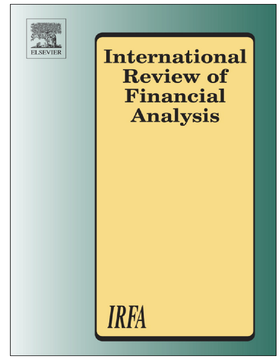
Accepted Manuscript



Dynamic trading volume and stock return relation: Does it hold out of sample?

Zijun Wang, Yan Qian, Shiwen Wang

PII:	S1057-5219(17)30128-X
DOI:	doi:10.1016/j.irfa.2017.10.003
Reference:	FINANA 1148
To appear in:	*International Review of Financial Analysis*
Received date:	1 February 2017
Revised date:	5 September 2017
Accepted date:	7 October 2017

Please cite this article as: Zijun Wang, Yan Qian, Shiwen Wang , Dynamic trading volume and stock return relation: Does it hold out of sample?. The address for the corresponding author was captured as affiliation for all authors. Please check if appropriate. Finana(2017), doi:10.1016/j.irfa.2017.10.003

Dynamic Trading Volume and Stock Return Relation: Does It Hold out of Sample?

Zijun Wang

School of Business, Suzhou University of Science and Technology

Yan Qian

School of Business, Suzhou University of Science and Technology

Shiwen Wang

School of Business, Suzhou University of Science and Technology

*Correspondence: Zijun Wang, School of Business, Suzhou University of Science and Technology, Suzhou City, Jiangsu Province 215009, China. E-mail: zijunwang@mail.usts.edu.cn

Dynamic Trading Volume and Stock Return Relation: Does It Hold out of Sample?

Abstract

This paper studies the dynamic relation between trading volume and stock returns from the perspective of out-of-sample stock return predictability. Evidence from the U.S. suggests that higher returns do follow more intensive trading, especially in the pre-2000 period. However, the ex-ante predictability delivers a small economic gain equivalent to an annual return of 0.73% for a risk-averse investor. This weak out-of-sample predictive power of volume is absent in most of the other major markets. Overall, investors are not likely to gain much financially by "riding the volume curve," at least at the levels of net profits suggested by our findings.

Keywords: Volume-return relation; Out-of-sample regression; High volume return premium

JEL number: G12; G15

Dynamic Trading Volume and Stock Return Relation: Does It Hold out of Sample?

1. Introduction

The relation between trading volume and stock returns has been an active area of research for many decades. The popularization of high-speed (high-frequency) trading, a conspicuous aspect of financial markets in the past two decades, has attracted increasing attention to the relation from both academicians and practitioners. Not surprisingly, it also figures prominently in the debate about the revived proposals to impose Tobin-type securities transaction taxes to reduce trading volume following the recent financial crisis. Better understanding of the volume-return relation clearly can also shed light on the ongoing debate about whether modern finance is too big (Cochrane, 2013; Greenwood and Scharfstein, 2013).

The questions at issue are whether there is any relation between trading volume and stock returns and, if the answer is yes, whether such a relation is economically significant. The latter is probably more important in the current debate. Market microstructure theory suggests that both trading volume and price changes (returns) are related to the arrival of information to the market. Thus volume and price movement may jointly depend on the intensity of information flow. Much of the early theoretical work on the volume-return relation therefore focuses primarily on the contemporaneous relation between volume and price changes (Karpoff, 1987; Gallant, Rossi, and Tauchen, 1992). However, considering the long-standing controversy about the simultaneous determination of price and quantity in economics, it is not surprising that such contemporaneous causality between volume and stock returns has proven difficult to sort out empirically given the observational nature of data.

Extending the early work but paying more attention to its dynamic nature, later research generally finds positive evidence on the volume-return relation under different assumptions. First, trading volume is a measure of liquidity, which is significantly related to future stock returns (Amihud and Mendelson, 1986; Datar, Naik, and Radcliffe, 1998; Lesmond, Ogden, and Trzcinka, 1999; Amihud, 2002; Lesmond, 2005; Liu, 2006). Second, trading volume indicates how investors trade on individual stocks to share risk or speculate on private information, which further induces different subsequent reversal or continuation patterns (Llorente, Michaely, Saar, and Wang, 2002). Third, trading volume describes investors' learning curve that leads to overconfidence and further affects future stocks returns (Gervais and Odean, 2001; Statman, Thorley, and Vorkink, 2006). Finally, trading volume is related to investor attention and reflects how investors react to the news of the firm (Hou, Peng, and Xiong, 2009). While many of these studies examine the cross sectional volume-return relation in individual stocks, there is another line of research explicitly investigating the dynamic relation between volume and stock returns via testing Granger (non-) causality since Hiemstra and Jones (1994). The intent is to determine whether including past volume information can help predict stock returns after controlling for past returns and other relevant information. Other important contributions in this sub-field include Easley, O’Hara, and Srinivas (1998), Chordia and Swaminathan (2000), Lee and Rui (2002), Malcolm and Stein (2004), Chuang, Kuang, and Lin (2009), and Chen (2012).

For the purpose of real time prediction and risk management, focusing on the dynamic relation between volume and returns is perhaps more informative than the often elusive contemporaneous causality. Nevertheless, almost all of these empirical studies conduct their analysis using in-sample regressions. And it is now well known that many commonly used variables have been found to have no or negligible out-of-sample forecasting ability despite their enormous in-sample predictive power for stock returns (see, for example, Welch and Goyal (2008)). The lack of robustness of the in-sample evidence may cast doubt on the real predictive power of these variables. The intended contribution of this paper is thus to re-examine the volume-return relation from the perspective of out-of-sample stock return predictability. Given the enormous interests from both practitioners and academician on whether stock returns can be predicted out-of-samples, it is interesting to find out if the volume-return relation which is well-established by in-sample regressions also hold in the out-of-sample forecasting regressions. Out-of-sample return predictability is economically important in itself from an asset allocation perspective. If, in addition, both in- and out-of-sample evidences derived from a predictive model are consistent, then the empirical model is less likely to be misspecified and the theory on which the model is based is more likely to be credible.[1]

Like most studies that consider out of sample forecasting regressions, we investigate the volume-return relation by studying the predictive power of aggregate time series measure of trading volume (turnover). However, as noted above, the volume-return relation has been studied mostly at the individual stock level, focusing on cross sectional variation in returns that is related to trading volume. We therefore also consider the out-of-sample forecasting performance of the high volume return premium for stock returns. The high volume return premium (or simply high volume premium, HVP), defined here as the return on a zero-cost portfolio that is long on stocks experiencing unusually high trading volume and short on low-volume stocks, has been studied before in regard to the interplay of short-run return autocorrelations and volume at the firm level (Chordia and Swaminathan, 2000; Gervais, Kaniel, and Mingelgrin, 2001; Kaniel, Ozoguz, Starks, 2012). Our contribution here is that we formally explore whether this cross-sectionally constructed variable has time series predictive power for future returns out of samples.

Like many empirical studies on the volume-return relation, we concentrate on the U.S. market. However, to examine how well the results obtained in the U.S. market hold out of samples in another

[1] Rapach, Strauss, Zhou (2013) is a recent example of studying dynamic relation in equity markets by performing Granger causality tests using both in- and out-of-sample regressions. Those authors' interests are whether and how U.S. stock returns lead markets in other industrialized countries.

sense, we also investigate major international equity markets. We examine evidence from six other developed countries in the Group of Seven (G-7) and an additional 12 countries of both developed and emerging economies which house large stock exchanges by market capitalization. The advantage of performing the same tests for many countries is that empirical findings may no longer be sample-specific. Nevertheless, it should be pointed out that, because the stock valuation processes vary across countries due to differences in institutions and information flows, the results from different countries may not be comparable.

Our main findings can be summarized as follows. First, for U.S. data from 1963 to 2010, out-of-sample regressions show mixed evidence of incorporating past volume information to predict returns. There is no gain in forecast accuracy for value-weighted portfolios. In contrast, compared to simple autoregression forecasts and historical average forecasts, models also using trading volume generate better forecasts for stock returns to equal-weighted portfolios by both root mean squared forecast errors and a forecast encompassing test. Nevertheless, for a typical risk-averse investor, the improvement in forecasts transforms to a mere utility gain of 0.73% per annum in terms of return rates. Interestingly, this estimate is close to French's (2008) estimate of 0.67% of the aggregate value of the market each year investors spend searching for superior returns over 1980 to 2006.

Further sub-sample analysis shows that trading volume's out-of-sample forecasting ability declines significantly over the last ten years.[2] One possible explanation for the weaker predictive power is that rapid growth of trading volume has been associated with cost saving and gains in market efficiency in the form of narrowing bid-ask price spreads observed during this period. Alternatively, the declining time series return predictability is also potentially consistent with recent trends in trading activity in which turnover has become more sensitive to return predictors and increased trading by

[2] Interestingly, Griffin, Nardari, and Stulz's (2007) also find that the return-volume relation weakens in recent years for U.S. and some high-income countries.

institutions has been accompanied by decreased cross-sectional return predictability (Chordia, Roll, and Subrahmanyam, 2011).

Our second empirical finding is derived from the 18 international markets. Canada is the only other G-7 country than the U.S. in which the high volume premium helps predict returns out of samples for equal-weighted portfolios by all three statistical and economic measures. Value-weighted, aggregate turnover shows out-of-sample predictive power in five out of the other 12 major markets. The high volume premium contains some additional predictive power for value-weighted portfolios only in India, and for equal-weighted portfolios only in China.

Motivated by the work of Lee and Rui (2002), we also study the spill-over effect of U.S. trading volume. Out-of-sample evidence shows that trading volume in the U.S. market in general does not contain additional information for forecasting returns in other markets after controlling for past returns, volume, and volatility information from domestic markets as well as past U.S. market returns. Because many international samples comprise predominantly recent data, the absence of a spillover effect of U.S. trading activity is consistent with its vanishing predictive power for U.S. market returns in the second sub-sample analysis. The volume effect, if present, is not globally integrated.

If the process by which prices adjust to information is not immediate, market statistics such as volume impound information that is not yet incorporated into the current market price. Our finding of a quantitatively small but statistically significant return forecastability by volume is consistent with this explanation and confirms the existence of a dynamic volume-return relation. However, our support for the published theoretical and empirical work is limited because we find that the out-of-sample predictive power of trading volume largely disappeared in the U.S. market in the recent period and that the evidence is scarce in international markets. Overall, it appears reasonable to conclude that investors cannot gain much financially by “riding the volume curve”, at least at the levels of net profits suggested by our findings.

2. Econometric Methodology

As discussed earlier, rather than seeking to establish a relation between trading volume and stock returns with a causal interpretation in the strict sense, we approach the issue in a less ambitious manner by considering the forecasting relation between the two variables, which is also known as Granger causality (Granger, 1969). Consider the following standard predictive regression model for variable Y,

$$y_t = \alpha + \sum_{m=1}^{L_y} \beta_m y_{t-m} + \sum_{n=1}^{L_x} \gamma_n x_{t-n} + \sum_{r=1}^{L_z} \lambda_r z_{t-r} + \varepsilon_t,\ t = 1, \ldots, T, \tag{1}$$

where y_t, x_t, and z_t are the realized values of Y, X, and Z at time period t, y_{t-m}, x_{t-n}, and z_{t-r} are the corresponding m-, n-, and r-period lagged values, ε_t is the error term, L_y, L_x, and L_z are the numbers of lags on the three variables, and α, β_m, γ_n, and λ_r are free parameters. In our later application, Y is daily return to the market portfolio, X is a variable measuring trading activity (aggregate turnover or the high volume premium, to be specific), and Z is market volatility. Note that the lagged returns are included in model (1) to capture generally small but statistically significant serial correlation resulting from, among other things, non-synchronous trading in the daily returns (e.g., Conrad and Kaul, 1988; Lo and MacKinlay, 1988). We thus rule out the possibility that the seemingly predictive power of volume simply reveals the well-documented autoregression in the return series. Model (1) could be motivated by the market microstructure literature, which explicitly takes the sequential nature of the trading process into account. For example, in Blume, Easley, and O'Hara's (1994) learning model, volume provides information on information quality that cannot be deduced from the price statistic, and traders who use information contained in price and volume statistics do better than traders who do not. Models similar to (1) have been used by Gallant, Rossi, and Tauchen (1992), Campbell, Grossman, and Wang (1993), and many of recent studies as reviewed in the Introduction.

