# The Friction-Free Weighted Price Contribution

David Abad
Universidad de Alicante
goliat@ua.es


Roberto Pascual[**]
Universitat de les Illes Balears
rpascual@uib.es

This version:
**January 2014**

# The Friction-Free Weighted Price Contribution

## Abstract

We introduce a methodology to obtain friction-free estimates of Barclay and Warner's (1993) Weighted Price Contribution (WPC). With this new approach, we verify recent simulation results suggesting that trading frictions may severely bias the WPC approach. We use high frequency data from a European electronic order-driven market to show that frictions generate a sizable downward bias in the WPC of non-aggressive small-size trades. The bias increases in periods of significant price discovery, and is due to both bid-ask bounce and serial correlation in the quote-midpoint changes. Our results are not market specific: we provide evidence of biased WPC estimates in US markets.

**Keywords:** Weighted price contribution, trading frictions, stealth trading, trade size, aggressiveness, bid-ask bounce, price formation, electronic markets, market microstructure.

## 1. Introduction

It is widely accepted that private information in financial markets is revealed through trading and that information-motivated traders, in an attempt to delay the full revelation of their information, look for ways to conceal their trading intentions.[1] Barclay and Warner (1993) (hereafter, BW93) study the trade-size choices of strategic informed traders. They introduce the Weighted Price Contribution (WPC) approach to test the so-called stealth trading hypothesis (STH).[2] The WPC measures how much of a stock's cumulative price change over a given time period is attributable to trades in particular trade-size categories. The WPC has become the standard tool to study strategic fragmentation of orders, and is frequently used as an alternative to parametric methods to measure price leadership (e.g., Hasbrouck, 1995).[3]

The WPC assumes that price changes are primarily information-driven. The existence of a noisy component in price changes, one of the milestones of market microstructure research (e.g., Hasbrouck, 2007), therefore challenges the WPC approach. BW93 (p. 300) state that the WPC allows for trading frictions, so long as the temporary components of the price change are not systematic. Some recent studies, however, cast some doubt on their claim. van Bommel (2011) studies the statistical properties of the WPC using simulated daily series. He finds that the WPC is an inconsistent and biased estimator of price discovery when prices deviate from a martingale process due to serial correlation. Likewise, Wang and Yang (2010) uses low-frequency data to show that the WPC approach deviates from the information share

---

[1] For example, strategic informed traders may obscure their positions by breaking up large trades and spreading them through time (Kyle, 1985); trading when liquidity-motivated volume is high (Admati and Pfleiderer, 1988); acting as passive liquidity providers (Harris, 1998, and Kaniel and Liu, 2006) or using undisclosed limit orders (e.g., Bessembinder, Panayides, and Venkataraman, 2009).

[2] The STH reads: "[…] if informed traders concentrate their trades in medium sizes, and stock-price movements are mainly due to private information revealed through these investors' trades, then most of the stocks cumulative price change will take place on medium size trades" (BW93, p. 282).

[3] A few examples include Cao, Ghysels, and Hatheway (2000), Huang (2002), Barclay and Hendershott (2003, 2008), and Ellul, Shin, and Tonks (2005).

approach of Hasbrouck (1995) in the presence of return serial correlation. At best, these studies stress the need to control for trading frictions when implementing the WPC approach.

In this paper, we propose a simple procedure to obtain friction-free WPC estimates using high (trade-by-trade) frequency data. The procedure consists of three steps: firstly, we replace trade prices by quote midpoints to control for the bid-ask bounce; secondly, we apply standard time series techniques to extract the friction-related dynamics in the quote-midpoint changes; finally, we use the estimated friction-unrelated component of quote-midpoint changes to obtain the friction-free WPC estimate ($\overline{\text{WPC}}$).[4] By means of the $\overline{\text{WPC}}$ estimates, we provide empirical evidence of the friction-related bias in the standard WPC when applied to high-frequency-data studies.

We begin by reporting a significant friction-driven downward bias in the standard WPC of small-size trades for the most liquid and active stocks of the electronic order-driven platform of the Spanish Stock Exchange (SSE) between July 2000 and December 2006. Ignoring trading frictions, the daily WPC of medium-size trades is 83%, while the daily WPC of small-size trades is -1.6%. In contrast, the $\overline{\text{WPC}}$ of medium (small) size trades is 55.4% (37.1%). Next, we test if this friction-related downward bias for small-size trades increases with the likelihood of information-motivated trading. Consistently, as we restrict ourselves to days with open-to-close returns above 1%, the downward bias of the small-size trades increases, from -38.7% to -72.71%. Negative WPCs for small size trades are frequently taken as evidence of underperformance by retail traders (Barber and Odean, 2000). Instead, our findings show that negative WPCs are due to trading frictions. We also document that the friction-related bias in WPC can significantly distort formal tests of the STH.

We provide some insights on the sources of the bias. Firstly, we estimate that 85.6% of the daily WPC downward bias is driven by the bid-ask bounce, but serial correlation significantly

---

[4] Here and in the rest of the text we use an overscore to indicate that a measure is friction-free.

adds to the bias. So, the friction-related bias of the standard WPC cannot be fully corrected by simply replacing the trade price by the quote midpoint. Secondly, trade aggressiveness helps to explain the bias. The downward bias of small-size trades only affects to non-aggressive trades (-49.1%) ─ the WPC of aggressive small-size trades is actually upward biased (11%).

We conduct a number of robustness tests. We start by showing that the magnitude of the WPC friction-related bias depends on the time resolution of the analysis. In particular, the downward bias in the WPC of small size trades increases when computed over monthly rather than daily intervals. Quite the opposite, $\overline{WPC}$s computed at different frequencies and/or minimum daily return cutoffs converge. We also show that our results are not market-specific. For two representative samples of 1995 and 2005 NYSE-listed US common stocks, we corroborate the downward bias in the WPC of small-size trades. Yet, the estimated bias for US markets is of smaller magnitude than for the SSE. In days with positive returns and with consolidated (NYSE-NASDAQ-Pacific/Arca) trades and quotes, we document a -12.35% bias in 1995 and a -16.78% bias in 2005. The friction-related bias decreases when we use NYSE trades and quotes only (-4.8% in 1995 and -12.83% in 2005).

