INSTITUTE
OF ECONOMIC STUDIES
Faculty of Social Sciences
Charles University

TRADING VOLUME AND STOCK RETURNS: A META-ANALYSIS

Josef Bajzik

IES Working Paper 45/2020

Institute of Economic Studies,
Faculty of Social Sciences,
Charles University in Prague

[UK FSV – IES]

Opletalova 26
CZ-110 00, Prague
E-mail : ies@fsv.cuni.cz
http://ies.fsv.cuni.cz

Institut ekonomických studií
Fakulta sociálních věd
Univerzita Karlova v Praze

Opletalova 26
110 00 Praha 1

E-mail : ies@fsv.cuni.cz
http://ies.fsv.cuni.cz

Bibliographic information:
Bajzik J. (2020): "Trading Volume and Stock Returns: A Meta-Analysis" IES Working Papers 45/2020. IES FSV. Charles University.

This paper can be downloaded at: http://ies.fsv.cuni.cz

Trading Volume and Stock Returns: A Meta-Analysis

Josef Bajzik[a]

[a]Charles University & Czech National Bank, Prague, Czech Republic

December 2020

Abstract:
I examine 468 estimates on the relationship between trading volume and stock returns reported in 44 studies. I deploy recent nonlinear techniques for detecting publication bias together with Bayesian and frequentist model averaging to evaluate the heterogeneity in the estimates. The results yield three key conclusions. First, publication bias distorts the findings of the primary studies. After this bias is corrected, the literature shows that with higher trading volume, returns decline in both effects in the contemporaneous and even in the dynamic one. Second, one cannot rely on any general conclusions about stock markets. The predictability of stock returns varies with different markets and stock types. Third, different data characteristics, structural variations and methodologies used drive the heterogeneity in the results of the primary articles. In particular, one should be cautious when using monthly data or VAR models.

JEL: G10, G12, G14
Keywords: Stock returns, trading volume, meta-analysis, Bayesian model averaging, publication bias

Acknowledgements: This work was supported by grant from the Grant Agency of Charles University (grant 1430220) and by Donatio Universitatis Carolinae.

1 Introduction

Does the efficient market hypothesis (EMH) for stock returns hold? Or is it simply a relic observation of early stock markets? These questions have attracted traders over the last century Fama (1970). Since to this day Malkiel (2003) no satisfying answer has been provided for people interested in stocks, it is no surprise that traders, researchers and investors are still looking for patterns to help them develop profitable trading strategies (Hu, 1997; Chen *et al.*, 2001).

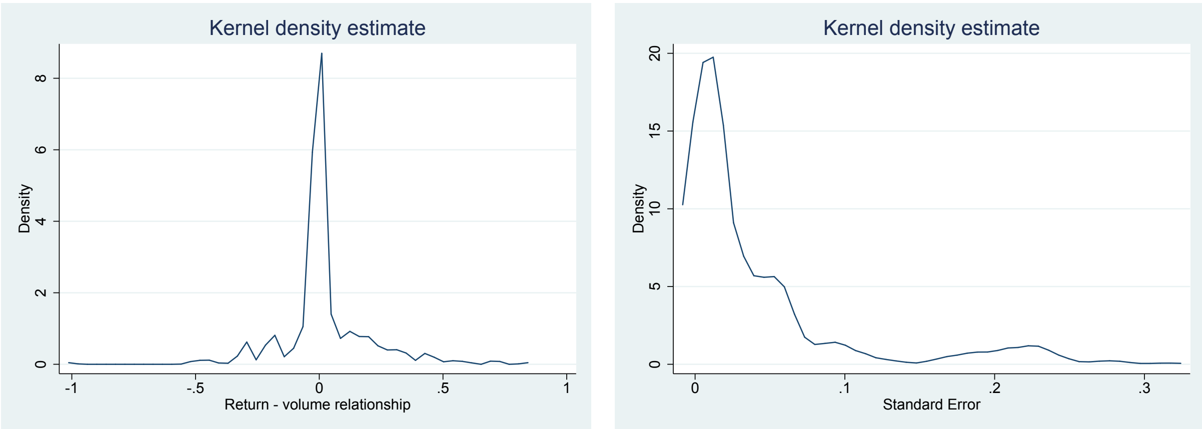
If stock returns evolve randomly (Lee, 1992) or if asset prices reflect all available information (Malkiel, 1989), there would be no place for developing profitable trading strategies. Since these premises may not hold, however, Jensen *et al.* (1972) and Gibbons (1982) came up with a capital asset pricing model. Then, asset pricing theory was developed Black & Scholes (1974). Two decades later Fama & French (1992, 1993, 1996) and Jegadeesh & Titman (1993, 1995) introduced their models and pointed out two non-risk factors explaining cross-sectional variation in expected stock returns. These factors are the well-known and thoroughly investigated book-to-market ratio and firm size (Astakhov *et al.*, 2019). In addition to these two factors, Amihud & Mendelson (1986) discussed another not-so-deeply explored factor that possibly affects stock markets – liquidity.

In addition to Amihud & Mendelson (1986), Brennan & Subrahmanyam (1996); Brennan *et al.* (1998), for instance, suggested that liquidity affects expected stock returns. Both articles found stock returns and liquidity to be negatively related. This relationship may be explained by the clientele effect hypothesis suggested by Merton *et al.* (1987) according to which "stocks with greater investor following should command lower expected return." The same was noted by, for example, Datar *et al.* (1998) about the liquidity premium. Moreover, Karpoff (1987) added several other reasons to study the liquidity-volume effect on stock returns. First, this type of research provides insight into financial markets' structure. Second, it is seminal for event studies. Third, the return-volume relationship has significant implications for futures markets researches. These suggestions make the findings related to this topic even more valuable.

Another question is *how* to measure liquidity. Since the bid-ask spread has been found to be a weak proxy for liquidity (Petersen & Fialkowski, 1994; Datar *et al.*, 1998) and, on the other hand, trading volume is considered a major determinant of liquidity measures (Stoll, 1978; Brennan & Subrahmanyam, 1995), I decide to thoroughly investigate the trading volume-stock return relationship through a meta-analysis. It is important to fill the research gap and shed light on this area. This study discusses what has actually been investigated, since seminal authors such as Fama & French (1996) omitted this factor, whereas others accepted it.

Altogether, I collect 468 estimates from 44 studies and 49 variables capturing the context in which the studies derived their findings. Figure 1 and Figure 2 provide an overview of the literature on trading volume and stock returns. Three observations can be drawn from these figures. First, the median value of the estimated relationship approaches zero, and most of the estimates appear close to this value. Second, the reported values seem to decline over time. In contrast, it remains unclear whether the improved data and sophisticated techniques used by recent studies cause the trend or whether some essential change has occurred. Third and last observation refers to the increasing variance of the estimates. Instead of converging toward some consensus, the estimates from the literature diverge over time.

Figure 1: Kernel densities of the return-volume relationship and corresponding standard errors

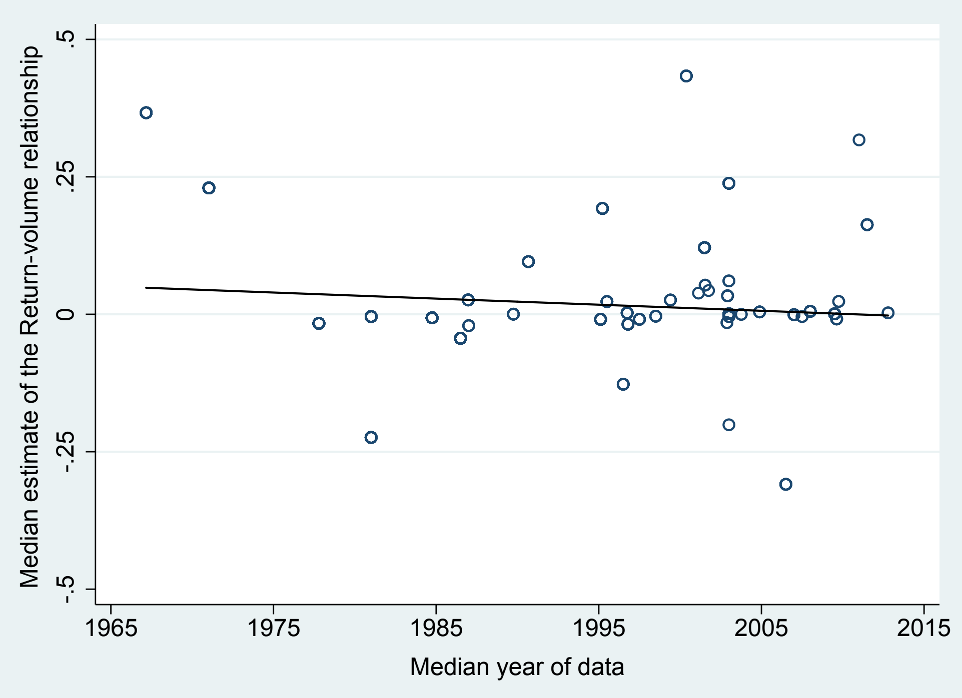


(a) (b)

Notes: The Figure 1 depicts kernel densities for return-volume relationship (on the left) and corresponding Standard Errors (on the right). Since primary studies employ many different estimation approaches, partial correlation coefficients normalize the estimates and the winsorization handles with the outliers.

The last observation about the increasing variance provides an additional reason to conduct a systematic assessment of all published results. The most suitable method for such an evaluation is a meta-analysis (Imai *et al.*, 2020). A meta-analysis addresses publication bias as well as model uncertainty issues. I follow seminal works such as Havranek & Irsova (2017) and employ the most modern techniques for correcting for publication bias together with Bayesian model averaging (Raftery *et al.*, 1997) and frequentist model averaging (Amini & Parmeter, 2012).

Figure 2: Increase in mean and variance of return-volume estimates over time



Notes: The Figure 2 captures median estimates per study of return-volume relationship at the vertical axis and measures the median year of the data used in particular studies at the horizontal axis.

Authors do not regularly publish insignificant results or results with the "wrong" sign (Stanley, 2001; Christensen & Miguel, 2018), and such decisions distort the literature. Since a focus on null results becomes secure because of the support of the efficient market hypothesis and because of Fama & French (1992), research on stock returns is prone to publication bias. Another reason for such a weakness is the workhorse procedures used in this area (Brennan *et al.*, 1998; Chordia *et al.*, 2001). Authors may follow up on findings of a negative correlation only.

To correct for publication bias, I start with the graphical visualization proposed by Egger *et al.* (1997). Then, I add simple formal tests using ordinary least squares (OLS), the between effect and weighted least squares (WLS) (Stanley & Doucouliagos, 2012). Furthermore, an extension of the formal tests is provided by means of the latest improvement suggested by Bom & Rachinger (2019). Moreover, newly developed nonlinear methods such as use of the weighted average of adequately powered estimates Ioannidis *et al.* (2017), the stem-based method (Furukawa, 2019) and the nonparametric method of Andrews & Kasy (2019) complement the investigation. Finally, the presence of publication bias at least in contemporaneous cases, is identified. The mean after correction for publication bias has a negligible value. Finally, a caliper test (Gerber *et al.*, 2008; Gerber & Malhotra, 2008) for p-hacking and HARKing (hypothesizing after the results are known) completes the estimations about biases. Even these kinds of selections affect the results from primary studies.

Moreover, other study-specific aspects affect the corrected mean.The results of both Bayesian model averaging (BMA) and frequentist model averaging (FMA) indicate that data characteristics, structural variation and different methodological approaches explain a large part of the inconsistency in the primary results. For example, usage of *Monthly* data or *VAR* models makes the effect of trading volume on returns substantially more negative. In addition, an association arises between data age and the magnitude of published estimates. Newer data yield lower results that are closer to the predictions of the efficient market hypothesis, as is clear even in Figure 2. Other causes of variation are the type of returns or type of stocks.

After the differences are controlled for, the overall implied estimates become negative for both the contemporaneous and the dynamic cases. Based on these findings, opportunities to predict the stock returns exist. This calls into question the relevance of the efficient market hypothesis. For instance, the country and type of stock matter. The trading volume of stocks has the opposite effect on returns in emerging countries as on returns in developed markets. Moreover, the stocks of firms follow the EMH more closely than any other stocks. Thus, one should bear in mind the specifics of each stock when forming a portfolio, calibrating a model, preparing a trading strategy or conducting research. Relying on overall conclusions is dangerous.

The rest of the article has the following structure. Section 2 describes the procedure for collecting the primary studies. Section 3 investigates the presence of publication bias in the literature. Section 4 addresses heterogeneity in the primary studies and provides implied estimates. Section 5 summarizes the paper. Appendix Appendix: provides additional data description and robustness checks.

2 Data Collection

The first studies discussing the price-volume relationship originated in the US in the 1960s (Granger & Morgenstern, 1963; Godfrey *et al.*, 1964). The focus on the US continued for the next decades (e.g., Crouch, 1970; Jain & Joh, 1988). By the turn of the millennium, researchers from every continent had started to show interest in the topic. For example, two decades ago, Lo & Wang (2000) found almost two hundred articles related to trading volume. The articles were from various fields – economics, finance, and accounting. Furthermore, during the last two decades, many more articles have been published. Thus, although the EMH and Fama & French (1992) suggest substantial results, these conclusions have come into question again with the increased number of articles.

For instance, Lee & Rui (2002) and Gurgul *et al.* (2007) supported the EMH with their finding of a small or even negligible relationship between trading volume and expected stock returns. On the other hand, Brennan *et al.* (1998) and Chordia *et al.* (2001) contradicted this conclusion. They were not alone in doing so: Mahajan & Singh (2009a) and Akpansung & Gidigbi (2015) provided overviews of the currently available literature related to the topic. While they pointed out reasons for the major differences in the existing literature, no consensus about the magnitude of the effect emerged.

The data collection itself follows the guidelines for meta-analyses in economics proposed by Havránek *et al.* (2020). In the first step, I search for all relevant studies. Based on related literature surveys conducted by Mahajan & Singh (2009a) and Akpansung & Gidigbi (2015) and the workhorse methods in this field (Brennan *et al.*, 1998; Chordia *et al.*, 2001), I design a search query for Google Scholar. The final query returns all relevant articles related to the volume-return relationship. The query is worded as follows: trade | trading and volume and "expected stock return" | "stock return" | "price changes". This search goes through the full text of the study regardless of the precise formulation of the title, abstract, and keywords Gechert *et al.* (2020). Reading of the abstracts leads to the removal of three-fourths of the articles. I then read the full text of the rest of them. The latest study, from March 2019, is added, and then the literature search is terminated.

The articles deploy four comprehensive and distinctive strategies for studying the trading volume-return relationship. First, authors such as Lee & Rui (2000); Statman *et al.* (2006); Chuang & Lee (2006); Gurgul *et al.* (2007) focus on the effect of lagged returns on current trading volume. They follow an intuitive logic, supposing that people invest in stocks that displayed profits in the last season. The results of these studies support this intuition. The authors find that most of their estimates are significant. The second group of articles tests Granger causality. Granger (1969) proposes this methodology to test for "a correlation between the current value of one variable and past values of other variables" (Brandle, 2010). VAR models serve as the baseline for these tests. In the context of return and volume, these models assess whether volume Granger-causes returns and vice versa (e.g., Mestel *et al.*, 2003; Akpansung & Gidigbi, 2015). The literature describes the relationship as weak or nonexistent. The third group of studies, starting with Ying (1966), investigates stock markets by trading volume growth and stock returns. Ying (1966) finds a significant and positive relationship between volume growth and corresponding returns using the S&P 500 composite index. Similar findings are obtained, for example, by Gervais *et al.* (2001) when expanding Ying's analysis and by Watkins (2007) when using monthly NASDAQ data. The fourth and last group of articles studies the trading volume-return relationship itself. This group is represented by, for example, the aforementioned Brennan *et al.* (1998); Chordia *et al.* (2001), who find a negative and significant relationship using US stock data.

Since the results based on these four approaches are mutually incomparable, I study the fourth group only. There are several reasons for this decision: First, the aim of Fama-French factor models and Amihud & Mendelson's approach is to determine stock returns based on trading volume, *not vice versa*. I thus eliminate the first group. Second, the discussion is not focused on the question of simply *whether* there is a relationship but on *how much* trading volume affects returns. This excludes the Granger-causality testing studies. Third, many meta-analysis articles observe the trading volume and return relationship. This is not the case for the previous groups, including the group studying growth in returns. Last but not least, even though some differences remain in the approaches to estimation in the fourth group, all of them can be accounted for with the meta-analytical tools described in the rest of this Section and Section 4.