The implementation of the out-of-sample forecasting evaluation using model (1) is straightforward.[3] Suppose there are a total of T observations. We estimate model (1) under both the null and the alternative hypotheses using the first R (in-sample) observations, where the H_0 model include the volume variable (x_{t-n}) and the H_A model does not. We then generate one-step-ahead recursive forecasts of y_t for the remaining P (out-of-sample) observations ($R + P = T$) from both H_0 and H_A models. Denote the corresponding forecast error series as $\hat{\varepsilon}_{0,t}$ and $\hat{\varepsilon}_{A,t}$, respectively. If the H_A model produces more accurate forecasts than H_0, or equivalently, if $\hat{\varepsilon}_{A,t}$ is smaller than $\hat{\varepsilon}_{0,t}$, then trading volume predict returns in the out-of-sample sense.

It is possible that, although two sets of forecasts are visually different from each other, they may not differ statistically due to sample variability. This can be a problem for studying daily stock returns, a large component of which are unexplained even at the aggregate level. Since the models we consider above are nested and they all may be misspecified, in this study we follow Corradi and Swanson (2006)'s recommendation to apply Clark and McCracken's (2001) encompassing test (ENC-NEW test) to formally compare forecast accuracy of the two rival models. Under the null hypothesis that forecasts from the H_0 model that excludes the trading volume encompass those of the H_A model, Clark and McCracken (2001) show that the following statistic has an asymptotic nonstandard distribution

$$\text{ENC-NEW} = P\frac{P^{-1}\sum_{t=(R+1)}^{T}(\hat{\varepsilon}_{0,t}^{2} - \hat{\varepsilon}_{0,t}\hat{\varepsilon}_{A,t})}{P^{-1}\sum_{t=(R+1)}^{T}\hat{\varepsilon}_{A,t}^{2}} \rightarrow_{d} \Gamma_1, \tag{2}$$

where $\Gamma_1 = \int_{\lambda}^{1} s^{-1}B'(s)$, $\lambda = (1+\pi)^{-1}$, π is the limit of P/R, the ratio of the out-of-sample size over the in-sample size, and $B(s)$ is a vector Brownian motion whose dimension equals that of x_t (namely L_x). If H_0 forecasts encompass model H_A, then H_A forecasts do not provide useful information absent from

[3] Since the standard predictive regression Model (1) is often used for testing in-sample Granger causality, the out-of-sample forecasting regressions using the model can therefore be interpreted as out-of-sample Granger causality tests.

forecasts from H_0. The encompassing test (2) has seen increasing application in finance (e.g., Butler, Grullon, and Weston, 2005; Welch and Goyal, 2008).

Like other statistical measures for forecast evaluations, the root-mean squared forecast errors and the encompassing test do not explicitly account for the risk borne by an investor in following portfolio recommendations from statistically preferred models (Rapach, Strauss, and Zhou, 2010). To address this limitation, in this paper we also provide evidence of economic significance of forecasting ability of trading activity for market returns. Specifically, following Marquering and Verbeek (2004) and many others, we calculate realized utility for a mean-variance investor on a real-time basis for the out-of-sample period. Specifically, we assume that the investor allocates her/his investment daily between stocks and risk-free bills. The standard portfolio allocation rule then stipulates that, conditional on information available at period t, the optimal weight of such an investor's portfolio on stocks at period (t + 1) is

$$s_{i,t+1} = \left(\frac{1}{\gamma}\right)\frac{E_t(R_{i,t+1})}{E_t(\sigma^2_{i,t+1})}, \tag{3}$$

where γ is the investor's relative risk aversion parameter, $E_t(R_{i,t+1})$ and $E_t(\sigma^2_{i,t+1})$ are the forecast for stock return and its variance based on forecasting model i. Correspondingly, the rest of the portfolio (1 – $s_{i,t+1}$) is invested in the risk-free bills. The realized average utility of the investor is given by

$$\hat{u}_i = \hat{\mu}_{i,p} - \frac{1}{2}\gamma\hat{\sigma}^2_{i,p}, \tag{4}$$

where $\hat{\mu}_{i,p}$ and $\hat{\sigma}^2_{i,p}$ are the out-of-sample mean and variance of the returns to the dynamic portfolio formed based on the above rule. Intuitively, the investor's utility increases with the average return but decreases with its volatility.

3. Data

We study stock return data from both U.S. and international markets. In this section, we describe the U.S. sample in detail. The international data are described briefly and analyzed in Section 6. For the U.S. sample, we consider all NYSE, Amex, and NASDAQ non-financial stocks for the period of July 1, 1963 through December 31, 2010 with July 1962-June 1963 reserved for pre-sample selection. They are obtained from CRSP monthly and daily stock securities files and events files.[4]

Volume can change simply because of (reverse) stock splits. Following Lo and Wang (2000), Griffin, Nardari, and Stulz (2007), and many other studies in the literature, we use turnover rather than raw volume data to measure trading intensity. We consider two types of aggregate time series estimates of turnover, one that is weighted by market capitalization of stocks (VWVOL) and the other that is equally weighted (EWVOL).[5] We exclude observations with missing price or volume data and stocks with less than one year of trading history. Also discarded are those delisted from the exchanges within one year due to merger, (partial) liquidation, and other capital events.[6] Figure 1 plots 100-day moving averages of both value-weighted and equal-weighted measures of volume (turnover), which clearly show an upward trend in both measures of aggregate trading activity over the sample period. In particular, turnover increased significantly starting early 2003 and appears to have tailed off by the end of the sample, largely coinciding with the development of high-frequency trading (e.g., Chordia, Roll, and Subrahmanyam 2011). As pointed out by Griffin, Nardari, and Stulz (2007), turnover may be influenced by trends in bid-ask spreads, commissions, availability of information, and other factors that

---

[4] Following the literature, we use daily data. The market microstructure literature also studies the relations between trading volume, stock returns, and market volatility. However, as Andersen (1996) points out, the focus of this area of research typically is on intraday rather than interday dynamics as we study here. Therefore, its theoretical predictions regarding the relations among these variables may not hold at the daily frequency due to the complicating effect of temporal aggregation on causality testing (Granger, 1988).

[5] For U.S. data, turnover is defined as trading volume divided by the number of outstanding shares (multiplied by 1,000 for presentation). For international data from Datastream, it is defined as traded value (price times volume) scaled by market capitalization. For convenience of exposition, we sometimes also refer turnover simply as volume throughout the paper.

[6] This is to follow Thornton and Valente (2012) in a partial attempt to control for possible reverse causality that investors anticipating better/worse future stock performance could be more likely to trade. Nevertheless, the main results hold without the data filtering.

might contribute to the general increase in trading activity through time. To remove this slowing moving average component, in all later analysis, we de-trend turnover by first taking its natural log and then subtracting its 100-trading-day (20 calendar weeks) trailing moving average (see, for example, Chen, Hong, and Stein (2001) for a similar treatment).

Panel A of Table 1 provides descriptive statistics of the value- and equal-weighted log-detrended trading volume (turnover) (VWVOL and EWVOL), and the corresponding excess returns on the market portfolio without dividends (VWMKT and EWMKT). The averages of the (detrended) volumes are positive for both measures because of the generally upward trend in the raw series. They are also serially correlated with first-order autocorrelations of 0.60 and 0.73, respectively. The value-weighted market portfolio returns are averaged at 0.010% on a daily basis and the equal-weighted returns at 0.061%.[7] Both return series also feature statistically significant serial correlation which is more evident in EWMKT than in VWMKT. The two measures of trading volume are contemporaneously correlated to the market returns (the details are not reported in the table).

To be comparable to popular return anomalies such as the value premium (HML), we slightly modify the construction of the high volume return premium (HVP) as implemented by Gervais, Kaniel, and Mingelgrin (2001) and Kaniel, Ozoguz, Starks (2012). We set the last trading day of each month as the portfolio formation period and define a stock as a low- (high-) volume stock if its trading volume on the one-day formation period is among the lowest (highest) ten percent out of its 50 daily volumes prior to the formation period (inclusive). We eliminate stocks for which the price or volume data are missing on the portfolio formation day. Stocks which are not traded for nine or more days or whose prices fall

[7] The averages of the two corresponding CRSP market portfolio returns without dividends are 0.009% and 0.052%. The correlations between our market portfolio returns and those of CRSP are 0.994 and 0.993, respectively, for the value and equal-weighted estimates.

below $5 out of the 50 trading days are also removed from the sample to alleviate the microstructure concerns associated with these securities.

We exclude the stocks with less than one year of trading history to mitigate backfilling biases, and those delisted from the exchanges one year prior to the formation date. We also delete observations with an earnings or dividend announcement during a three-day window around the formation date because the volume-return relation during announcement periods may be different than in non-announcement periods (e.g., Kandel and Pearson, 1995). The portfolios are rebalanced monthly by sorting all remaining stocks into ten low- and high-volume portfolios based on their volume classification at the end of each month (t). We then compute both value- and equal-weighted returns for each of the ten portfolios for all trading days in month ($t + 1$). The value-weighted high volume return premium (VWHVP) is the difference between the value-weighted portfolio return on the highest volume decile and the return on the lowest volume decile. The equal-weighted high volume return premium (EWHVP) is similarly defined.[8]

Rows 5 and 6 of Panel A provide descriptive statistics of the daily value- and equal-weighted high volume premiums (VWHVP and EWHVP), which are estimated using the sample period of July 1963 through December 2010. The daily average of the value-weighted HVP is 0.027%, which is slightly higher than the CRSP value-weighted market portfolio returns of 0.021%. The bottom line of Panel A shows that, not surprisingly, the alternative equal-weighted volume premium is more than twice as high as the value-weighted one (0.061%).[9]

---

[8] This volume classification follows from Gervais, Kaniel, and Mingelgrin (2001) and Kaniel, Ozoguz, and Starks (2012), although our methodology does differ from these two studies in the way portfolios' formation period and test period are chosen.

[9] Following Kaniel, Ozoguz, and Starks (2012), we also find that the high volume premiums are similar to the reported ones if a stock is eliminated from the portfolio if its price falls in the lowest 5 percent of the market during the 49-day reference period. The average premiums are 0.029% and 0.075% for the value- and equal- weighted high volume portfolios, respectively. Both are higher than their counterparts in Table 1. The in-sample and out-of-sample Granger causality test results based on these estimates are also very close to those benchmarks reported later in tables 2-4.

We also compute the volume premium on a monthly basis. Panel B of Table 1 presents monthly HVP estimates for the full sample as well as two sub-sample periods. The average monthly value-weighted HVP is a statistically significant value of 0.57%, which is close to the similarly defined value premium of 0.54%. The estimate for the equal-weighted HVP is 1.29%, reasonably close to the 20-day holding period returns of 1.12% for similarly constructed portfolios by Kaniel, Ozoguz, and Starks (2012, Table 2). Panel B also shows that the averages of value- and equal-weighted HVP decrease from 0.62% and 1.34% in the period July 1963-December 1999 to 0.43% and 1.13% in January 2000-December 2010 for VWHVP and EWHVP, respectively. The differential changes imply that the decrease in the high volume premium is more significant for large stocks than for small stocks.

4. Empirical Results

As a preliminary but intuitive way to evaluate the volume-return relation, we first sort all daily market portfolio returns into ten groups with equal numbers of observations based on the one-period lagged trading volume. Figure 2 plots the simple average returns of each of these ten groups from low to high. There is no clear lead-lag relationship between trading volume and subsequent stock returns for the value-weighted portfolios, possibly with the exception of the two high-volume ones. In contrast, the positive volume-return relation is nearly monotonic among the ten equally-weighted portfolios. Quantitatively, stocks in the lowest-volume decile on average have a next-day return of −0.086%, and those in the highest-volume decile have a return of 0.360%. The spread is a statistically significant 44.6 basis points.

To evaluate the hypothesis that (past) volume predicts stock returns using out-of-sample regressions, we obtain the one-step-ahead recursive forecasts from four specifications of Model (1): C, R, U, and W. The first model we consider is one with a constant as the sole explanatory variable (the C model). In this simple model, the one-step-ahead forecast for the market return on day (t + 1) is simply the up to day t historical average returns. We include model C as the benchmark because, as pointed out by Mayfield (2004), although a substantial body of research shows that expected returns vary over time, the naïve static approach of estimating the risk premium as the simple average of historical excess stock returns remains the most commonly employed method in practice. The R model is an autoregressive regression and includes past returns as the sole predictive variables. The U model includes both past returns and past trading volume. Many studies have suggested that trading volume is related to volatility[10]. Therefore, market volatility could be a confounding factor in testing for causality between volume and returns. To address this issue, we further extend model U to include past realized variance as an additional predictive variable for returns, which is denoted as model W. The realized variance is the sum of the squared daily returns in the past three months.