Overall, we conclude that the WPC approach can be seriously biased by the presence of the friction-related component in the time series of the trade price change. In contrast, the $\overline{WPC}$ approach provides accurate and reliable estimates.

The rest of the paper proceeds as follows. In Section 2, we summarize the stealth trading literature. In Section 3, we review the WPC approach. In Section 4, we describe the SSE database. In Section 5, we present $\overline{WPC}$ approach. In Section 6, we report our main findings. In Section 7, we provide robustness tests. In Section 8, we conclude.

## 2. Strategic fragmentation of orders

4

Studies about US markets during the 1980s and 1990s, spearheaded by BW93 seminal work, support the STH. With a sample of tender-offer target firms, BW93 find that 99% of the cumulative price change during the pre-announcement period occurs on medium sizes. Chakravarty (2001) uses TORQ data to show that nearly 80% of the cumulative price change is due to medium-size trades by institutional investors. Chakravarty, Chiyachantana, and Jiang (2008) report a disproportionately large WPC for mid-size trades in the 2-day interval immediately after (before) positive (negative) earnings surprises in the 1990s. Furthermore, institutional trading increases in the exact same intervals dominated by stealth traders.

More recently, price discovery in US markets has apparently shifted away from medium sizes and into small sizes. Using NASDAQ data, Choe and Hansch (2005) find support to the STH from 1993 to 1998, but from 1999 to 2003 about 70-85% of the cumulative price change is due to small trades. They attribute this reversal to microstructure changes.[5] Chakravarty, Van Ness, and Van Ness (2005) show that, after NYSE decimalization in 2001, the dollar adverse selection costs for medium-size trades decreased, which is compatible with informed traders shifting to smaller sizes. Using 2005 data from NASDAQ, Blau, Van Ness, and Van Ness (2009b) show that small short sales drive the short sales' predictability of negative next-day returns. Finally, O'Hara, Yao, and Ye (2014) use 2008-2009 NASDAQ data to show that more than 80% of the cumulative price change is accounted for trades of 100-shares or less. They point to the raise of high frequency trading as the most likely explanation. In our opinion, the 2000s evidence for US markets suggests that the lower bound of trade size for stealth trading has decreased over time and, therefore, it is consistent with the STH.[6]

All the studies above use BW93's WPC approach described next.

---

[5] Higher transparence, lower tick sizes, and enhanced competition have reduced transaction costs, most notably for small-size orders (e.g., Smith, 1998, Chung, Chuwonganant, and McCormick, 2004).

[6] There are a few studies about strategic fragmentation for non-US markets. Cai, Cai and Keasey (2006) find stealth trading in the Chinese stock market. Ascioglu, Comerton-Forde, and McInish (2011) show that small-size trades in the Tokyo Stock Exchange rule price formation except in high volatility days. Kalev and Pham (2009) find that in the Australian Stock Exchange stealth trading happens when liquidity is low. Finally, Menkhoff and Schmeling (2010) provide evidence of stealth trading on a foreign exchange electronic limit order market.

## 3. The WPC approach

Suppose we have $K$ size categories of trades and $S$ stocks. For each stock $s = \{1, \ldots, S\}$ there are $N^s$ trades on a given time interval $t$. The price contribution of the size category $k$ for stock $s$ on the interval $t$ is given by

$$PC_{k,t}^s = \frac{\sum_{n=1}^{N^s} \Delta p_n \delta_{n,k}}{\sum_{n=1}^{N^s} \Delta p_n}, \qquad [1]$$

where $\Delta p_n = p_n - p_{n-1}$ is the difference between the price of trade $n$ and the price of trade $n$-1, and $\delta_{n,k}$ equals one of the $n$-th trade belongs to the size category $k$ and zero otherwise.

Let $\Delta p_t^s$ be the cumulative price change for stock $s$ on interval $t$, the denominator of eq. [1]. The weighted cross-sectional average price contribution for the size category $k$ on the interval $t$ is computed as[7]

$$WPC_{k,t} = \sum_{s=1}^{S} w_t^s PC_{k,t}^s, \qquad [2]$$

and the weights in eq. [2] are given by

$$w_t^s = \frac{\left| \Delta p_t^s \right|}{\sum_{s=1}^{S} \left| \Delta p_t^s \right|}. \qquad [3]$$

In prior studies, $\Delta p_n$ in eq. [1] is defined either as the difference in prices (e.g., BW93) or as the log return (e.g., Chakravarty et al., 2005), and the length of $t$ varies from several months (e.g., Ascioglu et al., 2011) to a few hours (e.g., Blau, Van Ness, and Van Ness, 2009a). Some studies choose stocks with $\Delta p_t^s > 0$ (e.g., Chakravarty, 2001), while some others ignore the sign of $\Delta p_t^s$ (e.g., O'Hara et al., 2014). Finally, overnight returns can be included (e.g., Chakravarty, 2001) or excluded (e.g., Choe and Hansch, 2005). All studies,

---

[7] The weighting mitigates potential biases resulting from large cumulative price changes of volatile firms.

however, coincide in the way they define the price concession in eq. [1], the WPC in eq. [2], and the weights in eq. [3].

BW93 propose two alternatives to the STH. The public information hypothesis (PIH) claims that price changes are entirely due to public information. Accordingly, the price contribution in eq. [1] for each trade size category must be proportional to its share on total trades. The trading volume hypothesis (TVH) claims that large trades move prices more than small trades. Accordingly, the price contribution in eq. [1] for each trade-size category must be proportional to its share on total volume. BW93 propose testing the STH against the PIH using the following pooled weighted-least-squares regression

$$PC_{k,t}^s = \alpha_S D_{S,t}^s + \alpha_M D_{M,t}^s + \alpha_L D_{L,t}^s + \beta PT_{k,t}^s + \varepsilon_{k,t}^s, \qquad [4]$$

where $D_{S,t}^s$, $D_{M,t}^s$ and $D_{L,t}^s$ are dummies that equal 1 when $PC_{k,t}^s$ falls into the small-, medium- or large-size category, respectively. $PT_{k,t}^s$ is the proportion of trades in size category $k$ for stock $s$ on the interval $t$. To test the STH against the TVH, we replace $PT_{k,t}^s$ in eq. [4] by the proportion of volume in size category $k$ for stock $s$ on the interval $t$ ( $PV_{k,t}^s$ ). Regression weights are given by eq. [3]. Under the PIH or the TVH, the three $\alpha$ coefficients in eq. [4] must be zero and $\beta$ must be one.