A detailed description of the study selection path is provided in Figure A1. In addition to skipping the articles investigating the opposite relationship, Granger causality or growth in volume, I drop research without measurements of the uncertainty of the estimates. The test for the presence of publication bias requires either standard errors or other metrics derived from standard errors. This condition stops the inclusion of some key contributions such as Chordia & Swaminathan (2000).

The final set of estimates meeting all conditions for the meta-analysis consists of 468 observations from 44 studies. The majority of the studies focus on US stock markets (e.g., Crouch, 1970; Epps

& Epps, 1976). Nonetheless, numerous studies from recent years assess emerging markets (e.g., De Meiros & Van Doornik, 2008; Tapa & Hussein, 2016) and China (e.g., Shu *et al.*, 2004). The data include both published articles and working papers. While using only published studies offers reassurance of the quality of the estimates, the inclusion of unpublished papers does not negatively affect the results. Rusnak *et al.* (2013a) discuss the usage of working papers and their effect on publication bias and suggest, "Authors who would preferably publish some estimates would do it rationally in early stage of publication." The same idea is raised by Doucouliagos & Stanley (2013a) and their evidence on 87 meta-analyses. They conclude that there is "no difference in the magnitude of publications selection between unpublished and published studies". Moreover, the inclusion of both published and unpublished articles enables me to study the difference between these two subgroups and helps better reveal the drivers of heterogeneity. An overview of the articles used appears in Table 1. The oldest study in the collected sample is from 1970 and the newest from 2019; therefore, the data have almost 50 years of coverage.

Table 1: Studies included in the meta-analysis

Al-Jafari & Tliti (2013)	Chen *et al.* (2001)	Ochere *et al.* (2018)
Assogbavi *et al.* (2007)	Chordia *et al.* (2001)	De Meiros & Van Doornik (2008)
Brandle (2010)	Lee & Rui (2002)	Pisedtasalasai & Gunasekarage (2007)
Brennan *et al.* (1998)	Lee & Rui (2000)	Rotila *et al.* (2015)
Ciner (2002)	Lewellen (2015)	Saatcioglu & Starks (1998)
Ciner (2003)	Lin & Liu (2017)	Sheu *et al.* (1998)
Crouch (1970)	Long *et al.* (2018)	Shu *et al.* (2004)
Datar *et al.* (1998)	Louhichi (2012)	Tahir *et al.* (2016)
Devanadhen *et al.* (2010)	Loukil *et al.* (2010)	Tapa & Hussein (2016)
Epps & Epps (1976)	Mahajan & Singh (2009b)	Le & Mehmed (2009)
Hafner (2005)	Mahajan & Singh (2009a)	Tripathy (2011)
Han *et al.* (2018)	Mahajan & Singh (2008)	Yin & Liu (2018)
Sana Hsieh (2014)	Marshall & Young (2003)	Yonis (2014)
Hu (1997)	McGowan & Muhammad (2012)	Zhong *et al.* (2018)
Chang & Wang (2019)	Narayan & Zheng (2010)	

Despite the strict selection criteria for the articles, several inconsistencies remain. These relate to the measures of return and volume themselves. In the case of return measures, most of the studies employ returns or absolute returns. The definition of returns is as follows:

$$\Delta Ret = ln(P_t) - ln(P_{t-1}) = ln(P_t/P_{t-1}), \tag{1}$$

where P stands for price and t captures time horizon.

Some older papers, such as Crouch (1970), employ price changes instead of returns. Nevertheless, authors now prefer returns to price changes since returns can allow comparisons between different stocks, firms, or studies. Moreover, some authors use *Abnormal* returns instead of returns. Abnormal returns are above-average returns from the previous time frame (e.g., Yin & Liu, 2018). Other authors prefer *Excess* returns, considering only returns above the risk-free rate (e.g., Chordia *et al.*, 2001). In particular, the last method is widely applied in Fama-MacBeth types of models. Fama-MacBeth models adapt this measure from the capital asset pricing model (CAPM) and arbitrage pricing theory (APT) (Brennan *et al.*, 1998).

Measurement of trading volume has also evolved over time. Early authors such as Crouch (1970) and Epps & Epps (1976) employ the number of shares traded as their volume measure. However, the turn-of-the-century study of (Datar *et al.*, 1998) suggests, "The number of shares traded by itself is not a sufficient statistic for the liquidity of a stock since it does not take into account the differences in the number of shares outstanding or the shareholder base". These authors, together with Brennan *et al.* (1998), proposed two alternatives. First, the turnover rate is associated with the investor holding period. Second, the dollar trading volume is related to how long a dealer waits to turn around his position (Chordia *et al.*, 2001). Finally, Lo & Wang (2000) compare all these approaches and recommend turnover as the most natural proxy for trading volume in the stock market. Thus, turnover is the preferred measure in most studies today (e.g., Long *et al.*, 2018; Chang & Wang, 2019; Zhong *et al.*, 2018).

Last but not least, the authors differ even in their approach to return-volume relationship measurement. One group, represented by Brennan *et al.* (1998); Chordia *et al.* (2001), explores the effects of past volume on expected stock returns. This dynamic approach can be used to assess the efficient market hypothesis, but contrary to the EMH, the major authors in the field (Brennan *et al.*, 1998; Chordia *et al.*, 2001) suggest a negative effect of lagged volume on expected returns. The second, similarly sized group (e.g., Epps & Epps, 1976; Datar *et al.*, 1998) studies the relationship of volume and returns in the same time period. Unlike the dynamic relationship, the contemporaneous relationship between returns and volume clarifies information about trading volume asymmetry. Both the dynamic and contemporaneous approaches yield inconclusive results (Hu, 1997; Akpansung & Gidigbi, 2015; Poudel & Shrestha, 2019). Thus, this study incorporates both relationships and distinguishes them with a dummy.

The variability in the measures of returns and volume obliges me to employ the approach of Valickova *et al.* (2015). In their investigation of financial development and economic growth, these authors define four measures of economic growth (dependent variable) and ten variables for financial development (independent variable). The differences among the measures are captured by dummies, and partial correlation coefficients (*PCCs*) enable comparability of the estimates at the cost of losing some information. *PCCs* come from the t-statistic of the estimate and the residual degrees of freedom (Greene, 2008). The sign of a partial correlation coefficient is the same as the sign of the original coefficient:

$$PCC_{ij} = \frac{t_{ij}}{\sqrt{t_{ij}^2 + df_{ij}}}, \tag{2}$$

where r_{ij} stands for the partial correlation coefficient of the ith estimate of the jth study. t denotes the corresponding t-statistic, and df the degrees of freedom. In addition to the estimates themselves, the corresponding standard errors need recalculation. I again follow the approach adopted by Valickova *et al.* (2015), as suggested by Doucouliagos & Stanley (2013b). They adapt the formula from Fisher (1954):

$$SEr_{ij} = \frac{PCC_{ij}}{t_{ij}}, \tag{3}$$

where SEr_{ij} denotes the standard error of the particular partial correlation coefficient PCC_{ij}. t_{ij} expresses the t-statistic from the ith regression of the jth study. In regards to other authors employing the partial correlation coefficient in economic meta-analyses, I can mention, for instance, Doucouliagos (2005). An overview of the distribution of *PCCs* per study used in my research is provided in Figure 3.

The mean reported estimate of the return-volume relationship is 0.021. Winsorization at 2.5% helps to deal with some extreme outliers in the data. The mean remains the same after winsorization, and the results remain robust even under alternative winsorization at 1% and 5%. Obviously, the estimates vary greatly both between and even within the primary studies. Thus, 49 explanatory variables collected for each observation address this variance. The task of these variables is to clarify diverse data characteristics such as data type, methodology used or market size.

Figure 3: Variation in the estimates both across and within studies

Notes: The box plot of the estimates of return-volume relationship published in primary studies highlights both median and interquartile range (P25 – P75). The coverage of whiskers reaches from (P25 – 1.5*interquartile range) to (P75 + 1.5*interquartile range). The dots capture remaining outlying estimates. The winsorization handles with overall outliers before computational tasks.

Table 2: Mean return-volume estimates for different subsets of data

	No. of obs.	Mean	Stand. Dev.	95% conf. int.	
Temporal dynamics					
Contemporaneous	224	0.071	0.011	0.050	0.093
Dynamic	244	-0.026	0.008	-0.041	-0.010
Data characteristics					
Hourly data	52	0.182	0.030	0.121	0.242
Daily data	118	0.074	0.013	0.048	0.099
Weekly data	32	0.113	0.016	0.08	0.147
Monthly data	266	-0.045	0.006	-0.058	-0.033
Panel data	286	-0.038	0.006	-0.051	-0.025
Time series data	175	0.118	0.012	0.094	0.143
Cross-sectional data	7	-0.009	0.013	-0.041	0.023
Structural variation					
All stocks	220	-0.051	0.007	-0.029	-0.008
Indexed stocks	92	0.057	0.013	-0.050	0.010
NASDAQ stocks	9	-0.103	0.042	-0.200	-0.007
Banks stocks	18	0.042	0.023	-0.005	0.090
Firms stocks	129	0.123	0.015	0.092	0.154
Developing countries	136	0.058	0.013	0.031	0.084
OECD countries	332	0.006	0.008	-0.01	0.022
Publication status					
Published papers	367	0.024	0.009	0.007	0.042
Unpublished papers	101	0.008	0.007	-0.006	0.021
All estimates	468	0.021	0.007	0.007	0.035

Notes: Table 6 provides a complete description of the definitions of subsets. Winsorization at 2.5% and 97.5% levels deals with the outliers.

A glimpse of the heterogeneity examined more closely in Section 4 provides Table 2. It summarizes the mean values of the return-volume relationship for different subgroups of data. These subgroups consider temporal dynamics, data frequency, type of data, type of stocks and publication characteristics.

The dynamic estimates show negative and significant effects, as in Brennan *et al.* (1998); Chordia *et al.* (2001). On the other hand, the contemporaneous estimates display slightly positive effects, as in Hiemstra & Jones (1994). The dynamic relationship usually connects the panel data with monthly frequency; thus, the means for the panel and monthly data subgroups are negative, like the mean of the dynamic relationship subgroup. In contrast, time series data at higher frequencies exhibit substantially positive results. Different types of stocks provide the following information. Firm stocks tend to have the highest mean estimate. On the other hand, NASDAQ stocks remain the lowest by a substantial margin. This finding appears even in Brennan *et al.* (1998); Chordia *et al.* (2001). The distinction between developing and OECD countries appears dissimilar. According to Llorente *et al.* (2002), the magnitudes should be the opposite of what the data reveal. Last but not least, published and unpublished papers do not seem to differ significantly. In summary, this simple analysis proposes systematic differences among the reported estimates, but without correction for publication bias as in Section 3 and proper investigation of the sources of heterogeneity as in Section 4, any conclusions drawn will be misleading.

3 Publication Bias

The phenomenon of publication bias extensively affects economic literature. Ioannidis *et al.* (2017) find that estimates reported in the economics literature are typically exaggerated twofold because of publication bias. Authors naturally prefer a statistically significant estimate with the expected sign. From one point of view, this preference makes sense. One should not have to focus on evidently wrong estimates. On the other hand, substantial ignorance of statistically insignificant estimates with the "wrong" sign distorts the literature as a whole. Addressing this subject, Nansen McCloskey & Ziliak (2019) discussed the Lombard effect. Like speakers who raise their voice in the presence of noise, researchers particularly augment their efforts to find a significant effect in the case of noisy data or poor estimation techniques. Statistically significant estimates at the 5% level with the "correct" sign are nearly always possible to reach in economics in the presence of the freedom to choose from among a large number of different specifications. On the other hand, statistically significant results gained in this manner no longer reflect the primary theoretical purpose of conducting statistical tests.

A classical perspective from which to study the return-volume relationship is in relation to the efficient market hypothesis. The assumptions that returns are not predictable and that the stock price already incorporates all available information affect the thinking of the financial community as a whole. Fama (1970) concluded the following: "The evidence in support of the efficient markets model is extensive, and (somewhat uniquely in economics) contradictory evidence is sparse. Nevertheless, we certainly do not want to leave the impression that all issues are closed". Despite this strong declaration, the possibility of predicting the market drives research on stock exchange returns. Just few years later, Basu (1977) observed, "While there is substantial empirical evidence supporting the efficient market hypothesis, many still question its validity". The same holds even today. For instance, Malkiel (2003) emphasizes that "pricing irregularities and even predictable patterns in stock returns can appear over time and even persist for short periods". Thus, some authors may find or try to find no significant effect, with null estimates, or make excessive efforts to find significant results (the Lombard effect). Thus, studying how estimates of the trading-volume relationship are obtained is a compelling topic to scrutinize. It is important to determine whether estimates differ simply because of different economic and data backgrounds (Section 4) or because of selection by authors.

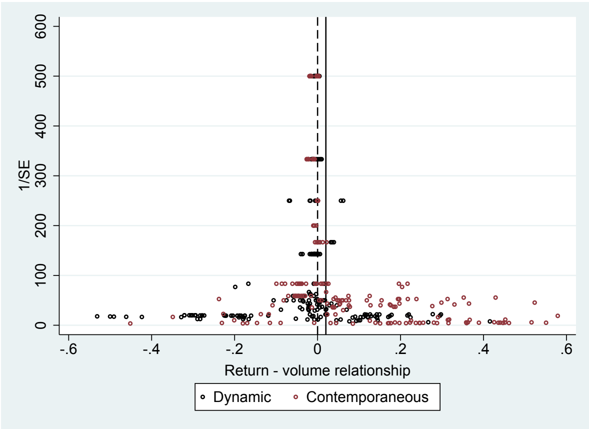
A common tool for detecting the extent of publication selection is the so-called funnel plot, first proposed by Egger *et al.* (1997). A funnel plot depicts the magnitude of the estimated effect on the horizontal axis. The vertical axis then captures the precision, measured by the inverse of the estimated standard error. Since the studies on the return-volume relationship provide standard errors with a symmetrical distribution (usually a t-distribution), the estimates should have a symmetrical distribution around the true mean effect regardless of their magnitude and precision. The estimates become further from the true effect as precision decreases. Thus, the estimates form a symmetrical inverted funnel. In the presence of publication bias, the funnel plot should be asymmetrical or hollow. The discarding of estimates of a particular sign or magnitude would cause this asymmetry, while the rejection of statistically insignificant estimates would cause the hollowness. The worst case arises when the funnel plot is both asymmetrical and hollow (Egger *et al.*, 1997).

Figure 4, which presents contemporaneous and dynamic estimates separately, gives a clear message. The depicted funnel plots show that the dynamic estimates are distributed more or less equally around zero. The same holds for the median estimates for each study. This result offers evidence for the EMH. In contrast, the contemporaneous estimates are skewed to the left, which indicates the possible presence of publication bias in this case.

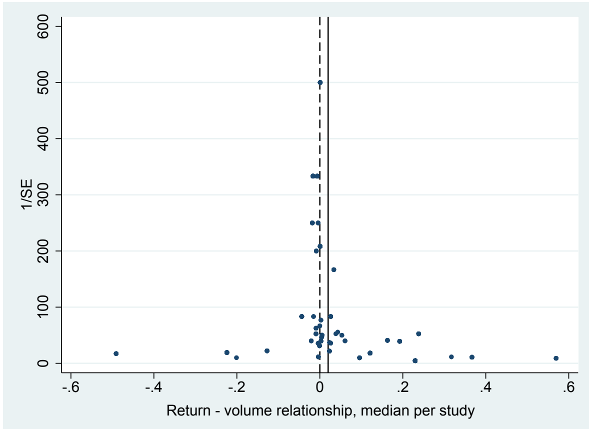
Nonetheless, the funnel plot represents a simple visual test only. Regression-based funnel asymmetry tests offer a more reliable way to check for publication selection. The following base regression (Stanley & Doucouliagos, 2012) explores the correlation between the return-volume relationship and its standard error $SE(r_{ij})$:

$$r_{ij} = \beta_0 + \beta_1 SE(r_{ij}) + e_{ij}, e_{ij} \sim N(0, \sigma^2), \tag{4}$$

Figure 4: Funnel plot: Little evidence of publication bias in this field



(a)



(b)

Notes: Without the publication bias the scatter plot seems like an inverted funnel symmetrical around the most precise estimates. The left panel depicts all estimates distinguished by the time dynamics. The right panel shows median estimates per study. The solid line stands for overall mean relationship. The dashed line is set at zero. The computational tasks includes even outliers in winsorized form, but for the ease of exposition the funnels excludes them.

where r_{ij} stands for the ith estimate of the partial correlation coefficient between expected stock returns and trading volume from study j. β_0 expresses the mean underlying effect beyond publication selection bias, and the coefficient β_1 reveals the strength of publication bias. The aforementioned Lombard effect or discarding of estimates with the "wrong" sign may cause the correlation. If $\beta_1 = 0$, publication bias is not present in the field. Otherwise, publication bias is present.