Basic Results

As the starting point, we estimate these four models using daily observations from the first ten years in our U.S. sample (July 1963-June 1973). To determine the lag orders in model (1) (L_y , L_x, and L_z), we use a model selection approach via the Bayesian information criterion (BIC), assuming that the maximum lag order is 22 (approximately the number of trading days in a month). The first set of one-step-ahead out-of-sample forecasts for the market portfolio returns are generated using the estimated coefficients and observed values of the predictive variables. The models are then re-estimated and new forecasts are generated after each daily observation is sequentially added to the estimation sample for the remaining 37 years of data. The forecast errors are formed by the differences between observed returns and the four forecasted returns.

Table 2 summarizes the performance of the four forecasting models where trading intensity is represented by the value-weighted turnover (VWVOL) in the left panels and the equal-weighted turnover (EWVOL) in the right panels. For forecast evaluations, we first consider root mean squared

[10] A short list includes Lamoureux and Lastrapes (1990), Andersen (1996), Gomes (2005), Li and Wu (2006), He and Velu (2014), and Do et al. (2014).

forecast errors (RMSFE) of the four competing models. Panel A shows that model U that includes trading volume VWVOL and the model R that does not include the variable generate essentially the same forecast errors (1.063% after rounding) for the full sample period 1963-2010. However, both models underperform the simple historical average forecasts which have a RMSFE of 1.060%. The more complicated model W that also includes market volatility carries even larger forecast errors of 1.064%. In the middle of Panel A, we further test if the forecasts from the four models are statistically different from each other by employing the ENC-NEW encompassing test. The null hypothesis is that the forecasts from models in the first column (H_0) encompass those from models in the first row (H_A). Note that model W nests U, which in turn nests R. The C model is nested by all other three. We are interested in three null hypotheses: C encompasses U, R encompasses U, and U encompasses W. The null hypotheses that C encompasses U is strongly rejected, which contradicts the RMSFE measures. The hypothesis that R encompasses U is also rejected at the 5% level, meaning that lagged trading volume contains additional useful information about the next day's excess market return relative to the pure autoregressive model R. In line with the simple RMSFE measure, the test for model U encompassing W also concludes that market volatility has no significant predictive power beyond what is captured by lagged returns and the trading volume. Finally, we report in the bottom of Panel A realized utility levels associated with the four forecasting models. Model R has an annualized utility of 10.322%, which is slightly higher than that of the U model (10.187%).[11] Therefore, the rankings of models R and U by monetary gains are in line with those based on the RMSFE metric. Similarly, the realized utility of model W (9.835%) is lower than those of both models R and U, which is also consistent with the ranking by RMSFE and the encompassing test. However, the benchmark historical average, while delivering the smallest average forecast errors, attains the lowest level of utility.

[11] Following the literature, we set the risk aversion parameter r at 3. We also constrain the equity share in the optimal portfolio to the closed interval [0, 1], excluding short sales. Varying these parameters changes the magnitudes of the computed utility estimates but generally does not alter the models' rankings.

Panel B presents quite a different picture of model rankings for the equal-weighted portfolio. The unrestricted model U generates the smallest average forecast errors followed by models W, R, and C. The hypotheses that model R and C encompass U are rejected at any conventional level, further confirming that trading volume helps predict returns to the equal-weighted portfolio. Realized utility based on model U's forecasts is 0.73% per annum higher than that of the R model. Both U and R models beat the historical average forecasts in economic gains by large margins. There is also some gain in utility (0.28%) by model W which further adds information on market volatility for forecasting. Nevertheless, in support of the RMSFE measure, the test of model U encompassing W has a statistic of 0.243 which is insignificant.

In panels C through F we compare the forecast performance of the four competing models for the two sub-sample periods. For the value-weighted portfolio, Panel C shows no significant evidence that model U generates better forecasts than model R during the period 1973-1999, although both appear to perform somewhat better than the historical average. Market volatility shows no additional predictive power in addition to past returns and the trading volume by all three evaluation statistics. As in Panel B for the equal-weighted portfolio, Panel D provides strong evidence that forecasts from model U are more accurate than those from model R. Although model W has the same RMSFE as model U, both the realized utility measure and the encompassing test result suggest that market volatility does contain some useful information for future stock return not captured by trading volume during the first sub-sample period.

In striking contrast to the results in panels C and D, panels E and F clearly show that, during the more recent 2000-2010 period, the simple forecasts based on the historical averages are more accurate for returns to both types of portfolios than those generated from the other three competitors.[12] This

[12] Recall that forecasts are recursively generated. As for the full sample analysis, the initial estimation sample for the first sub-sample analysis includes first ten years of data (1963-1973). And the initial estimation sample for the second sub-sample forecasting exercises spans a longer period of 1963-1999. This is why the null hypothesis of

result based on the RMSFE measure is supported by the other two forecast evaluation methods for value-weighted portfolios. It however contradicts the rankings by the latter two for equal-weighted portfolios. The realized utility of model U is 24.561%, which is about 1% higher than that of the R model and more than 10% higher than that of the historical average. Furthermore, neither model R nor C encompasses model U.

Replacing the direct measures of trading volume VWVOL and EWVOL with the high-volume premiums VWHVP and EWHVP, we re-estimate and generate forecasts from the four models we have just examined.[13] Their performances are summarized in Table 3. The main findings from Table 2 all hold true in Table 3. In particular, the additional predictive power of the high volume return premium is significant only for the equal-weighted portfolio. More specifically, the U model performs better than the other three models by all three measures reported in Panel B for the full sample and in Panel D for the first sub-sample. Note, however, that the annualized utility gains relative to the R model are small in both cases (0.4% and 0.5%, respectively). Finally, as in Table 2, there is no consistent evidence of a positive relationship between EWHVP and future market returns during the 2000-2010 forecasting period. Model U with EWHVP has a larger RMSFE than model C (1.215% vs. 1.195%), although it still contains additional information useful for forecasting according to the encompassing test. The utility gain of model U over model R is merely 0.12%, essentially nonexistent.

Robustness

---

model R encompassing model U is rejected in Panel C only at the 10% level with a statistic of 1.488, while it is rejected in Panel E at the 5% level with a smaller statistic of 1.219.

[13] In studying the predictive power of aggregate trading volume, we have followed the literature and used market portfolio returns without dividends (capital gain only). Furthermore, we use our own estimates rather than CRSP estimates because stocks in NASDQ are not included in our market portfolio until November 1982 due to missing volume. However, literature focusing on stock return predictability especially in the out-of-sample context often use total returns. Therefore, to be comparable to the existing evidence, when examining the predictive power of the high volume premium we use CRSP market portfolio returns with dividends (VWRET and EWRET), assuming that high volume premium estimates based on NYSE/ASE stocks are also representative of those based on NASDQ stocks for the 1982 and earlier period (the summary statistics for VWRET and EWRET are reported in the bottom two rows of Panel A, Table 1). Nevertheless, our basic findings hold if returns without dividends are used instead.

Studies focusing on out-of-sample forecast errors rather than within sample pricing errors have found that empirical results from dynamic models may be sensitive to the choice of predictive variables, assets, and in-sample window lengths (e.g., Cooper and Gulen, 2006). To guard against possible data snooping bias, we re-examine the performance of the four forecasting models, doubling the length of the initial estimation sample from 10 years (as used in tables 2 and 3) to 20 years (July 1963-July 1983). The new results are summarized in Appendix tables 1 and 2 for turnover and the high volume return premium, respectively. Note that because forecasts are generated recursively, the change in the initial in-sample length has no impact on the second forecasting period (January 2000-December 2010). Comparing the results in the two appendix tables to their counterparts in tables 2 and 3, we can easily see that changing the in-sample window length does not affect any of our earlier findings. In short, turnover and the volume premium display additional predictive power for future returns only in equal-weighted portfolios by both statistical and monetary measures. Utility gains for the unrestricted model U are in the range of 0.65%~0.78% when using equal-weighted turnover as the extra predictor, and 0.47%~0.71% when using the volume premium. These estimates are similar in magnitudes to the initial estimation using the first ten years of data.

We have so far found from tables 2 and 3 that trading volume consistently helps predict future returns only for the equal-weighted portfolio. This suggests that the lead/lag relationship may exist more prominently in small stocks.[14] To shed further light on this issue, we also study the predictive power of aggregate trading volume for returns to portfolios of various sizes. Specifically, we sort stocks into small, medium, and large portfolios based on the breakpoints for the low 30%, medium 40%, and high 30% of the ranked values of market capitalization. The value- and equal-weighted returns are computed for each of these three size portfolios. To save space, Appendix Table 3 only presents the

[14] One possible explanation from microstructure theory is that small stocks are generally less liquid than their large counterparts. A sell or buy order of the same magnitude can have much larger price impacts on small stocks.

performance of the four models in forecasting the daily returns to the small and large portfolios. The results for small stocks are tabulated in panels A, C, and E in the left half of the table, and those for large stocks in panels B, D, and F in the right half. By all three forecast evaluation criteria, model U with the extra variable EWVOL performs better than model R and the historical average for small stocks in the full and the first sub-samples. The forecasting ability of the equal-weighted volume premium is also significant in the recent sub-sample period by the two statistical measures. However, incorporating this variable in the forecasting information set decreases an investor's utility by about 0.4% compared to model R. There is also no economic gain in forecasting small stock returns by using past market volatility in either the full sample or the two sub-samples. Consider the three panels in the right half of the table, we find that model U does not outperform model R for large stocks by the average forecast errors and the realized utility measures, although the encompassing test results continue to show that volume information relevant for forecasting future returns is not fully captured by historical returns. Overall, the predictive power of trading volume recorded in Table 2 is indeed mainly driven by small stocks.

Both Campbell, Grossman, and Wang (1993) and Llorente et al. (2002) predict a nonlinear relation between trading volume and stock returns. Hiemstra and Jones (1994), Chuang, Kuang, and Lin (2009), Chen (2012), and Ciner (2015) also present empirical evidence of nonlinear between the two variables. However, model (1) does not allow for such nonlinearity in the test. To address the potential misspecification issue, we adopt the robust method of quantile regressions (Koenker, 2005). The lag structures for the quantile regressions are the same as those used in the linear models in tables 2 and 3. In deriving forecasts from the quantile regressions, we estimate a total of 99 regressions for quantiles 0.01, 0.02 .., 0.99. We then generate one-step-ahead return forecasts for each quantile and form the mean forecast by taking a simple average over these 99 estimates. We find that the quantile models perform slightly better than the linear ones for the equal-weighted portfolios during the 1973-1999 sub-sample period when the predictive power of trading volume is most noticeable. Overall the restricted and unrestricted quantile regressions provide strong evidence for the predictive power of trading volume as in tables 2 and 3 using the linear models.

In sum, the out-of-sample forecasting results provide partial support for the findings of the in-sample regressions. The additional predictive power of information on lagged trade activity is corroborated by the out-of-sample tests for future stock returns to equal-weighted portfolios but not for returns to value-weighted ones.

5. International Evidence

In this section we extend our analysis on the U.S. equity market to other developed and emerging markets with two goals. Our first goal is to examine whether the high volume return premium documented by Kaniel, Ozoguz, and Starks (2012) up to year 2001 continues to exist in the international markets after we include data from the past decade. Our second goal is to examine whether the out-of-sample predictive power of trading volume on stock returns found in the U.S. market is also present in international markets. We are interested in whether the burst of the technology bubble and the economic recession around the new millennium, and the latest financial crisis and the ensuing economic recession have had significant impact on the volume premium and the volume-return relationship in the international markets. To develop perspective on whether the trading volume effect is integrated across regions, we further examine whether turnover and the volume premium in the U.S. market have predictive power for stock returns in the other markets. This part of the analysis is motivated by such empirical evidence as the spillover effect documented in the literature of idiosyncratic volatility (e.g., Guo and Savicks, 2008) and the leading role for the U.S. with respect to monthly international excess return predictability reported in Rapach, Strauss, and Zhou (2013).

Data

For international analysis, we first consider stock markets in six non-U.S. Group of Seven (G-7) countries: Canada, France, Germany, Italy, Japan, and the United Kingdom (U.K.). We also study another 12 countries whose primary exchanges make up the top 20 worldwide major stock exchanges by market capitalization. These 12 countries/regions, including both developed and developing economies, are Australia, Brazil, China, Hong Kong, India, Korea, Russia, Singapore, South Africa, Spain, Switzerland, and Taiwan. Our international data on firm-level daily returns, trading volume and monthly market capitalization, both for currently trading and defunct securities, are obtained from Thomson-Reuter Datastream. Appendix 1 summarizes data filtering and the estimation of the high volume premium in the international markets.

Table 4 reports the start date and the average number of firms that the samples comprise for the G-7 countries (excluding the U.S.) in Panel A and for the other 12 countries in Panel B. Although the end date is December 2010 for all 18 countries, the effective sample start date ranges from January 1977 to September 2005. The average number of firms considered also varies considerably from 61 of Brazil to 1712 of Japan, with an average of 486 and a median of 410.