## 4. The dataset

We use trade and quote data from the SSE.[8] Our database has several advantages over the most accessible database for US markets, the Monthly TAQ of the NYSE: all trades and quotes are reported, while the TAQ does not report odd lots (O'Hara et al., 2014); trades and quotes are already matched, while TAQ users face an imprecise matching because of the treatment of millisecond time stamps (e.g., Holden and Jacobsen, 2014); we do not need of

---

[8] In 2010, the World Federation of Exchanges ranked the SSE as the 5th largest stock exchange in Europe by market capitalization, and the 4th by total value of share trading.

inaccurate trade classification algorithms to determine the direction of trades (e.g., Easley, López de Prado, and O'Hara, 2013); we can accurately measure trade aggressiveness, and, finally, we avoid problems that arise with market fragmentation (e.g., Holden and Jacobsen, 2014) by using pre-MiFID (effective since November 1st, 2007) data.[9]

We start with all the common stocks handled by the electronic order-driven platform of the SSE. Its continuous session spans from 9:00 a.m. to 5:30 p.m., and is preceded by a 30-minute opening call auction and followed by a 5-minute closing call auction. There are no designated market makers. Limit orders are stored in an open limit order book following the usual price-time priority rule. Every trade consumes liquidity either at the ask side or the bid side of the book. Price improvements are not possible.[10] Finally, the tick size is 0.01€.

Our sample period extends from July 2000 to December 2006. During this period there were no remarkable microstructure changes in the SSE. From the whole population of SSE-listed stocks, we keep 55 that, during the 6½ years considered, were IBEX-35 constituents for 1 year or more, never delisted from the electronic continuous platform, and traded for at least 3 consecutive years.[11]

The data consists of all the updates of the limit order book (LOB). Trades are easily located because they augment the accumulated volume field; change (at least) the best quotes and/or the depth at the best quotes, and the trade price field is non-zero. For each trade, we know the price, size, and time stamp (in seconds). While many trades may occur within the same second, the order of execution is known because of an internal sequence indicator. Quotes and depth prevailing right before each trade are used to classify trades as aggressive or

---

[9] Even nowadays, the fragmentation of the SSE is limited. In December 2013, the SSE accounts for 83.96% of the trading volume on SSE-listed stocks, whilst Bat-ChiX has a 14.35% share and Turquoise a 1.59% share. Source: LiquidMetrix (http://www.if5.com/LiquidMetrix/Battlemap).

[10] Iceberg orders are allowed, but not fully hidden orders (see Pardo and Pascual, 2012).

[11] The IBEX-35 is the official index. Regularly revised every semester, extraordinary revisions are also common. Index constituents represent 90.2% of the trading volume of the SSE within our sample period.

non-aggressive. We exclude data from the opening and closing call auctions, and from intraday rule-based 5-minute volatility auctions.[12]

In Table I, we provide descriptive sample statistics. We form four equally-sized portfolios based on market capitalization, whose composition is revised at the end of each year. Table I shows that the stocks in the upper quartile ("C1") are much more active and liquid than the other stocks in the sample.

[Table I]

## 5. Methodological details: the friction-free WPC

In our empirical application, we obtain daily WPCs for our sample of SSE-listed stocks.[13] We define $p_n$ in eq. [2]-[3] as the *marginal* price of trade $n$, that is, the price of the last share transferred with the $n^{\text{th}}$ trade. We control for outliers in $PC_{k,t}^s$ in eq. [1] by applying the rule of thumb (e.g., Hawkins, 1980) of dropping observations with a Z-score equal or larger than 3 in any trade-size category.[14] We use the same trade-size cutoffs than in previous studies: 500 and 10,000 shares. In the SSE there are not round lots as in US markets. Trades below 100 shares represent 14.75% of all trades in our sample.[15]

We obtain daily $\overline{\text{WPC}}$s as follows. Firstly, we use quote midpoints instead of prices to control for bid-ask bounce effects. The midpoint change is $\Delta q_n = q_n - q_{n-1}$, where $q_n$ stands for the average of the inside quotes right after trade $n$.[16] Notice that a bid-ask-bounce-driven price change ($\Delta p_n \neq 0$) results in a zero midpoint change ($\Delta q_n = 0$). We assume that $q_n$ can be decomposed into an efficient price ($m_n$) and a transitory component ($s_n$): $q_n = m_n + s_n$

---

[12] For details on the short-lived rule-based trading halts of the SSE, see Abad and Pascual (2010).

[13] In the robustness section, we report results with a monthly resolution.

[14] With a cutoff of 1.5, our results are the same.

[15] We have also considered stock-specific trade-size cutoffs derived from the empirical distribution of price changes of each particular stock in our sample. Our conclusions do not vary.

[16] We also considered log prices and log quotes. Our findings and conclusions are the same.

(e.g., Stoll, 2000). We define $m_n = E[\mathfrak{I}_T | \Phi_n]$ as the expected true value of the asset at some distant future conditional on the public information available right after trade $n$. The unexpected component of $\Delta q_n$ is $\Delta m_n = \Delta q_n - E[\Delta s_n | \Phi_{n-1}]$.