I estimate Equation 4 with four different estimation methodologies. First, I use simple OLS with standard errors clustered at the level of individual studies and countries. The two-way clustering follows the suggestion of Cameron *et al.* (2012). Second, I run a panel data regression employing between effects. Third, I follow Stanley & Doucouliagos (2012) and Astakhov *et al.* (2019) in multiplying Equation 4 by $1/SE(r_{ij})$. This assigns more weight to more precise studies and directly deals with heteroskedasticity. Therefore, the weight $1/SE(r_{ij})$ is called *Precision*. In the fourth specification, instead of *Precision*, I use the inverse number of estimates per study as a weight.

Table 3 presents the results of these four specifications. The first row shows the baseline result of the OLS regression of the partial correlation coefficient on its standard error. The β_1 coefficients

Table 3: Formal tests on the presence of publication bias

	All	Contemporaneous	Dynamic
PANEL A: Unweighted estimations			
OLS			
SE (publication bias)	0.867***	0.876***	-0.169
	(0.091)	(0.100)	(1.247)
	[-0.487; 2.349]	[-0.844; 4.775]	[-3.729; 2.714]
Constant (effect beyond bias)	-0.013	0.027	-0.021
	(0.014)	(0.018)	(0.013)
	[-0.063; 0.042]	[-0.021; 0.106]	[-0.094;-0.004]
Between effects			
SE (publication bias)	1.069**	1.229**	1.436**
	(0.483)	(0.601)	(0.621)
	-	-	-
Constant (effect beyond bias)	-0.002	0.038	-0.057*
	(0.026)	(0.033)	(0.029)
	-	-	-
PANEL B: Weighted OLS estimations			
Weighted by the inverse of the number of estimates reported per study			
SE (publication bias)	0.965**	0.960*	0.860
	(0.478)	(0.531)	(0.799)
	[-1.091; 2.995]	[-2.644; 4.613]	[-1.405; 2.454]
Constant (effect beyond bias)	0.001	0.044*	-0.041**
	(0.012)	(0.021)	(0.018)
	[-0.025; 0.026]	[0.000; 0.090]	[-0.083; -0.007]
Weighted by the inverse of the standard error			
SE (publication bias)	0.771**	1.672***	-0.807
	(0.376)	(0.444)	(1.255)
	[-2.330; 2.498]	[-6.747; 7.385]	[-3.585; 2.019]
Constant (effect beyond bias)	-0.009**	-0.014	-0.003
	(0.004)	(0.010)	(0.003)
	[-0.037; -0.002]	[-0.148; 0.019]	[-0.122; 0.070]
Observations	468	224	244

Notes: The uncorrected mean of the estimates is 0.021. Panels A and B report the results of the regression $r_{ij} = \beta_0 + \beta \cdot SE(r_{ij}) + u_{ij}$, where r_{ij} is the ith estimate of the partial correlation coefficient between expected stock returns and trading volume from study j. All = entire dataset. Contemporaneous = immediate effect of the trading volume-stock return relationship. Dynamic = lagged trading volume-stock return relationship. SE = standard error. Standard errors reported in parentheses are clustered at the study and country level (except between effects; the usage of two-way clustering follows Cameron *et al.* 2012). The square brackets report 95% confidence intervals from wild bootstrap clustering and Rademacher weights with 999 replications (except between effects; the implementation follows Roodman 2020). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

indicating the possible presence of publication bias are both positive and significant. This suggests a strong selective reporting bias. The negative but insignificant constant represents the underlying mean partial correlation coefficient corrected for reporting bias. Hence, the baseline result suggests negligible evidence for a return-volume relationship in the data. Taking a closer look, one can distinguish the presence of bias in the contemporaneous estimates. On the other hand, the dynamic estimates do not show this bias. Both cases display a insignificant effect beyond bias. It seems that authors make efforts to find an effect of trading volume on stock returns.

The second part of Panel A of Table 3 shows the results of the panel data regression with between effects. The between effects indicate an even stronger selective reporting bias than that found in the case of OLS. The corrected partial correlation coefficient again appears insignificant. In the between effects estimation, publication bias is present even in the dynamic cases, where a negative and significant effect beyond bias emerges. However, even in this case, the effect is not substantial: Doucouliagos (2011), in his guidelines on partial correlation coefficients, considers such an effect not even "small". He defines a small effect as one ranging from 0.07 to 0.17 in absolute value, a medium effect as one ranging between 0.17 and 0.33, and a large effect as one above 0.33. Moving on, Panel B reports the analysis of the WLS estimation with the precision and inverse number of estimates per study as weights. The findings derived from these two specifications simply accentuate the findings from Panel A. They confirm the presence of bias in the contemporaneous case and a negligible effect beyond bias.

In addition to commonly used and widely known publication bias detection techniques, I employ four recently developed advanced techniques. Estimating β_0 from Equation 4 yields an unbiased estimate of the mean corrected for publication bias only if publication selection is proportional to the standard error. Nevertheless, in practice, I am dealing with an unknown functional form of the publication selection procedure. Therefore, first, I employ the advanced estimator introduced by Andrews & Kasy (2019). Their estimator addresses the detected problem and remains unbiased under probably any form of publication selection (Havranek & Sokolova, 2019). Furthermore, the Andrews & Kasy (2019) specification works especially well with a small corrected effect, as in the case of my dataset. Thus, I consider this specification the most suitable method for my research.

The next approach was recently released by Furukawa (2019). This approach, known as the stem-based method, works only with the most precise estimates, optimizing the number of estimates for investigating publication bias by minimizing the mean squared error of the estimates. This conservative, fully data-dependent, nonparametric method robustly alleviates publication bias under various assumptions. The third technique was introduced by Ioannidis *et al.* (2017) and Stanley *et al.* (2017). The weighted average of the adequately powered (WAAP) estimator defines only the most precise estimate to be "adequately" powered for testing publication bias. WAAP estimation is the dominant method in studies with numerous high- and low-powered estimates. This characteristic does not hold in my case, and even one estimation cannot be run given the lack of precise observations. Finally, the kinked method of Bom & Rachinger (2019) is used. It improves on the precision effect estimate with standard error (PEESE) test for publication bias and outperforms the WAAP method.

Table 4 summarizes the results of the nonlinear techniques and shows ambiguous findings on the significance and sign of the overall effect. The same holds for the contemporaneous effect. On the other hand, the dynamic effect is negative and significant after publication bias is corrected. These results line up with the findings of, for instance, Brennan *et al.* (1998).

Table 4: Results of nonlinear techniques support previous findings

	All	Contemporaneous	Dynamic
PANEL C: Non-linear estimations			
Selection model (Andrews & Kasy, 2019)			
Effect beyond bias	0.010	0.062***	-0.031***
	(0.007)	(0.013)	(0.008)
Stem-based method (Furukawa, 2019)			
Effect beyond bias	-0.006**	0.001	-0.006***
	(0.003)	(0.006)	(0.001)
Weighted average of adequately powered (Ioannidis *et al.*, 2017)			
Effect beyond bias	-0.012	-0.010	-
	(0.013)	(0.008)	-
Kinked method (Bom & Rachinger, 2019)			
Effect beyond bias	-0.011***	-0.023***	-0.003***
	(0.002)	(0.004)	(0.001)

Notes: Andrews & Kasy (2019) proposed the nonlinear estimation technique. The approaches of Furukawa (2019), Ioannidis *et al.* (2017) and Bom & Rachinger (2019) work with the most precise estimates. For the Ioannidis *et al.* (2017) approach, the constant denominator of 1.84 instead of 2.8 is preferred. Standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Moreover, I use the caliper test proposed by Gerber *et al.* (2008); Gerber & Malhotra (2008); Bruns *et al.* (2019) to supplement the inspection of publication bias. This test focuses on different results selection stages called p-hacking and HARKing. In general, in cases of publication bias, authors simply do not publish results with insignificant estimates, but in the case of p-hacking, authors include only models with significant estimates in the study. In the case of HARKing, authors set their hypothesis after the results are already known Bruns *et al.* (2019). No test can distinguish between p-hacking and HARKing themselves, but the two can be distinguished from the first kind of bias. The caliper test aims at uncovering p-hacking and HARKing.

The test does not reveal anything about the corrected effect. It is based on the study of the break in reported t-statistics, where the break around the usual significance threshold indicates selective reporting. When authors do not report selectively, the distribution of t-statistics remains even around the usual significance thresholds of 1.96, 1.645, and 1.

Table 5: Caliper test

		All		Contemporaneous		Dynamic	
	Caliper size	0.1	0.2	0.1	0.2	0.1	0.2
90%	Above C	20.0%	26.1%	-	40.0%	22.2%	22.2%
	Below C	80.0%	73.9%	-	60.0%	77.8%	77.8%
	p-value	0.051	0.018	-	0.704	0.095	0.014
95%	Above C	75.0%	70.0%	66.7%	66.7%	85.7%	73.3%
	Below C	25.0%	30.0%	33.3%	33.3%	14.3%	26.7%
	p-value	0.041	0.026	0.347	0.207	0.047	0.068
99%	Above C	43.8%	53.8%	66.7%	50.0%	30.0%	55.6%
	Below C	56.2%	46.2%	33.3%	50.0%	70.0%	44.4%
	p-value	0.633	0.703	0.465	1.000	0.223	0.651

Notes: The table provides caliper tests following Bruns *et al.* (2019) for caliper sizes 0.1 and 0.2 and for the hypothesis of a 50:50 distribution. The numbers express the share of observations in a given interval around the significance threshold. The test parameter follows $C = \frac{n_{oc}}{n_{oc}+n_{uc}}$, where n_{oc} and n_{uc} stand for the number of observations with t-statistics in the interval above and below the threshold.

The results summarized in Table 5 suggest no selection around the 90% and 99% levels. On the other hand, breaks at the middle level – around the 95% interval – indicate bias. The results hold for different caliper sizes. This means that authors push their estimates above the 95% level but do not do so at the 10% and 1% levels. Moreover, contrary to the conclusions on publication bias, this type of bias distorts the dynamic estimates.

In summary, the sample indicates the presence of publication bias and seemingly even p-hacking or HARKing. Various tests inform these conclusions. First, the contemporaneous effects are biased, but the dynamic ones are not. Second, authors are particularly likely to provide biased estimates around the higher confidence intervals. Third, the corrected mean has a negligible value. On the other hand, aspects other than publication bias may influence the value of the corrected mean. The Section 4 deals with these heterogeneity drivers.

4 Drivers of the Relationship

Five different explanations of the difference in the estimates of the return-volume relationship have been repeatedly mentioned in the literature. The first three of them – the different measures of volume and returns and the difference between contemporaneous and dynamic estimates – I have already discussed in Section 2. As a fourth reason, the existing literature suggests an effect of data frequency. The use of lower-frequency data yields a positive relationship between returns and volume. This conclusion is supported by, for instance, Lee & Swaminathan (2000), who use monthly data, and Comiskey *et al.* (1987), who employ annual data. The fifth explanation is that the differences may spring from data aggregation. Studies such as Jain & Joh (1988) and Lee & Rui (2000) use aggregate data from stock markets. Their findings again indicate a positive relationship between return and volume. In contrast, Chordia *et al.* (2001) distinguish *NASDAQ* stocks, since the trading volume-stock return effect appears more negative for this exchange than for the NYSE or AMEX (which belong to the *All* group). Other groups include *Index* stocks Han *et al.* (2018), *Banks* stocks (Rotila *et al.*, 2015; Al-Jafari & Tliti, 2013) and *Firms* stocks (Tahir *et al.*, 2016; Datar *et al.*, 1998). The findings from the studies focused on these groups vary.

These five explanations represent only a few reasons why the published estimates might differ among themselves. I present my first attempt to clarify the origins of the heterogeneity in Table 2. To inspect the heterogeneity between the estimates of the return-volume relationship, I capture 49 features of the individual study design and expand Equation 4 by adding these features as independent variables. All the classified variables, with their definitions, mean, standard deviation and median lists the Table 6. The variables are divided into subgroups for ease of exposition. Twenty-two aspects relate to data characteristics, 13 to structural variation, 11 to estimation techniques, and the last three to publication characteristics.

Table 6: Description and summary statistics of the regression variables

Label	*Description*	*Mean*	*SD*	*Median*
SE	Estimates of standard errors of return-volume relationship (winsorized at 1% level)	0.04 (0.04)	0.06 (0.06)	0.02 (0.02)
Data characteristics				
Contemporaneous	=1 if the return-volume relationship is contemporaneous	0.48	0.50	0.00
Dynamic	=1 if the return-volume relationship is dynamic	0.52	0.50	1.00
Returns	=1 if the returns in any form are estimated	0.96	0.20	1.00
Price change	=1 if the price change is estimated instead of returns	0.04	0.20	0.00
Normal	=1 if the returns or price change itself are used	0.42	0.50	0.00
Absolute returns	=1 if the returns or price change are in absolute terms	0.12	0.32	0.00
Abnormal returns	=1 if the returns are defined as abnormal	0.03	0.16	0.00
Excess returns	=1 if the returns are defined as excess	0.43	0.50	0.00
Turnover	=1 if the volume is expressed as the number of shares traded during a time period divided by the number of shares outstanding at the end of the time period	0.55	0.50	1.00
Dollar volume	=1 if the volume is expressed in terms of dollar volume of the trade	0.13	0.34	0.00
Shares traded	=1 if the volume is expressed in terms of shares traded	0.32	0.47	0.00
Detrended series	=1 if the volume series was detrended	0.11	0.31	0.00
Data period	Length of time period	14.53	10.81	11.5
Data size	Total of observation (in logarithms)	8.61	3.27	8.19
Midyear	The logarithm of the mean year of the data used minus the earliest mean year in our data plus one	2.99	0.75	3.13
Hourly data	=1 if the data were collected hourly or more frequently	0.11	0.32	0.00
Daily data	=1 if the data were collected daily	0.25	0.44	0.00
Weekly data	=1 if the data were collected weekly	0.07	0.25	0.00
Monthly data	=1 if the data were collected monthly	0.57	0.50	1.00
Panel	=1 if the panel data were used	0.61	0.49	1.00
Time series	=1 if the time series data were used	0.37	0.48	0.00
Cross-section	=1 if the cross-sectional data were used	0.02	0.12	0.00
Structural variation				
All	=1 if the research relies on the data for the whole stock-exchange at least	0.47	0.50	0.00
Index	=1 if the cumulative returns value for stocks from particular index was used	0.20	0.40	0.00
NASDAQ	=1 if the cumulative returns value for NASDAQ stocks is used	0.02	0.14	0.00
Banks	=1 if the returns relate only to banking sector	0.04	0.19	0.00
Firms	=1 if the returns relate to firms stocks (i. e. do not relate to the banks)	0.28	0.45	0.00
Developing country	=1 if the estimate is for developing country	0.29	0.46	0.00
OECD	=1 if the estimate is for OECD country	0.71	0.46	1.00
Market size	Market size in terms of GDP (billions of dolars) in midyear of data (in logarithms)	7.90	1.39	7.11
North America	= 1 if the observation is linked to the North America	0.47	0.50	0.00
Asia	= 1 if the observation is linked to Asia	0.36	0.48	0.00
Europe	= 1 if the observation is linked to Europe	0.13	0.34	0.00
Australia	= 1 if the observation is linked to Australia	0.02	0.15	0.00
Other Continents	= 1 if the observation is linked to Latin America or Africa	0.02	0.15	0.00
Estimation technique				
Fama-Macbeth	=1 if the Fama-Macbeth model is used	0.48	0.50	0.00
VAR	=1 if the VAR model is used	0.24	0.42	0.00
Simple model	=1 if the simple linear model is used	0.22	0.42	0.00

Continued on next page

Table 6: Description and summary statistics of the regression variables (continued)

Label	*Description*	*Mean*	*SD*	*Median*
GARCH	=1 if the ARIMA with GARCH in error term is used	0.06	0.24	0.00
Monday	=1 if effect of Monday or January trading is considered	0.03	0.17	0.00
Trimmed	=1 if the primary dataset was trimmed	0.09	0.29	0.00
January excluded	=1 if all months but January are included in the primary dataset	0.08	0.27	0.00
OLS	= 1 if OLS estimation method is employed	0.43	0.50	0.00
MLE	= 1 if MLE estimation method is employed	0.04	0.20	0.00
GMM	=1 if GMM estimation method is employed	0.29	0.45	0.00
Other methods	=1 if other types of estimation is employed	0.24	0.43	0.00
Publication characteristics				
Impact factor	Discounted recursive impact factor from RePEc IDEAS	0.64	0.97	0.05
Citations	The logarithm of the number of Google Scholar citations normalized by the number of years since the publication year plus one	1.96	1.49	1.60
Published	=1 if the article was published	0.78	0.41	1.00

Notes: This table shows the mean, standard deviation and median for each variable used in the estimation. The effects in the brackets for standard error shows values for the unwinsorized estimates. The average partial effect method (Wooldridge, 2015) is used for the means of all variables in logarithms. Market sizes are collected from the World Bank database. (The World Bank: GDP [online]. Infogram: © 2016 [cit. 18. 2. 2019]. Available from https://data.worldbank.org/indicator/NY.GDP.MKTP.CD.) GDP is in billions of US dollars at the midyear point of the data. For Taiwan only, data are obtained from the National Statistics Republic of China (National Statistics Republic of China [online]. Infogram: @ 2019 [cit. 18. 2. 2019]. Available from:https://eng.stat.gov.tw/point.asp?index=1), since the World Bank does not provide information for Taiwan. These values are recalculated to US dollars based on the midyear NTD-USD exchange rate according to the Federal Reserve Bank. (Federal Reserve Bank: Real Effective Exchange Rates [online]. Infogram: ©2016 [cit. 18. 2. 2019]. Available from: https://fred.stlouisfed.org/graph/?id=AEXTAUS.) For each midyear, I use the year-end exchange rate to recalculate the current value. For 1981 only, I use the exchange rate from 31 December 1983, since earlier data are not available. The impact factor is downloaded from RePEc, and the number of citations is downloaded from Google Scholar. The rest of the variables are collected from studies investigating the return-volume relationship.