Table 5 provides mean statistics for the monthly market portfolio returns and the two types of high volume return premiums for the non-U.S. G-7 countries in Panel A and for the 12 other countries in Panel B. The third and the fourth columns of the table present our estimates of value- and equal-weighted market portfolio returns. The value-weighted estimates are close to Datastream's total market estimates for the respective countries in Column 2. The equal-weighted market portfolio returns are higher than the value-weighted ones with the exceptions of Germany, South Africa, Spain, and Switzerland. Column 5 reports the average value-weighted high volume premium. The premium is positive for all non-U.S. G-7 countries and statistically different from zero for France, Japan, and UK at the 5% level. The equal-weighted premium is both positive and statistically significant in all six countries. These results are similar to those reported in Kaneil, Ozoguz, and Starks (2012) who find that significant volume effects exist in all G-7 countries but Italy. The magnitudes of our estimates are also similar to their counterparts in Kaneil, Ozoguz, and Starks (2012) despite using quite different sample periods. For the remaining 12 countries, the value-weighed premium is positive in all but three markets, China, Korea, and Spain. Nevertheless, the premium is statistically significant in Hong Kong only. The equal-weighted premium is estimated more precisely and larger compared to the value-weighted one. It is significant in eight countries, although the estimate is again negative for the Korean market.

Out-of-sample forecast evaluation

In conducting out-of-sample forecasting analysis, we include first ten years of data in the initial samples for most countries. However, the sample start dates of Germany, Brazil, and Russia are 1999 or later. Therefore, the in-samples only contain the first five years of data for Germany and Brazil, and the first three years of data for Russia. Although we consider the same set of forecasting models for each of the 18 markets as we did for the U.S. market, to save space, we only report the results for the following three models: model C with a constant only, model RZ with past returns and market volatility, and model W with all three sets of predictive variables (past returns, measures of trading intensity, and market volatility). Obviously, model W nests RZ, which in turn nests C. To test if past trading volume and the high volume premium contain useful information in forecasting current returns, we consider the null hypothesis that forecasts from the restricted model RZ encompass those from the unrestricted model W. If this hypothesis is rejected we further test if the forecasts of the simple historical averages encompass those from model W.

We first present the test results for the six G-7 countries in Table 6.[15] As shown in Panel A, model W that uses value-weighted turnover as a predictor performs no better than model RZ that does not include the variable in terms of the root mean squared forecast errors (RMSFE) of five markets. It achieves relatively high realized utility only in the French market (2.4% per annum vs. 0.74% by model C and −0.75% by RZ). The encompassing test results are consistent with the rankings by the RMSFE measure. The U.K. appears to be an exception where turnover used in model W has additional predictive power by all three measures in comparison to the RZ model. We also reject the null hypothesis that the historical average forecasts contain all of the useful information that model W does. The gain in utility by including volume information is about 1% relative to the RZ model, and 0.7% higher than forecasts from model C. In Panel B we use equal-weighted turnover as a proxy for trade intensity. The evidence is more consistent in the sense that turnover does not have additional predictive power for future returns for any market by all three evaluation methods after past returns and market volatility are controlled for in the regressions.

The results tabulated in Panel C suggest that the unrestricted model W only performs better than both C and RZ models by the two statistical measures in the German market. It however obtains lower utility than the restricted RZ model by 0.9%. The evidence is similar in the U.K. market. If proxied by the equal-weighted volume premium in forecasting equations (Panel D), trading volume shows consistent evidence of forecasting ability for returns in the Canadian market by all three evaluation measures. The volume premium also appears to have predictive power for German stock returns according to the utility gain and the encompassing test result. However, the W model with the volume premium produces relatively larger average forecast errors than the historical average forecasts.

[15] The lag orders of the predictive variables (past returns, volume and volatility) in model (1) are determined by minimizing the BIC information criterion for each country. To save space, they are not reported.

Tables 7 and 8 summarize the results of the out-of-sample forecast comparisons for the remaining 12 countries using turnover and the high volume premium, respectively. Given that the sample sizes are generally small for the international markets (in terms of the period spanned and/or the number of stocks included), inferences drawn from the statistical measures may be more likely to diverge from those drawn from the economic measure. For this reason and for brevity as well, from now on we define the two trading activity proxies as having predictive power for stock returns if (1) the unrestricted model W that uses this information achieves higher utility than both model RZ and the simple historical average forecasts (model C); and (2) the latter two models have larger forecast errors on average and do not encompass the unrestricted model W at the 5% significance level. Applying these rules we can see from Panel A of Table 7 that model W with the variable of value-weighted turnover performs better than the C and RZ models in three markets, India, Russia, and Switzerland. Panel B presents results for the aggregate measure of turnover when it is constructed on an equal-weight basis. Although the equal-weighted turnover EWVOL helps forecast returns by the encompassing test in Australia, Hong Kong, and Taiwan, there is no economic gain in doing so since the realized utility of the unrestricted model is slightly lower than that of the restricted model RZ in all three cases.

As shown in Panel A of Table 8, the only market in which the value-weighted volume premium VWHVP helps forecast future returns by all three evaluation methods is India. Still, the economic benefit is admittedly small, an increases of less than 0.1% per annum relative to the model without the volume information. And there is no consistent evidence that the cross-sectionally constructed volume premium helps forecast returns in any of the other 11 countries.

Like the value-weighted measure, the equal-weighted high volume premium does not have predictive power by three performance measures for all but one country (Panel B). The exception is the Chinese stock market where the unrestricted model W performs better than both the historical average and the restricted RZ model by both statistical and economic measures. Economically, an investor would be better off with additional annual returns of 1.1% if volume information is used in rebalancing her/his portfolio. That trading volume shows significant out-of-sample predictive power for stock returns in the Chinese market may be related to its unique institutional arrangement. During a significant portion of the sample period analyzed in the paper, a Chinese company may issue both A-shares in mainland China and H-shares in Hong Kong. This market segmentation may cause illiquidity and under-reaction. In addition, non-institutional investors have played an important part of daily trading in the Chinese market, which may also have contributed to Chinese market dynamics different than those in other markets.

Finally, we briefly discuss the empirical results on whether trading volume from the U.S. market contains additional information for forecasting returns in the 18 other markets controlling for volume and volatility information from the domestic markets. We consider one-step-ahead forecasts from the following four models. Model C includes an intercept only. Model W contains past returns, volume, and volatility of a domestic market. The other two models, WR and WRV, augment model W with U.S. market information. Specifically, WR adds one lag of U.S. market returns, and WRV adds both one lag of U.S. market returns and one lag of trading volume. Based on the four sets of return forecasts, we compute annualized utility levels according to equations (3) and (4). Appendix Table 4 presents the estimated economic gains associated with model C, W, WR, and WRV for each market when trading activity is approximated by two aggregate turnover series. To facilitate the presentation, we again define U.S. trading volume information as having predictive power for stock returns on another market if model WRV that includes this information (1) achieves higher utility than the more restricted models C, W, and WR models, and (2) the latter three models have larger forecast errors and do not encompass model WRV at the 5% significance level. Although past U.S. market returns contain substantial information for predicting current returns to the other markets, the value-weighted turnover of the U.S. market only provides additional information useful for predicting the Indian market. The added economic gain is 2.8% per annum. The equal-weighted turnover shows predictive power in three more markets (Canada, Japan, and Korea). Nevertheless, the gains in utility are smaller, ranging from 0.12% to 0.74%.

Judged on all three statistical and economic criteria, we also find little evidence that U.S. market trading activity carries additional predictive power for international markets when it is represented by the high volume return premium. We see from Appendix Table 5 that when the value-weighted volume premium is included in the models, investors' welfare improves in the French and Russian markets. For the equal-weighted portfolio investment, the improvement is found in one market only (Hong Kong). And in all three cases, the economic gains are small (0.16~0.41%).

6. Concluding Remarks

We provide a comprehensive reexamination of the lead-lag relationship between trading volume and stock returns. Our contribution to the literature rests importantly in the paper's emphasis being on detailing out-of-sample evidence, thereby complementing in-sample findings in earlier empirical studies. In the U.S. market, higher trading volume, whether measured by aggregate time series of turnover or by the cross-sectionally constructed high volume return premium, is indeed followed by higher stock returns. However, such predictive power of trading volume should be interpreted with caution in that the associated economic gain is quantitatively small for the market as a whole. Furthermore, the predictive power of trading volume becomes insignificant even statistically in the more recent period featuring high-profile high-speed trading. Similarly, with only a few exceptions, the predictive power of trading volume for stock returns fail to pass the rigorous statistical and economic tests in out-of-sample regressions for most of the non-U.S. markets.

The lack of significant out-of-sample predictive power of trading volume for stock returns is in stark contrast to the existing in-sample analyses that have often found that a dynamic volume-return relationship exists. This empirical finding suggests that we may need rethink about the theoretical models which predict that trading volume is significantly related to future stock returns as reviewed earlier in the paper. Our finding is probably more important for practitioners who might otherwise consider exploiting the relation for timing the market and forming aggressive trading strategies.

References

Amihud, Y., 2002. Illiquidity and stock returns: Cross-section and time-series effects. Journal of Financial Markets 5, 31-56.

Amihud, Y., Mendelson, H., 1986. Asset pricing and the bid-ask spread. Journal of Financial Economics 17, 223-249.

Andersen, T.G. 1996. Return Volatility and Trading Volume: An Information Flow Interpretation of Stochastic Volatility. Journal of Finance 51, 169-204.

Blume, L., D. Easley, and M. O'Hara. 1994. Market Statistics and Technical Analysis: The Role of Volume. Journal of Finance 49, 153-81.

Butler, A.W., G. Grullon, and J.P. Weston. 2005. Can Managers Forecast Aggregate Market Returns? Journal of Finance 60, 963-986.

Campbell, J.Y., S.J. Grossman, and J. Wang. 1993. Trading Volume and Serial Correlation in Stock Returns. Quarterly Journal of Economics 108, 905-939.

Chen, S.-S. 2012. Revisiting the Empirical Linkages between Stock Returns and Trading Volume. Journal of Banking & Finance 26, 1781-1788.

Chen, J., H. Hong, and J. Stein. 2001. Forecasting Crashes: Trading Volume, Past Returns and Conditional Skewness in Stock Prices. Journal of Financial Economics 61, 345-381.

Chordia, T., R. Roll, and A. Subrahmanyam. 2011. Recent Trends in Trading Activity and Market Quality. Journal of Financial Economics 101, 243-263.

Chordia, T., and B. Swaminathan. 2000. Trading Volume and Cross-Autocorrelations in Stock Returns. Journal of Finance 55, 913-935.

Chuang, C.-C., C.-M. Kuan, and H.-Y., Lin. 2009. Causality in Quantiles and Dynamic Stock Return-Volume Relations. Journal of Banking & Finance 33, 1351-1360.

Ciner, C. 2015. Time Variation in Systematic Risk, Returns and Trading Volume: Evidence from Precious Metals Mining Stocks. International Review of Financial Analysis 42, 277-283.

Clark, T.E., and M.W. McCracken. 2001. Tests of Equal Forecast Accuracy and Encompassing for Nested Models. Journal of Econometrics 105, 85-110.

Cochrane, J.H. 2013. Finance: Function Matters, not Size. Journal of Economic Perspectives 27, 29-50.

Conrad, J., A. Hameed, and C. Niden. 1994. Volume and Autocovariances in Short-Horizon Security Returns. Journal of Finance 49, 1305-1329.

Conrad, J., and G. Kaul. 1988. Time-Variation in Expected Returns. Journal of Business 61, 409-425.

Cooper, M., and H. Gulen. 2006. Is Time-Series-Based Predictability Evident in Real Time? Journal of Business 79, 1263-1292.

Corradi, V., and N.R. Swanson. 2006. Predictive Density Evaluation. In Elliott, G., C.W.J. Granger, and A. Timmermann (eds.) *Handbook of Economic Forecasting*, Vol. 1, Elsevier, Amsterdam, pp. 197-284.

Datar, V.T., Naik, N.Y., Radcliffe, R., 1998. Liquidity and asset returns: An alternative test. Journal of Financial Markets 1, 203-220.

Do, H.X., R. Brooks, S. Treepongkaruna, and E. Wu. 2014. How Does Trading Volume Affect Financial Return Distributions. International Review of Financial Analysis 35, 190-206.

Easley, D., M. O'Hara, and P.S. Srinivas. 1998. Option Volume and Stock Prices: Evidence on Where Informed Traders Trade. Journal of Finance 53, 431-465.