Secondly, we follow Hasbrouck (1991) in assuming that the relevant information in $\Phi_{n-1}$ is the history of mid-quote revisions and trades up to trade *n-1*. The trading process is summarized by the trade sign ($x_n$), which equals 1 for buyer-initiated trades and -1 for seller-initiated trades; the signed trade size ($v_n$) in shares, and the interaction of $\{x_n, v_n\}$ with the prevailing bid-ask spread ($sp_n$) and the trade duration ($d_n$) – time in seconds between consecutive trades. The rational for including trade durations builds upon Easley and O'Hara's (1992) model, in which trades of any size executed at short durations, should have a higher average price impact. Dufour and Engle (2000) provide supporting evidence. The inclusion of the spread is justified by the extensive adverse-selection costs literature.[17] Empirically, Hasbrouck (1991) finds that trades that occur when the spread is wide have a relatively higher price impact. We assume that $E[\Delta s_n | \Phi_{n-1}]$ is a stable-over-time linear function of the history of trades and quotes,

$$\Delta q_n = \sum_{j=1}^{r}\left(\alpha_j \Delta q_{n-j} + \beta_j^x x_{n-j} + \beta_j^v v_{n-j}\right) + \sum_{j=1}^{r}\left(\beta_j^{xsp} x_{n-j} sp_{n-j} + \beta_j^{vs} v_{n-j} sp_{n-j}\right) + $$
$$+ \sum_{j=1}^{r}\left(\beta_j^{xd} x_{n-j} d_{n-j} + \beta_j^{vd} v_{n-j} d_{n-j}\right) + \Delta m_n \tag{5}$$

For each stock-month, we estimate eq. [5] by OLS with White-robust standard errors. We discard the first *r* trades each day, so that no lag reaches back to the previous day. We consider different options for *r*, from 5 to 15 lags. Our conclusions are robust to the choice of *r*, but the average first order autocorrelation of $\Delta m_n$ decreases up to *r=*10. So, we report

---

[17] See Copeland and Galai (1983), Glosten and Milgrom (1985), Glosten (1994), and Handa, Schwartz, and Tiwari (2003), among others.

findings exclusively for $r$=10. The average Durbin Watson statistic for $\Delta m_n$ across stock-monthly estimates is 1.99 (std. 0.0217).

Finally, we compute the $\overline{WPC}$ of a given trade-size category by means of equations [1] to [3], but replacing $\Delta p_n$ by the estimated $\Delta m_n$.

## 6. Empirical findings

### 6.1. The friction-related bias

In Table II, we present the average WPC in eq. [2] and its friction-free version ($\overline{WPC}$), computed using the time series of $\Delta m_n$ in eq. [5]. We also report the weighted average daily proportion of trades (%T) and volume (%V) using the weights in eq. [3].

[Table II]

According to the WPC, medium-size trades account for more than 83% of the price discovery in the SSE between 2000 and 2006, overcoming their %T (about 39%) and %V (about 61%). Large-size trades' contribution is about 18.5%, largely above their %T (2.2%) but below its %V (23.4%). In contrast, the WPC of small-size trades is negative (-1.6%), far below their %T (58.7%) and %V (15.7%). Overall, the SSE displays similar stealth trading patterns than the US markets during the 1980s and 1990s.

The $\overline{WPC}$ approach shows a markedly different scenario. While medium-size trades account for a more modest 55.4% of the cumulative price change, the $\overline{WPC}$ of small-size trades achieves a remarkable 37.1%. In other words, we report a 38.71% friction-related downward bias in the daily WPC of SSE small-size trades between 2000 and 2006. While still below their %T, the $\overline{WPC}$ of small-size trades is more than twice their %V.

11

Chakravarty (2001) argues that, because of short-selling restrictions, informed sales in declining stocks are less likely than informed purchases in rising stocks. Hence, the chances of detecting stealth trading should increase if we pick days that display a significant price increase. In Table II, we provide separated results for days with open-to-close log returns ($R$) >1%. With this subset of days, the friction-related downward bias in the WPC of small-size trades increases to -72.7%. The friction-related bias in WPC is therefore more prominent in periods of significant price discovery, precisely when information-motivated trading is more likely.

Table II reports negative WPC for small-size trades, a frequent finding in the literature.[18] Some studies (e.g., Choe and Hansch, 2005) associate negative WPCs with underperformance by retail traders. Noticeably, our findings attribute negative WPCs to noise in the time series of price changes. By construction, only trades that generate non-zero $\Delta p_n$ contribute to the WPC. A negative WPC therefore implies that trades are pushing prices systematically in the "wrong" direction. We label a trade as being "wrong" if it causes a positive (negative) price change whilst the daily cumulative price change is negative (positive). If the underperformance explanation applies, we should expect to find that the initiators of small-size trades are systematically wrong. In our sample, there are 11,337,755 wrong trades, representing 21.5% of all trades. Nearly 46% of them are small-sized and 47% are medium-sized. Moreover, 22.1% of the small-sized trades are wrong, rather close to the 21.2% (19.8%) of medium- (large-) size trades. It seems unlikely that these tiny differences could explain neither the negative WPC of small-size trades, nor the huge gap in WPC found between small- and medium-size trades. So, we can conclude that our results do not support a performance-based explanation for negative WPCs.

---

[18] See Barclay and Warner (1993, Table 1), Chakravarty (2001, Table 1), Choe and Hansch (2005, Table 1), Chakravarty et al. (2008, Table 3), Ascioglu et al. (2011, Table 1).

We follow a pooled regression approach to provide evidence on the statistical significance of the friction-related bias in WPC. The magnitude of the bias for a particular stock $i$, at day $d$, and trade-size category $s$ is computed as $Bias_{ids} = PC_{ids} - \overline{PC}_{ids}$, where $PC_{ids}$ is the price contribution in eq. [1], and $\overline{PC}_{ids}$ is its friction-free version, obtained using $\Delta m_t$ and eq. [5]. We study both the direction (signed) bias and the magnitude (absolute) bias. Our stock-daily pooled regression model is of the following form:

$$Bias_{ids} = \alpha + \beta_1 M_{ids} + \beta_2 L_{ids} + \beta_3 LCaps_{ids} + \sum_{z=1}^{6} \theta_z Y_{ids}^z + \sum_{k=1}^{54} \phi_k S_{ids}^k + \varepsilon_{ids}, \quad [6]$$

where $M_{ids}$ ($L_{ids}$) is a dummy variable that equals 1 if the trade-size category is medium (large). Thus, the intercept captures the effect of the omitted category: small-size trades. As control variables, we include the following dummies: $LCap_{ids}$ equals 1 if stock $i$'s market capitalization at the beginning of year $z$ is above the upper quartile across stocks; $Y^z$ equals 1 if the observation corresponds to year $z$, with $z = \{2000, ..., 2006\}$, and $S^k$ equals 1 if the observation corresponds to stock $k$, with $k=\{1, ..., 54\}$. With $Y^z$ and $S^k$, we control for time- and stock-specific fixed effects. We estimate eq. [6] by weighted LS, with heteroskedasticity-adjusted standard errors. Weights are given by eq. [3]. We consider days with $R>1\%$ only. We report the estimated coefficients in Table III.