Data characteristics. In addition to the different proxies for return and volume measures mentioned in Section 2, other study-invariant characteristics arose during data collection. For example, the studies are divided by data type – that is, whether the data are *Cross-sectional*, *Time series* or *Panel*. In terms of the direction of the effect of any of these characteristics, there is no prior knowledge in this field (Akpansung & Gidigbi, 2015). In addition to the previously mentioned distinctions, I discern the number of observations, *length* of the time period in years and *Midyear* of the data used. I considered using the year of publication (*Pubyear*) to capture differences in publishing data, but its correlation of above 85% with *Midyear* led me to discard this idea. Then, I considered adding the squares of *Midyear*, but again, the correlation with its linear term of above 97% did not allow me to do so. According to the findings of Schürenberg-Frosch (2015), linear terms suffice. Furthermore, one variable from each mentioned group of variables is dropped due to the dummy variable trap.

Structural variation. In addition to the different volume and return measures, I investigate and capture the research area in each article with dummies. Beyond the previously mentioned categories of stock types, the distinction of whether a country belongs to the *OECD* deserves attention. The logic of the variable is to determine whether more advanced markets display different effects on returns. Emerging markets may exhibit higher volatility and a greater probability of large price changes than developed markets (De Santis *et al.*, 1997). In addition, there is a similar intuition behind the differentiation of continents. North America, Asia, Europe and Australia each have their own dummies. Furthermore, *Market size*, measured in terms of GDP at the midyear of the data, helps distinguish larger markets from smaller ones on each continent. These variables together are sufficient to capture diversity in the origins of the scrutinized stocks.

Estimation techniques. Several different approaches to estimating the trading volume relationship have evolved over time. The current workhorse model in this field is the *Fama-MacBeth* methodology. It has the same basis as the Fama-French models, and Chordia *et al.* (2001), Lewellen (2015) or Brandle (2010), for instance, have promoted it significantly. The *Fama-MacBeth* approach dominates among the estimation techniques in my sample, with 43% usage among the primary articles. The baseline equation for this group of models is $Ret = \alpha_0 + \alpha_1 Vol + \alpha_2 Size + \alpha_3 BM + \alpha_4 Price + \alpha_5 Ret_{2-3} + \alpha_6 Ret_{4-6} + \alpha_7 X + e, e \sim N(0, \sigma^2),$, where *Ret* represents excess return, *Vol* stands for trading volume, *Size* expresses the natural logarithm of the firm's market value of equity and *BM* indicates the natural logarithm of the book value of equity to the market value of equity. Moreover, Ret_{2-3} and Ret_{4-6} record the returns in previous periods, and X covers a set of other variables added into the model, such as yield, firm beta, market beta and firm size. I initially included the X variables in my estimation as well, but they appear insignificant, so for the sake of parsimony, I do not include them in the main results and move them into the appendix. See Table A2 and A4. Sometimes either *Size* or *Price* are removed from the baseline model, but again this makes no significant difference, so I move these estimations into the appendix as well.

Bivariate VAR approaches comprise the second group of models. They again have small differences among themselves with respect to the number of lags included. All the VAR models belong to one group in the baseline estimation, and a more granular division is used in the robustness checks (Table A2 and A4). In addition to other authors, Saatcioglu & Starks (1998), Lee & Rui (2002) and Ciner (2002) use VAR models in their estimation. The baseline VAR equation is as follows:

$$\begin{aligned} Ret_t &= \alpha_0 + \alpha_1 Vol_t + \alpha_2 Vol_{t-1} + \alpha_3 Ret_{t-1} + e_t, e_t \sim N(0, \sigma^2), \\ Vol_t &= \beta_0 + \beta_1 Ret_t + \beta_2 Ret_{t-1} + \beta_3 Vol_{t-1} + u_t, u_t \sim N(0, \sigma^2). \end{aligned} \tag{5}$$

The *Simple* model group is a similarly large group, used in 21% of all the estimates. This group uses the following equation: $Ret_t = \alpha_0 + \alpha_1 Vol_t + e_t, e_t \sim N(0, \sigma^2)$. The *simple* model employ, for example, Shu *et al.* (2004) or Tapa & Hussein (2016). In some cases, GARCH improves the variance equation by capturing heteroskedasticity: $h_t = \sigma_t^2 = \beta_0 + \beta_1 e_{t-1}^2 + \beta_2 h_{t-1}$. These models, used by Sana Hsieh (2014) and Tahir *et al.* (2016), among others, represent the fourth and last model group.

Another differentiation is based on the estimation methodology. Epps & Epps (1976) suggest that *OLS* estimates may have an upward bias; thus, they estimate the equations not only by *OLS* but also by maximum likelihood estimation (*MLE*). Beyond these two estimation techniques, newer articles prefer generalized method of moments (*GMM*) estimation. Among these are the papers of, for instance, Lee & Rui (2002) and Ciner (2003). Some articles do not mention the estimation technique; thus, these three groups are supplemented by the *Other methods* category when the method used is not clear or specified.

Another differentiation comes from the estimation methodology. Epps & Epps (1976) suggests that OLS estimates may have upward bias, thus they estimate the equations not only by OLS, but even with MLE. Besides these two estimation techniques newer articles prefer GMM estimation. Among these authors belongs, for instance, Lee & Rui (2002) and Ciner (2003). Articles do not mention other estimation technique, thus these three groups supplements the category *Other methods*, when the method used is not clear or specified.

Moreover, some estimates related to Monday trading only (Pisedtasalasai & Gunasekarage, 2007). After weekends, stock markets are supposedly calmer (French, 1980; Gibbons & Hess, 1981). Therefore, these estimates are designated with the dummy variable *Monday*. A similar effect relates to January estimates in monthly data (Hu, 1997). Thus, the January effect is joined with the Monday effect into one dummy. On the other hand, *January excluded* estimates are also distinguished. Finally, I control for whether the primary data are *Trimmed*.

Publication characteristics. Last but not least among the variables capturing differences in the primary estimates are those related to publications. The impact of the quality of any publication may be studied from several perspectives. I employ three of them. One may expect these factors to be correlated with the unobserved features of the paper. With the dummy variable *Published*, I obtain a systematic overview of whether published studies display different results from those in unpublished articles.

Furthermore, the quality of the outlet distinguishes the *Impact factor* variable. This represents the discounted recursive RePEc impact factor of the primary study. This variable has been used in previous meta-analyses, for example, Valickova *et al.* (2015) or Rusnak *et al.* (2013b). The advantage of its use arises from its availability for both working paper series and journals. As the last publication characteristic, I choose the logarithm of the number of Google Scholar *Citations* normalized by the number of years since the first version of the study appeared. This reflects each article's relevance in the literature. In addition, these three variables are useful for the detection of potential publication bias. This fact increases their importance for the study.

4.1 Model Averaging

Model averaging techniques account for the outlined heterogeneity. Namely, the analysis deploys Bayesian model averaging in several specifications together with frequentist model averaging. Model averaging approaches have several merits in comparison with best-model approaches. First, they address model uncertainty in a systematic manner. Second, they deal with potential problems arising from mental conflict when one is faced with several competing model specifications. Third, they treat omitted variable bias methodically.

Raftery (1995) and Raftery *et al.* (1997) pioneered the deployment of BMA in social sciences. The widespread usage of the BMA approach, even in economics, is testified to in a summarizing article written by Moral-Benito (2015). In contrast, the usage of FMA in economics does not have such a long history. This branch of techniques was thoroughly described just a decade ago by Magnus *et al.* (2010) and Amini & Parmeter (2012). In economics, its usage has developed only recently (e.g., Havranek *et al.*, 2017; Steel, 2020; Gechert *et al.*, 2020; Bajzik *et al.*, 2020; Ehrenbergerova & Bajzik, 2020).

The base equation for both model averaging techniques consists of regressing an estimate on its standard error plus on the set of all control variables. Equation 6 clarifies the approach:

$$r_{ij} = \alpha_0 + \beta_0 SE + \sum_{k=1}^{39} \beta_k X_{k,ij} + e_{ij}, \tag{6}$$

where SE represents the standard error of the primary estimate and $X_{k,ij}$ the value of the kth explanatory variable for the ith estimate from the jth study.

Based on the definition, model averaging techniques do not exclude any variables in advance. This fact is of considerable importance when the aim is to explain heterogeneity among the studies. In my case, model averaging could potentially imply running 2^{39} regressions stemming from all the possible model combinations. Since such a process would be time consuming, BMA deploys a Markov chain Monte Carlo process with the Metropolis-Hastings algorithm (Zeugner & Feldkircher, 2015) to avoid it. This algorithm walks through the most probable models and assigns a posterior model probability (PMP) to each of them. The PMP expresses the probability of employment of the particular model. Based on the different PMPs, the posterior inclusion probability (PIP) of each variable arises. The PIP is a weighted average of the estimated coefficient of the variable, where the weights are the PMPs of the models. In comparison, FMA uses the orthogonalization of the covariate space (Amini & Parmeter, 2012). The easily interpretable posterior model probabilities and posterior inclusion probabilities make BMA preferable to FMA (Steel, 2020). These statistics show more information than the simple point estimates with confidence intervals from FMA. Moreover, the scale at which one performs BMA does not matter based on the transformation invariance of the approach. This represents the second distinctive advantage of BMA over FMA (Fletcher, 2018).

Both model averaging techniques require the setting of some prior knowledge. In the baseline BMA setting, I prefer the unit-information g-prior suggested by Eicher *et al.* (2011). It assigns each model the same prior weight and hence provides a convenient setting when there is a lack of knowledge of the parameter values. In addition to the g-prior, BMA requires model prior setting. Due to the small sample size, I prefer the dilution prior recommended by George *et al.* (2010) and by Hasan *et al.* (2018). It multiplies the prior model probabilities by the determinant of

the model's correlation matrix. When the considered model is highly collinear, the determinant goes to zero. Thus, the model is given a small weight. For models with little collinearity, the opposite holds. Thus, the dilution prior deals with potential collinearity problems.
The robustness checks examine several BMA setting alternatives. Namely, I examine the data by combining a uniform g-prior with a uniform model prior and BRIC g-prior with a beta-binomial random model prior Fernandez *et al.* (2001); Ley & Steel (2009). Last, Mallow's criterion for model averaging (Hansen, 2007) is used to deal with prior knowledge in the FMA setting.

4.2 Results

Turning to the results, an early presentation of the BMA conclusions appears in Figure 5. The model ordering goes from left to right, from the most significant to the least significant. The PMP of each model is captured by the corresponding column width. In a similar manner, the rows sort the explanatory variables. The variables with the highest PIP appear at the top. The cells at the nexus of the rows and the columns capture the effect that a variable has in a particular model. Red indicates a negative effect on the coefficient of interest and blue a positive effect, and the cell remains blank when the model does not include the variable. Hence, the stable red and blue variables are the ones of main interest.
Table 7 converts the graph into numbers. The PIP is now expressed in decimal numbers. According to Eicher *et al.* (2011), decisive variables are those with a PIP between 0.99 and 1, strong variables those with a PIP between 0.95 and 0.99, substantial variables those with a PIP from 0.75 to 0.95 and weak variables those with a PIP in the range of 0.5 to 0.75. Furthermore, the OLS frequentist check on at least weak variables completes the table.

Finally, I classify seven variables (without intercept) as decisive. This finding indicates that BMA is the proper choice for estimation. Moreover, these variables display high stability, as is clear in Figure 5. Of decisive importance for the trading volume-stock return relationship are the *Abnormal returns*, *Data size*, *Midyear*, *Monthly* data frequency, *VAR* model, *Other methods* and *Other continents* variables. Altogether, variables from three of the four categories (data characteristics, structural variation and estimation technique) drive the differences across the estimated coefficients. The main discussion of the results focuses on these difference-making variables.

Table 7: Explaining heterogeneity – BMA dilution prior and frequentist check

	BMA – Dilution prior			Frequentist check – OLS		
Response variable:	Post. Mean	Post. SD	PIP	Estimate	SE	p-value
Constant	0.155	NA	1.000	0.137	0.017	0.000
Winsorized SE	-0.001	0.029	0.016			
Data characteristics						
Contemporaneous	0.000	0.002	0.017			
Price change	0.000	0.004	0.014			
Absolute returns	0.040	0.037	0.598	0.074	0.024	0.002
Abnormal returns	0.197	0.039	1.000	0.188	0.107	0.080
Excess returns	-0.073	0.030	0.908	-0.081	0.044	0.064
Dollar volume	-0.005	0.015	0.122			
Shares traded	-0.005	0.015	0.118			
Detrended series	-0.023	0.031	0.407			
Data period	0.000	0.000	0.014			
Data size	0.020	0.003	1.000	0.020	0.003	0.000
Midyear	-0.054	0.012	1.000	-0.049	0.011	0.000
Daily data	0.000	0.005	0.023			
Weekly data	0.016	0.034	0.226			
Monthly	-0.166	0.028	1.000	-0.166	0.042	0.000
Time series	0.024	0.036	0.358			
Cross section	0.023	0.052	0.194			
Structural variation						
Index	-0.005	0.018	0.087			
NASDAQ	0.000	0.005	0.014			
Banks	-0.001	0.010	0.031			
Firms	0.067	0.023	0.943	0.080	0.024	0.001
Developing	0.000	0.003	0.022			
Market size	-0.001	0.004	0.121			
Asia	0.001	0.004	0.036			
Europe	0.000	0.006	0.024			
Australia	0.000	0.006	0.017			
Other Continents	0.272	0.045	1.000	0.293	0.049	0.000
Estimation technique						
VAR	-0.117	0.026	0.998	-0.134	0.027	0.000
Simple model	0.000	0.003	0.018			
GARCH	0.000	0.006	0.018			
Monday	0.000	0.006	0.020			
Trimmed	0.011	0.024	0.210			
January Excluded	0.001	0.007	0.041			
MLE	0.001	0.009	0.037			
GMM	0.000	0.006	0.021			
Other methods	-0.091	0.020	0.993	-0.098	0.015	0.000
Publication characteristics						
Impact factor	0.000	0.003	0.036			
Citations	-0.001	0.005	0.082			
Published	0.000	0.003	0.019			
Studies	44			44		
Observations	468			468		

Notes: Response variable = partial correlation between trading volume and stock returns. SD = standard deviation. PIP = posterior inclusion probability. SE = standard error. The UIP g-prior and dilution model prior are deployed in BMA, as suggested by George *et al.* (2010). The frequentist check (OLS) includes only variables with PIP>0.5 to form the best model. SEs are clustered at the study and country levels, as proposed by (Cameron *et al.*, 2012). Table 6 describes all variables used.

