability.

6.1 Tug-of-War: Persistence of Overnight and Intraday Returns

Lou et al. (2019) empirically shows that expected overnight (intraday) returns are positively predicted by current overnight (intraday) returns, while negatively predicted by intraday (overnight) returns. They attribute such 'self-momentum, cross-reversal' effect to the 'tug-of-war' of specific demand from different investor clienteles. We show in this subsection that the existence of high-volume overnight premium and intraday discount is independent to the past realizations of overnight/intraday returns.

As for the empirical exercise, we double sort stocks into 25 portfolio independently based on quntiles of trading volume and past overnight/intraday returns. We compute the value-weighted average overnight/intraday returns within each portfolio and plot their time-series average in Figure 8.

[Insert Figure 8 here]

The top-left panel of Figure 8 exhibits the future overnight returns of 5×5 portfolios sorted by trading volume and current overnight returns. We can see that the positive self-predictability of overnight returns holds in each volume quintile. More importantly, regardless of the five quntiles of current overnight return (bars with same color), there is always a monotonic relation between future overnight returns and trading volume. The monthly overnight spread between extreme volume quintiles is very significant, ranging from 1.08% to 1.87% among quintiles of current overnight returns. The bottom-left panel plots the average overnight returns of portfolios sorted by trading volume and current intraday returns. Aligned with 'tug-of-war' story, within each volume quintile, the future overnight returns are negatively predicted by of current intraday returns. Meanwhile, positive overnight return-volume relation still persists given a certain quintile of current intraday returns. As for the spread between extreme volume quintiles, the corresponding overnight returns vary from 1.32% to 2.30%, across intraday return quintiles.

The right side of Figure 8 provides evidence on the universal existence of the negative relation of trading volume with future intraday returns. The bottom-right panel of Figure 8 first confirm the positive self-predictability of intraday returns across volume quintiles. Similarly, trading volume is negatively related to future expected returns given any quintiles of current intraday returns, among which the the monthly intraday spreads between extreme volume quintiles vary from -1.03% to -1.97%. Similar arguments can be given to expected intraday returns of portfolios sorted by trading volume and overnight returns, consistent with the logic of 'tug-of-war' story, with significant spreads between extreme volume deciles (from -1.35% to -2.30%).

Overall, trading volume captures a different dimension on the dynamics of expected overnight and intraday returns, distiniguished from the 'tug-of-war' implications (persistence of night/day returns).

6.2 Systematic Risk (Betas) and Mispricing (Alphas)

We show below that the predictability of trading volume on the cross-section of overnight and intraday returns is not subsumed by two popular explanations in the previous literature: systematic risk (Hendershott et al., 2020) and mispricing (Bogousslavsky, 2021). In addition, we show that these alternative explanations are more concentrated in the group of high-volume stocks, especially stronger for risk (mispricing)-based measures when predicting overnight (intraday) returns.

Systematic risk. Hendershott et al. (2020) establish the robust evidence that cross-sectional risk-return tradeoff is strongly positive (negative) in overnight and intraday, explaining a "flat" security market line when using CTC returns. [22] Specifically, portfolios are sorted into quintiles of market beta.[23] In online appendix, we document that the overnight returns of high-beta firms (quintile group with highest market beta) outperform the low-beta ones by a monthly average of 1.46% (t-stat: 4.86). This number becomes -1.20% (t-stat: -3.32) for intraday results. These results are robust to adjustment of Fama and French (2015) five-factors (FF5), Hou et al. (2015) investment-based four factors (q^4), and Stambaugh and Yuan (2017) mispricing four factors (SY4).

We now explore whether trading volume plays a role in risk-return relation. At the beginning of each month, the five-by-five portfolios are formed based on the independent double sort on trading volume and market beta. All portfolios are value-weighted and rebalanced at the beginning of next month. We refer to $\beta_{\text{high}} - \beta_{\text{low}}$ spread as the spread between the portfolios with the highest and lowest quintile of market beta, within a certain quintile of trading volume. The magnitude of H-L β is interpreted as the strength of risk-return tradeoff under this setting.

[Insert Table VII here]

Interstingly, shown in Panel A of Table VII, the relation of CTC returns with the systematic risk turns out to be significantly positive, concentrated in the stocks with higher trading volume. $\beta_{\text{high}} - \beta_{\text{low}}$ spread (in CTC returns) is rather weak for both high and low-volume stocks, even though the difference (H-L: $\beta_{\text{high}} - \beta_{\text{low}}$) is marginally positive (0.80%, t-stat: 1.88). Appealingly, the relation is substantially stronger in overnight period for high-volume stocks, where the H-L β spread is 1.28% (t-stat: 4.07). This is significantly larger than the analogue of low-volume stocks, with a difference of 1.03% (t-stat: 3.00). The results are robust to the adjustment of FF5, q^4 and SY4 models. A quick conclusion is that *high trading volume strengthens risk-return tradeoff* (*in overnight*). To understand this, we can interpret trading volume as the speed that information is incorporated into asset prices. Note that public information explains more idiosyncratic variation in overnight returns (Boudoukh et al., 2019). These might help explain why high-volume stocks are compensated more in overnight.

For intraday period, there is no significant difference (-0.38%, t-stat: 1.10) on the $\beta_{\text{high}} - \beta_{\text{low}}$ spread between high and low-volume stocks, even though there is a marginally negative tradeoff (-0.65%, t-stat: -1.79) in high volume group. In fact, the results of double sort suggest that trading volume actually subsumes the power of market beta to explain variation in the cross-section of intraday returns. Concretely, the high-volume intraday discount remains very strong among any quintile of market beta.

Mispricing. Mispricing is well-corrected in overnight period but worsens in the intraday, especially at the end of the trading session (Bogousslavsky, 2021). It simply implies more underpriced stocks have

[22] We confirm this fact by showing the univariate sort results on market betas in online appendix.

[23] Recall in section 3.1, market beta is estimated by CAPM model using CTC returns in a 60-month rolling window. This is different from Hendershott et al. (2020), where they estimate β using daily overnight returns.

better (worse) performance in overnight (intraday) period. [24] We sort stocks into quintiles of mispricing measure $MISP$[25] constructed by Stambaugh et al. (2015). In terms of CTC returns, the UMO spread is on average 0.66% (t-stat: 2.65) per month. This mispricing primarily comes from the intraday period (1.14%, t-stat: 4.30) and is significantly corrected (-0.61%, t-stat: -4.91) in overnight with a smaller magnitude.

Similarly, it is natural to see how trading volume affects mispricing in overnight and intraday, which is indeed an extension to Han et al. (2022). We form five-by-five portfolios based on independent sort on trading volume and mispricing measure $MISP$. Panel B of Table VII presents the results. There is no substantial mispricing (UMO: -0.16%, t-stat: 0.72) for low-volume stocks. Consistent to the main finding in Han et al. (2022), for the quintile group with the highest volume, underpriced stocks outperform significantly than overpriced stocks (UMO: 1.82%, t-stat: 5.03). Mispricing is enlarged by 1.99% (t-stat: 5.46) per month in the high-volume group, i.e. *trading volume amplifies mispricing* in Han et al. (2022). Importantly, *the volume amplification effect on mispricing principally arises from the intraday period*, when the difference in UMO spread is 2.24% (t-stat: 5.96) per month between the high and low volume groups. Particularly, this is dominated by the leg of overpriced stocks. This effect is robust to adjustment of leading factor models (FF5, q^4, and SY4) and its significance persists for at least 24 months after portfolio formation. Meanwhile, mispricing is corrected in the overnight period by a much marginal magnitude (-0.39%, t-stat: 1.78) and becomes insignificant after adjustment of factor models.

One may argue that $MISP$, as an average rank, might cancel out the noise via information aggregation. We repeat the exercise above by replacing $MISP$ with firm characteristics[26] of anomalies studied in Stambaugh et al. (2012). For simplicity, we only report the UMO spread (in CTC, overnight, intraday) of the quintile group with the highest (high volume: UMO), the lowest (low volume: UMO) volume, and their difference (H-L:UMO). We show in Table VIII that the volume amplification effect exists in the majority (7 out of 11) of mispricing-related anomalies, all of which originate from the effect in intraday period. The intraday effects of these anomalies exhibit substantial economic significance, from 0.80% (net operating assets) to 2.30% (failure probability), and *mostly stem from the high volume stocks*. The only exception is momentum, which reduces mispricing for low-volume stocks in intraday period. This may be explained by investors' underreaction to information from other investors' trade (Barardehi et al., 2023). Noticeably, 4 anomalies also significantly mitigates mispricing in overnight period, from 0.59% (O Score) to 1.19% (failure probability).

[24] This finding is reconfirmed in our data sample in Online Appendix.

[25] $MISP$ is a percentile rank averaging over 11 FF3-robust asset pricing anomalies. For each anomaly, each firm-month observation is assigned with a percentile rank, in that a higher (lower) rank is linked to a lower (higher) average abnormal return in line with previous literature, i.e. overpricing (underpricing).

[26] Details on constructing 11 characteristics can be referred to Appendix of Stambaugh and Yuan (2017). We replicate Ohlson's Oscore using data in Annual Fundamentals from WRDS Compustat. The remaining characteristics are from direct download in open source asset pricing.

[Insert Table VIII here]

Why does volume amplify mispricing in intraday period? At the close of market, arbitrageurs usually reduce their position to trade on mispricing due to institutional constraints and overnight risk (Bogousslavsky, 2021). Therefore, the price movements are more impacted by speculation reasons (Hong and Wang, 2000), which is more aggressive for stocks with high disagreement, potentially proxied by high trading volume (Han et al., 2022). Sufficiently large disagreement and the short-selling constraint will thus lead to overpricing.

6.3 Day-of-the-Week Effect

Anomalies are heterogeneous in performance on different days of the week, probably due to time-varying investor mood (Birru, 2018). Bogousslavsky (2021) also shows that mispricing worsens most at the end of trading sessions on Fridays, then mitigated in the overnight period preceding Mondays. Based on these arguments, checking whether high-volume overnight premium (intraday discount) also differentiates across weekdays is of necessity. We repeat the univariate portfolio sort in subsection 3.2 and report the overnight and intraday returns on specified weekdays accumulated in a month.

[Insert Figure 9 here]

Figure 9 presents the results. Across all weekdays, average portfolio overnight (intraday) returns monotonically increase (decrease) with trading volume. At the right of the figure, we also plot the corresponding average spread of the high-minus-low (H-L) trading volume portfolio on corresponding weekdays. H-L spreads in night and day are highly significant across weekdays, except for intraday returns on Wednesdays (H-L spread: -0.16%, t-stat: -1.54). Surprisingly, Tuesdays are the most pronounced days for the return-volume relation. The monthly overnight H-L spread on Tuesdays is 0.68% (t-stat: 6.48) on average, accounting for 28% of the overnight premium. As for intraday period, the average H-L spread on Tuesdays is -0.71% (t-stat: -6.77), making up 31% of intraday discount. Results above are still mainly due to high-volume stocks. Therefore, the high-volume overnight premium and intraday discount are not driven by certain day-of-the-week. In other words, the main finding can't be interpreted as a calendar/seasonality effect.