[Table III]

The intercept of the (signed) bias regression, which captures the average daily bias for small-size trades, is significantly negative at the 1% level. Hence, we confirm the statistical significance of the downward bias of small size trades reported in Table II. The coefficients for $M_{ids}$ and $L_{ids}$ are positive and larger than the intercept, confirming the upward-bias in the contributions of medium- and large-size trades. According to the absolute bias regression, the daily average magnitude of the bias is much larger for small-size trades than for the other

13

trade-size categories. Moreover, the magnitude of the daily bias tends to decrease with market capitalization.

How much of the friction-related bias in WPC is due to bid-ask bounce? We address this question by estimating the cross-sectional average $WPC_{k,t}$ in eq. [2] using quote midpoint changes ($\Delta q_n$) instead of trade price changes ($\Delta p_n$). Computed in this manner, the WPC accounts for bid-ask bounce, but ignores serial correlation in $\Delta q_n$. Table IV reports the resulting decomposition.

[Table IV]

The bid-ask bounce effect explains 85.3% of the bias across trade-size categories (89.6% for days with open-to-close returns >1%). Hence, replacing trade prices by quote midpoints in any empirical analysis should correct most of the bias, but still, $\overline{\text{WPC}}$ will provide more accurate estimates since transitory dynamics in $\Delta q_n$ markedly add to the bias.

Overall, the $\overline{\text{WPC}}$ approach provides empirical support to the simulation-based conclusion of van Bommel (2011) for low frequency data analyses, but using actual high frequency data: trading frictions can severely bias the WPC approach.

*6.2. Trade aggressiveness and the WPC friction-related bias*

Size-based classifications of trades, like those employed in stealth trading analyses, ignore order aggressiveness. In order-driven markets, aggressive buy (sell) orders walk up (down) the book. In contrast, non-aggressive trades do not consume all the depth available at the best opposite quote. Consequently, aggressive trades carry on higher price concessions than non-aggressive trades (e.g., Biais, Hillion, and Spatt, 1995). In Table V - Panel A, we provide sample statistics about trade aggressiveness in the SSE. Non-aggressive trades outnumber

aggressive trades by more than 3.5 to 1. They are also more common among small-size trades (82.8%) than among any other trade size category.

[Table V]

We use pooled regressions to get some insight on the distinct role that aggressive and non-aggressive trades may play on the WPC statistic. For each stock $i$, day $d$, and trade-size category $s$, we compute the proportion of both aggressive and non-aggressive trades that: (a) contribute to the WPC ($\Delta p_n \neq 0$); (b) contribute with a noise-driven price change ($\Delta p_n \neq 0$ and $\Delta q_n = 0$); (c) are wrong, and (d) are wrong and noisy.[19] Our pooled regression model is

$$
\begin{aligned}
y_{idsa} = \alpha + \beta_1 M_{idsa} + \beta_2 L_{idsa} + \beta_3 AS_{idsa} + \beta_4 AM_{idsa} + \beta_5 AL_{idsa} + \\
+ \beta_6 LCaps_{idsa} + \sum_{z=1}^{6} \theta_z Y_{idsa}^z + \sum_{k=1}^{54} \phi_k S_{idsa}^k + \varepsilon_{idsa}
\end{aligned}
\qquad [7]
$$

where sub-index $a$ represents the aggressiveness category (i.e., aggressive v. non-aggressive). The dependent variable $y$ is one of the (a) to (d) proportions defined above; $M$, $L$, $AS$, $AM$, and $AL$ are all dummy variables: $M$ ($L$) equals 1 if trades are medium- (large-) sized and non-aggressive; $AS$, $AM$ and $AL$ equal 1 if trades are aggressive and small-sized, medium-sized, and large-sized, respectively. Thus, the omitted category is small-size non-aggressive trades. We use the same control variables than in eq. [6]. Table V - Panel B reports the coefficients of eq. [7] estimated by weighted LS with robust standard errors.

We find that aggressive trades, no matter the size, contribute more often to the WPC than non-aggressive trades and are less often associated with bid-ask bounce driven price changes. Initiators of aggressive trades are more often wrong but, even then, price changes are less likely to be bid-ask bounce related. Besides, non-aggressive small-size trades have more chances to cause friction-related contributions to WPC than non-aggressive medium-size trades.

---

[19] Wrong trades are defined as in the prior subsections. Notice that all wrong trades have non-zero price changes.

Results in Table V suggest that non-aggressive trades may contribute relatively more often to the WPC without enhancing price discovery. It is therefore possible that the downward friction-related bias in the WPC of small-size trades may be partially explained by the higher proportion of non-aggressive trades for small-size trades. In Table VI, we report the WPC and $\overline{\text{WPC}}$ of our three trade-size categories but this time we control for aggressiveness. We report results for all days and days with $R>1\%$.

[Table VI]

We find that the friction-related downward bias in the WPC of small-size trades concentrates on non-aggressive trades. The WPC of aggressive small-size trades is, in effect, upward biased. In days with $R>1\%$, the downward bias of small-size non-aggressive trades rises to -70.4%, whilst non-aggressive medium-size trades experience the highest upward bias. In days with $R>1\%$, small- and medium-size aggressive trades display $\overline{\text{WPC}}$s of 22.7% and 50.55%, respectively, more than proportional to their %T and %V. Markedly, the $\overline{\text{WPC}}$ of non-aggressive medium-size trades reveals that these trades do not use to conceal informed traders.