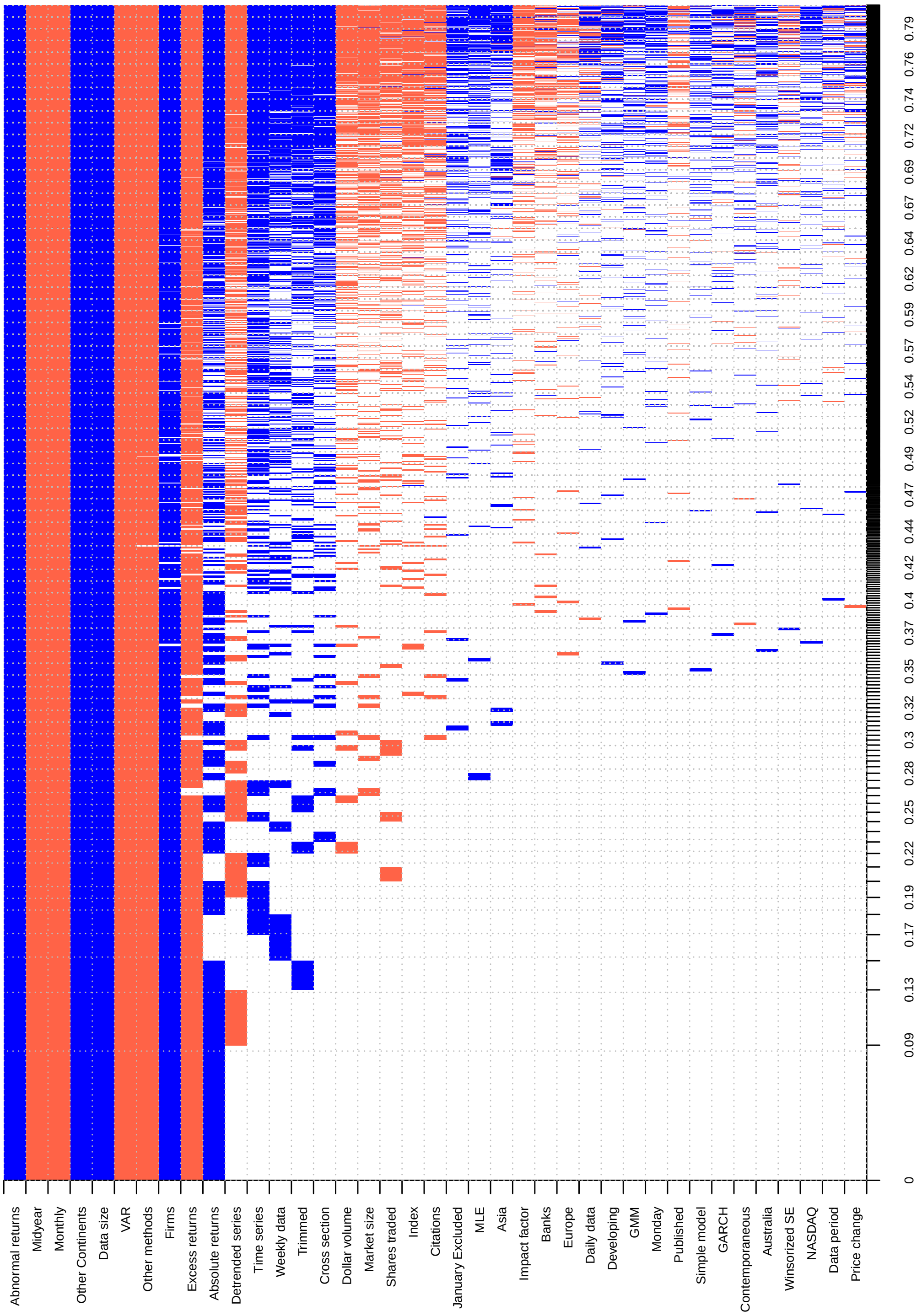


Figure 5: Bayesian model averaging

Data characteristics. The category most represented among the decisive variables is that of data characteristics. A higher volatility of *Abnormal returns* caused by differences in trading volume is anticipated by Lin & Liu (2017). The same holds for the sign and magnitude of *Absolute returns*, which are completely in line with the predictions of (Lin & Liu, 2017; Mahajan & Singh, 2009a). On the other hand, *Excess returns* by definition should behave similarly to *Absolute returns*, but the results contradict this idea. Indeed, Table 7 suggests the opposite: the *Excess* variable even has a negative impact. On the other hand, Brennan *et al.* (1998); Chordia *et al.* (2001) support this finding.

Next, the decreasing value of the estimates with the newer data has already been indicated in Figure 2. Therefore, the most important findings relate to data frequency and data type. The estimates originating from *Monthly* data are significantly lower than those based on higher-frequency (*Hourly, Daily, Weekly*) data. This result completely contradicts the expectations of Lee & Swaminathan (2000); Comiskey *et al.* (1987). The difference between the average *Hourly* and *Monthly* estimates is, ceteris paribus, -0.168 (0.028). This outcome reveals data frequency to be the major driver of heterogeneity. The effect of the *Monthly* variable acts as a counterweight to *Data size* in the case of panel estimates, albeit not completely. The findings on data frequency clarify the previously unexplained variation in Akpansung & Gidigbi (2015). The same holds for conclusions about the data dimensions: in the end, the dimensions of the data do not matter Mahajan & Singh (2009a).

Structural variation. In line with the intuition of De Santis *et al.* (1997), the results show that emerging markets, especially those in Latin America and Africa, exhibit higher volatility and a greater probability of large price changes. The coefficients of 0.272 and 0.045 emphasize this effect. On the other hand, differences in *Market size* do not affect the relationship of interest. Moreover, stock markets across North America, Europe and Australia display no differences at all. This indicates that stock exchanges in emerging markets behave differently from those in developed markets regardless of the size of the market.

Furthermore, the results contradict the conclusions of Chordia *et al.* (2001) about the stronger effect of trading volume on stock returns on the *NASDAQ* stock exchange. *NASDAQ* stocks, together with *Banks* and *Index* stocks, exhibit similar results to those based on *All* stocks from a given stock exchange. On the other hand, *Firms* stocks rise significantly higher (0.067). This finding uncovers previously hidden dynamics.

Estimation techniques. Another potential explanation of the heterogeneity in the estimates from the primary studies relates to the use of the various estimation techniques. *VAR* models provide substantially lower results (-0.117) than other estimation techniques. This outcome should be taken into account by anyone considering employing a VAR model in a future analysis. Similarly, the usage of *other methods* produces more strongly negative estimates than those produced by *OLS*, *MLE*, or *GMM*. The results do not support questions over the reliability of OLS estimates and thus contradict the concerns raised by Epps & Epps (1976). Moreover, neither trimming the primary dataset nor considering the Monday or January effect impacts the primary findings, contradicting the predictions of French (1980); Gibbons & Hess (1981); Hu (1997).

Publication characteristics. The results indicate no strong association between publication characteristics and the magnitude of the reported results. The number of citations, the impact factor of the series and publication in a peer-reviewed journal do not substantially affect the results. This conclusion does credit to journals as well as authors.

4.3 Robustness check

The stability of the results underscores the complex robustness checks. Table 7 provides a glimpse of the robustness of the results. I run simple OLS regressions on the variables with a PIP over 0.5 in the BMA baseline results. The OLS coefficients accord with the baseline results in both sign and magnitude. Moreover, the variables remain significant according to their t-statistics as indicated by the PIPs in BMA.

In addition, the robustness check deploys several different BMA specifications. The combinations of the UIP g-prior with the uniform model prior and the BRIC g-prior with the random model prior are chosen. A comparison of these two specifications with respect to the baseline setting of the UIP g-prior with dilution model prior depicts Figure 6. The results further captured numerically in Table A3 indicate stable PIPs across the priors.

The UIP estimation removes the significance of *Absolute returns* and elevates the PIP of the

Detrended series and *time series* data above 0.5. This indicates that the use of *Detrended* series lowers the estimates by -0.035. On the other hand, the use of *Time series* data increases the estimates by 0.042. This means that detrended series yield a less pronounced relationship between returns and volume, which is what one should expect from detrending. The *Time series* effect acts as a counterweight to *Data size*, the effect of which grows with *Panel* data. The BRIC estimation duplicates the baseline dilution conclusions entirely.

Table A3 provides the last robustness check, the FMA. Even the results of this robustness check line up with the baseline results. They also suggest the significance of *Detrended* series (a negative effect), *Trimmed* data (a positive effect) and the *January excluded* variable (a negative effect). The negative coefficient of the *Trimmed* variable indicates the presence of negative outliers in the data of the primary articles. The significance of *January excluded* captures investor sentiment in January. This result contradicts (Datar *et al.*, 1998)'s finding of no effect of sentiment in January.

This model averaging technique suggests the same conclusions as BMA except that the importance in the data characteristics category shifts from the data type variable to that of measurement of returns and volume. The main findings regarding the significance of the standard error effect and the effects of *Abnormal returns*, *Excess returns*, usage of *Monthly* data, *Data size*, *Midyear* of the data, *VAR* models, *Other methods* and *Other continents* remain the same. This validates the robustness of the baseline model.

Last, I run BMA on a larger sample of variables. This broader dataset distinguishes the estimation technique characteristics in the primary studies in more detail. Table A2 summarizes the broader set of estimation technique variables, and Table A4 captures the complex results. In brief, the main findings remain stable. These additional results again support the robustness of the primary findings. Only the effect of *VAR* models is diminished, and, for instance, the *Illiquidity* effect in the primary models comes to the forefront. Taken together, these results support the original aggregation of the variables because the more granular view shows insignificant differences between the component variables.

In summary, the robustness checks testify to the decisive importance of the choice of type data and country of interest. Moreover, the results provide substantial evidence of the decisiveness of the effects related to *Abnormal returns*, *Other continents*, *VAR* models, and *Other methods.*

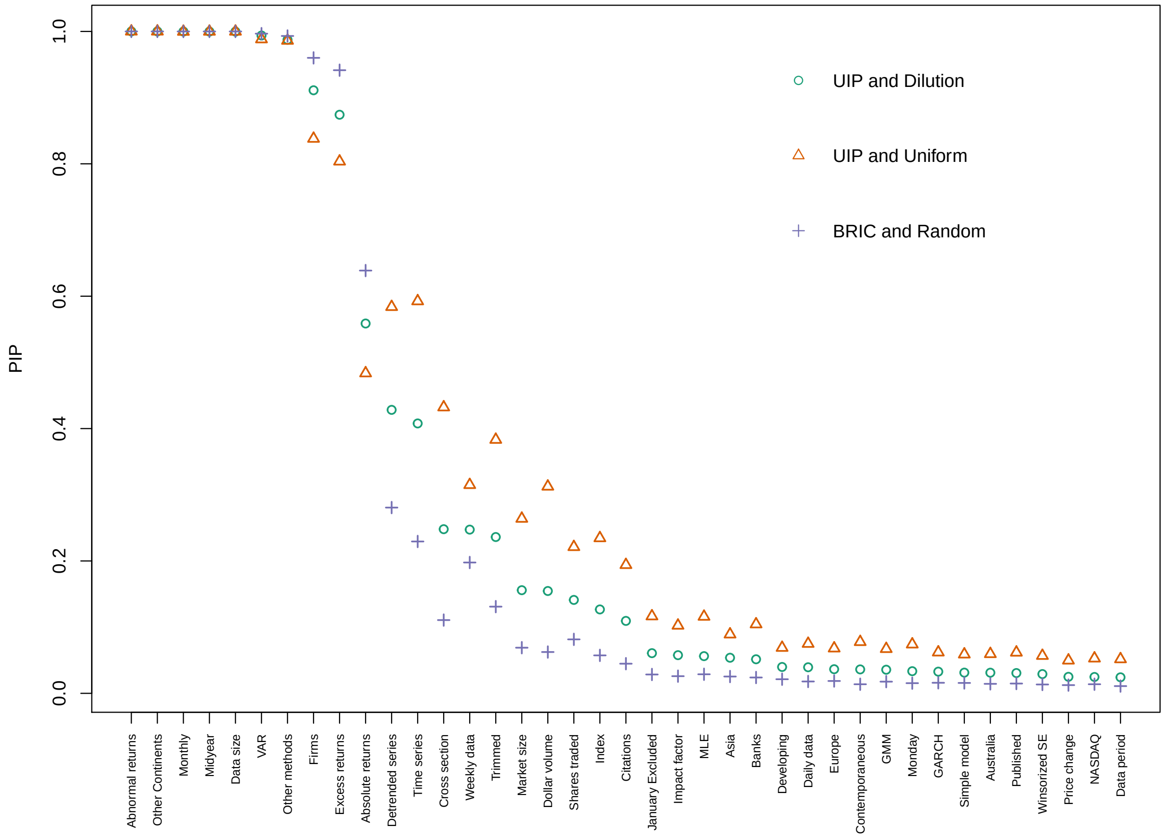


Figure 6: Bayesian model averaging - a comparison of the different priors

All these findings are evident across the models and specifications.

4.4 Implied Effect of Trading Volume on Stock Returns

What are the implications for the effect of trading volume on stock returns? Although the estimation unveils several drivers of heterogeneity in the estimates, the question about the "true" underlying effect remains open. Since the conducted estimations suggest several key factors determining stock returns, I can propose a preferred estimation specification for future research based on current knowledge.

The presented results indicate the following: i) publication bias affects the results on the contemporaneous effect but not those on the dynamic effect; ii) the results differ based on the type of stocks and continent of origin; and iii) different data employed and estimation methodology used cause some differences in findings. I consider all of these major findings when suggesting the "true" effect.

First, I choose workhorse models in this area to provide the baseline model and data specification settings. I choose the models of Brennan *et al.* (1998) and Chordia *et al.* (2001) for the dynamic estimates and that of Datar *et al.* (1998) for the contemporaneous ones. These are perhaps not the newest studies, but they remain seminal in this area of research, and their findings have not yet been overturned. Furthermore, they have been published in journals with high-quality peer review (the first two in the Journal of Financial Economics and the third in the Journal of Financial Markets).

The implied estimates captured in Table 8 and 9 are based on a linear combination of the model characteristics from these three papers except in regard to the coefficient of the standard error, the variables distinguishing between dynamic and contemporaneous effects and the factors related to structural heterogeneity. In the case of the standard error, the coefficient is zero, indicating no presence of publication bias. The characteristics from Brennan *et al.* (1998), Chordia *et al.* (2001) and Datar *et al.* (1998) correspond to the dynamic and contemporaneous effects, and the estimates by continent indicate whether the continent is developing (*Asia* and *other continents*, that is, Africa and Latin America) or developed (*Europe*, *North America*, and *Australia*). The 90th percentile settings for *market size* and *midyear* indicate that newer datasets are preferred. Furthermore, stock characteristics serve as the last but not the least important input in the implied estimates.