7 Conclusions

As the fundamental dual in asset pricing, the returns and trading volume are documented with a mixed (even weak) relation empirically. We disentangle this issue by uncovering its heterogeneity in overnight and intraday periods. We document a high-volume overnight premium (intraday discount): in the cross-section of stocks, overnight (intraday) returns in the future are positively (negatively) related to trading volume. This result is remarkably large in magnitude and highly robust, in worldwide equity markets and persistent for at least three years. Our result is also not explained by leading explanations of the cross-sectional behavior of overnight and intraday returns, such as persistent component of night/day returns, systematic risk and mispricing.

We explain this robust premium/discount with a model of heterogeneous investors. The key spirit is that uninformed investors, usually late-informed overestimate their signal precision relative to informed ones, who are early-informed of cash-flow news at market open. This points to the larger dispersion in investor beliefs as market open. With sufficient belief dispersion, overreacting uninformed investors lead to both high volume and overpricing at the market open. Mispricing is corrected by early-informed investors (as an arbitrageur), in a cyclic fashion, at market close. Thus, high volume is associated with a high (low) overnight intraday return. We test and empirically support the model implications, especially documenting that the cross-sectional dispersion in systematic risk is decreasing over the trading day. This phenomenon is especially strong for the high-volume stocks. Meanwhile, the retail trading activity dominates the direction of volume's relation with expected returns in night and day, consistent to previous literature.

References

Aleti, S. (2022). The high-frequency factor zoo. *Duke University*.

Andersen, T. G., Riva, R., Thyrsgaard, M., and Todorov, V. (2023). Intraday cross-sectional distributions of systematic risk. *Journal of Econometrics*, 235(2):1394–1418.

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131(4):1593–1636.

Barardehi, Y. H., Bogousslavsky, V., and Muravyev, D. (2023). What drives momentum and reversal? evidence from day and night signals. Technical report, Available at SSRN 4069509.

Barber, B. M. and Odean, T. (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. *The Journal of Finance*, 55(2):773–806.

Berkman, H., Koch, P. D., Tuttle, L., and Zhang, Y. J. (2012). Paying attention: overnight returns and the hidden cost of buying at the open. *Journal of Financial and Quantitative Analysis*, 47(4):715–741.

Birru, J. (2018). Day of the week and the cross-section of returns. *Journal of Financial Economics*, 130(1):182–214.

Boehmer, E., Jones, C. M., Zhang, X., and Zhang, X. (2021). Tracking retail investor activity. *The Journal of Finance*, 76(5):2249–2305.

Bogousslavsky, V. (2021). The cross-section of intraday and overnight returns. *Journal of Financial Economics*, 141(1):172–194.

Boudoukh, J., Feldman, R., Kogan, S., and Richardson, M. (2019). Information, trading, and volatility: Evidence from firm-specific news. *The Review of Financial Studies*, 32(3):992–1033.

Campbell, J. Y., Grossman, S. J., and Wang, J. (1993). Trading volume and serial correlation in stock returns. *The Quarterly Journal of Economics*, 108(4):905–939.

Chen, A. Y. and Zimmermann, T. (2022). Open source cross-sectional asset pricing. *Critical Finance Review*, 27(2):207–264.

Chordia, T. and Swaminathan, B. (2000). Trading volume and cross-autocorrelations in stock returns. *The Journal of Finance*, 55(2):913–935.

Conrad, J. S., Hameed, A., and Niden, C. (1994). Volume and autocovariances in short-horizon individual security returns. *The Journal of Finance*, 49(4):1305–1329.

Daniel, K., Hirshleifer, D., and Subrahmanyam, A. (1998). Investor psychology and security market under-and overreactions. *the Journal of Finance*, 53(6):1839–1885.

Dumas, B., Kurshev, A., and Uppal, R. (2009). Equilibrium portfolio strategies in the presence of sentiment risk and excess volatility. *The Journal of Finance*, 64(2):579–629.

Fama, E. F. and French, K. R. (1997). Industry costs of equity. *Journal of Financial Economics*, 43(2):153–193.

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

Fleming, J. and Kirby, C. (2011). Long memory in volatility and trading volume. *Journal of Banking & Finance*, 35(7):1714–1726.

Gao, X. and Ritter, J. R. (2010). The marketing of seasoned equity offerings. *Journal of Financial Economics*, 97(1):33–52.

Gervais, S., Kaniel, R., and Mingelgrin, D. H. (2001). The high-volume return premium. *The Journal of Finance*, 56(3):877–919.

Gervais, S. and Odean, T. (2001). Learning to be overconfident. *The Review of Financial Studies*, 14(1):1–27.

Griffin, J. M., Kelly, P. J., and Nardari, F. (2010). Do market efficiency measures yield correct inferences? a comparison of developed and emerging markets. *The Review of Financial Studies*, 23(8):3225–3277.

Grossman, S. J. and Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American economic review*, 70(3):393–408.

Gu, M., Hu, Y., and Xiong, Z. (2023). Dissecting the lottery-like anomaly: Evidence from china. *Available at SSRN 4433510*.

Han, Y., Huang, D., Huang, D., and Zhou, G. (2022). Expected return, volume, and mispricing. *Journal of Financial Economics*, 143(3):1295–1315.

Hendershott, T., Livdan, D., and Rösch, D. (2020). Asset pricing: A tale of night and day. *Journal of Financial Economics*, 138(3):635–662.

Hirshleifer, D. (2020). Presidential address: Social transmission bias in economics and finance. *The Journal of Finance*, 75(4):1779–1831.

Hirshleifer, D., Subrahmanyam, A., and Titman, S. (1994). Security analysis and trading patterns when some investors receive information before others. *The Journal of finance*, 49(5):1665–1698.

Hong, H. and Wang, J. (2000). Trading and returns under periodic market closures. *The Journal of Finance*, 55(1):297–354.

Hou, K., Xue, C., and Zhang, L. (2015). Digesting anomalies: An investment approach. *The Review of Financial Studies*, 28(3):650–705.

Hvidkjaer, S. (2008). Small trades and the cross-section of stock returns. *The Review of Financial Studies*, 21(3):1123–1151.

Israeli, D., Kaniel, R., and Sridharan, S. A. (2022). The real side of the high-volume return premium. *Management Science*, 68(2):1426–1449.

Jensen, T. I., Kelly, B., and Pedersen, L. H. (2023). Is there a replication crisis in finance? *The Journal of Finance*, 78(5):2465–2518.

Johnson, D. D. and Fowler, J. H. (2011). The evolution of overconfidence. *Nature*, 477(7364):317–320.

Jones, C. S., Pyun, S., and Wang, T. (2022). Return extrapolation and day/night effects. Technical report, University of Southern California.

Kaniel, R., Ozoguz, A., and Starks, L. (2012). The high volume return premium: Cross-country evidence. *Journal of Financial Economics*, 103(2):255–279.

Lee, C. M. and Swaminathan, B. (2000). Price momentum and trading volume. *The Journal of Finance*, 55(5):2017–2069.

Llorente, G., Michaely, R., Saar, G., and Wang, J. (2002). Dynamic volume-return relation of individual stocks. *The Review of financial studies*, 15(4):1005–1047.

Lou, D., Polk, C., and Skouras, S. (2019). A tug of war: Overnight versus intraday expected returns. *Journal of Financial Economics*, 134(1):192–213.

Lu, Z., Malliaris, S., and Qin, Z. (2023). Heterogeneous liquidity providers and night-minus-day return predictability. *Journal of Financial Economics*, 148(3):175–200.

Luo, J., Subrahmanyam, A., and Titman, S. (2021). Momentum and reversals when overconfident investors underestimate their competition. *The Review of Financial Studies*, 34(1):351–393.

Nagel, S. (2012). Evaporating liquidity. *The Review of Financial Studies*, 25(7):2005–2039.

Odean, T. (1998). Volume, volatility, price, and profit when all traders are above average. *The Journal of Finance*, 53(6):1887–1934.

Pan, L., Tang, Y., and Xu, J. (2016). Speculative trading and stock returns. *Review of Finance*, 20(5):1835–1865.

Qiao, K. and Dam, L. (2020). The overnight return puzzle and the "t+ 1" trading rule in chinese stock markets. *Journal of Financial Markets*, 50:100534.

Saadon, Y. and Schreiber, B. Z. (2023). Newspapers tone and the overnight-intraday stock return anomaly. *Journal of Financial Markets*, page 100838.

Stambaugh, R. F., Yu, J., and Yuan, Y. (2012). The short of it: Investor sentiment and anomalies. *Journal of Financial Economics*, 104(2):288–302.

Stambaugh, R. F., Yu, J., and Yuan, Y. (2015). Arbitrage asymmetry and the idiosyncratic volatility puzzle. *The Journal of Finance*, 70(5):1903–1948.

Stambaugh, R. F. and Yuan, Y. (2017). Mispricing factors. *The Review of Financial Studies*, 30(4):1270–1315.

Sun, K., Wang, H., and Zhu, Y. (2023). Salience theory in price and trading volume: Evidence from china. *Journal of Empirical Finance*, 70:38–61.

Wang, J. (1994). A model of competitive stock trading volume. *Journal of Political Economy*, 102(1):127–168.

Wang, Z. (2021). The high volume return premium and economic fundamentals. *Journal of Financial Economics*, 140(1):325–345.