We use the pooled regression model in eq. [7] to test the statistical significance of the friction-related bias in WPC. In this case, the dependent variable is either the signed bias ($Bias_{idsa} = PC_{idsa} - \overline{PC}_{idsa}$) or the absolute bias per trade-size/aggressiveness category. We report the weighted LS estimates in Table VII.

[Table VII]

The intercept of the (signed) bias regression is significantly negative at the 1% level, whilst the coefficient for the other categories is positive and larger than the intercept. Therefore, we confirm a statistically significant downward bias in the WPC of small-size non-aggressive trades. Medium-size aggressive and non-aggressive trades report the highest significant

upward biases. The |bias| regression confirms that the average magnitude of the bias is much larger for small-size non-aggressive trades than for the other categories. Moreover, it is significantly lower among large caps (at the 5% level of statistical significance).

*6.3. Trading frictions and formal tests of the STH*

Next, we use the $\overline{WPC}$ approach to provide evidence on the robustness of the BW93's methodology to test the STH. In Table VIII - Panel A, we report the estimated pooled weighted LS regression in eq. [4] using either $PC_{ids}$ or $\overline{PC}_{ids}$ as the explanatory variable. We focus on stock-days with $R>1\%$ to increase the likelihood of detecting stealth trading.

[Table VIII]

The public information (PIH) and the trading volume (TVH) hypotheses are rejected across the board. The nulls that the coefficient of *PT* or *PV* equals one and that all dummy variables have zero coefficients are rejected at the 1% level. Ignoring trading frictions, our results suggest that informed traders concentrate exclusively on medium-size trades ($\alpha_M > 0$), whilst small-size trades play no role in price discovery ($\alpha_S < 0$). Without trading frictions, however, $\alpha_S$ becomes positive and significant in the TVH regression. In fact, the null $\alpha_S = \alpha_M$ cannot be rejected at the 1% level.

In Panel B of Table VIII, we extend eq. [4] to control for trade aggressiveness. If we ignore trading frictions, the regressions do reveal the significant role small-size aggressive trades play in price discovery in the SSE. However, the PIH regression assigns a comparable role to medium-size non-aggressive trades, which contradicts the $\overline{WPC}$ results in Table VI. Once we filter the price changes for trading frictions, our results in all regressions reveal disproportional contributions for only small- and medium-size aggressive trades.

Overall, our results cast some doubts on the robustness of the formal tests of the STH to trading frictions.

## 7. Robustness

### 7.1. Monthly WPC

Do our results depend on the time resolution of the WPC analysis? In Table IX, we present the average $\overline{WPC}$ across stocks and months, together with the estimated friction-related bias in WPC, and the weighted average daily proportion of trades (%T) and volume (%V). We focus on monthly observations with $R>2.5\%$ to increase the likelihood of observing stealth trading. We also control for trade aggressiveness.

[Table IX]

We find that the magnitude of the friction-related bias depends on the time resolution. Namely, the bias increases with time aggregation. As in the daily analysis, small-size aggressive trades' WPC is downward biased, whilst medium-size trades' WPC experiences the highest upward bias. Once we control for trading frictions, however, the resulting $\overline{WPC}$s are close to those in Table VI. Hence, the $\overline{WPC}$ approach provides not only more accurate estimates than the WPC approach, but also more robust and reliable.

### 7.2. US markets

Is the friction-related bias of the WPC a market-specific phenomenon? In this section, we test if it extends to US markets, where the stealth trading research has been more active.

Our US data analysis suffers from limitations inherent to the database we employ, the Monthly TAQ of the NYSE: (i) The TAQ database does not report odd lots. O'Hara et al. (2014) show that small-size odd lots remarkably contribute to price discovery. (ii) We cannot measure trade aggressiveness because the TAQ does not provide order-level data (e.g.,

Bacidore, Ross and Sofianos, 2003). (iii) The TAQ does not include a trade-initiator flag. Thus, we need to align trades and quotes to assess the direction of trades. This task is particularly problematic in modern high frequency markets.[20] For all these reasons, our findings must be carefully interpreted.

We avoid dealing with high frequency markets using 1995 and 2005 data (whole years). By choosing these particular periods, we are also able to test if the friction-related bias in WPC is independent of the lower bound of trade size for stealth trading (e.g., Choe and Hansch, 2005).

We form a representative sample of common stocks by market capitalization as follows. We start with the CRSP universe of common stocks (SHRCLS=. or A) and restrict it to NYSE-listed stocks (PRIMEXCH=N). We eliminate stocks with SIC code greater than 9000 (Government owned), between 8880 and 8888 (ADRs and foreign Governments), and between 4800 and 4999 (regulated utilities). We also drop stocks from non-US companies, with security status other than "regular" (SECSTAT≠R), with dual shares, with trade prices below $2 or above $200, or with no daily price or volume records.

We rank and divide the remaining assets in four groups by daily average market capitalization. We discard those in the bottom quartile (smallest stocks). For each of the top three groups, we retain the 200 largest stocks, sort them by ticker symbol, and then pick every 4th stock. This procedure results in 150 randomly selected stocks (50 for each size group) with a significant size difference between groups.

We use trades and quotes during regular market hours. We clean the trade and quote files with the filters in Hendershott and Moulton (2011, p. 578) and Chakrabarty and Moulton

---

[20] Holden and Jacobsen (2013) show that the proliferation of fleeting orders and the TAQ treatment of millisecond time stamps cause significant distortions in methodologies that rely in the alignment of trades and quotes. Trade classification algorithms are a notorious example (e.g., Odders-White, 2000). Recently, Easley et al. (2013) argue that the growing speed, volume and fragmentation of current US markets pose serious challenges to traditional classification algorithms.