The implied estimate results reveal two points. When one considers only overall best practice estimates, there is a negative contemporaneous and even a negative dynamic effect of trading volume on stock returns. This observation contradicts the efficient market hypothesis. Moreover, after a closer look at particular elements of structural heterogeneity, important findings for developing profitable trading strategies appear. First, there is a negative effect of trading volume on stock *Indexes* and *Banks* stocks that ranges from -0.152 to -0.132. This effect is even stronger in the dynamic case (ranging from -0.180 to -0.159), when the effect appears negative even for *Firms* stocks. In contrast, the estimates for *NASDAQ* stocks do not appear more pronounced than those for other exchanges, despite what Chordia *et al.* (2001) proposes. Second, even the conclusions on individual continents vary. The developing continents, especially South America and Africa, display an effect of trading volume on stocks with the opposite sign to

Table 8: Implied estimates – contemporaneous effect

Continent	All	Index	Stock type Banks	Firms	NASDAQ
North America	-0.101**	-0.143***	-0.152***	-0.079***	-0.076**
	(0.039)	(0.038)	(0.056)	(0.030)	(0.039)
Europe	-0.090*	-0.132***	-0.141**	-0.068	
	(0.050)	(0.049)	(0.056)	(0.043)	
Australia	-0.090	-0.132*	-0.140**	-0.067	
	(0.066)	(0.069)	(0.068)	(0.061)	
Asia	-0.092*	-0.133***	-0.142**	-0.069*	
	(0.048)	(0.047)	(0.061)	(0.042)	
Other Continents	0.217**	0.175**	0.166	0.239***	
	(0.096)	(0.086)	(0.108)	(0.088)	
Best practice	-0.099**				
	(0.039)				

Notes: The values suggest the best practice estimates of trading-volume relationship across different continents and stocks implied by study design of Datar *et al.* (1998). Standard errors reported in parentheses are derived from OLS estimates and clustered at the study and country level as suggested Cameron *et al.* (2012)). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 9: Implied estimates – dynamic effect

Continent	All	Index	Stock type Banks	Firms	NASDAQ
North America	-0.129***	-0.171***	-0.180***	-0.107***	-0.104***
	(0.035)	(0.046)	(0.057)	(0.032)	(0.033)
Europe	-0.118**	-0.160***	-0.168***	-0.095**	
	(0.047)	(0.056)	(0.057)	(0.045)	
Australia	-0.118*	-0.159**	-0.168**	-0.095	
	(0.070)	(0.080)	(0.075)	(0.069)	
Asia	-0.119***	-0.161***	-0.170***	-0.097**	
	(0.044)	(0.053)	(0.062)	(0.044)	
Other Continents	0.189**	0.147*	0.139	0.211**	
	(0.090)	(0.085)	(0.104)	(0.084)	
Best practice	-0.126***				
	(0.034)				

Notes: The values suggest the best practice for estimating the trading volume relationship across different continents and stocks, as implied by the study design of Brennan *et al.* (1998) and Chordia & Swaminathan (2000). Standard errors, reported in parentheses, are derived from OLS estimates and clustered at the study and country level, as suggested by Cameron *et al.* (2012)). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

that in the previous results. The effect is positive in both the contemporaneous and dynamic cases across all stock types. The implied estimates are all significant and range from 0.139 to 0.239. This indicates that these stock markets are still evolving and provide more arbitrage opportunities.

These findings clarify why the results in the literature on stock returns and trading volume diverge. Authors use different data from different countries, and naturally this leads to different conclusions. Some studies support the efficient market hypothesis, while others do not. Moreover, conclusions from influential papers of Lee & Rui (2002) and Chen (2012) are overturned. Thus, one should bear it in mind when proposing conclusions that for different countries in varying circumstances, findings differ widely.

5 Conclusions

The first quantitative synthesis of the broad economic literature on the relationship between trading volume and stock returns comes into existence. This relationship, crucial for building profitable trading strategies, conducting event studies, futures markets investigation or for confirming efficient market hypothesis, subdues thorough examination. A total of 468 estimates collected from 44 studies reveal several significant outcomes.

First of all, I investigate the publication bias by common approaches such as a funnel plot and OLS and WLS estimations. Then I deploy recently developed nonlinear estimators proposed by (Furukawa, 2019), Bom & Rachinger (2019), or Andrews & Kasy (2019). The study indicates the presence of publication bias, at least in cases when the investigated relationship is contemporaneous. The mean effect after correction for publication bias has a negligible value. Furthermore, based on a a caliper test (Gerber *et al.*, 2008; Gerber & Malhotra, 2008), even the p-hacking and HARKing affect the results from primary studies.

Moreover, other study-specific factors affect the corrected mean. For this heterogeneity investigation, Bayesian (Raftery, 1995) and frequentist (Amini & Parmeter, 2012) model averaging is deployed. The results show that data characteristics, structural variation and different methodological approaches explain a large part of the inconsistency in the primary results. Concretely, one has to be cautious in using *Monthly* data or *VAR* models. These variables are associated with a substantially more negative effect of trading volume on returns. In addition, there is an association between data age and the magnitude of published estimates. Newer data tend to produce smaller results that are closer to the predictions of the efficient market hypothesis. On the other hand, journal quality and the significance of the article do not affect the results.

After controlling for the differences, I find overall implied estimates of a negative effect in both the contemporaneous and dynamic cases. This calls into question the efficient market hypothesis since it implies opportunities to predict stock returns. For instance, the country and type of stock matter. The effect of trading volume on stocks in emerging countries has the opposite sign to that on returns in developed markets. Moreover, the stocks of firms follow the predictions of the EMH more closely than any other stock type. Thus, one should bear in mind the specifics of each stock type and avoid dangerously relying on some overall conclusion. All these results can serve as a baseline for model calibration or can directly help in traders' strategies.

Bibliography

Akpansung, A. O. & M. O. Gidigbi (2015): "The relationship between trading volumes and returns in the nigerian stock market." *International Research Journal of Finance and Economics* **132**: pp. 150–163.

Al-Jafari, M. K. & A. Tliti (2013): "An empirical investigation of the relationship between stock return and trading volume: Evidence from the jordanian banking sector." *Journal of Applied Finance and Banking* **3(3)**: pp. 45–64.

Amihud, Y. & H. Mendelson (1986): "Asset pricing and the bid-ask spread." *Journal of Financial Economics* **17(2)**: pp. 223 – 249.

Amini, S. M. & C. F. Parmeter (2012): "Comparison of model averaging techniques: Assessing growth determinants." *Journal of Applied Econometrics* **27(5)**: pp. 870–876.

Andrews, I. & M. Kasy (2019): "Identification of and correction for publication bias." *American Economic Review* **109(8)**: pp. 2766–94.

Assogbavi, T., J. Schell, & S. Fagnisse (2007): "Equity price-volume relationship on the russian stock exchange." *International Business and Economics Research Journal* **6(9)**: pp. 107–116.

Astakhov, A., T. Havranek, & J. Novak (2019): "Firm size and stock returns: A quantitative survey." *Journal of Economic Surveys* **33(5)**: pp. 1463–1492.

Bajzik, J., T. Havranek, Z. Irsova, & J. Schwarz (2020): "Estimating the armington elasticity: The importance of data choice and publication bias." *Journal of International Economics* **127**: p. 103383.

Basu, S. (1977): "Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis." *The journal of Finance* **32(3)**: pp. 663–682.

Black, F. & M. Scholes (1974): "The effects of dividend yield and dividend policy on common stock prices and returns." *Journal of financial economics* **1(1)**: pp. 1–22.

Bom, P. R. & H. Rachinger (2019): "A kinked meta-regression model for publication bias correction." *Research Synthesis Methods* **10(4)**: pp. 497–514.

Brandle, A. (2010): "Volume based portfolio strategies - analysis of the relationship between trading activity and expected returns in the cross-section of swiss stocks." *Technical report*, University of St. Gallen.

Brennan, M. J., T. Chordia, & A. Subrahmanyam (1998): "Alternative factor specifications, security characteristics, and the cross-section of expected returns." *Journal of Financial Economics* **49**: pp. 345–373.

Brennan, M. J. & A. Subrahmanyam (1995): "Investment analysis and price formation in securities markets." *Journal of financial economics* **38(3)**: pp. 361–381.

Brennan, M. J. & A. Subrahmanyam (1996): "Market microstructure and asset pricing: On the compensation for illiquidity in stock returns." *Journal of financial economics* **41(3)**: pp. 441–464.

Bruns, S. B., I. Asanov, R. Bode, M. Dunger, C. Funk, S. M. Hassan, J. Hauschildt, D. Heinisch, K. Kempa, J. König *et al.* (2019): "Reporting errors and biases in published empirical findings: Evidence from innovation research." *Research Policy* **48(9)**: p. 103796.

Cameron, A. C., J. B. Gelbach, & D. L. Miller (2012): "Robust inference with multiway clustering." *Journal of Business & Economic Statistics* .

Chang, S. S. & F. A. Wang (2019): "Informed contrarian trades and stock returns." *Journal of Financial Markets* **42**: pp. 75–93.

Chen, G.-m., M. Firth, & O. M. Rui (2001): "The dynamic relation between stock returns, trading volume, and volatility." *Financial Review* **36(3)**: pp. 153–174.

Chen, S.-S. (2012): "Revisiting the empirical linkages between stock returns and trading volume." *Journal of Banking & Finance* **36(6)**: pp. 1781–1788.

Chordia, T., A. Subrahmanyam, & V. R. Anshuman (2001): "Trading activity and expected stock returns." *Journal of financial Economics* **59(1)**: pp. 3–32.

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

Christensen, G. & E. Miguel (2018): "Transparency, reproducibility, and the credibility of economics research." *Journal of Economic Literature* **56(3)**: pp. 920–80.

Chuang, W.-I. & B.-S. Lee (2006): "An empirical evaluation of the overconfidence hypothesis." *Journal of*

Banking and Finance **30(9)**: pp. 2489–2515.

Ciner, C. (2002): "The stock price-volume linkage on the toronto stock exchange: Before and after automation." *Review of Quantitative Finance and Accounting* **19(4)**: pp. 335–349.

Ciner, C. (2003): "Dynamic linkages between trading volume and price movements: Evidence for small firm stocks." *The Journal of Entrepreneurial Finance* **8(1)**: pp. 87–102.

Comiskey, E. E., R. A. Walkling, & M. A. Weeks (1987): "Dispersion of expectations and trading volume." *Journal of Business Finance and Accounting* pp. 229–239.

Crouch, R. L. (1970): "The volume of transactions and price changes on the new york stock exchange." *Financial Analysts Journal* **26(4)**: pp. 104–109.

Datar, V. T., N. Y. Naik, & R. Radcliffe (1998): "Liquidity and stock returns: An alternative test." *Journal of Financial Markets* **1(2)**: pp. 203–219.

De Meiros, O. R. & B. F. N. Van Doornik (2008): "The empirical relationship between stock returns, return volatility and trading volume in the brazilian stock market." *Brazilian Business Review (English Edition)* **5(1)**.

De Santis, G. *et al.* (1997): "Stock returns and volatility in emerging financial markets." *Journal of International Money and finance* **16(4)**: pp. 561–579.

Devanadhen, K., K. Srnivasan, & M. Deo (2010): "Price changes, trading volume and time-varying conditional volatility: Evidence from asia pacific stock market." *International Review of Applied Financial Issues and Economics* **2(2)**: p. 379.

Doucouliagos, C. (2005): "Publication bias in the economic freedom and economic growth literature." *Journal of Economic Surveys* **19(3)**: pp. 367–387.

Doucouliagos, C. (2011): "How large is large? preliminary and relative guidelines for interpreting partial correlations in economics." *Economics series working paper 05*, Deakin University.

Doucouliagos, C. & T. D. Stanley (2013a): "Are all economic facts greatly exaggerated? theory competition and selectivity." *Journal of Economic Surveys* **27(2)**: pp. 316–339.

Doucouliagos, H. & T. D. Stanley (2013b): "Are all economic facts greatly exaggerated? theory competition and selectivity." *Journal of Economic Surveys* **27(2)**: p. 316–339.

Egger, M., G. D. Smith, & C. Minder (1997): "Bias in meta-analysis detected by a simple, graphical test." *Journal of Economic Surveys* **316**: p. 629–634.

Ehrenbergerova, D. & J. Bajzik (2020): "The effect of monetary policy on house prices – how strong is the transmission?" *CNB WP* .

Eicher, T. S., C. Papageorgiou, & A. E. Raftery (2011): "Default priors and predictive performance in bayesian model averaging, with application to growth determinants." *Journal of Applied Econometrics* **26(1)**: pp. 30–55.

Epps, T. W. & M. L. Epps (1976): "The stochastic dependence of security price changes and transaction volumes: Implications for the mixture-of-distributions hypothesis." *Econometrica: Journal of the Econometric Society* pp. 305–321.

Fama, E. F. (1970): "Efficient capital markets: A review of theory and empirical work." *The journal of Finance* **25(2)**: pp. 383–417.

Fama, E. F. & K. R. French (1992): "The cross-section of expected stock returns." *The Journal of Finance* **47(2)**: pp. 427–465.

Fama, E. F. & K. R. French (1993): "Common risk factors in the returns on stocks and bonds." *Journal of Financial Economics* **33**: pp. 3–56.

Fama, E. F. & K. R. French (1996): "Multifactor explanations of asset pricing anomalies." *The Journal of Finance* **51(1)**: pp. 55–84.

Fernandez, C., E. Ley, & M. F. J. Steel (2001): "Model uncertainty in cross-country growth regressions." *Journal of Applied Econometrics* **16(5)**: pp. 563–576.

Fisher, R. (1954): *Statistical Methods for Research Workers, 12th edn.* Edinburgh, UK: Oliver and Boyd.

Fletcher, D. (2018): "Bayesian model averaging." In "Model Averaging," pp. 31–55. Springer.

French, K. R. (1980): "Stock returns and the weekend effect." *Journal of financial economics* **8(1)**: pp. 55–69.

Furukawa, C. (2019): "Publication bias under aggregation frictions: Theory, evidence, and a new correction method." *Technical report*, ZBW – Leibniz Information Centre for Economics, Kiel, Hamburg.

Gechert, S., T. Havranek, Z. Irsova, & D. Kolcunova (2020): "Measuring capital-labor substitution: The importance of method choices and publication bias." *CNB WP* pp. 1–49.

George, E. I. *et al.* (2010): "Dilution priors: Compensating for model space redundancy." In "Borrowing Strength: Theory Powering Applications–A Festschrift for Lawrence D. Brown," pp. 158–165. Institute of Mathematical Statistics.

Gerber, A., N. Malhotra *et al.* (2008): "Do statistical reporting standards affect what is published? publication bias in two leading political science journals." *Quarterly Journal of Political Science* **3(3)**: pp. 313–326.

Gerber, A. S. & N. Malhotra (2008): "Publication bias in empirical sociological research: Do arbitrary significance levels distort published results?" *Sociological Methods & Research* **37(1)**: pp. 3–30.

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

Gibbons, M. R. (1982): "Multivariate tests of financial models: A new approach." *Journal of financial economics* **10(1)**: pp. 3–27.

Gibbons, M. R. & P. Hess (1981): "Day of the week effects and asset returns." *Journal of business* pp. 579–596.

Godfrey, M. D., C. W. Granger, & O. Morgenstern (1964): "The random-walk hypothesis of stock market behavior." *Kyklos* **17(1)**: pp. 1–30.

Granger, C. W. (1969): "Investigating causal relations by econometric models and cross-spectral methods." *Econometrica: Journal of the Econometric Society* pp. 424–438.

Granger, C. W. & O. Morgenstern (1963): "Spectral analysis of new york stock market prices 1." *Kyklos* **16(1)**: pp. 1–27.

Greene, W. H. (2008): *Econometric Analysis.* Upper Saddle River, New Jersey: Pearson International in Prentice-Hall International Editions.

Gurgul, H., P. Majdosz, & R. Mestel (2007): "Price volume relations of dax companies." *Financial Markets and Portfolio Management* **21**: pp. 353–379.

Hafner, C. M. (2005): "Durations, volume and the prediction of financial returns in transaction time." *Quantitative Finance* **5(2)**: pp. 145–152.

Han, Y., D. Huang, D. Huang, & G. Zhou (2018): "Volume and return: The role of mispricing." *Ssrn*, University of North Carolina.

Hansen, B. E. (2007): "Least squares model averaging." *Econometrica* **75(4)**: p. 1175–1189.

Hasan, I., R. Horvath, & J. Mares (2018): "What type of finance matters for growth? bayesian model averaging evidence." *The World Bank Economic Review* **32(2)**: pp. 383–409.

Havranek, T. & Z. Irsova (2017): "Do borders really slash trade? a meta-analysis." *IMF Economic Review* **65(2)**: pp. 1–32.

Havranek, T., M. Rusnak, & A. Sokolova (2017): "Habit formation in consumption: A meta-analysis." *European Economic Review, Elsevier* **95(C)**: pp. 142–167.

Havranek, T. & A. Sokolova (2019): "Do consumers really follow a rule of thumb? three thousand estimates from 144 studies say 'probably not'." *Working paper*, Czech National Bank and Charles University, Prague.

Havránek, T., T. Stanley, H. Doucouliagos, P. Bom, J. Geyer-Klingeberg, I. Iwasaki, W. R. Reed, K. Rost, & R. van Aert (2020): "Reporting guidelines for meta-analysis in economics." *Journal of Economic Surveys* .