French, K. 2008. The Cost of Active Investing. Journal of Finance 63, 1537-1573.

Gallant, A.R., P.E. Rossi, and G. Tauchen. 1992. Stock Prices and Volume. Review of Financial Studies 5, 199-242.

Gervais, S., R. Kaniel, and D. Mingelgrin. 2001. The High Volume Return Premium. Journal of Finance 56, 877-919.

Gervais S., Odean, T., 2001. Learning to be overconfident. Review of Financial Studies 14, 1-27.

Gomes, F.J. 2005. Portfolio Choice and Trading Volume with Loss-Averse Investors. Journal of Business 78, 675-706.

Granger, C.W.J. 1969. Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. Econometrica 36, 424-438.

Greenwood, R., and D.S. Scharfstein. 2013. The Growth of Finance. Journal of Economic Perspectives 27, 3-28.

Griffin, J.M. 2002. Are the Fama and French Factors Global or Country-Specific? Review of Financial Studies 15, 783-803.

Griffin, J.M., F. Nardari, and R.M. Stulz. 2007. Do Investors Trade More When Stocks Have Performed Well? Evidence from 46 Countries. Review of Financial Studies 23, 905-951.

Guo, H., and R. Savickas. 2008. Average Idiosyncratic Volatility in G7 Countries. Review of Financial Studies 21, 1259-1296.

He, X., and R. Velu. 2014. Volume and Volatility in a Common Factor Mixture of Distributions Model. Journal of Financial and Quantitative Analysis 49, 33-49.

Hiemstra, C., and J.D. Jones. 1994. Testing for Linear and Nonlinear Granger Causality in the Stock Price-Volume Relation. Journal of Finance 49, 1639-1664.

Hou, K., Peng, L., Xiong, W., 2009. A tale of two anomalies: The implications of investor attention for price and earnings momentum. SSRN working paper.

Ince, O.S., and R.B. Porter. 2006. Individual Equity Return Data from Thomson Datastream: Handle with Care! Journal of Financial Research 29, 463-479.

Kandel, E., and N.D. Pearson. 1995. Differential Interpretation of Public Signals and Trade in Speculative Markets. Journal of Political Economy 103,831-872.

Kaniel, R., A. Ozoguz, and L. Starks. 2012. The High Volume Return Premium: Cross Country Evidence. Journal of Financial Economics 103, 255-279.

Karpoff, J. 1987. The Relation between Price Changes and Trading Volume: A Survey. Journal of Financial and Quantitative Analysis 22, 109-126.

Koenker, R. 2005. Quantile Regression. Cambridge University Press, Cambridge.

Lamoureux, C.G., and W.D. Lastrapes. 1990. Heteroskedasticity in Stock Return Data: Volume versus GARCH Effects. Journal of Finance 45, 221-229.

Lee, B.-S., and O.M. Rui. 2002. The Dynamic Relationship between Stock Returns and Trading Volume: Domestic and cross-Country Evidence. Journal of Banking & Finance 26, 51-78.

Lesmond, D.A., 2005. Liquidity of emerging markets. Journal of Financial Economics 77, 411-452.

Lesmond, D.A., Ogden, J.P., Trzcinka, C.A., 1999. A new estimate of transaction costs. Review of Financial Studies 12, 1113-1141.

Li, J., and C. Wu. 2006. Daily Return Volatility, Bid-Ask Spreads, and Information Flow:

Analyzing the Information Content of Volume. Journal of Business 79, 2697-2739.

Liu, W., 2006. A liquidity-augmented capital asset pricing model. Journal of Financial Economics 82, 631-671.

Llorente, G., R. Michaely, G. Saar, and J. Wang. 2002. Dynamic Volume-Return Relation of Individual Stocks. Review of Financial Studies 15, 1005-1047.

Lo, A.W., and C. MacKinlay. 1988. Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test. Review of Financial Studies 1, 41-66.

Lo, A.W., and J. Wang. 2000. Trading Volume: Definitions, Data Analysis, and Implications of Portfolio Theory. Review of Financial Studies 13, 257-300.

Lo, A.W., and J. Wang. 2006. Trading Volume: Implications of an Intertemporal Capital Asset Pricing Model. Journal of Finance 61, 2805-2840.

Malcolm, B., and J.C. Stein. 2004. Market Liquidity as a Sentiment Indicator. Journal of Financial Markets 7, 271-299.

Marquering, W., and M. Verbeek. 2004. The Economic Value of Predicting Stock Index Returns and Volatility. Journal of Financial and Quantitative Analysis 39, 407-429.

Mayfield, E.S. 2004. Estimating the Market Risk Premium. Journal of Financial Economics 73, 465-496.

Rapach, D.E., J.K. Strauss, and G. Zhou. 2010. Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy. Review of Financial Studies 23, 821-862.

Rapach, D.E., J.K. Strauss, and G. Zhou. 2013. International Stock Return Predictability: What is the Role of the United States? Journal of Finance 68, 1633-1662.

Statman, M., Thorley, S., Vorkink, K., 2006. Investor overconfidence and trading volume. Review of Financial Studies 19, 1531-1565.

Thornton, D. L., and G. Valente. 2012. Out-of-Sample Predictions of Bond Excess Returns and Forward Rates: An Asset Allocation Perspective. Review of Financial Studies 25, 31413168.

Welch, I., and A. Goyal. 2008. A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. Review of Financial Studies 21, 1455-1508.

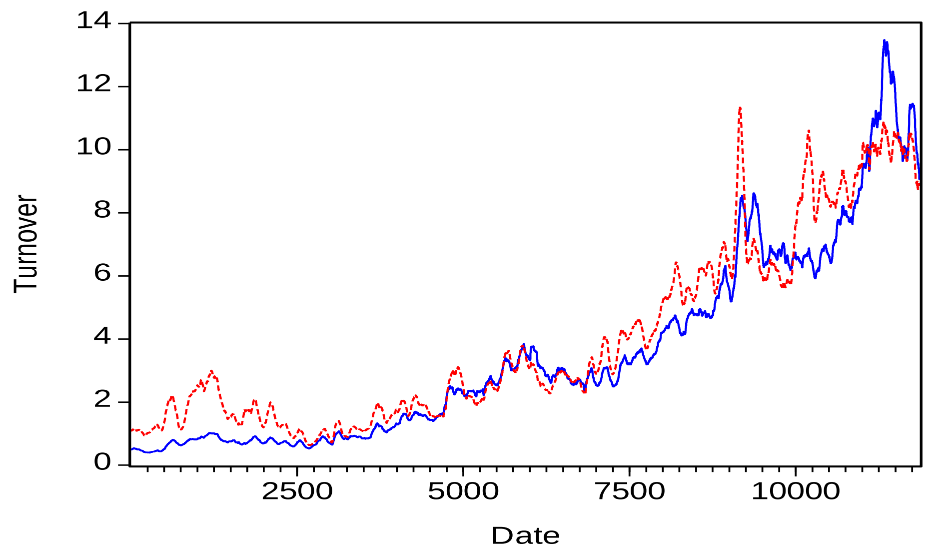


Figure 1. Daily value-weighted and equally-weighted turnover in the U.S. market (raw measures)

The solid and broken lines are 100-day moving averages of value- and equal-weighted turnover ratios, respectively. They are estimated using daily data from July 1, 1963 to December 31, 2010. The first observation on the horizontal axis corresponds to November 20, 1963 and the last observation 11860 corresponds to December 31, 2010.

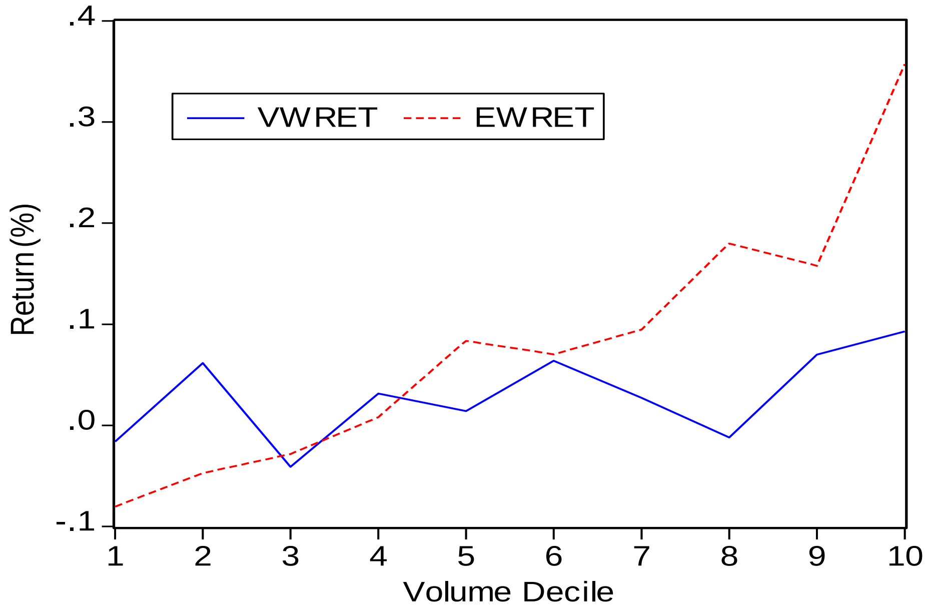


Figure 2. Average daily market returns when sorted on the lagged volume

This figure reports on subset daily U.S. market returns from July 1963 to December 2010, which are sorted into ten decile groups based on the one-day lagged trading volume. Group 1 is the lowest volume decile and Group 10 the highest decile.

Table 1. Descriptive statistics of volume and returns in the U.S. market

VWRET and EWRET are value- and equal-weighted market portfolio returns, and VWVOL and EWVOL are the corresponding turnover. VWHVP and EWHVP are value- and equal-weighted high volume return premiums. They are estimated using U.S. market daily data from July 1, 1963 through December 31, 2010, of which the first 100 observations are used for detrending the trading volume series (therefore, the effective sample for the first four rows starts at November 21, 1963). VWMKT and EWMKT are CRSP market portfolio returns in excess of the risk free rate.

Note that our VWRET and EWRET are not comparable to those of the CRSP counterparts. Because the volume data of NASDAQ stocks are not included in the database until November 1982, these stocks are excluded from the estimation of the two portfolio returns during this period. Differences in other stock selection criteria for the whole sample period also contribute to the differences between CRSP's estimates and ours.

The mean and standard deviation are both in percentage forms. AR(1) is the first-order autocorrelation coefficient. Q(1) is the Ljung-Box Q-statistic for the null hypothesis that there is no autocorrelation up to order 1, which follows the χ^2 distribution with one degree-of-freedom. Symbols *, **, and *** indicate significant at the 10%, 5%, and 1% levels, respectively.

Panel A. Daily volume and returns (July 01, 1963-December 30, 2010)						
Variable	Mean	Std. dev.	Skewness	Kurtosis	AR(1)	Q(1)
VWVOL	1.197**	0.217	–0.252***	2.959***	0.604	>100***
EWVOL	0.879	0.237	0.048**	1.614***	0.734	>100***
VWMKT	0.010	0.991	–0.522***	17.474***	0.064	48.482***
EWMKT	0.061***	0.921	–0.529***	12.092***	0.234	>100***
VWHVP	0.027***	0.727	–0.135***	18.428***	0.059	41.395***
EWHVP	0.061***	0.452	0.779***	8.525***	0.088	92.205***
VWRET	0.021**	0.982	−0.533***	17.118***	0.069	57.233***
EWRET	0.060***	0.831	−0.591***	14.564***	0.237	>100***
Panel B. Monthly volume premiums						
	July 1963-Dec. 2010		July 1963-Dec. 1999		Jan. 2000- Dec. 2010	
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
VWHVP	0.574***	3.134	0.618***	2.865	0.430	1.197
EWHVP	1.294***	2.568	1.343***	2.541	1.132***	4.483

Table 2. Forecasting performance of trading volume for stock returns in the U.S. market (trading volume proxied by turnover)

The general form of the forecasting model is $y_t = \alpha + \sum_{m=1}^{L_y} \beta_m y_{t-m} + \sum_{n=1}^{L_x} \gamma_n x_{t-n} + \sum_{r=1}^{L_z} \lambda_r z_{t-r} + \varepsilon_t$, (1)

where y_t is value-weighted (equal-weighted) market portfolio returns VWMKT (EWMKT), x_t is the corresponding market turnover VWVOL (EWVOL), and z_t is market volatility. Model C includes a constant α only. Model R includes α and lagged y_t. Model U includes α and lags of y_t and x_t. Model W includes α and lags of y_t, x_t, and z_t.

Root mean squared forecast errors (RMSFE) and the annualized realized utility have been multiplied by 100 for ease of presentation. All the statistics in the table are based on one-step-ahead recursive forecast errors from the above models for the daily excess returns to the U.S. stock market portfolio.