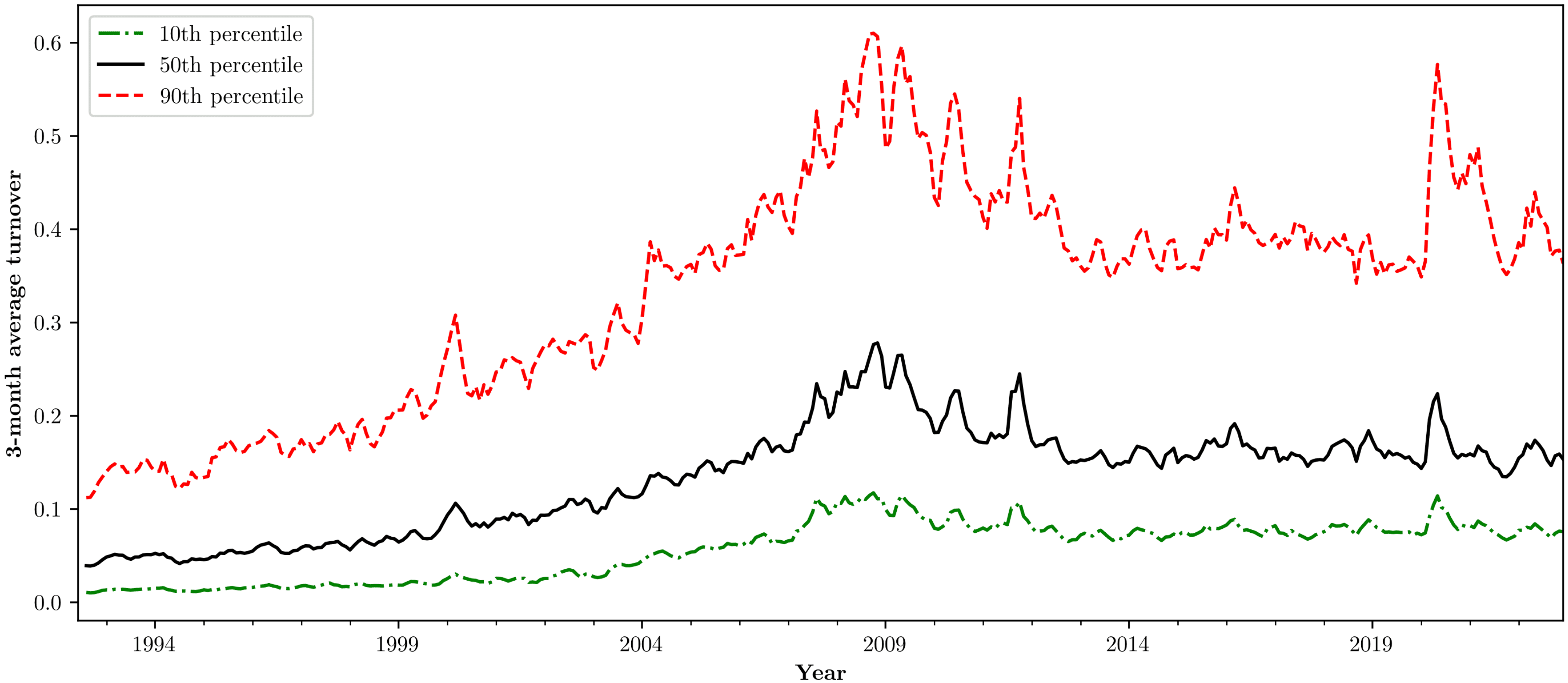


Figure 2 Cross-sectional distribution of trading volume over time

This figure plots time series of empirical percentiles (10%, 50%, 90%) of the cross-sectional distribution of trading volume. Trading volume is defined as average turnover (share volume to shares outstanding) in past 3 months. The sample period covers from June 1992 to December 2022, excluding stocks with month-end prices lower than \$5 and market equity lower than 20th NYSE breakpoints at portfolio formation.

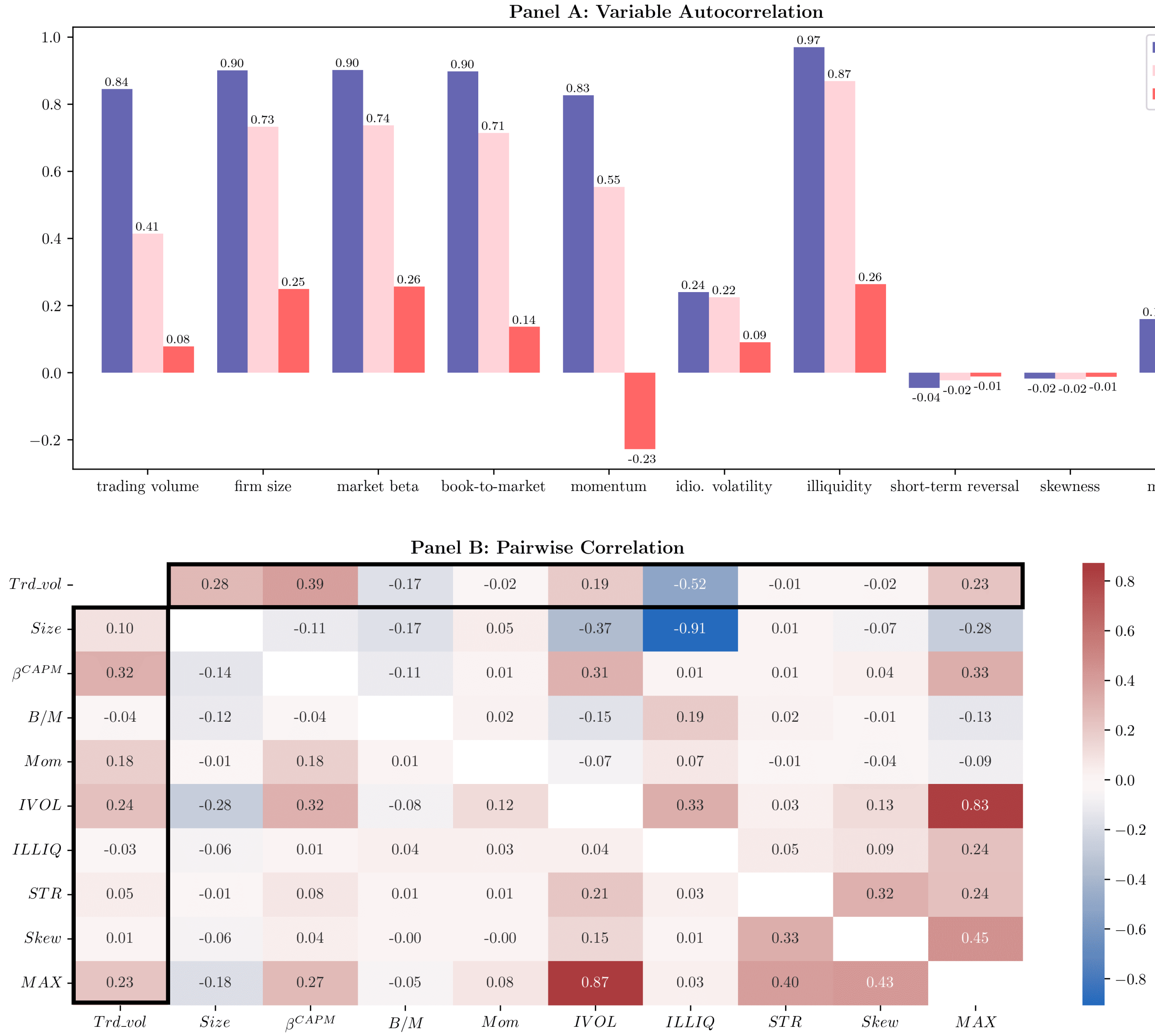


Figure 3 Summary Statistics

This figure plots the summary statistics of key variables used in this paper. Detailed definitions can be seen in Table A.I. Panel A plots the autocorrelation with lags of 1 (monthly), 3 (quarterly) and 12 (yearly). Panel B presents the pairwise correlations of these variables. Numbers below (above the diagonal) are correlations based on Pearson's (Spearman's) method.

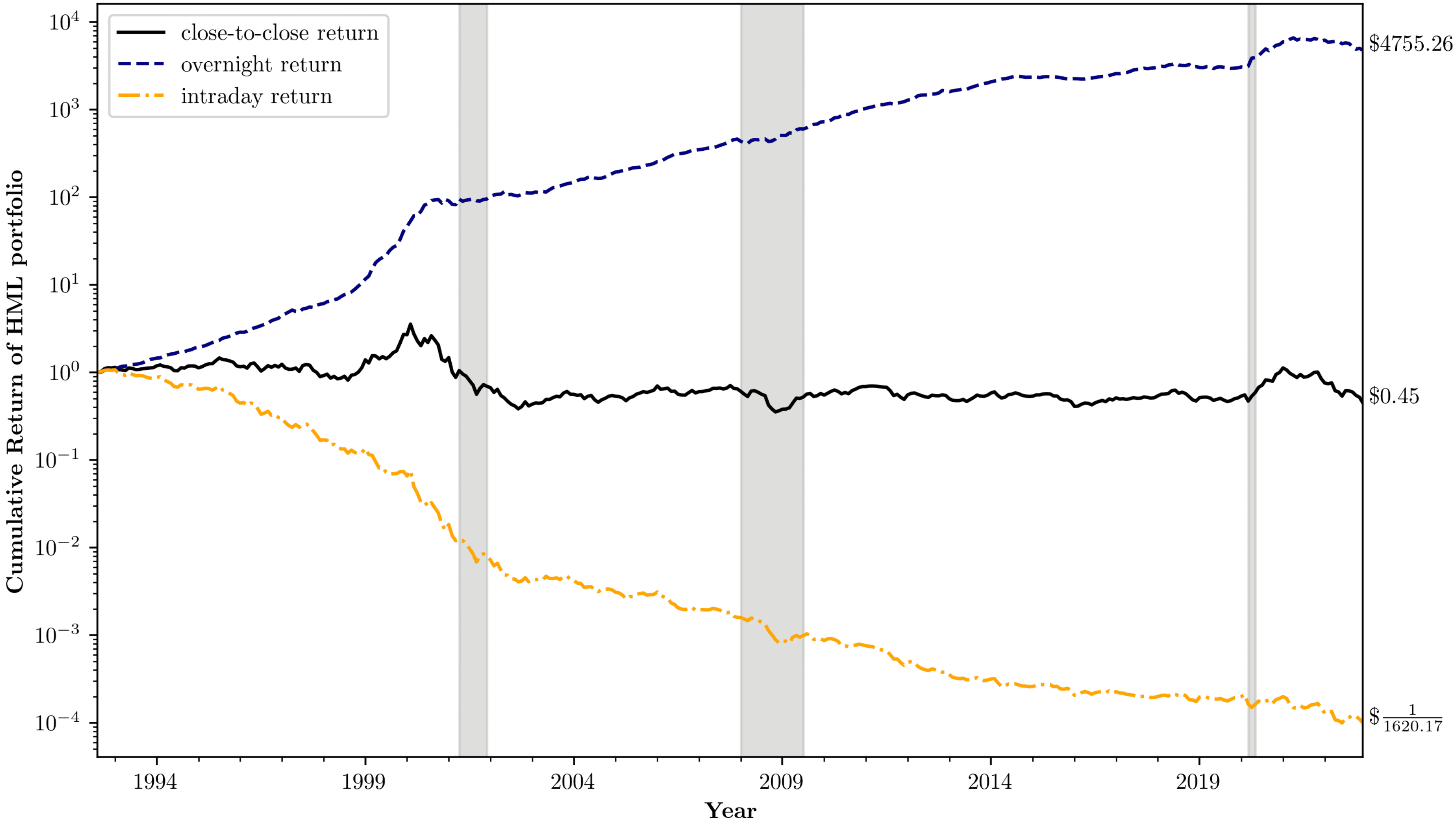


Figure 4 Cumulative performance: high-volume overnight premium and intraday discount

This figure plots the common logarithm of cumulative dollar performance (close-to-close, overnight and intraday) of high-minus-low trading volume (HML) portfolio. $1 is invested at the start of sample, which covers from July, 1992 to December, 2022. Performance realized at the end of sample for all strategies are annotated at the right of figure. NBER recession periods are indicated as grey shaded bars.