(2012, p. 14). To avoid problems of stale quotes, we construct two BBO series for each stock, one from NYSE quotes only (NYSE BBO) and the other with quotes from NYSE, Pacific Exchange/Arca and NASDAQ (NBBO).[21]

Regarding the alignment of trades and quotes, we make no allowance of trade reporting lags.[22] NYSE BBO quotes are aligned with NYSE trades only. The BBO constructed from the N/P/T/Q quotes is aligned with trades originating from those exchanges. We use three alternative trade classification algorithms: the tick rule, Lee and Ready (1991), and Ellis, Michaely, and O'Hara (2000).[23] Overall, our findings are robust to the classification rule.

In Table X, we provide the WPC for our sample of 1995 (Panel A) and 2005 (Panel B) NYSE-listed stocks. Using the NBBO, the 1995 WPC of small-size trades is 11.54%, whilst in 2005 raises to 60.23%.[24] Using NYSE data only, the 1995 WPC of small-size trades is higher (19.18%) but only slightly lower in 2005 (57.65%). In days with strictly positive open-to-close returns, the 2005 WPC of small-size trades falls to 48.53% with NBBO and 36% with NYSE BBO. Overall, results in Table X illustrate the increasing role of small-size trades in price discovery previously reported for US stock exchanges.

[Table X]

In Table X, we also report the estimated friction-related bias ($\mathrm{WPC} - \overline{\mathrm{WPC}}$) using Lee and Ready (1991) as trade classification rule. As in the SSE case, the overall WPC of small-size trades is downward biased both in 1995 and 2005, but the magnitude of the bias is smaller in

---

[21] We discard data from other exchanges. They represent 4.4% of all trades in 2005 and 22% in 1995.

[22] We obtain similar results with a 1-second trade reporting lag. In the 1980s and early 1990s, the rule was to allow for a 5-second trade reporting lag (e.g., Lee and Ready, 1991). Using 1998 data, Bessembinder (2003) concludes the 5-second rule is no longer necessary. Since then, the common practice is the one we follow (e.g., Hendershott and Moulton, 2011).

[23] Odders-White (2000), Chakrabarty, Li, Nguyen and Van Ness (2007), Easley et al. (2013), and Chakrabarty, Pascual and Shkilko (2012) offer reviews of trade classification algorithms.

[24] For NASDAQ-listed stocks, O'Hara et al. (2014) report a 60.9% WPC for 2008-2010 trades with size [100, 500) (Table 6, Panel B), whilst Choe and Hansch (2005) report a 77.5% WPC for 2000-2003 trades and a -1% WPC for 1993-1996 trades with size [100, 500) (Table 1).

US markets. The WPC bias is larger with consolidated (N/P/T/Q) trades and quotes (1995: -12.35% to -18.51%; 2005: -11.08% to -16.78%) than with just NYSE trades and quotes (1995: -4.8% to -5.38%; 2005: -7.77% to -12.83%). Markedly, the friction-related bias has not vanished with the fall in the trade size lower bound for stealth trading during the 2000s.

We test the statistical significance of the bias for each trade-size category by means of the following regression model,

$$Bias_{ids} = \alpha + \beta_1 M_{ids} + \beta_2 L_{ids} + \beta_3 N_{ids} + \sum_{k=1}^{150} \phi_k S_{ids}^k + \varepsilon_{ids} \,, \qquad [8]$$

where $Bias_{ids} = PC_{ids} - \overline{PC}_{ids}$. The dummies $M$ and $L$ equal 1 if $s$ is 'medium' and 'large', respectively. As controls, we include the log number of daily trades ($N$) and 150 dummies for stock fixed effects ($S^k$).[25] We estimate eq. [8] by weighted LS with heteroskedasticity-adjusted standard errors. We report the estimated $\alpha$, $\beta_1$, and $\beta_2$ coefficients in Table X.

We find positive and statistically significant (at the 1% level) $\beta_1$ and $\beta_2$ coefficients across the board, corroborating the upward bias WPC for medium- and large-size trades. $\beta_1$ and $\beta_2$ tend to be larger using the NBBO than the NYSE BBO. The small-size trades' coefficient ($\alpha$) is negative and highly significant for the 1995 consolidated trades and quotes. For the 2005 sample, the sign of $\alpha$ is consistent with the downward bias in WPC but only significant with NYSE BBO and on days with strictly positive returns.

## 8. Summary and conclusions

Introduced by Barclay and Warner (1993), the WPC approach has been extensively used to evaluate strategic fragmentation of orders and price leadership. Recently, van Bommel (2011) uses simulated data to question the unbiasedness of the WPC approach in the presence of

---

[25] We exclude the large-cap dummy in eq. [7] because of perfect multicollinearity with the stock-specific dummies. Separated models for the 50 largest stocks in each subsample, 1995 and 2005, generate weaker but similar results. These analyses are available upon request.

serial correlation in the price changes. van Bommel focuses on the use of the WPC in low-frequency analyses about price leadership. In this paper, we extend this line of research. Firstly, we introduce a simple methodology to obtain friction-free WPC estimates using high-frequency data. Secondly, we generalize van Bommel's result by showing that the WPC approach generates biased estimates in the presence of a general friction-related component in price changes.

Our friction-free WPC approach ($\overline{\text{WPC}}$), builds on Hasbrouck (1991). In a first step, we use quote midpoints instead of trade prices to get rid of the bid-ask bounce effect. In a second step, we use a time-series econometric model to extract the friction-related dynamics in the time series of the quote midpoint change. The resulting times series of friction-free quote midpoint changes is then used to compute the $\overline{\text{WPC}}$.

To illustrate the usefulness of our approach, we study strategic fragmentation of orders in the electronic trading platform of the SSE applying both the traditional WPC approach and the new $\overline{\text{WPC}}$ approach. We use high-frequency data from July 2000 to December 2006. We report a significant -38.7% friction-related bias in the WPC of small-size trades in the SSE. The bias increases to -72.71% if we restrict our analysis to days with open-to-close returns above 1%. We therefore show that the WPC is not robust to trading frictions and the bias is more notorious when stealth trading is more likely to occur.