ENC-NEW is the encompassing test statistic of Clark and McCracken (2001), in which the associated null hypothesis is that forecasts from the model in the first column (H_o) encompass those from the model in the second row (H_A). Thus, the H_o model is preferred to H_A. The critical values of the ENC-NEW test are linearly interpolated from the unpublished Appendix of Clark and McCracken (2001). The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Lag orders in models using value-weighted volume and returns				Lag orders in models using equal-weighted volume and returns			
	C	R	U	W	C	R	U	W
L_y	0	2	2	2	0	3	3	3
L_x	0	0	1	1	0	0	1	1
L_z	0	0	0	1	0	0	0	1
	Panel A. VWVOL & VWMKT, 1973-10				Panel B. EWVOL & EWMKT, 1973-10			
	RMSFE				RMSFE			
	1.060	1.063	1.063	1.064	0.956	0.939	0.935	0.936
	Encompassing test statistic				Encompassing test statistic			
C			149.933***				839.127***	
R			3.112**				60.535***	
U				−8.551				0.243
	Realized Utility				Realized Utility			
	5.074	10.322	10.187	9.835	13.733	32.722	33.448	33.730
	Panel C. VWVOL & VWMKT, 1973-99				Panel D. EWVOL & EWMKT, 1973-99			
	RMSFE				RMSFE			
	0.901	0.898	0.898	0.899	0.744	0.704	0.700	0.700
	Encompassing test statistic				Encompassing test statistic			

C			221.754***				1099.131***	
R			1.488*				64.656***	
U				−3.830				5.005***
	Realized Utility				Realized Utility			
	6.800	16.792	16.693	16.208	15.203	36.562	37.188	37.627
	Panel E. VWVOL & VWMKT, 2000-10				Panel F. EWVOL & EWMKT, 2000-10			
	RMSFE				RMSFE			
	1.366	1.380	1.379	1.381	1.329	1.341	1.337	1.339
	Encompassing test statistic				Encompassing test statistic			
C			−4.842				108.919***	
R			1.219**				11.874***	
U				−3.459				−1.252
	Realized Utility				Realized Utility			
	0.971	−5.042	−5.263	−5.298	10.238	23.598	24.561	24.470

Table 3. Forecasting performance of trading volume for stock returns in the U.S. market

(trading volume proxied by the high volume premium)

The general form of the forecasting model is $y_t = \alpha + \sum_{m=1}^{L_y} \beta_m y_{t-m} + \sum_{n=1}^{L_x} \gamma_n x_{t-n} + \sum_{r=1}^{L_z} \lambda_r z_{t-r} + \varepsilon_t,$ (1)

where y_t is value-weighted (equal-weighted) market portfolio returns VWMKT (EWMKT), x_t is the corresponding high volume return premiums VWHVP (EWHVP), and z_t is market volatility. Model C includes a constant α only. Model R includes α and lagged y_t. Model U includes α and lags of y_t and x_t. Model W includes α and lags of y_t, x_t, and z_t.

Root mean squared forecast errors (RMSFE) and the annualized realized utility have been multiplied by 100 for ease of presentation. All the statistics in the table are based on one-step-ahead recursive forecast errors from the above models for the daily excess returns to the U.S. stock market portfolio.

ENC-NEW is the encompassing test statistic of Clark and McCracken (2001), in which the associated null hypothesis is that forecasts from the model in the first column (H_o) encompass those from the model in the second row (H_A). Thus, the H_o model is preferred to H_A. The critical values for linear models of the ENC-NEW test are linearly interpolated from the unpublished Appendix of Clark and McCracken (2001). The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Lag orders in models using value-weighted volume and returns				Lag orders in models using equal-weighted volume and returns			
	C	R	U	W	C	R	U	W
L_y	0	2	2	2	0	5	5	5
L_x	0	0	1	1	0	0	1	1
L_z	0	0	0	1	0	0	0	3
	Panel A. VWHVP & VWRET, 1973-10				Panel B. EWHVP & EWRET, 1973-10			
	RMSFE				RMSFE			
	1.057	1.059	1.059	1.060	0.857	0.842	0.841	0.847
	Encompassing test statistic				Encompassing test statistic			
C			166.687***				851.692***	
R			2.395*				10.720***	
U				−10.216				−20.798
	Realized Utility				Realized Utility			
	4.277	13.023	12.895	12.817	13.439	31.283	31.676	31.409
	Panel C. VWHVP & VWRET, 1973-99				Panel D. EWHVP & EWRET, 1973-99			
	RMSFE				RMSFE			
	0.886	0.879	0.879	0.880	0.666	0.622	0.621	0.627
	Encompassing test statistic				Encompassing test statistic			
C			261.705***				1219.171***	
R			0.058				11.402***	
U				−3.772				−7.251
	Realized Utility				Realized Utility			
	5.801	19.535	19.269	19.218	15.062	35.618	36.124	35.457
	Panel E. VWHVP & VWRETD, 2000-10				Panel F. EWHVP & EWRETD, 2000-10			

	RMSFE				RMSFE			
	1.381	1.398	1.398	1.400	1.195	1.216	1.215	1.221
	Encompassing test statistic				Encompassing test statistic			
C			−7.901				89.519***	
R			1.351**				2.157***	
U				−4.369				−8.089
	Realized Utility				Realized Utility			
	0.633	−2.524	−2.322	−2.466	9.558	20.929	21.050	21.739

Table 4. Basic information on international samples of stock returns and trading volume

This table reports the sample start dates and the average numbers of included stocks for international data which are obtained from Datastream. All samples end in December 2010. The numbers of stocks used to estimate aggregate turnover are generally higher than those used in estimating the cross-sectional high volume return premiums as reported in this table.

Country	Start date	Average number of stocks
Panel A G-7 countries excluding the U.S.		
Canada	January 1977	453
France	July 1991	452
Germany	January 1999	384
Italy	April 1994	194
Japan	December 1990	1712
UK	January 1991	702
Pane B Twelve other countries		
Australia	January 1984	436
Brazil	August 2003	61
China	March 1996	930
Hong Kong	June 1988	384
India	January 1995	1118
Korea	September 1987	811
Russia	September 2005	72
Singapore	January 1984	176
South Africa	January 1996	185
Spain	January 1991	105
Switzerland	May 1990	128
Taiwan	April 1991	448

Table 5. Descriptive statistics for international markets

TOTMKT is Datastream total market returns for the country. VWRET and EWRET are our estimates of value- and equal-weighted market portfolio returns, and VWHVP and EWHVP are value- and equal-weighted high volume return premiums. All three market portfolio returns are in percentage. The symbols *, **, and *** denote that the entry (i.e., the mean statistic) is significant at the 10%, 5%, and 1% levels, respectively, based on heteroskedasticity-and-autocorrelation consistent (HAC) errors.

Countries	TOTMKT	VWRET	EWRET	VWHVP	EWHVP
Panel A G-7 countries excluding the U.S.					
Canada	1.049***	1.073***	1.602***	0.189	1.041***
France	0.843**	0.796**	0.879**	0.548**	1.266***
Germany	0.444	0.504	0.304	0.656	0.841**
Italy	0.629	0.624	0.578	0.297	0.647***
Japan	−0.040	−0.050	0.206	0.479**	0.980***
UK	0.826***	0.833***	0.885*	0.641**	1.494***
Pane B Twelve other countries					
Australia	1.104***	1.052***	1.215***	0.273	1.603***
Brazil	1.970**	2.370***	2.728**	0.579	1.604***
China	1.267	1.458*	1.960**	−0.519	−0.630
Hong Kong	1.393***	1.388***	1.534**	0.858***	1.500***
India	1.547**	1.623**	2.435**	0.585	0.187
Korea	1.199**	1.074*	1.687**	−0.469	−0.894***
Russia	1.280	1.492	2.745	0.535	1.566
Singapore	0.826*	0.952**	1.204*	0.385	0.625**
South Africa	1.421***	1.334***	1.243***	0.432	0.688***
Spain	0.945**	0.934**	0.859*	−0.250	0.655**
Switzerland	0.824**	0.827**	0.814*	0.285	0.593**
Taiwan	0.872	0.700	0.907	0.461	0.381

Table 6. Forecasting performance of trading volume for stock returns in G-7 countries excluding the U.S.

The general form of the forecasting model is $y_t = \alpha + \sum_{m=1}^{L_y} \beta_m y_{t-m} + \sum_{n=1}^{L_x} \gamma_n x_{t-n} + \sum_{r=1}^{L_z} \lambda_r z_{t-r} + \varepsilon_t$, (1)

where y_t is value-weighted (equal-weighted) market portfolio returns VWMKT (EWMKT), x_t is the corresponding aggregate turnover VWVOL (EWVOL) or the high volume return premiums VWHVP (EWHVP), and z_t is market volatility.

Model C includes a constant α only. Model RZ includes α and lags of y_t and z_t. Model W includes α and lags of y_t, x_t, and z_t.

Root mean squared forecast errors (RMSFE) and the annualized realized utility have been multiplied by 100 for ease of presentation. All the statistics in the table are based on one-step-ahead recursive forecast errors from the above models.

ENC-NEW is the encompassing test statistic of Clark and McCracken (2001), in which the associated null hypothesis is that forecasts from model H_o (model RZ or C) encompass those from H_A (model W). Thus, the H_o model is preferred to H_A. The critical values of the ENC-NEW test are linearly interpolated from the unpublished Appendix of Clark and McCracken (2001). The symbols ** and *** denote significance at the 5% and 1% levels, respectively.

Country	RMSFE			ENC-NEW		Realized Utility		
	C	RZ	W	RZ vs. W	C vs. W	C	RZ	W
	Panel A $VWMKT_t = \alpha + \sum_{m=1}^{L_y} \beta_m VWMKT_{t-m} + \sum_{n=1}^{L_x} \gamma_n VWVOL_{t-n} + \sum_{r=1}^{L_z} \lambda_r \sigma_{t-r}^2 + \varepsilon_t$							
Canada	0.941	0.965	0.966	0.789		3.604	14.814	14.166
France	1.356	1.357	1.357	0.940		0.741	−0.753	2.387
Germany	1.307	1.308	1.308	−0.830		2.809	6.556	5.915
Italy	1.381	1.373	1.373	0.226		−2.475	5.825	5.250
Japan	1.431	1.440	1.440	−0.292		−0.673	1.359	1.657
UK	1.294	1.295	1.294	2.506**	19.667***	3.276	2.940	3.946
	Panel B $EWMKT_t = \alpha + \sum_{m=1}^{L_y} \beta_m EWMKT_{t-m} + \sum_{n=1}^{L_x} \gamma_n EWVOL_{t-n} + \sum_{r=1}^{L_z} \lambda_r \sigma_{t-r}^2 + \varepsilon_t$							
Canada	0.889	0.862	0.862	−2.233		28.244	46.720	46.783
France	0.712	0.677	0.677	−0.189		23.053	40.029	39.940
Germany	0.963	0.923	0.923	−0.395		63.138	74.020	73.833
Italy	1.017	0.998	0.998	−0.198		−3.199	20.589	20.110
Japan	1.181	1.187	1.188	−0.830		3.675	24.242	23.263
UK	0.736	0.673	0.673	0.296		6.545	36.676	36.619
	Panel C $VWRET_t = \alpha + \sum_{m=1}^{L_y} \beta_m VWRET_{t-m} + \sum_{n=1}^{L_x} \gamma_n VWHVP_{t-n} + \sum_{r=1}^{L_z} \lambda_r \sigma_{t-r}^2 + \varepsilon_t$							
Canada	0.945	0.969	0.969	−0.160		3.846	15.117	14.785
France	1.360	1.361	1.361	−0.538		0.949	−0.428	0.232
Germany	1.346	1.346	1.344	6.071***	25.868***	2.050	6.355	5.471

Italy	1.369	1.361	1.362	−0.410		−1.362	5.990	5.828
Japan	1.431	1.440	1.438	4.309***	−0.999	−0.620	1.508	3.009
UK	1.315	1.317	1.316	2.455**	19.111***	2.334	1.846	1.668
	Panel D $EWRET_t = \alpha + \sum_{m=1}^{L_y} \beta_m EWRET_{t-m} + \sum_{n=1}^{L_x} \gamma_n EWHVP_{t-n} + \sum_{r=1}^{L_z} \lambda_r \sigma_{t-r}^2 + \varepsilon_t$							
Canada	0.873	0.862	0.861	12.615***	290.92***	23.724	37.032	38.323
France	0.742	0.717	0.717	0.078		7.146	27.503	27.651
Germany	1.148	1.162	1.163	4.040***	59.470***	9.190	28.439	30.185
Italy	1.019	1.001	1.001	0.433		−1.295	18.543	18.452
Japan	1.208	1.211	1.212	−0.873		1.461	20.294	18.300
UK	0.829	0.788	0.788	0.164		3.725	30.915	31.052

Table 7. **Forecasting performance of trading volume for stock returns in 12 other countries (trading volume proxied by turnover)**

The general form of the forecasting model is $y_t = \alpha + \sum_{m=1}^{L_y} \beta_m y_{t-m} + \sum_{n=1}^{L_x} \gamma_n x_{t-n} + \sum_{r=1}^{L_z} \lambda_r z_{t-r} + \varepsilon_t$, (1)

where y_t is value-weighted (equal-weighted) market portfolio returns VWMKT (EWMKT), x_t is the corresponding aggregate turnover VWVOL (EWVOL) and z_t is market volatility.