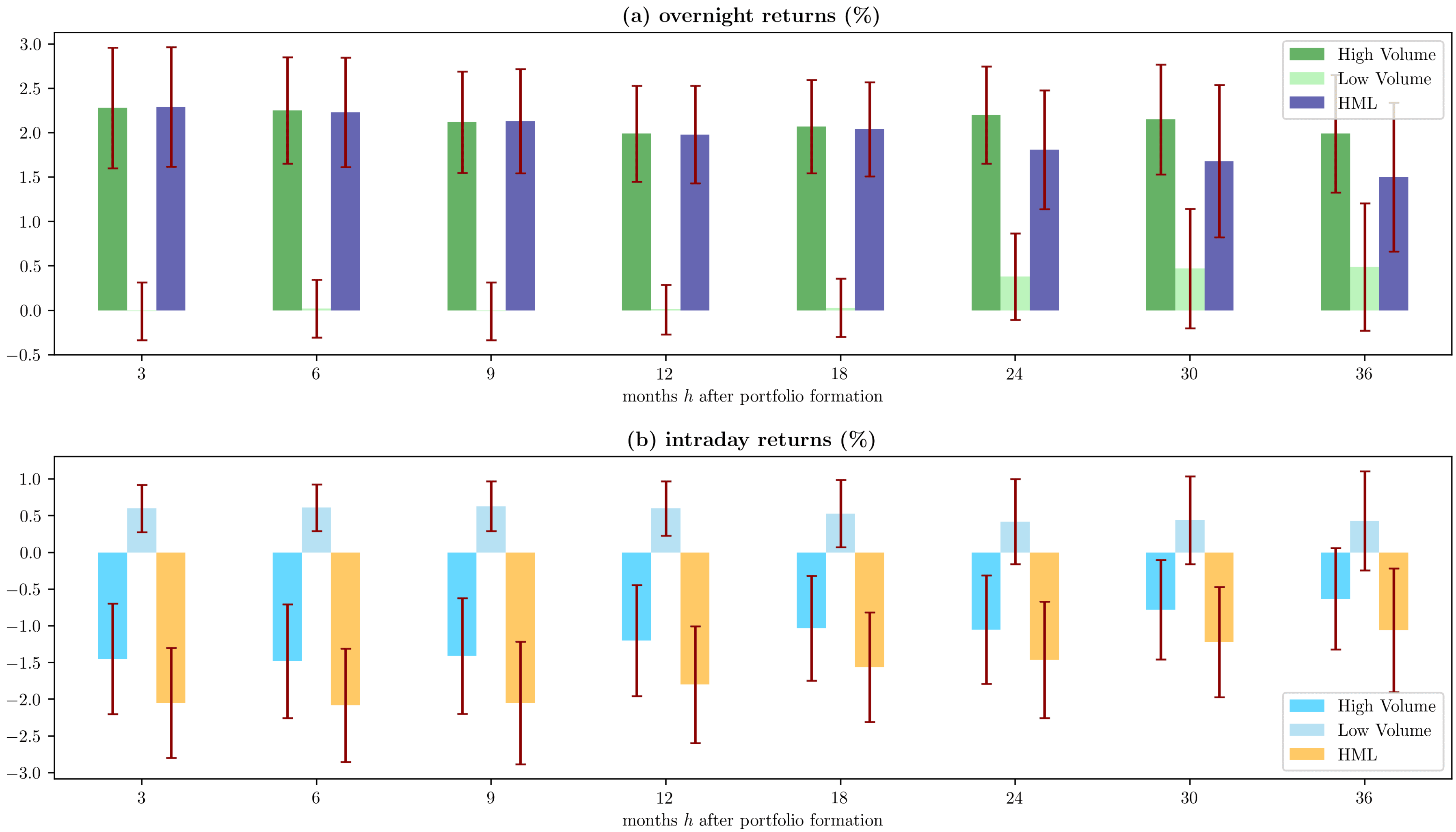


Figure 5 Duration of high-volume overnight premium and intraday discount

This figure plots average monthly returns (overnight, intraday) in future h months after portfolio formation, adjusted by Fama-French 5 factors. Portfolios are sorted into deciles by trading volume, defined as 3-month average turnover. 'High (low) volume' respectively refers to the portfolio with highest (lowest) decile in trading volume. Portfolios are value-weighted, rebalanced each month and held for 60 months. 'HML' refers to the spread of high-minus-low volume portfolio . The sample period covers from June 1992 to December 2022. 95% confidence intervals are displayed.

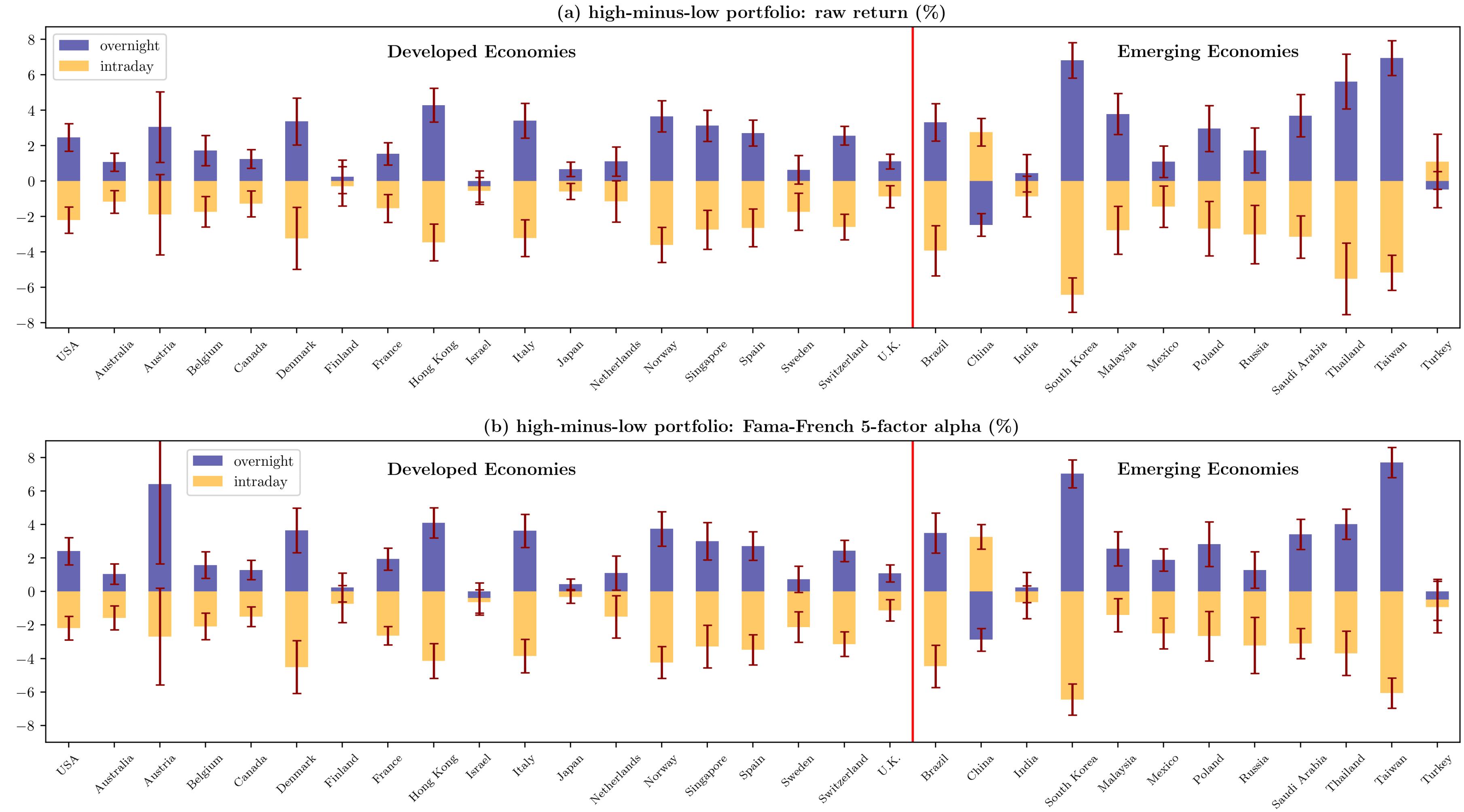


Figure 6 High-volume overnight premium and intraday discount: global evidence

This figure plots average monthly returns (overnight, intraday) of high-minus-low (HML) trading volume portfolio across global markets, adjusted by Fama-French 5 factors in respective region. Trading volume is defined as monthly turnover. Portfolios are value-weighted and rebalanced each month. For each market, we exclude stock-month observation when month-end price is lower than $1 or market equity is lower than 5th NYSE breakpoint at portfolio formation. 95% confidence intervals are displayed.

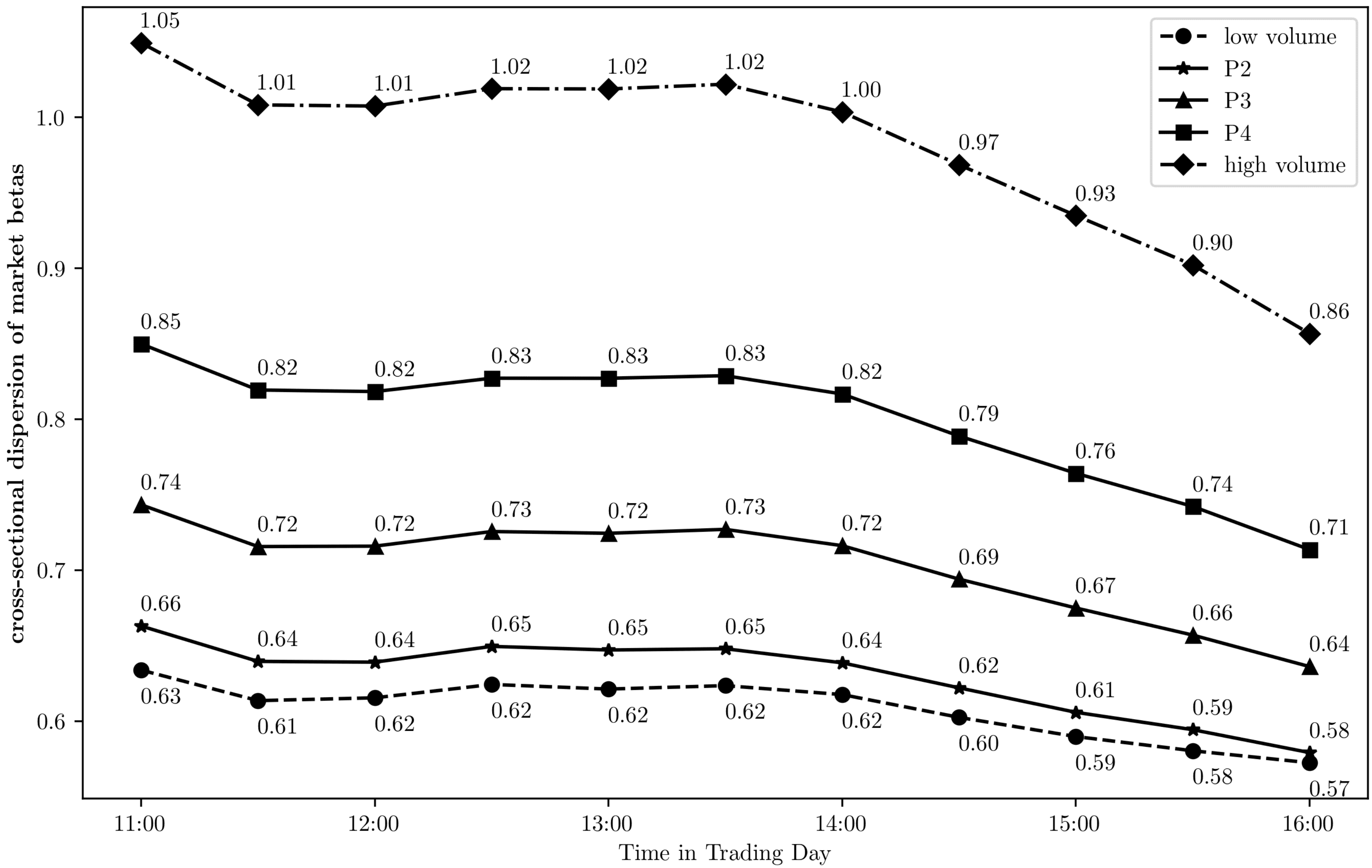


Figure 7 Time-variation in dispersion of intraday systematic risks: role of trading volume

This figure plots the dispersion of intraday systematic risk over the trading day, grouped by stocks based on quintiles of trading volume. Intraday systematic risk is computed based on CAPM model using 5-minute level returns of individual asset and market in a 90-minute rolling window. The dispersion is computed as the standard deviation of high-frequency market betas, given certain quintile of trading volume and fixed point of time in the day.