Hiemstra, C. & J. D. Jones (1994): "Testing for linear and nonlinear granger causality in the stock price-volume relation." *The Journal of Finance* **49(5)**: pp. 1639–1664.

Hu, S.-Y. (1997): "Trading turnover and expected stock returns: The trading frequency hypothesis and evidence from the tokyo stock exchange." *Working paper*, University of Chicago.

Imai, T., T. A. Rutter, & C. Camerer (2020): "Meta-analysis of present-bias estimation using convex time budgets." *Economic Journal* **forthcoming**.

Ioannidis, J. P., T. D. Stanley, & H. Doucouliagos (2017): "The power of bias in economics research." *The Economic Journal* **127(605)**: p. F236–F265.

Jain, P. C. & G.-H. Joh (1988): "The dependence between hourly prices and trading volume." *The Journal of Financial and Quantitative Analysis* **23**: pp. 269–283.

Jegadeesh, N. & S. Titman (1993): "Returns to buying winners and selling losers: Implications for stock market efficiency." *The Journal of Finance* **48(1)**: pp. 65–91.

Jegadeesh, N. & S. Titman (1995): "Overreaction, delayed reaction, and contrarian profits." *The Review of Financial Studies* **8(4)**: pp. 973–993.

Jensen, M. C., F. Black, & M. S. Scholes (1972): "The capital asset pricing model: Some empirical tests." *Technical report*, Praeger Publishers Inc.

Karpoff, J. M. (1987): "The relation between price changes and trading volume: A survey. journal of financial and quantitative analysis." *Journal of Financial and quantitative Analysis* **22(1)**: pp. 109–126.

Le, Q. T. & M. Mehmed (2009): "The relationship between trading volume, stock index returns and volatility: Empirical evidence in nordic countries." *Master thesis in finance*, Lund University.

Lee, C. & O. Rui (2000): "Does trading volume contain information to predict stock returns? evidence from china's stock markets." *Review of Quantitative Finance and Accounting* **14**: pp. 341–360.

Lee, C. & O. Rui (2002): ""the dynamic relationship between stock returns and trading volume: Domestic and cross-country evidence." *Journal of Banking and Finance* **26**: pp. 51–78.

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

Lee, U. (1992): "Do stock prices follow random walk?:: Some international evidence." *International Review of Economics & Finance* **1(4)**: pp. 315–327.

Lewellen, J. (2015): "The cross section of expected stock returns." *Critical Financial Review* **4(1)**: pp. 1–44.

Ley, E. & M. F. Steel (2009): "On the effect of prior assumptions in bayesian model averaging with applications to growth regression." *Applied Econometrics* **24**: pp. 651–674.

Lin, T.-C. & X. Liu (2017): "Skewness, individual investor preference, and the cross-section of stock returns." *Review of Finance* pp. 1–36.

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

Lo, A. W. & J. Wang (2000): "Trading volume: Definitions, data analysis, and implications of portfolio theory." *The Review of Financial Studies* **13**: pp. 257–300.

Long, H., Y. Jinag, & Y. Zhu (2018): "Idiosyncratic tail risk and expected stock returns: Evidence from the chinese stock markets." *Finance Research Letters* **24**: pp. 129–136.

Louhichi, W. (2012): "Does trading activity contain information to predict stock returns? evidence from euronext paris." *Applied Financial Economics* **22(8)**: pp. 625–632.

Loukil, N., M. B. Zayani, & A. Omri (2010): "Impact of liquidity on stock returns: an empirical investigation of the tunisian stock market." *Macroeconomics and Finance in Emerging Market Economies* **3(2)**: pp. 261–283.

Magnus, J. R., O. Powell, & P. Prufer (2010): "A comparison of two model averaging techniques with an application to growth empirics." *Journal of econometrics* **154(2)**: pp. 139–153.

Mahajan, S. & B. Singh (2008): "An empirical analysis of stock price-volume relationship in indian stock market." *Vision* **12(3)**: pp. 1–13.

Mahajan, S. & B. Singh (2009a): "The empirical investigation of relationship between return, volume and volatility dynamics in indian stock market." *Eurasian Journal of Business and Economics* **2(4)**: pp. 113–137.

Mahajan, S. & B. Singh (2009b): "Return-volume dynamics in indian stock market." *Asia-Pacific Business Review* **5(3)**: pp. 63–70.

Malkiel, B. G. (1989): "Efficient market hypothesis." *Finance* pp. 127–134.

Malkiel, B. G. (2003): "The efficient market hypothesis and its critics." *Journal of economic perspectives* **17(1)**: pp. 59–82.

Marshall, B. R. & M. Young (2003): "Liquidity and stock returns in pure order-driven markets: evidence from the australian stock market." *International Review of Financial Analysis* **12(2)**: pp. 173–188.

McGowan, C. B. & J. Muhammad (2012): "The relationship between price and volume for the russian trading system." *The International Business and Economics Research Journal (Online)* **11(9)**: pp. 963–970.

Merton, R. C. *et al.* (1987): "A simple model of capital market equilibrium with incomplete information." *Sloan School of Management, Massachusetts Institute of Technology* .

Mestel, R., H. Gurgul, & P. Majdosz (2003): "The empirical relationship between stock returns, return volatility and trading volume on the austrian stock market." *Research paper*, University of Graz, Institute of Banking and Finance.

Moral-Benito, E. (2015): "Model averaging in economics: An overview." *Journal of Economic Surveys* **29(1)**: pp. 46–75.

Nansen McCloskey, D. & S. T. Ziliak (2019): "What quantitative methods should we teach to graduate students? a comment on swann's "is precise econometrics an illusion?"." *The Journal of Economic Education* **50(4)**: pp. 356–361.

Narayan, P. K. & X. Zheng (2010): "Market liquidity risk factor and financial market anomalies: Evidence from the chinese stock market." *Pacific-Basin finance journal* **18(5)**: pp. 509–520.

Ochere, O. G., M. T. Nasieku, & T. O. Olweny (2018): "Trading volume and fama-french three factor model on excess return. empirical evidence from nairobi security exchange." *European Scientific Journal* **14(22)**: pp. 276–289.

Petersen, M. A. & D. Fialkowski (1994): "Posted versus effective spreads: Good prices or bad quotes?" *Journal of Financial Economics* **35(3)**: pp. 269–292.

Pisedtasalasai, A. & A. Gunasekarage (2007): "Causal and dynamic relationships among stock returns, return volatility and trading volume: Evidence from emerging markets in south-east asia." *Asia-Pacific Financial Markets* **14(4)**: pp. 277–297.

Poudel, R. B. & S. R. Shrestha (2019): "Stock return and trading volume relation in nepalese stock market: An ardl approach." *SEBON JOURNAL* .

Raftery, A. E. (1995): "Bayesian model selection in social research." *Sociological methodology* pp. 111–163.

Raftery, A. E., D. Madigan, & J. A. Hoeting (1997): "Bayesian model averaging for linear regression models." *Journal of the American Statistical Association* **92(437)**: pp. 179–191.

Roodman, D. (2020): "Boottest: Stata module to provide fast execution of the wild bootstrap with null imposed." .

Rotila, D.-M., M. Onofrei, & A. M. Andries (2015): "The relation between stock returns, trading volume and return volatility of the cee banks." *Transformations in Business and Economics* **14**.

Rusnak, M., T. Havranek, & R. Horvath (2013a): "How to solve the price puzzle? a meta-analysis." *Journal of Money, Credit and Banking* **45(1)**: pp. 37–70.

Rusnak, M., T. Havranek, & R. Horvath (2013b): "How to solve the price puzzle? a meta-analysis." *Journal of Money, Credit and Banking* **45(1)**: p. 37–70.

Saatcioglu, K. & L. T. Starks (1998): "The stock price–volume relationship in emerging stock markets: the case of latin america." *International Journal of forecasting* **14(2)**: pp. 215–225.

Sana Hsieh, H.-C. (2014): "The causal relationships between stock returns, trading volume, and volatility: Empirical evidence from asian listed real estate companies." *International Journal of Managerial Finance* **10(2)**: pp. 218–240.

Schürenberg-Frosch, H. (2015): "We could not care less about armington elasticities–but should we? a meta-sensitivity analysis of the influence of armington elasticity misspecification on simulation results." *Working paper 594*, Ruhr Economic Papers.

Sheu, H.-J., S. Wu, & K.-P. Ku (1998): "Cross-sectional relationships between stock returns and market beta, trading volume, and sales-to-price in taiwan." *International Review of Financial Analysis* **7(1)**: pp. 1–18.

Shu, P.-G., Y.-H. Yeh, & Y.-C. Huang (2004): "Stock price and trading volume effects associated with changes in the msci free indices: evidence from taiwanese firms added to and deleted from the indices." *Review of Pacific Basin Financial Markets and Policies* **7(04)**: pp. 471–491.

Stanley, T. D. (2001): "Wheat from chaff: Meta-analysis as quantitative literature review." *Journal of economic perspectives* **15(3)**: pp. 131–150.

Stanley, T. D. & H. Doucouliagos (2012): *Meta-regression analysis in economics and business.* New York, USA: Routledge.

Stanley, T. D., H. Doucouliagos, & J. P. Ioannidis (2017): "Finding the power to reduce publication bias." *Statistics in medicine* **36(10)**: pp. 1580–1598.

Statman, M., S. Thorley, & K. Vorkink (2006): "Investor overconfidence and trading volume." *The Review of Financial Studies* **19(4)**: pp. 1531–1565.

Steel, M. F. (2020): "Model averaging and its use in economics." *Journal of Economic Literature* **forth-**

coming.

STOLL, H. R. (1978): "The pricing of security dealer services: An empirical study of nasdaq stocks." *The journal of finance* **33(4)**: pp. 1153–1172.

TAHIR, S. H., F. ALL, N. GHAFFAR, & H. M. SABIR (2016): "Causal relationship among trading volume, returns and stock volatility: evidence from an emerging market." *Technical report*, University Faisalabad.

TAPA, A. & M. HUSSEIN (2016): "The relationship between stock return and trading volume in malaysian ace market." *International Journal of Economics and Financial Issues* **6(75)**: pp. 271–278.

TRIPATHY, N. (2011): "The relation between price changes and trading volume: A study in indian stock market." *Interdisciplinary Journal of Research in Business* **1(7)**: pp. 81–95.

VALICKOVA, P., T. HAVRANEK, & R. HORVATH (2015): "Financial development and economic growth: A meta-analysis." *Journal of Economic Surveys* **29(3)**: pp. 506–526.

WATKINS, B. D. (2007): "The economic and predictive value of trading volume growth: A tale of three moments." *Applied Financial Economics* **17**: pp. 1489–1509.

WOOLDRIDGE, J. M. (2015): *Introductory econometrics: A modern approach.* Scarborough, Canada: Nelson Education.

YIN, Y. & Y. LIU (2018): "Information of unusual trading volume." *Emerging Markets Finance and Trade* **54(11)**: pp. 2409–2432.

YING, C. C. (1966): "Market prices and volumes of sales." *Econometrica* **34**: pp. 676–685.

YONIS, M. (2014): "Trading volume and stock return: Empirical evidence for asian tiger economies." *Technical report*, UMEA Universitet.

ZEUGNER, S. & M. FELDKIRCHER (2015): "Bayesian Model Averaging Employing Fixed and Flexible Priors: The BMS Package for R." *Journal of Statistical Software* **68(4)**: pp. 1–37.

ZHONG, A., D. CHAI, B. LI, & M. CHIAH (2018): "Volume shocks and stock returns: An alternative test." *Pacific-Basin Finance Journal* **48**: pp. 1–16.

Appendix: BMA Diagnostics and Robustness Checks

Figure A1: Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram

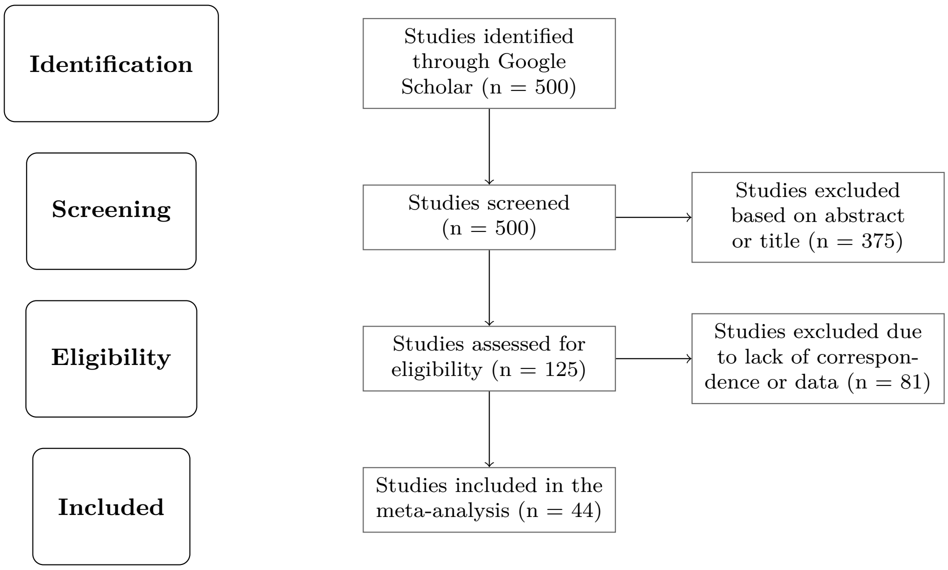


Table A1: Indication of publication bias

	Log-log cases	Log-level cases
PANEL A: Unweighted estimations		
OLS		
SE (publication bias)	0.186	1.467***
	(0.190)	(0.240)
Constant (effect beyond bias)	-0.036	-0.089
	(0.065)	(0.158)
Between effects		
SE (publication bias)	0.463***	0.737***
	(0.155)	(0.271)
Constant (effect beyond bias)	-0.032	0.177
	(0.046)	(0.164)
PANEL B: Weighted OLS estimations		
Weighted by the inverse of the number of estimates reported per study		
SE (publication bias)	0.312	0.999***
	(0.237)	(0.285)
Constant (effect beyond bias)	-0.011	0.103
	(0.038)	(0.104)
Weighted by the inverse of the standard error		
SE (publication bias)	-0.222	1.206***
	(0.429)	(0.257)
Constant (effect beyond bias)	-0.001***	-0.000
	(0.000)	(0.000)
Observations	217	231

Notes: The table above displays the results of the regression $S_{it} = S_0 + \sigma * SE(S_{it}) + \epsilon_{it}$, where S_{it} denotes the ith estimate of the effect size in study j and $SE(SE_{it})$ stands for the respective standard error. Specification (1) uses OLS. Specification (2) employs a panel data regression with between effects. The estimates in specification (3) use WLS with precision weights. Similarly, specification (4) uses the reciprocal of the number of estimates reported per study as the weights. Standard errors are in parentheses and clustered at the country and study level (except between effects; the usage of two-way clustering follows Cameron *et al.* 2012). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A2: Description and summary statistics of the additional variables explaining heterogeneity across the primary studies

Label	*Description*	*Mean*	*SD*	*Median*
Estimation technique				
Fama-Macbeth	=1 if in the specification of the Fama-Macbeth model is next to trading volume used even firm size, book-to-market ratio and price of the stock	0.22	0.42	0.00
Fama-Macbeth II	=1 if in the specification of the Fama-Macbeth model is next to trading volume used even book-to-market ratio and one of the two firm size, or price of the stock	0.15	0.36	0.00
VARa	=1 if in the VAR equation return and volume are with one lag at maximum	0.10	0.30	0.00
VARb	=1 if in the VAR equation return and volume are with two lags at maximum, one of the second lags at one VAR equation at one moment	0.07	0.25	0.00
VARc	=1 if in the VAR equation return and volume are with two and more lags at the same time	0.07	0.25	0.00
Simple model	=1 if the simple model regression of returns on volume by OLS is used	0.33	0.47	0.00
GARCH	=1 if the ARIMA with GARCH in error term is used	0.06	0.24	0.00
Monday	=1 if effect of Monday or January trading is considered	0.03	0.16	0.00
Yield	=1 if the dividend yield as measured by the sum of all dividends paid over the previous 12 months, divided by the share price at the end of the second to last month is incorporated in the primary study	0.14	0.35	0.00
Standard deviation	=1 if some measure of standard deviation is added in the primary model	0.14	0.35	0.00
Market Beta	=1 if variable represents market beta is included in the primary model	0.07	0.25	0.00
Illiquidity	=1 if the average ratio of the absolute daily stock returns to its dollar trading volume is included	0.07	0.26	0.00
Accrual	=1 indicates inclusion of variable measuring change in non-cash net working capital minus depreciation in the prior fiscal year in the primary model	0.02	0.14	0.00
Sales-to-price ratio	=1 if sales to price ratio is added in the primary model	0.04	0.19	0.00
Firm Beta	=1 if firm or portfolio beta is included in the primary model	0.07	0.26	0.00
Firm size	=1 if natural logarithm of firm's market capitalization is included in the primary model	0.03	0.16	0.00
Trimmed	=1 if the primary dataset was trimmed	0.09	0.29	0.00
January excluded	=1 if all months but January are included in the primary dataset	0.08	0.27	0.00
Informed	=1 if measure of the probability of information-based trading in the previous year is included in the primary model	0.03	0.18	0.00
Idiosyncratic volatility	=1 if idiosyncratic volatility is explained variable in primary study	0.03	0.17	0.00
OLS	= 1 if OLS estimation method was used	0.43	0.50	0.00
MLE	= 1 if MLE estimation method was used	0.04	0.20	0.00
GMM	=1 if GMM estimation method was used	0.29	0.45	0.00
Other methods	=1 if other types of estimation were used	0.24	0.43	0.00

Notes: SD = standard deviation. The broader set of variables describing estimation techniques used by the primary studies. All of the variables are collected from the studies investigating the return-volume relationship.