Model C includes a constant α only. Model RZ includes α and lags of y_t and z_t. Model W includes α and lags of y_t, x_t, and z_t.

Root mean squared forecast errors (RMSFE) and the annualized realized utility have been multiplied by 100 for ease of presentation. All the statistics in the table are based on one-step-ahead recursive forecast errors from the above models.

ENC-NEW is the encompassing test statistic of Clark and McCracken (2001), in which the associated null hypothesis is that forecasts from model H_o encompass those from H_A. Thus, the H_o model is preferred to H_A. The critical values of the ENC-NEW test are linearly interpolated from the unpublished Appendix of Clark and McCracken (2001). The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Country	RMSFE			ENC-NEW		Realized Utility		
	C	RZ	W	RZ vs. W	C vs. W	C	RZ	W
	Panel A $VWRET_t = \alpha + \sum_{m=1}^{L_y} \beta_m VWRET_{t-m} + \sum_{n=1}^{L_x} \gamma_n VWVOL_{t-n} + \sum_{r=1}^{L_z} \lambda_r \sigma_{t-r}^2 + \varepsilon_t$							
Australia	0.965	0.968	0.968	1.063		3.134	4.267	4.742
Brazil	2.138	2.173	2.173	−0.278		−7.076	−8.195	−7.966
China	2.340	2.342	2.344	−0.515		3.679	5.672	2.732
Hong Kong	1.533	1.532	1.532	1.664*	8.041***	4.363	11.312	12.011
India	1.811	1.802	1.795	6.832***	25.705***	10.384	24.699	26.854

Korea	1.942	1.939	1.938	0.300		4.823	15.203	16.690
Russia	1.850	1.853	1.849	4.633***	4.117***	12.436	7.577	24.703
Singapore	1.262	1.265	1.265	0.441		1.110	13.124	13.314
South Africa	1.372	1.374	1.373	0.780*	10.109***	5.154	15.786	15.693
Spain	1.350	1.356	1.356	-0.100		1.874	−4.921	−3.051
Switzerland	1.223	1.222	1.221	2.869**	18.038***	−1.313	3.247	3.752
Taiwan	1.435	1.435	1.436	−0.233		2.646	5.296	6.561
	Panel B $EWRET_t = \alpha + \sum_{m=1}^{L_y} \beta_m EWRET_{t-m} + \sum_{n=1}^{L_x} \gamma_n EWVOL_{t-n} + \sum_{r=1}^{L_z} \lambda_r \sigma_{t-r}^2 + \varepsilon_t$							
Australia	1.020	0.975	0.975	3.800***	4.575*	36.358	55.501	55.430
Brazil	1.626	1.632	1.632	0.145		4.531	23.010	22.070
China	2.580	2.574	2.575	0.004		13.185	38.649	34.278
Hong Kong	1.463	1.410	1.410	2.958**	216.31***	20.396	60.671	60.091
India	1.654	1.532	1.531	0.626		8.173	68.212	66.970
Korea	1.798	1.778	1.779	−0.534		8.933	47.012	47.271
Russia	1.470	1.451	1.451	−0.092		82.960	85.736	85.558
Singapore	1.460	1.424	1.424	−0.709		8.379	43.366	42.714
South Africa	0.810	0.794	0.795	−0.206		27.910	38.183	38.556
Spain	0.900	0.895	0.895	0.235		6.249	18.365	18.574
Switzerland	0.791	0.765	0.765	0.008		8.442	28.649	28.499
Taiwan	1.427	1.417	1.417	3.857***	27.054***	4.857	22.081	21.996

Table 8. Forecasting performance of trading volume for stock returns in other 12 countries

(trading volume proxied by the high volume premium)

The general form of the forecasting model is $y_t = \alpha + \sum_{m=1}^{L_y} \beta_m y_{t-m} + \sum_{n=1}^{L_x} \gamma_n x_{t-n} + \sum_{r=1}^{L_z} \lambda_r z_{t-r} + \varepsilon_t$, (1)

where y_t is value-weighted (equal-weighted) market portfolio returns VWMKT (EWMKT), x_t is the corresponding high volume return premiums VWHVP (EWHVP), and z_t is market volatility.

Model C includes a constant α only. Model RZ includes α and lags of y_t and z_t. Model W includes α and lags of y_t, x_t, and z_t.

Root mean squared forecast errors (RMSFE) and the annualized realized utility have been multiplied by 100 for ease of presentation. All the statistics in the table are based on one-step-ahead recursive forecast errors from the above models. ENC-NEW is the encompassing test statistic of Clark and McCracken (2001), in which the associated null hypothesis is that forecasts from model H_o encompass those from H_A. Thus, the H_o model is preferred to H_A. The critical values of the ENC-NEW test are linearly interpolated from the unpublished Appendix of Clark and McCracken (2001). The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Country	RMSFE			ENC-NEW		Realized Utility		
	C	RZ	W	RZ vs. W	C vs. W	C	RZ	W
	Panel A							

	$VWRET_t = \alpha + \sum_{m=1}^{L_y} \beta_m VWRET_{t-m} + \sum_{n=1}^{L_x} \gamma_n VWHVP_{t-n} + \sum_{r=1}^{L_z} \lambda_r \sigma_{t-r}^2 + \varepsilon_t$							
Australia	0.969	0.972	0.972	−1.072		2.675	3.743	3.540
Brazil	2.161	2.196	2.198	−0.681		−2.923	-5.341	−5.363
China	2.322	2.323	2.323	−0.001		7.326	6.150	7.773
Hong Kong	1.549	1.548	1.548	−0.721		3.666	10.721	11.099
India	1.802	1.792	1.790	3.196***	22.461***	11.904	26.373	26.455
Korea	1.968	1.964	1.963	7.630***	19.816***	5.141	15.289	15.083
Russia	1.750	1.753	1.753	−0.092		11.409	10.356	9.535
Singapore	1.273	1.276	1.276	1.498*	41.085***	1.053	12.900	12.161
South Africa	1.374	1.377	1.377	−0.275		5.231	14.517	14.348
Spain	1.363	1.369	1.368	3.002***	−5.407	0.442	−0.963	−0.894
Switzerland	1.225	1.224	1.224	0.097		−1.513	2.932	3.055
Taiwan	1.453	1.453	1.454	0.425		2.847	5.559	3.340
	Panel B $EWRET_t = \alpha + \sum_{m=1}^{L_y} \beta_m EWRET_{t-m} + \sum_{n=1}^{L_x} \gamma_n EWHVP_{t-n} + \sum_{r=1}^{L_z} \lambda_r \sigma_{t-r}^2 + \varepsilon_t$							
Australia	1.021	1.001	0.999	10.014***	176.27***	9.018	32.342	31.472
Brazil	1.673	1.691	1.692	−0.188		8.332	18.146	17.433
China	2.563	2.560	2.558	0.763*	4.300***	13.650	30.964	32.082
Hong Kong	1.507	1.474	1.475	−0.681		8.913	46.052	45.831
India	1.700	1.588	1.590	−1.331		7.882	67.834	67.756
Korea	1.862	1.838	1.839	−1.299		6.754	43.844	43.314
Russia	1.315	1.312	1.312	−0.079		45.353	51.751	51.600
Singapore	1.495	1.471	1.472	−1.669		0.570	33.649	32.957
South Africa	0.835	0.826	0.827	−0.517		10.279	24.833	24.854
Spain	0.938	0.933	0.934	−1.000		2.796	13.850	12.645
Switzerland	0.837	0.815	0.814	3.797***	140.73***	3.058	24.840	24.213
Taiwan	1.445	1.436	1.436	0.447		2.847	5.559	3.340

Appendix 1. Filters in the Datastream Data

Given the potential data errors or outliers in the Datastream as identified by previous research (e.g., Ince and Porter, 2006), we implement two sets of filtering rules on the raw data in addition to those sampling requirements applied to the U.S. data in forming volume-based portfolios. We first follow Kaniel, Ozoguz, and Starks (2012) and remove stocks whose local currency prices fall below the lowest five percentile of stock prices in the country's sample for that year. We then follow Guo and Savickas (2008) and set the daily return on a stock to a missing value if the recorded return is greater than 300% on that day. If the price of the stock falls by more than 90% in a day *and* it has increased by more than 200% within the previous 20 trading days, we set all daily returns between the two dates to missing values. Similarly, if the price of a stock increases by more than 100% in a day *and* it has decreased by more than 200% within the previous 20 trading days, we also set all daily returns between the two dates to missing values. The price we pay for these more stringent data-filtering rules is that, compared to Kaniel, Ozoguz, and Starks (2012), we work with a smaller number of stocks and shorter samples for some countries during the overlapping periods.

Based on the filtered data, we form the estimates of daily value- and equal-weighted aggregate turnover for each of the18 countries. We also follow the same strategies used in previous sections for the U.S. data in forming the volume portfolios and estimating the high volume return premium. Because of the limited availability of the volume data, the number of trading stocks meeting all the selection criteria is small for many countries, particularly in the earlier years of the samples. Therefore, for the international markets, we sort all stocks into quintiles rather than centiles as we did for the U.S. market. The value- (equal-)weighted high volume return premium is defined as the difference between the value- (equal-)weighted portfolio returns on the top volume quintile and the returns on the bottom volume quintile. Conceivably, HVP would be higher were it based on centiles rather than on quintiles. Different stock exchanges within the same country may have different trade-volume dynamics due to differences in the institutional details. To reduce this type of heterogeneity and its possible impact on the estimation, we follow the literature and only study the primary exchanges in each country. The list of countries that have multiple primary exchanges includes Canada, China, Germany, India, Korea, Russia, and Spain.

Appendix Table 1. Forecasting performance of trading volume for stock returns in the U.S. market (trading volume proxied by turnover, in-sample length 20 years)

The general form of the forecasting model is $y_t = \alpha + \sum_{m=1}^{L_y} \beta_m y_{t-m} + \sum_{n=1}^{L_x} \gamma_n x_{t-n} + \sum_{r=1}^{L_z} \lambda_r z_{t-r} + \varepsilon_t$, (1)

where y_t is value-weighted (equal-weighted) market portfolio returns VWMKT (EWMKT), x_t is the corresponding market turnover VWVOL (EWVOL), and z_t is market volatility. Model C includes a constant α only. Model R includes α and lagged y_t. Model U includes α and lags of y_t and x_t. Model W includes α and lags of y_t, x_t, and z_t. The numbers of lags of each variable (L_y, L_x, and L_z) are the same as in Table 2.

Root mean squared forecast errors (RMSFE) and the annualized realized utility have been multiplied by 100 for ease of presentation. All the statistics in the table are based on one-step-ahead recursive forecast errors from the above models for the daily excess returns to the U.S. stock market portfolio.

ENC-NEW is the encompassing test statistic of Clark and McCracken (2001), in which the associated null hypothesis is that forecasts from the model in the first column (H_o) encompass those from the model in the second row (H_A). Thus, the H_o model is preferred to H_A. The critical values of the ENC-NEW test are linearly interpolated from the unpublished Appendix of Clark and McCracken (2001). The symbols * and *** denote significance at the 10% and 1% levels, respectively.