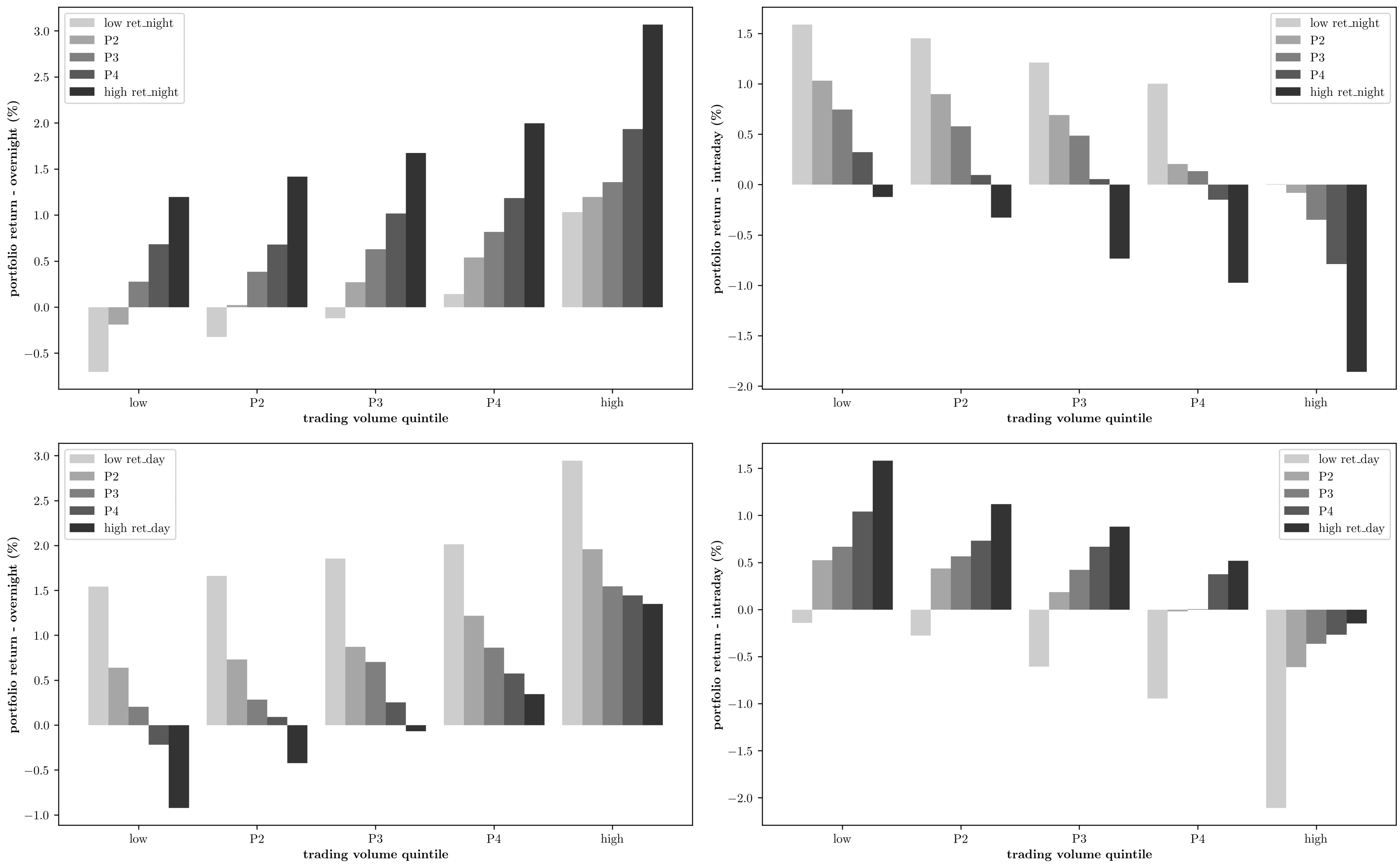


Figure 8 Portfolios sorted by trading volume and contemporaneous night/day returns

This figure plots the average monthly (overnight, intraday) returns of porfolios independently sorted by quntiles of trading volume and contemporaneous overnight/intraday returns. Portfolios are value-weighted and rebalanced each month. The sample period covers from June 1992 to December 2022, excluding stocks with month-end prices lower than $5 and market equity lower than 20th NYSE breakpoints at portfolio formation.

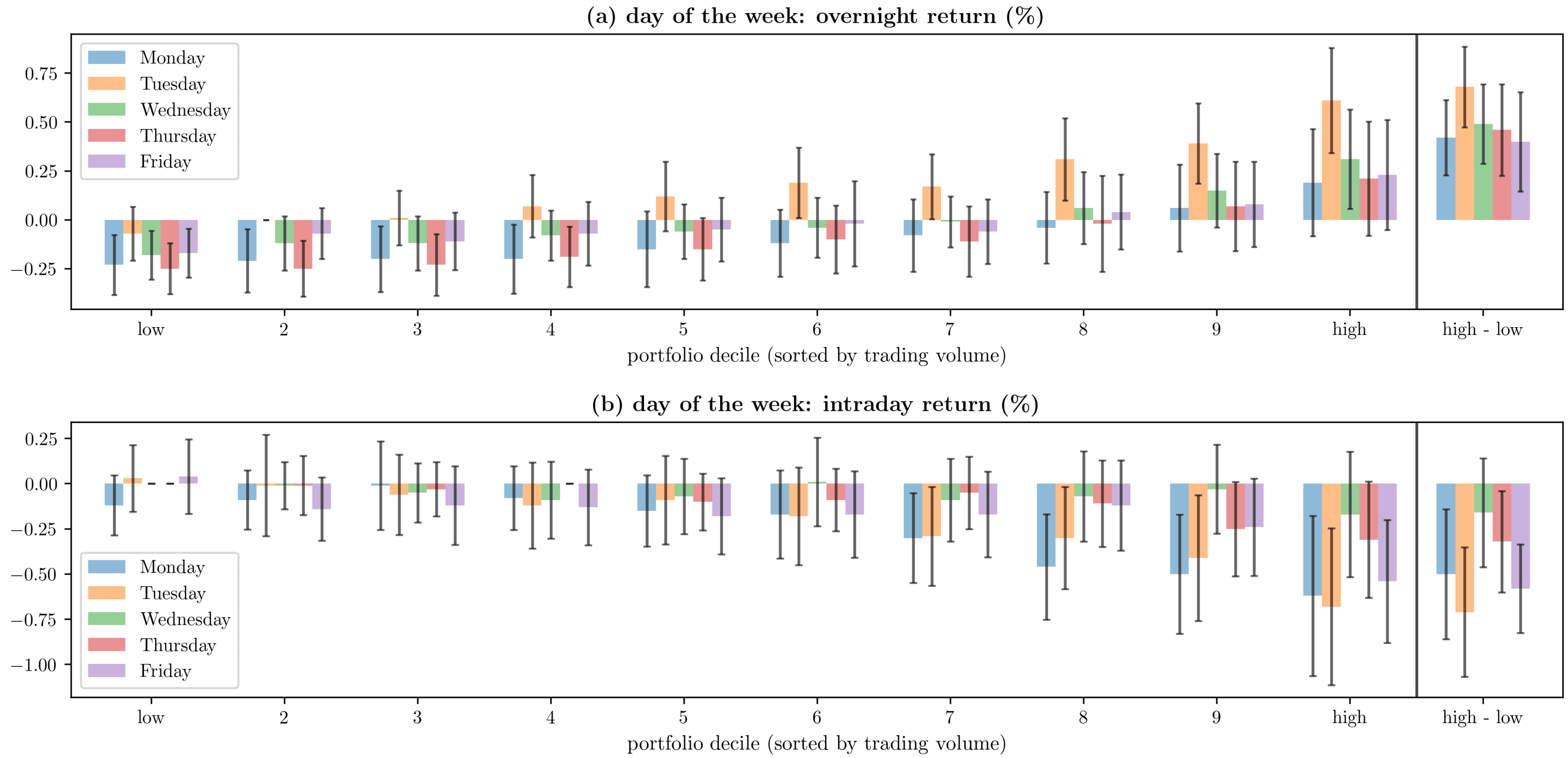


Figure 9 High-volume overnight premium & intraday discount: across days of the week

This figure plots monthly average returns (overnight, intraday), realized on the specific weekdays, of portfolios sorted into deciles of trading volume, defined as 3-month average turnover. Portfolios are value-weighted and rebalanced each month. Ths sample period covers from June 1992 to December 2022. 95% confidence intervals are displayed.

Table I Portfolios sorted by trading volume: average returns in night and day

This table reports the monthly average excess returns (close-to-close, overnight and intraday) of decile portfolios sorted by trading volume. H-L, α_{FF5} refer to the excess return and FF5-alphas of H-L (high-minus-low trading volume) portfolio. Panel (b) reports the results of industry-level demeaned trading volume. Reported in parentheses are t-statistics, computed based on Newey-West standard errors with a maximum lag of 6 months. Portfolios are value-weighted and rebalanced each month. The sample period covers from June 1992 to December 2022.