Table A3: Why the estimates vary – various robustness checks

Response variable:	BMA – UIP prior			BMA – BRIC prior			Frequentist model averaging		
Estimate of the Armington elasticity	Post. mean	Post. SD	PIP	Coef.	SE	PIP	Coef.	SE	p-value
Constant	0.175	NA	1.000	0.151	NA	1.000	0.209	0.101	0.039
Winsorized SE	-0.004	0.067	0.057	0.000	0.028	0.013	-0.304	0.395	0.441
Data characteristics									
Contemporaneous	0.001	0.006	0.078	0.000	0.002	0.014	0.027	0.019	0.154
Price change	0.000	0.007	0.050	0.000	0.003	0.012	0.028	0.040	0.480
Absolute returns	0.029	0.034	0.484	0.045	0.037	0.639	0.026	0.030	0.393
Abnormal returns	0.207	0.044	1.000	0.194	0.037	1.000	0.207	0.045	0.000
Excess returns	-0.062	0.038	0.804	-0.076	0.027	0.941	-0.068	0.037	0.065
Dollar volume	-0.014	0.024	0.312	-0.002	0.011	0.062	-0.044	0.026	0.088
Shares traded	-0.009	0.020	0.221	-0.003	0.013	0.082	-0.034	0.030	0.266
Detrended series	-0.035	0.034	0.584	-0.016	0.028	0.281	-0.058	0.031	0.059
Data period	0.000	0.000	0.052	0.000	1.000	0.011	0.001	0.001	0.437
Data size	0.020	0.003	1.000	0.020	0.003	1.000	0.018	0.005	0.000
Midyear	-0.057	0.013	1.000	-0.053	0.011	1.000	-0.058	0.018	0.001
Daily data	0.001	0.010	0.075	0.000	0.005	0.018	0.030	0.040	0.449
Weekly data	0.022	0.037	0.315	0.015	0.033	0.198	0.085	0.055	0.125
Monthly	-0.171	0.034	1.000	-0.166	0.026	1.000	-0.184	0.054	0.001
Time series	0.042	0.042	0.592	0.015	0.031	0.229	0.057	0.037	0.121
Cross section	0.052	0.070	0.432	0.012	0.040	0.111	0.118	0.056	0.033
Structural variation									
Index	-0.013	0.027	0.235	-0.003	0.015	0.057	-0.042	0.034	0.223
NASDAQ	0.001	0.009	0.053	0.000	0.005	0.014	0.025	0.037	0.500
Banks	-0.005	0.020	0.104	-0.001	0.009	0.024	-0.051	0.047	0.283
Firms	0.054	0.029	0.838	0.072	0.022	0.960	0.021	0.037	0.561
Developing	0.001	0.006	0.069	0.000	0.004	0.021	-0.035	0.027	0.196
Market size	-0.003	0.006	0.264	-0.001	0.003	0.069	-0.008	0.008	0.300
Asia	0.001	0.006	0.089	0.000	0.003	0.025	0.034	0.033	0.309
Europe	-0.001	0.009	0.068	0.000	0.005	0.019	-0.005	0.034	0.879
Australia	0.001	0.011	0.059	0.000	0.005	0.014	-0.012	0.051	0.808
Other Continents	0.263	0.048	1.000	0.277	0.043	1.000	0.324	0.064	0.000
Studies	44			44			44		
Observations	468			468			468		

Response variable = partial correlation between trading volume and stock returns. SD = standard deviation. PIP = posterior inclusion probability. SE = standard error. In BMA uniform g-prior with uniform model prior and BRIC g-prior with random model prior are deployed (Eicher *et al.* 2011, Fernandez *et al.* (2001)). FMA employs Mallow's criterion for model averaging using the orthogonalization of the parameter space (Hansen 2007, Amini & Parmeter 2012). All the variables describes Table 6.

Table A3: Why the estimates vary – various robustness checks (continued)

Response variable:	BMA – UIP prior			BMA – BRIC prior			Frequentist model averaging		
Estimate of the Armington elasticity	Post. mean	Post. SD	PIP	Coef.	SE	PIP	Coef.	SE	p-value
Estimation technique									
VAR	-0.110	0.030	0.988	-0.120	0.026	0.997	-0.155	0.075	0.040
Simple model	0.001	0.005	0.059	0.000	0.003	0.016	0.019	0.024	0.411
GARCH	0.001	0.013	0.062	0.000	0.006	0.016	-0.070	0.074	0.344
Monday	0.002	0.013	0.074	0.000	0.005	0.015	0.017	0.041	0.681
Trimmed	0.021	0.031	0.383	0.007	0.019	0.131	0.053	0.027	0.052
January Excluded	0.003	0.012	0.116	0.001	0.006	0.028	0.040	0.023	0.077
MLE	0.005	0.017	0.116	0.001	0.008	0.029	0.050	0.031	0.111
GMM	0.001	0.012	0.067	0.000	0.005	0.018	0.065	0.061	0.286
Other methods	-0.081	0.023	0.986	-0.094	0.018	0.993	-0.058	0.022	0.009
Publication characteristics									
Impact factor	-0.001	0.005	0.102	0.000	0.003	0.026	0.017	0.022	0.438
Citations	-0.003	0.007	0.194	-0.001	0.004	0.045	-0.010	0.012	0.419
Published	0.000	0.005	0.062	0.000	0.002	0.015	0.019	0.028	0.499
Studies	44			44			44		
Observations	468			468			468		

Response variable = partial correlation between trading volume and stock returns. SD = standard deviation. PIP = posterior inclusion probability. SE = standard error. In BMA uniform g-prior with uniform model prior and BRIC g-prior with random model prior are deployed (Eicher *et al.* 2011, Fernandez *et al.* (2001)). FMA employs Mallow's criterion for model averaging using the orthogonalization of the parameter space (Hansen 2007, Amini & Parmeter 2012). All the variables describes Table 6.

Table A4: Why the estimates vary – more heterogeneity components

Response variable	*Post. Mean*	*Post. SD*	*PIP*	Response variable	*Post. Mean*	*Post. SD*	*PIP*
Constant	0.349	NA	1.000	Studies	44		
Winsorized SE	-0.266	0.409	0.324	Observations	468		
Data characteristics				*Estimation technique*			
Contemporaneous	0.013	0.022	0.310	FamMacB	-0.001	0.005	0.017
Price change	0.000	0.004	0.007	VARa	-0.002	0.013	0.026
Absolute returns	0.002	0.011	0.040	VARb	-0.012	0.029	0.183
Abnormal returns	0.196	0.048	0.995	VARc	-0.003	0.018	0.042
Excess returns	-0.005	0.019	0.089	Simple model	0.000	0.004	0.018
Dollar volume	0.000	0.002	0.005	GARCH	0.023	0.036	0.319
Shares traded	0.000	0.006	0.023	Monday	0.000	0.004	0.007
Detrended series	-0.090	0.034	0.974	Yield	-0.003	0.016	0.048
Data period	0.000	0.000	0.009	Standard deviation	-0.033	0.040	0.449
Data size	0.011	0.008	0.699	Market Beta	0.000	0.004	0.007
Midyear	-0.083	0.012	1.000	Illiquidity	0.196	0.060	0.998
Daily data	-0.001	0.006	0.016	Accrual	0.020	0.050	0.154
Weekly data	0.010	0.028	0.129	Sales-to-price ratio	0.000	0.004	0.008
Monthly data	-0.218	0.041	1.000	Firm Beta	-0.004	0.018	0.058
Time series	0.018	0.037	0.230	Firm size	0.000	0.004	0.006
Cross section	0.152	0.068	0.899	Trimmed	0.047	0.041	0.641
Structural variation				January Excluded	0.000	0.002	0.007
Index	-0.016	0.033	0.227	Informed	0.000	0.002	0.003
NASDAQ	0.000	0.006	0.011	Idiosyncratic volatility	0.007	0.027	0.084
Banks	-0.017	0.038	0.185	MLE	0.001	0.010	0.028
Firms	0.057	0.036	0.769	GMM	-0.002	0.011	0.036
Developing	0.014	0.029	0.212	Other methods	-0.056	0.027	0.873
Market size	-0.010	0.011	0.505	*Publication characteristics*			
Asia	0.036	0.040	0.494	Impact factor	0.000	0.004	0.018
Europe	0.000	0.004	0.011	Citations	0.006	0.014	0.193
Australia	0.052	0.060	0.475	Published	-0.034	0.046	0.446
Other Continents	0.254	0.060	1.000				

Notes: Response variable = partial correlation between trading volume and stock returns. SD = standard deviation. PIP = posterior inclusion probability. The UIP g-prior and dilution model prior are deployed in BMA, as suggested by George *et al.* (2010). The estimation follows Table 7 but with the broader set of variables explaining the heterogeneity in estimation techniques presented in Table A2.

IES Working Paper Series

2020

1. Tomas Kucera: *Cognitive Bias Mitigation: How to Make Decision-Making Rational?*
2. Tomas Kucera: *Are Employment Effects of Minimum Wage the Same Across the EU? A Meta-Regression Analysis*
3. Petr Hanzlik, Petr Teply: *Institutional and Other Determinants of the Net Interest Margin of US and European Banks in a Low Interest Rate Environment*
4. Michal Hlavacek, Ilgar Ismayilov, Ayaz Zeynalov: *Reassessment of the Fiscal Multiplier in Developing Countries: Regime-Switching Model*
5. Evzen Kocenda, Karen Poghosyan: *Nowcasting Real GDP Growth: Comparison between Old and New EU Countries*
6. Diana Zigraiova, Tomas Havranek, Jiri Novak: *How Puzzling Is the Forward Premium Puzzle? A Meta-Analysis*
7. Barbora Malinska: *Time-Varying Pricing of Risk in Sovereign Bond Futures Returns*
8. Shahriyar Aliyev, Evzen Kocenda: *ECB Monetary Policy and Commodity Prices*
9. Roman Kalabiska, Michal Hlavacek: *Regional Determinants of Housing Prices in the Czech Republic*
10. Boris Fisera, Roman Horvath: *Are Exchange Rates Less Important for Trade in a More Globalized World? Evidence for the New EU Members*
11. Jana Votapkova: *The Effect of Inpatient User Charges on Inpatient Care*
12. Lenka Slegerova: *Using 'Costs States' in a Semi-Markov Model to Estimate Cost-Effectiveness with an Illustration for Metastatic HER2+ Breast Cancer in the Czech Republic*
13. Periklis Brakatsoulas, Jiri Kukacka: Credit Rating Downgrade Risk on Equity Returns
14. Roman Horvath: *Natural Catastrophes and Financial Development: An Empirical Analysis*
15. Vit Machacek: *Globalization of Science: Evidence from Authors in Academic Journals by Country of Origin*
16. Nino Buliskeria, Jaromir Baxa: *Do Rural Banks Matter That Much? Burgess and Pande (AER, 2005) Reconsidered*
18. Kseniya Bortnikova: *Beauty and Productivity: A Meta-Analysis*
19. Radomir Mach, Milan Scasny, Jan Weinzettel: *The Importance of Retail Trade Margins for Calculating the Carbon Footprint of Consumer Expenditures: A Sensitivity Analysis*
20. Javier Garcia-Bernardo, Petr Jansky, Thomas Tørsløv: *Multinational Corporations' Effective Tax Rates: Evidence from Orbis*

21. Petr Jansky, Andres Knobel, Markus Meinzer, Tereza Palanska, Miroslav Palansky: *Country-by-Country Reporting and Other Financial Transparency Measures Affecting the European Union*
22. Marek Sedivy: *Mortality shocks and household consumption: The case of Mexico*
23. Lydia Chikumbi, Milan Scasny, Edwin Muchapondwa, Djiby Thiam: *Premium Price For Natural Preservatives In Wine: A Discrete Choice Experiment*
24. Roman Horvath: *Peer Effects in Central Banking*
25. Nicholas Tyack, Milan Scasny: *Estimating the Social Value of Specific Crop Diversity Conservation Plans: Do Czechs Care More About Conserving Hop, Wine or Fruit Tree Varieties?*
26. Salim Turdaliev: *Labor Force Participation of Married Woman in Russia*
27. Jaromir Baxa, Michal Paulus: *Exchange rate misalignments, growth, and institutions*
28. Michal Paulus, Jaromir Baxa, Eva Michalikova: *Does Enforcement Of the Rules Against Foreign Bribery Discourage Exports? A Case of the OECD Anti-Bribery Convention*
29. Tomas Havranek, Zuzana Irsova, Lubica Laslopova, Olesia Zeynalova: *Skilled and Unskilled Labor Are Less Substitutable than Commonly Thought*
30. Levan Bezhanishvili, William Appleman, Zurab Abramishvili: *Was the Georgian Policy Shifting Public Sector Working Hours by One Hour "Family Friendly" and Did It Increase Female Labor Participation?*
31. Fan Yang: *A Survey of Empirical Literature on Hedge Fund Performance*
32. Ali Elminejada, Tomas Havranek, Roman Horváth: *A Meta-Analysis of the Frisch Extensive Margin Elasticity*
33. Petra Landovská: *Business cycle sensitivity of Statutory Health Insurance: Evidence from the Czech Republic*
34. Natalia Li: *Estimating the Relationship Between Resource Intensity and Occupational Health and Safety in Kazakhstan*
35. Sophio Togonidze, Evžen Kočenda: *Macroeconomic Responses of Emerging Market Economies to Oil Price Shocks: Analysis by Region and Resource Profile*
36. Olena Chornaa, Lucas van der Veldeb : *Do Women Benefit from Minimum Wages?*
37. Sarah Godar and Petr Janský*: Corporate Profit Misalignment: Evidence from German Headquarter Companies and Their Foreign Affiliates*
38. Leyla Ates, Alex Cobham, Moran Harari, Petr Janský, Markus Meinzer, Lucas Millan-Narotzky, Miroslav Palanský: *The Corporate Tax Haven Index: A New Geography of Profit Shifting*
39. Ichiro Iwasaki, Evžen Kočenda, Yoshisada Shida: *Institutions, Financial Development, and Small Business Survival: Evidence from European Emerging Markets*
40. Laure de Batz, Evžen Kočenda: *Financial Crime and Punishment: A Meta Analysis*

41. Petr Janský: *Corporate Effective Tax Rates for Research and Policy*
42. Svatopluk Kapounek, Zuzana Kučerová, Evžen Kočenda: *Selective Attention in Exchange Rate Forecasting*
43. Vědunka Kopečná, Milan Ščasný and Lukáš Rečka: *Estimating Elasticity of Substitution in CES Production Function: Examining Different Nesting Structures and EU Regions*
44. Barbora Stepankova: *Consistency of Banks' Internal Probability of Default Estimates*
45. Josef Bajzik: *Trading Volume and Stock Returns: A Meta-Analysis*

All papers can be downloaded at: http://ies.fsv.cuni.cz

Univerzita Karlova v Praze, Fakulta sociálních věd
Institut ekonomických studií [UK FSV – IES] Praha 1, Opletalova 26
E-mail: ies@fsv.cuni.cz http://ies.fsv.cuni.cz