	Forecasting models				Forecasting models			
	C	R	U	W	C	R	U	W
	Panel A. VWVOL & VWMKT, 1983-10				Panel B. EWVOL & EWMKT, 1983-10			
	RMSFE				RMSFE			
	1.113	1.121	1.121	1.123	1.000	0.996	0.992	0.994
	Encompassing test statistic				Encompassing test statistic			
			38.608***				437.42***	
			1.724*				40.698***	
				−7.512				−2.361
	Realized Utility				Realized Utility			
	3.894	6.908	6.305	6.046	14.955	29.502	30.277	30.249
	Panel C. VWVOL & VWMKT, 1983-99				Panel D. EWVOL & EWMKT, 1983-99			
	RMSFE				RMSFE			
	0.899	0.903	0.903	0.905	0.691	0.665	0.661	0.661
	Encompassing test statistic				Encompassing test statistic			
C			70.747***				540.91***	
R			−0.187				43.198***	
U				-3.507				−0.201
	Realized Utility				Realized Utility			
	5.890	15.079	14.214	13.801	18.178	33.535	34.182	34.197

Appendix Table 2. Forecasting performance of trading volume for stock returns in the U.S. market (trading volume proxied by the volume premium, in-sample length 20 years)

The general form of the forecasting model is $y_t = \alpha + \sum_{m=1}^{L_y} \beta_m y_{t-m} + \sum_{n=1}^{L_x} \gamma_n x_{t-n} + \sum_{r=1}^{L_z} \lambda_r z_{t-r} + \varepsilon_t,$ (1)

where y_t is value-weighted (equal-weighted) market portfolio returns VWMKT (EWMKT), x_t is the corresponding market turnover VWVOL (EWVOL), and z_t is market volatility. Model C includes a constant α only. Model R includes α and lagged y_t. Model U includes α and lags of y_t and x_t. Model W includes α and lags of y_t, x_t, and z_t. The numbers of lags of each variable (L_y, L_x, and L_z) are the same as in Table 2.

Root mean squared forecast errors (RMSFE) and the annualized realized utility have been multiplied by 100 for ease of presentation. All the statistics in the table are based on one-step-ahead recursive forecast errors from the above models for the daily excess returns to the U.S. stock market portfolio.

ENC-NEW is the encompassing test statistic of Clark and McCracken (2001), in which the associated null hypothesis is that forecasts from the model in the first column (H_o) encompass those from the model in the second row (H_A). Thus, the H_o model is preferred to H_A. The critical values of the ENC-NEW test are linearly interpolated from the unpublished Appendix of Clark and McCracken (2001). The symbols ** and *** denote significance at the 5% and 1% levels, respectively.

	Forecasting models				Forecasting models			
	C	R	U	W	C	R	U	W
	Panel A. VWHVP & VWRET, 1983-10				Panel B. EWHVP & EWRET, 1983-10			
	RMSFE				RMSFE			
	1.110	1.119	1.119	1.121	0.903	0.903	0.903	0.910
	Encompassing test statistic				Encompassing test statistic			
C			44.677***				418.73***	
R			2.017**				5.306***	
U				−8.963				−21.609
	Realized Utility				Realized Utility			
	4.579	9.220	9.162	8.975	14.058	27.427	27.898	27.589
	Panel C. VWHVP & VWRET, 1983-99				Panel D. EWHVP & EWRET, 1983-99			
	RMSFE				RMSFE			
	0.880	0.881	0.881	0.882	0.633	0.607	0.607	0.616
	Encompassing test statistic				Encompassing test statistic			
C			91.911***				568.59***	

R			−0.149				3.104***	
U				−3.455				−15.380
	Realized Utility				Realized Utility			
	7.251	17.178	16.944	16.729	17.104	31.828	32.536	31.550

Appendix Table 3. Forecasting performance of trading volume for stock returns in the U.S. market, by firm size (trading volume proxied by turnover)

We sort CRSP stocks into small, medium and large portfolios based on the breakpoints for the low 30%, medium 40%, and high 30% of the ranked values of market capitalization. For each portfolio, we calculate equal-weighted portfolio returns (EWRET) and the corresponding aggregate measure of turnover (EWVOL) for the sample period July 1963 to December 2010.

The general form of the forecasting model is $$y_t = \alpha + \sum_{m=1}^{L_y} \beta_m y_{t-m} + \sum_{n=1}^{L_x} \gamma_n x_{t-n} + \sum_{r=1}^{L_z} \lambda_r z_{t-r} + \varepsilon_t, \quad (1)$$

where y_t is equal-weighted market portfolio returns EWMKT, x_t is the corresponding market turnover EWVOL, and z_t is market volatility. Model C includes a constant α only. Model R includes α and lagged y_t. Model U includes α and lags of y_t and x_t. Model W includes α and lags of y_t, x_t, and z_t. L_y, L_x, and L_z are the numbers of lags on stocks returns, turnover, and market volatility, respectively. They are selected by minimizing Schwarz's Bayesian information criterion (BIC).

Root mean squared forecast errors (RMSFE) and the annualized realized utility have been multiplied by 100 for ease of presentation. All the statistics in the table are based on one-step-ahead recursive forecast errors from the above models.

ENC-NEW is the encompassing test statistic of Clark and McCracken (2001), in which the associated null hypothesis is that forecasts from the model in the first column (H_o) encompass those from the model in the second row (H_A). Thus, the H_o model is preferred to H_A. The critical values of the ENC-NEW test are linearly interpolated from the unpublished Appendix of Clark and McCracken (2001). The symbol *** denotes significance at the 1% level.

	Forecasting models				Forecasting models			
	C	R	U	W	C	R	U	W
L_y	0	3	3	3	0	3	3	3
L_x	0	0	1	1	0	0	1	1
L_z	0	0	0	3	0	0	0	1
	Panel A. Small stocks, 1963-10				Panel B. Large stocks, 1963-10			

	RMSFE				RMSFE			
	0.838	0.763	0.756	0.756	1.082	1.079	1.079	1.080
	Encompassing test statistic				Encompassing test statistic			
C			2430.5***				378.63***	
R			163.57***				8.817***	
U				45.942***				−9.342
	Realized Utility				Realized Utility			
	43.881	60.167	60.711	60.168	4.743	19.838	19.643	20.106
	Panel C. Small stocks, 1963-99				Panel D. Large stocks, 1963-99			
	RMSFE				RMSFE			
	0.718	0.648	0.640	0.642	0.845	0.822	0.822	0.823
	Encompassing test statistic				Encompassing test statistic			
C			1942.4***				583.20***	
R			147.82***				5.552***	
U				14.119***				-3.610
	Realized Utility				Realized Utility			
	46.655	59.380	60.324	59.793	5.812	25.524	25.674	25.739
	Panel E. Small stocks, 2000-10				Panel F. Large stocks, 2000-10			
	RMSFE				RMSFE			
	1.070	0.984	0.978	0.975	1.502	1.525	1.524	1.527
	Encompassing test statistic				Encompassing test statistic			
C			620.44***				20.018***	
R			34.425***				2.801***	
U				21.493***				−3.628
	Realized Utility				Realized Utility			
	37.288	62.038	61.629	61.060	2.201	6.335	5.320	6.726

Appendix Table 4. The impact of U.S. market turnover on the international markets

The general form of the forecasting model is $y_t = \alpha + \sum_{m=1}^{L_y} \beta_m y_{t-m} + \sum_{n=1}^{L_x} \gamma_n x_{t-n} + \sum_{r=1}^{L_z} \lambda_r z_{t-r} + \varepsilon_t ,$ (1)

where y_t is value-weighted (equal-weighted) market portfolio returns VWMKT (EWMKT), x_t is the corresponding market turnover VWVOL (EWVOL), and z_t is market volatility. Model C includes a constant α only. Model W includes α and lags of y_t, x_t, and z_t. Model WR augments Model W with one lag of the U.S. market returns. Model WRV augments Model W with one lag of the U.S. market returns and aggregate turnover.

Each entry is the annualized utility level a risk-averse investor can attain by following the above models' stock return forecasts in allocating her/his investment daily between stocks and risk-free bills. These realized utility estimates are computed according to equations (3) and (4) in the text and have been multiplied by 100 for ease of presentation. Numbers in bold indicate that (1) the unrestricted model WRV achieves higher utility than models C, W, and WR; and (2) the latter three models have larger forecast errors and do not encompass model WRV at the 5% significance level.

	Value-weighted portfolios				Equal-weighted portfolios			
	C	W	WR	WRV	C	W	WR	WRV
Canada	3.60	14.17	18.10	17.76	28.24	46.78	49.02	**49.62**
France	0.74	2.39	36.00	35.99	23.05	39.94	47.17	47.20
Germany	2.81	5.92	18.44	18.73	63.14	73.83	78.93	78.53
Italy	−2.48	5.25	19.38	18.21	-3.20	20.11	17.98	20.04
Japan	−0.67	1.66	53.34	53.61	3.68	23.26	53.44	**53.85**
UK	3.28	3.95	36.31	36.26	6.55	36.62	43.74	43.75
Australia	3.13	4.74	46.49	46.49	36.36	55.43	74.39	74.41
Brazil	−7.08	−7.97	5.54	4.33	4.53	22.07	18.94	19.47
China	3.68	2.73	35.88	37.68	13.19	34.28	51.97	52.72
Hong Kong	4.36	12.01	64.39	64.63	20.40	60.09	77.74	**78.48**
India	10.38	26.85	53.43	**56.21**	8.17	66.97	74.79	**74.91**
Korea	4.82	16.69	63.21	62.65	8.93	47.27	70.14	69.30
Russia	12.44	24.70	54.33	50.69	82.96	85.56	96.74	96.36
Singapore	1.11	13.31	41.32	41.50	8.38	42.71	54.44	54.39
South Africa	5.15	15.69	47.14	47.01	27.91	38.56	55.42	55.35
Spain	1.87	−3.05	22.02	21.90	6.25	18.57	28.98	29.03
Switzerland	−1.31	3.75	31.80	30.98	8.44	28.50	39.61	39.49
Taiwan	2.65	6.56	47.35	47.07	4.86	22.00	42.90	43.01

Appendix Table 5. The impact of U.S. market volume premium on the international markets

The general form of the forecasting model is $y_t = \alpha + \sum_{m=1}^{L_y} \beta_m y_{t-m} + \sum_{n=1}^{L_x} \gamma_n x_{t-n} + \sum_{r=1}^{L_z} \lambda_r z_{t-r} + \varepsilon_t$, (1)

where y_t is value-weighted (equal-weighted) market portfolio returns VWMKT (EWMKT), x_t is the corresponding volume premium VWHVP (EWHVP), and z_t is market volatility. Model C includes a constant α only. Model W includes α and lags of y_t, x_t, and z_t. Model WR augments Model W with one lag of the U.S. market returns. Model WRV augments Model W with one lag of the U.S. market returns and volume premium.

Each entry is the annualized utility level a risk-averse investor can attain by following the above models' stock return forecasts in allocating her/his investment daily between stocks and risk-free bills. These realized utility estimates are computed according to equations (3) and (4) in the text and have been multiplied by 100 for ease of presentation. Numbers in bold indicate that (1) the unrestricted model WRV achieves higher utility than models C, W, and WR; and (2) the latter three models have larger forecast errors and do not encompass model WRV at the 5% significance level.

	Value-weighted portfolios				Equal-weighted portfolios			
	C	W	WR	WRV	C	W	WR	WRV
Canada	3.85	14.79	18.41	18.55	23.72	38.32	41.33	41.32
France	0.95	0.23	35.68	**35.84**	7.15	27.65	36.00	36.05
Germany	2.05	5.47	16.68	19.94	9.19	30.19	40.90	40.88
Italy	−1.36	5.83	21.41	22.31	−1.30	18.45	28.90	28.89
Japan	−0.62	3.01	51.36	51.13	1.46	18.30	50.33	49.98
UK	2.33	1.67	34.62	34.66	3.73	31.05	41.76	41.74
Australia	2.68	3.54	46.20	46.63	9.02	31.47	55.01	55.16
Brazil	−2.92	−5.36	9.00	7.71	8.33	17.43	19.40	20.46
China	7.33	7.77	35.81	34.36	13.65	32.08	53.64	52.20
Hong Kong	3.67	11.10	64.61	65.06	8.91	45.83	66.77	**66.95**
India	11.90	26.46	48.11	46.86	7.88	67.76	77.23	76.95
Korea	5.14	15.08	62.80	62.47	6.75	43.31	64.41	64.69
Russia	11.41	9.54	57.63	**58.04**	45.35	51.60	74.14	72.95
Singapore	1.05	12.16	41.32	40.92	0.57	32.96	47.18	46.40
South Africa	5.23	14.35	45.90	45.40	10.28	24.85	44.22	44.11
Spain	0.44	−0.89	23.80	23.69	2.80	12.65	27.01	27.32
Switzerland	−1.51	3.06	28.75	28.65	3.06	24.21	38.57	38.72
Taiwan	2.85	3.34	47.34	47.18	4.30	22.03	44.37	44.06

Highlights

• Dynamic trading volume and stock return relation is studied from the perspective of out-of-sample stock return predictability

• Evidence from the U.S. suggests that higher returns follow more trading in the pre-2000 period

• The weak out-of-sample predictive power of volume is absent in most of the other major markets

• Our results contradict findings by many in-sample studies and do not support a significant volume and return relation predicted by some theoretical models