	Low	2	3	4	5	6	7	8	9	High	H-L	α_{FF5}
Panel A: Sorted by 3-month average turnover												
close-to-close	0.64	0.66	0.65	0.68	0.67	0.86	0.56	0.80	0.81	0.73	0.09	0.17
	(3.64)	(3.63)	(3.12)	(3.05)	(2.72)	(3.26)	(1.94)	(2.38)	(2.10)	(1.44)	(0.20)	(0.74)
overnight	-0.15	0.09	0.09	0.27	0.44	0.65	0.65	1.08	1.49	2.29	**2.44**	**2.35**
	(-0.92)	(0.64)	(0.57)	(1.70)	(2.34)	(3.66)	(3.32)	(4.55)	(5.46)	(5.99)	(6.44)	(6.16)
intraday	0.71	0.47	0.45	0.30	0.13	0.13	-0.17	-0.33	-0.68	-1.56	**-2.26**	**-2.09**
	(4.39)	(3.16)	(3.23)	(1.76)	(0.70)	(0.64)	(-0.80)	(-1.38)	(-2.32)	(-4.16)	(-6.18)	(-6.97)
Panel B: Sorted by 3-month average turnover, demeaned by industry												
close-to-close	0.64	0.63	0.70	0.74	0.82	0.66	0.65	0.79	0.79	0.76	0.12	0.05
	(3.09)	(2.73)	(3.11)	(3.22)	(3.86)	(2.85)	(2.65)	(2.75)	(2.13)	(1.59)	(0.35)	(0.24)
overnight	0.34	0.34	0.30	0.45	0.46	0.34	0.47	0.72	1.19	2.14	**1.80**	**1.67**
	(2.01)	(1.80)	(1.95)	(2.78)	(2.98)	(2.11)	(3.16)	(4.04)	(4.99)	(6.21)	(7.48)	(6.83)
intraday	0.21	0.20	0.32	0.21	0.27	0.21	0.10	-0.02	-0.47	-1.38	**-1.59**	**-1.54**
	(1.34)	(1.18)	(1.92)	(1.28)	(1.53)	(1.17)	(0.52)	(-0.08)	(-1.82)	(-3.84)	(-5.53)	(-6.04)

Table II **Portfolios sorted by trading volume: returns adjusted by factors realized in corresponding period**

This table reports average monthly overnight and intraday returns of portfolios across trading volume deciles. Trading volume is defined as 3-month average turnover. Portfolios are value-weighted and rebalanced each month. $\alpha_{CAPM}, \alpha_{FF5}$ are average overnight and intraday returns adjusted by market (or FF5) factors realized in corresponding periods . Reported in parentheses are t-statistics, computed based on Newey-West standard errors with a maximum lag of 6 months. The sample period covers from January 1996 to December 2020.

	Low	2	3	4	5	6	7	8	9	High	H-L
Panel A: overnight returns											
excess	-0.10	0.13	0.10	0.32	0.52	0.73	0.70	1.18	1.66	2.55	**2.65**
	(-0.54)	(0.85)	(0.55)	(1.79)	(2.43)	(3.65)	(3.27)	(4.49)	(5.51)	(5.90)	(6.13)
$\alpha_{CAPM,night}$	-0.54	-0.40	-0.54	-0.32	-0.18	0.03	-0.01	0.40	0.86	1.55	**2.09**
	(-4.29)	(-4.19)	(-5.82)	(-3.50)	(-3.13)	(0.41)	(-0.07)	(3.76)	(5.61)	(6.08)	(6.00)
$\alpha_{FF5,night}$	-0.22	-0.05	-0.14	-0.10	-0.15	0.01	-0.21	0.04	0.32	0.74	**0.96**
	(-1.74)	(-0.73)	(-2.05)	(-1.48)	(-2.30)	(0.16)	(-2.59)	(0.53)	(2.47)	(3.58)	(3.41)
Panel B: intraday returns											
excess	0.62	0.46	0.45	0.30	0.11	0.09	-0.18	-0.43	-0.77	-1.70	**-2.32**
	(3.38)	(2.92)	(2.90)	(1.55)	(0.55)	(0.41)	(-0.70)	(-1.53)	(-2.25)	(-3.98)	(-5.44)
$\alpha_{CAPM,day}$	0.55	0.39	0.38	0.22	0.02	-0.01	-0.29	-0.54	-0.90	-1.86	**-2.41**
	(3.64)	(3.56)	(4.04)	(3.08)	(0.24)	(-0.09)	(-2.98)	(-3.87)	(-5.31)	(-6.43)	(-6.39)
$\alpha_{FF5,day}$	0.23	0.00	0.03	0.10	-0.11	0.02	-0.19	-0.12	-0.35	-0.84	**-1.07**
	(1.51)	(0.00)	(0.32)	(1.11)	(-1.10)	(0.25)	(-1.89)	(-1.05)	(-3.48)	(-4.62)	(-4.38)

Table III High-volume overnight premium and intraday discount: robustness

This table reports the monthly average excess returns and FF5 alphas of high-minus-low (H-L) trading volume portfolios in overnight and intraday periods and examines its robustness. Panel A consider various trading volume measures as alternative sorting variable. Panel B considers the H-L spread controlled by correlated characterstics (definition in Table A.I) as control variables. In particular, stocks are first sorted into deciles based on one control variable, and then are independently sorted into deciles based on trading volume. We report H-L spread averaged across the ten control groups. Panel C split sample in time series into two equal halves. Panel D split the sample based on the exchange where stocks are traded. Panel E compare the performance of H-L spread inperiods of economic recession and expansions. Panel F considers alternative portfolio breakpoints and sample filters (using only microcaps, or excluding financial and utility firms). Portfolios are value-weighted and rebalanced each month. Reported in parentheses are t-statistics, computed based on Newey-West standard errors with a maximum lag of 6 months. The sample period covers from June 1992 to December 2022.

	H-L (overnight)		H-L (intraday)			H-L (overnight)		H-L (intraday)	
	Raw	α_{FF5}	Raw	α_{FF5}		Raw	α_{FF5}	Raw	α_{FF5}
Panel A: Alternative measures of trading volume									
Baseline Result	**2.44**	**2.35**	**-2.26**	**-2.09**	dollar volume	**1.00**	**0.99**	**-0.96**	**-0.86**
	(6.44)	(6.16)	(-6.18)	(-6.97)		(5.32)	(5.11)	(-4.05)	(-5.53)
6-month turnover	**2.37**	**2.26**	**-2.15**	**-1.98**	abnormal volume	**0.20**	**0.21**	0.00	-0.10
	(6.55)	(6.27)	(-5.91)	(-6.52)		(2.24)	(2.41)	(0.01)	(-0.60)
growth of trades	**0.57**	**0.53**	**-0.71**	**-0.63**	3-month turnover	**2.46**	**2.37**	**-2.24**	**-2.07**
	(4.88)	(4.29)	(-3.39)	(-2.97)	(*excluding eDay*)	(6.63)	(6.36)	(-6.32)	(-6.92)
Panel B: Bivariate portfolio sort controlled by correlated characteristics									
$Size$	**2.01**	**1.95**	**-1.78**	**-1.59**	$IVOL$	**1.54**	**1.43**	**-1.22**	**-1.18**
	(6.10)	(5.81)	(-5.25)	(-5.74)		(7.00)	(6.59)	(-5.40)	(-6.28)
β^{CAPM}	**1.82**	**1.69**	**-1.62**	**-1.62**	$ILLIQ$	**2.30**	**2.25**	**-1.90**	**-1.72**
	(7.45)	(6.71)	(-6.59)	(-7.89)		(6.53)	(6.29)	(-5.19)	(-5.98)
Mom	**1.76**	**1.64**	**-1.65**	**-1.55**	MAX	**1.84**	**1.73**	**-1.45**	**-1.40**
	(7.24)	(7.07)	(-5.84)	(-7.56)		(7.61)	(7.12)	(-5.68)	(-6.71)
Panel C: Subsamples of time periods					**Panel D: Subsamples divided by listed exchanges**				
1992-2008	**3.32**	**3.56**	**-3.27**	**-2.74**	NASDAQ	**3.37**	**3.35**	**-3.10**	**-2.83**
	(5.80)	(6.27)	(-5.88)	(-5.86)		(6.78)	(6.60)	(-6.36)	(-6.65)
2009-2022	**1.40**	**1.02**	**-1.10**	**-1.16**	NYSE/AMEX	**1.80**	**1.68**	**-1.71**	**-1.78**
	(3.63)	(3.77)	(-3.06)	(-3.79)		(7.08)	(6.64)	(-5.54)	(-7.23)
Panel E: Recession v.s. Expansions					**Panel F: Alternative portfolio formation or filters**				
Recession	2.16	**2.67**	-2.87	**-2.53**	NYSE breakpoints	**2.21**	**2.10**	**-1.89**	**-1.78**
	(1.92)	(2.71)	(-1.69)	(-2.11)		(6.60)	(6.39)	(-5.36)	(-6.47)
Expansion	**2.46**	**2.32**	**-2.22**	**-2.06**	only microcaps	**3.92**	**3.86**	**-3.24**	**-3.63**
	(11.05)	(9.68)	(-6.27)	(-7.83)		(8.34)	(7.73)	(-8.78)	(-9.98)
Difference	-0.30	0.35	-0.66	-0.47	exclude fin. & util.	**2.69**	**2.59**	**-2.31**	**-2.22**
	(-0.26)	(0.34)	(-0.38)	(-0.38)		(6.99)	(6.76)	(-6.03)	(-6.59)

Table IV **Fama-MacBeth regressions**

This table reports the results of Fama-MacBeth regression of 1-month ahead excess returns (close-to-close, overnight, intraday) on trading volume (3-month average turnover). Control variables are defined in Table A.I. In Columns (3) (6) (9), industries dummies classified by 48 categories in Fama and French (1997) are included. All independent variables are winsorized at 1% and 99% level and normalized cross-sectionally to have zero mean and standard deviation of one. Reported in parentheses are t-statistics, computed based on Newey-West standard errors with a maximum lag of 6 months. The sample period covers from June 1992 to December 2022. *, **, *** denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	close-to-close			**overnight**			**intraday**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Trd_vol	-0.06	-0.01	-0.04	0.74***	0.40***	0.39***	-0.77***	-0.45***	-0.48***
	(-0.58)	(-0.26)	(-0.92)	(5.97)	(7.28)	(6.95)	(-6.82)	(-7.01)	(-7.63)
$Size$		-0.13***	-0.13***		0.12***	0.09***		-0.26***	-0.24***
		(-2.58)	(-2.76)		(3.69)	(2.83)		(-4.79)	(-4.65)
β^{CAPM}		-0.03	-0.06		0.13***	0.16***		-0.13	-0.19**
		(-0.26)	(-0.68)		(2.67)	(3.22)		(-1.27)	(-1.99)
B/M		-0.02	-0.00		-0.02	-0.09***		-0.01	0.07
		(-0.35)	(-0.03)		(-0.74)	(-3.55)		(-0.26)	(1.52)
Mom		0.21**	0.19**		0.33***	0.28***		-0.13*	-0.11*
		(2.53)	(2.42)		(8.95)	(7.72)		(-1.88)	(-1.69)
$ILLIQ$		-0.07**	-0.05**		-0.05**	-0.06**		0.02	0.04
		(-2.52)	(-2.09)		(-2.16)	(-2.52)		(0.51)	(1.16)
$IVOL$		-0.13	-0.15		0.84***	0.84***		-0.84***	-0.85***
		(-0.75)	(-0.95)		(5.73)	(5.64)		(-4.88)	(-5.56)
STR		0.08	0.12		-0.09	-0.05		0.22*	0.22**
		(0.58)	(1.16)		(-1.17)	(-0.79)		(1.88)	(2.20)
MAX		-0.08	-0.16		-0.49***	-0.53***		0.33***	0.30***
		(-0.87)	(-1.60)		(-5.91)	(-6.15)		(3.62)	(3.64)
$Skew$		0.01	0.05		0.09***	0.11***		-0.06	-0.04
		(0.28)	(1.28)		(4.21)	(4.89)		(-1.58)	(-1.13)
Observation	724,935	642,431	642,431	724,935	642,431	642,431	724,935	642,431	642,431
Within R^2 (%)	0.01	0.03	0.03	0.11	0.58	0.56	0.15	0.29	0.28
Industry FEs	No	No	Yes	No	No	Yes	No	No	Yes

Table V **Time variation in high-volume overnight premium and intraday discount**

This table reports FF5 alphas (close-to-close, overnight and intraday) of portfolios sorted by trading volume in different economy regimes, indicated by high and low periods of several indices. Panel A and B show the results using investor sentiment and economic policy uncertainty as the economic regime of aggregate market. Trading volume is defined as 3-month average turnover. FF5 alphas in high and low periods a_H, a_L are estimated from regression below:

$$R_{p,t} = a_H d_{H,t} + a_L d_{L,t} + \gamma' \mathbf{FF5}_t + \varepsilon_{p,t}$$

where $R_{p,t}$ is excess return of portfolio, $d_{H,t}, d_{L,t}$ are dummy variables indicating high (low) periods, when index in previous month is above (below) the median of sample, spanning from June 1992 to December 2022. Reported in parentheses are t-statistics, computed based on Newey-West standard errors with a maximum lag of 6 months. *, **, *** denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	close-to-close			**overnight**			**intraday**		
	low volume	high volume	H-L spread	low volume	high volume	H-L spread	low volume	high volume	H-L spread
Panel A: investor sentiment									
high period	0.01	0.08	0.07	-0.42	2.70	3.12	0.30	-2.62	-2.92
	(0.04)	(0.28)	(0.20)	(-3.18)	(7.96)	(9.20)	(1.81)	(-7.47)	(-8.02)
low period	-0.03	0.23	0.26	-0.30	1.28	1.58	0.22	-1.06	-1.28
	(-0.21)	(0.85)	(0.76)	(-1.66)	(4.21)	(6.03)	(0.95)	(-3.31)	(-3.85)
high-minus-low	0.04	-0.15	-0.19	-0.11	**1.42*** **	**1.54*** **	0.08	**-1.56*** **	**-1.64*** **
	(0.19)	(-0.39)	(-0.38)	(-0.55)	(3.23)	(3.68)	(0.30)	(-3.37)	(-3.46)
Panel B: economic policy uncertainty									
high period	0.12	0.07	-0.05	-0.20	1.34	1.54	0.34	-1.25	-1.59
	(0.86)	(0.24)	(-0.15)	(-1.05)	(4.25)	(5.40)	(1.42)	(-3.74)	(-4.51)
low period	-0.14	0.23	0.37	-0.51	2.58	3.09	0.19	-2.37	-2.56
	(-1.01)	(0.90)	(1.08)	(-4.09)	(7.91)	(9.77)	(1.14)	(-6.97)	(-7.36)
high-minus-low	0.26	-0.17	-0.42	0.31	**-1.24*** **	**-1.54*** **	0.15	**1.12** **	**0.97** **
	(1.34)	(-0.44)	(-0.89)	(1.50)	(-2.82)	(-3.73)	(0.56)	(2.38)	(2.03)

Table VI **Fama-MacBeth regressions: role of retail investors**

This table reports the results of Fama-MacBeth regression of 1-month ahead excess returns (close-to-close, overnight, intraday) on trading volume (3-month average turnover). Retail volume is the 3-month average turnover of retail trading identified as Boehmer et al. (2021). Non-retail volume is difference between trading volume and retail volume. Control variables are defined in Table A.I. In Columns (3) (6) (9), industries dummies classified by 48 categories in Fama and French (1997) are included. All independent variables are winsorized at 1% and 99% level and normalized cross-sectionally to have zero mean and standard deviation of one. Reported in parentheses are t-statistics, computed based on Newey-West standard errors with a maximum lag of 6 months. The sample period covers from October 2006 to December 2021. *, **, *** denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	close-to-close			**overnight**			**intraday**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Trd_vol	-0.01			0.26***			-0.29***		
	(-0.12)			(5.77)			(-4.29)		
Rt_vol		0.06	-0.06		0.44***	0.42***		-0.36***	-0.44***
		(0.66)	(-0.81)		(6.06)	(5.08)		(-3.33)	(-4.61)
$NonRt\_vol$		-0.06	-0.01		-0.10**	-0.10**		0.01	0.07
		(-0.77)	(-0.08)		(-2.23)	(-2.12)		(0.13)	(1.00)
$Size$	-0.13**	-0.14**	-0.10*	0.05	0.01	0.01	-0.16**	-0.14**	-0.09
	(-1.97)	(-2.06)	(-1.65)	(1.41)	(0.27)	(0.26)	(-2.44)	(-2.17)	(-1.53)
β^{CAPM}	0.03	0.03	0.01	0.10***	0.09**	0.09**	-0.08	-0.07	-0.09
	(0.30)	(0.30)	(0.05)	(2.68)	(2.36)	(2.28)	(-0.74)	(-0.64)	(-1.00)
B/M	-0.12	-0.12	-0.10*	0.04	0.04	0.02	-0.16**	-0.17**	-0.12**
	(-1.52)	(-1.51)	(-1.67)	(1.28)	(1.63)	(1.17)	(-2.55)	(-2.67)	(-2.09)
Mom	0.06	0.06	0.07	0.26***	0.24***	0.19***	-0.21**	-0.19**	-0.13
	(0.60)	(0.54)	(0.66)	(7.52)	(6.44)	(4.88)	(-2.10)	(-1.98)	(-1.40)
$ILLIQ$	-0.12**	-0.12**	-0.07	0.05**	-0.01	0.00	-0.14***	-0.09**	-0.05
	(-2.42)	(-2.46)	(-1.49)	(2.19)	(-0.47)	(0.01)	(-2.93)	(-2.18)	(-1.18)
$IVOL$	-0.04	-0.05	-0.08	0.34***	0.31***	0.28***	-0.35*	-0.34*	-0.31*
	(-0.16)	(-0.22)	(-0.40)	(2.84)	(2.62)	(2.69)	(-1.77)	(-1.72)	(-1.69)
STR	-0.11	-0.11	-0.06	-0.01	-0.01	0.03	-0.06	-0.06	-0.08
	(-0.63)	(-0.61)	(-0.58)	(-0.10)	(-0.13)	(0.40)	(-0.48)	(-0.44)	(-0.78)
MAX	0.06	0.07	0.05	-0.14**	-0.14**	-0.16***	0.16	0.16	0.18
	(0.48)	(0.51)	(0.36)	(-2.38)	(-2.35)	(-2.79)	(1.24)	(1.26)	(1.45)
$Skew$	-0.07	-0.07	-0.05	0.01	0.01	0.02	-0.06	-0.06	-0.05
	(-1.18)	(-1.18)	(-0.94)	(0.39)	(0.60)	(1.04)	(-0.99)	(-1.06)	(-1.04)
Observation	286,414	286,414	286,414	286,414	286,414	286,414	286,414	286,414	286,414
Adjusted R^2 (%)	7.90	8.16	14.13	4.14	4.55	6.27	8.88	9.27	15.79
Industry FEs	No	No	Yes	No	No	Yes	No	No	Yes

Table VII Portfolios sorted by trading volume and risk/mispricing

This table reports average monthly returns (close-to-close, overnight, intraday) of five-by-five portfolios independently sorted by trading volume trading volume and CAPM market β (mispricing measure $MISP$) in Panel A (Panel B). Trading volume is defined as 3-month average turnover. Market β is estimated based on a 60-month rolling window using CAPM model. $MISP$ is a percentile ranking on mispricing aggregated from 11 robust anomalies as in Stambaugh et al. (2015). $\beta_{\text{high}} - \beta_{\text{low}}$ (UMO) is the spread portfolio between highest and lowest quintile of CAPM market β (mispricing measure $MISP$), conditional on certain quintile of trading volume. H-L: $\beta_{\text{high}} - \beta_{\text{low}}$ (H-L: UMO) refers to the difference of H-L β (UMO) between highest and lowest quintile of trading volume. Reported in parentheses are t-statistics, computed based on Newey-West standard errors with a maximum lag of 6 months. Portfolios are value-weighted and rebalanced each month. The sample period covers from June 1992 to December 2022.

Panel A: sorted by trading volume and market beta β

	low volume				high volume				HML: $\beta_{\text{high}} - \beta_{\text{low}}$			
	low β	medium	high β	H-L β	low β	medium	high β	H-L β	raw	α_{FF5}	α_{q^4}	α_{SY4}
close-to-close	0.66	0.75	0.38	-0.27	0.38	0.92	0.91	0.53	**0.80**	**1.13**	**0.99**	**1.26**
	(3.97)	(2.98)	(0.99)	(-0.84)	(1.11)	(2.10)	(1.68)	(1.15)	(1.88)	(2.66)	(2.36)	(2.71)
overnight	-0.08	-0.10	0.17	0.25	0.89	1.60	2.17	**1.28**	**1.03**	**1.07**	**1.08**	**1.26**
	(-0.53)	(-0.47)	(0.69)	(1.15)	(4.39)	(5.34)	(5.84)	(4.07)	(3.00)	(3.04)	(2.75)	(3.24)
intraday	0.62	0.80	0.35	-0.27	-0.59	-0.67	-1.25	**-0.65**	-0.38	-0.08	-0.21	-0.10
	(3.89)	(3.62)	(0.97)	(-0.70)	(-2.27)	(-2.24)	(-3.18)	(-1.79)	(-1.10)	(-0.24)	(-0.60)	-0.23

Panel B: sorted by trading volume and mispricing score $MISP$

	low volume				high volume				HML: UMO			
	overpriced	medium	underpriced	UMO	overpriced	medium	underpriced	UMO	raw	α_{FF5}	α_{q^4}	α_{SY4}
close-to-close	0.75	0.63	0.58	-0.16	-0.44	0.80	1.38	**1.82**	**1.99**	**1.70**	**1.44**	**0.96**
	(2.54)	(2.25)	(3.24)	(-0.72)	(-0.74)	(1.72)	(3.28)	(5.03)	(5.46)	(4.23)	(3.87)	(2.98)
overnight	-0.09	-0.13	-0.27	-0.19	2.11	1.46	1.53	**-0.58**	**-0.39**	-0.20	-0.37	-0.37
	(-0.56)	(-0.90)	(-1.97)	(-1.30)	(6.27)	(4.46)	(4.53)	(-2.55)	(-1.78)	(-0.84)	(-1.37)	(-1.40)
intraday	0.79	0.60	0.73	-0.06	-2.41	-0.71	-0.23	**2.18**	**2.24**	**1.78**	**1.63**	**1.20**
	(2.65)	(2.69)	(4.30)	(-0.25)	(-4.56)	(-1.92)	(-0.71)	(5.93)	(5.96)	(4.76)	(4.34)	(3.43)

Table VIII Portfolios sorted by trading volume and mispricing anomalies

This table reports average monthly value-weighted returns (close-to-close, overnight and intraday) of portfolios sorted by trading volume and 11 mispricing anomalies documented in Stambaugh et al. (2012). Overpriced (underpried) refers to the quntile group of stocks with anomalies associated with lowest (highest) average return. UMO (H-L) is the spread portfolio of underpriced-minus-overpriced (high-minus-low trading volume) stocks. Reported in parentheses are t-statistics, computed based on Newey-West standard errors with a maximum lag of 6 months.

	low volume: UMO			**high volume: UMO**			**H-L: UMO**		
	close-to-close	overnight	intraday	close-to-close	overnight	intraday	close-to-close	overnight	intraday
Failure probability	-0.12	0.06	**-0.65**	**0.77**	**-1.12**	**1.65**	**0.89**	**-1.19**	**2.30**
	(-0.41)	(0.28)	(-1.90)	(2.21)	(-3.64)	(4.56)	(2.35)	(-3.24)	(4.75)
Ohlson's O Score	-0.09	**0.33**	**-0.53**	**0.61**	-0.27	**0.73**	**0.70**	**-0.59**	**1.26**
	(-0.45)	(2.34)	(-2.80)	(2.07)	(-1.14)	(2.67)	(2.20)	(-2.15)	(3.57)
Net stock Issues	-0.14	**-0.52**	**0.35**	0.30	**-1.40**	**1.66**	0.44	**-0.88**	**1.31**
	(-0.82)	(-3.60)	(1.77)	(0.80)	(-7.20)	(4.78)	(1.13)	(-3.78)	(3.71)
Composite Equity Issues	-0.03	0.13	-0.17	0.36	**0.64**	-0.27	0.39	**0.50**	-0.11
	(-0.11)	(0.74)	(-0.57)	(0.98)	(2.62)	(-0.93)	(1.12)	(1.66)	(-0.32)
Total Accruals	0.17	0.01	0.15	-0.04	0.04	-0.10	-0.21	0.04	-0.24
	(0.83)	(0.03)	(0.63)	(-0.14)	(0.20)	(-0.35)	(-0.63)	(0.13)	(-0.72)
Net Operating Assets	0.12	0.07	0.02	**0.91**	0.15	**0.82**	**0.79**	0.09	**0.80**
	(0.64)	(0.47)	(0.13)	(3.15)	(0.83)	(3.05)	(2.52)	(0.39)	(2.47)
Momentum	0.08	**0.86**	**-0.83**	**1.31**	**1.22**	-0.01	**1.24**	0.36	**0.82**
	(0.23)	(4.72)	(-2.36)	(3.31)	(5.13)	(-0.03)	(3.57)	(1.39)	(2.55)
Gross Profitability	-0.04	**-0.47**	**0.45**	**0.76**	**-1.45**	**2.05**	**0.80**	**-0.98**	**1.60**
	(-0.24)	(-3.01)	(2.19)	(2.58)	(-4.82)	(5.46)	(2.76)	(-2.91)	(4.23)
Asset Growth	-0.11	**-0.60**	**0.53**	0.47	**-0.41**	**0.89**	0.58	0.19	0.36
	(-0.59)	(-3.37)	(2.47)	(1.41)	(-2.01)	(2.85)	(1.51)	(0.78)	(1.09)
Return on Assets	-0.27	-0.13	-0.26	**0.79**	-0.11	**0.73**	**1.06**	0.03	**0.99**
	(-1.17)	(-0.83)	(-1.08)	(2.54)	(-0.49)	(2.41)	(3.71)	(0.11)	(2.88)
Investment-to-Assets	0.10	0.27	-0.15	**0.75**	-0.12	**0.86**	**0.65**	-0.39	**1.00**
	(0.47)	(1.53)	(-0.74)	(2.56)	(-0.63)	(4.02)	(1.85)	(-1.52)	(3.41)

Appendix A Variable Definitions

The table below indicates t as the generic period that targets return prediction. Specifically, to predict returns in month t, the variable is formed using the information only before the end of month $t-1$.

Table A.I Variable Definitions

Variable	**Definition**	**Data Source**
Panel A: volume-related variables		
Trd_vol	Trading volume, i.e. average turnover over the period from $t-3$ to $t-1$. For each month, turnover is computed as the share volume (vol) scaled by number of shares outstanding ($shrout$).	CRSP
Rt_vol	Retail component of Trd_vol.	TAQ
$NonRt\_vol$	Non-retail component of Trd_vol, simply the difference between Trd_vol and Rt_vol	"
Panel B: firm-level characteristics		
$Size$	Market equity. Natural logarithm of product between month-end price (prc) by number of shares outstanding ($shrout$) at $t-1$	CRSP
β^{CAPM}	Market beta. Regression coefficient of monthly excess stock returns on market excess return, using a rolling window between $t-60$ and $t-1$.	OSAP
B/M	Book-to-market ratio. Natural Logarithm of tangible book equity in most recent December, scaled by market equity in month $t-1$.	"
Mom	Momentum. Cumulative stock return between months $t-12$ and $t-2$	"
$IVOL$	Idiosyncratic volatility. Standard deviation of residuals from Fama-French three factor regressions using daily returns in month $t-1$.	"
$ILLIQ$	Amihud illiquidity. Ratio of daily absolute return over dollar volume, averaged at daily level over period from months $t-12$ to $t-1$.	"
STR	Short-term reversal. Stock return in month $t-1$.	"
$Skew$	Return skewness. Skewness of daily returns in month $t-1$.	"
MAX	Maximum of daily returns in month $t-1$.	"

Note: OSAP is short for Open Source Asset Pricing, a webiste that provides replicated results of a wide range of variables used in asset pricing.

Appendix B Proofs

Proof of Proposition 2.1. In period $t = 4$, since the signal s_2 is effectively public, conjecture the following price as a function of $s_2 = \theta_2 + \epsilon_2$ and the supply shock z_4,

$$p_4 = v_0 + \theta_1 + s_2 + Dz_4, \tag{B.1}$$

where the signal is with precision $\tau_\theta + \tau_\epsilon$, thereby $\mathrm{Var}(v|\mathcal{F}_4) = (\tau_\theta + \tau_\epsilon)^{-1} + \tau_\xi^{-1}$. Since the price is normally distributed, the investor's optimization problem (2.2) solves

$$x_4^* = \gamma[\mathrm{Var}(v|\mathcal{F}_4)]^{-1}[\mathrm{E}(v|\mathcal{F}_4) - p_4] = \frac{\gamma\tau_\xi(\tau_\theta + \tau_\epsilon)}{\tau_\xi + \tau_\theta + \tau_\epsilon}(v_0 + \theta_1 + s_2 - p_4). \tag{B.2}$$

Substituting into the market clearing condition 2.3, we obtain

$$p_4 = v_0 + \theta_1 + \theta_2 + \epsilon_2 - \frac{z_4}{A_1}, \quad \text{where} \quad A_1 = \frac{\gamma\tau_\xi(\tau_\theta + \tau_\epsilon)}{\tau_\xi + \tau_\theta + \tau_\epsilon}. \tag{B.3}$$

The above solution also verifies the price conjecture (B.1).

Consider period $t = 3$. θ_1 is realized, while θ_2 is not. Consider the price conjecture

$$p_3 = A + Bs_2 + Cz_3. \tag{B.4}$$

For the early-informed investor E, $\mathrm{E}(v|\mathcal{F}_{3,E}) = \mathrm{E}(v|s_2) = v_0 + \theta_1 + s_2$. Then the optimal allocation $x_{3,E}^*$ solves

$$x_{3,E}^* = \gamma[\mathrm{Var}(v|s_2)]^{-1}[\mathrm{E}(v|s_2) - p_3] = \frac{\gamma\tau_\xi(\tau_\theta + \tau_\epsilon)}{\tau_\xi + \tau_\theta + \tau_\epsilon}(v_0 + \theta_1 + s_2 - p_3). \tag{B.5}$$

The late-informed investor I extracts the signal from the price, and believes the signal has a noise with variance $(m\tau_\epsilon)^{-1}$. Therefore,

$$x_{3,I}^* = \gamma[\mathrm{Var}(v|p_3)]^{-1}[\mathrm{E}(v|p_3) - p_3] = \frac{\gamma\tau_\xi(\tau_\theta + m\tau_\epsilon)}{\tau_\xi + \tau_\theta + m\tau_\epsilon}\left(v_0 + \theta_1 + \frac{p_3 - A - Cz_3}{B} - p_3\right). \tag{B.6}$$

The market clearing condition reads

$$(1 - s)x_{3,E}^* + sx_{3,I}^* = z_3. \tag{B.7}$$

Denote

$$A_1 = \frac{\gamma\tau_\xi(\tau_\theta + \tau_\epsilon)}{\tau_\xi + \tau_\theta + \tau_\epsilon}, \quad A_2 = \frac{\gamma\tau_\xi(\tau_\theta + m\tau_\epsilon)}{\tau_\xi + \tau_\theta + m\tau_\epsilon}. \tag{B.8}$$

Substituting (B.5) and (B.6) into (B.7) and rearranging the equation, we obtain

$$\begin{aligned} p_3 = & \frac{[(1-s)A_1 + sA_2](v_0+\theta_1) - sA_2A/B}{(1-s)A_1 + sA_2(1-1/B)} + \frac{(1-s)A_1}{(1-s)A_1 + sA_2(1-1/B)} s_2 \\ & + \frac{-(1+sA_2C/B)}{(1-s)A_1 + sA_2(1-1/B)} z_3. \end{aligned} \tag{B.9}$$

Comparing to (B.4), we verify the conjecture. Then by comparing the coefficients, we have a system of equations for $\{A, B, C\}$. The solution reads

$$\begin{aligned} & A = v_0 + \theta_1, \quad B = 1 + \frac{1}{A_1 + s(A_2 - A_1)}, \quad C = -\frac{1}{A_1 + s(A_2 - A_1)}. \\ \Rightarrow & p_3 = v_0 + \theta_1 + \left(1 + \frac{1}{A_1 + s(A_2 - A_1)}\right)(\theta_2 + \epsilon_2) - \frac{1}{A_1 + s(A_2 - A_1)} z_3. \end{aligned} \tag{B.10}$$

Similarly, p_1 and p_2 can be solved, of which the solving process is similar to solving p_3 and p_4, respectively. Finally, note that A_2 increases in m, $\forall m \geq 1$. When $m = 1$, $A_2 = A_1$. Since $m > 1$, we have $A_2 > A_1$.

Proof of Proposition 2.2. In period 2, $\forall i \in [0, 1]$, $x_{i,2} = z_2$. Consider period 1. The early-informed investor E's optimal allocation (2.4), we obtain

$$(1-s)x^*_{1,E} = \frac{(1-s)A_1}{A_1 + s(A_2 - A_1)} \left[z_1 - (\theta_1 + \epsilon_1)\right]. \tag{B.11}$$

Together with the market clearing condition, we have

$$sx^*_{1,I} = \frac{sA_2z_1}{A_1 + s(A_2 - A_1)} + \frac{(1-s)A_1}{A_1 + s(A_2 - A_1)}(\theta_1 + \epsilon_1). \tag{B.12}$$

Substituting into (2.12),

$$\mathrm{E}(V^{\text{day 1}}) = \frac{2s(1-s)z(A_2 - A_1)}{A_1 + s(A_2 - A_1)}. \tag{B.13}$$

Combining with Corollary 1 directly yields (2.14).