Most of the friction-related bias (85.6%) is driven by the bid-ask bounce, but bid-ask-bounce-unrelated serial correlation markedly adds to the bias. Therefore, although replacing price changes by quote midpoint changes may correct most of the friction-related bias, the $\overline{\text{WPC}}$ approach guarantees more precise estimates. Moreover, the magnitude of the bias depends on the time resolution of the WPC analysis. In contrast, the $\overline{\text{WPC}}$ estimates obtained for alternative time resolutions are alike.

We show that negative WPCs for small-size trades, commonly found in US and non-US studies during the 80s and 90s, can be attributed to trading frictions rather than underperformance by retail traders, as it is often assumed. We also document that the friction-related bias can distort formal tests of the stealth trading hypothesis.

We study the role that trade-aggressiveness plays in explaining the downward bias in the WPC of small-size trades in the SSE. We find that non-aggressive trades, which are more common among small size trades, frequently add noise to the WPC. In fact, the downward bias only affects to small-size non-aggressive trades (-49.1%). In days with strictly positive returns, small- and medium-size aggressive trades disproportionally contribute to price discovery.

Finally, we show that the friction-related bias in WPC is not market-specific. Using 1995 and 2005 NYSE-listed US stocks, we find a downward-biased WPC for small-size trades, but of smaller magnitude than for the SSE. The bias is larger if we use consolidated trades and quotes rather than NYSE trades and quotes alone. Because of inherent limitations of the Monthly TAQ database, however, we must treat the US findings with caution.

# References

Abad, D. and R. Pascual, 2010. Switching to a temporary call auction in times of high uncertainty. *Journal of Financial Research* 33, 45-75.

Admati, A. R. and P. Pfleiderer, 1988. A theory of intraday trading patterns: Volume and price variability. *Review of Financial Studies* 1, 3-40.

Ascioglu, A., C. Comerton-Forde and T. H. McInish, 2011. Stealth trading: The case of the Tokyo Stock Exchange. *Pacific-Basin Finance Journal* 19, 194-207.

Bacidore, J., K. Ross and G. Sofianos, 2003. Quantifying market order execution quality at the New York Stock Exchange. *Journal of Financial Markets* 6, 281-307.

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

Barclay, M. J. and J. B. Warner, 1993. Stealth trading and volatility: Which trades move prices? *Journal of Financial Economics* 34, 281-305.

Barclay, M. J. and T. Hendershott, 2003. Price discovery and trading after hours. *Review of Financial Studies* 16, 1041–1073.

Barclay, M. J. and T. Hendershott, 2008. A comparison of trading and non-trading mechanisms for price discovery. *Journal of Empirical Finance* 15, 839-849.

Bessembinder, H., 2003. Issues in assessing trade execution costs. *Journal of Financial Markets* 6, 233-257.

Bessembinder, H., M. Panayides and K. Venkataraman, 2009. Hidden liquidity: an analysis of order exposure strategies in electronic stock markets. *Journal of Financial Economics* 94, 361-383.

Biais, B., P. Hillion and C. Spatt, 1995. An empirical analysis of the limit order book and the order flow in the Paris Bourse. *Journal of Finance*, 50, 1655-1689.

Blau, B. M., B. F. Van Ness and R. A. Van Ness, 2009a. Intraday stealth trading: which trades move prices during periods of high volume? *Journal of Financial Research* 32, 1-21.

Blau, B. M., B. F. Van Ness and R. A. Van Ness, 2009b. Information and trade sizes: The case of short sales. *Quarterly Review of Economics and Finance* 49, 1371-1388.

Cai, B. M., C. X. Cai and K. Keasey, 2006. Which trades move prices in emerging markets?: Evidence from China's stock market. *Pacific-Basin Finance Journal* 14, 453-466.

Cao, C., E. Ghysels, and F. Hatheway, 2000. Price discovery without trading: evidence from the Nasdaq pre-opening. *Journal of Finance* 55, 1339–1365.

Chakrabarty, B., B. Li, V. Nguyen, and R. Van Ness, 2007. Trade classification algorithms for electronic communications network trades. *Journal of Banking and Finance* 31, 3806-3821.

Chakrabarty, B. and P.C. Moulton, 2012. Earnings announcements and attention constraints. *Journal of Accounting and Economics* 52, 612-634.

Chakrabarty, B., R. Pascual and A. Shkilko, 2012. Trade classification algorithms: A horse race between the bulk-based and the tick-based rules. SSRN Working *Paper Series*.

Chakravarty, S., 2001, Stealth trading: Which traders' trades move stock prices? *Journal of Financial Economics*, 61, 289-307.

Chakravarty, S., B. F. Van Ness and R. A. Van Ness, 2005. The effect of decimalization on trade size and adverse selection costs. *Journal of Business, Finance and Accounting* 32, 1063-1081.

Chakravarty, S., C. N. Chiyachantana, and C. Jiang, 2008. Earnings Surprises and Stealth Trading, in *Stock Market Liquidity: Implications for Market Microstructure and Asset Pricing*, F.S. Lhabitant and G.N. Gregoriou (Eds.), John Wiley and Sons, New York.

Choe, H., and O. Hansch, 2005. Which trades move stock prices in the internet age? *Working paper*, Pennsylvania State University.

Chung, K. H., C. Chuwonganant, and D. T. McCormick, 2004. Order preferencing and market quality on NASDAQ before and after decimalization. *Journal of Financial Economics* 71, 581-612.

Copeland, T. E., and D. Galai, 1983. Information Effects on the Bid-Ask Spread. *Journal of Finance* 38, 1457–1469.

Dufour, A. and R. F. Engle, 2000. Time and the price impact of a trade. *Journal of Finance* 55, 2467-2498.

Easley, D., M. López de Prado and M. O'Hara, 2013. Bulk classification of trading activity. *Johnson School Research Paper Series #8-2012.*

Easley, D., and M. O'Hara, 1992. Time and the process of security price adjustment. *Journal of Finance* 47, 577-605.

Ellis, K., R. Michaely and M. O'Hara, 2000. The accuracy of trade classification rules: Evidence from Nasdaq. *Journal of Financial and Quantitative Analysis* 35, 529-551.

Ellul, A., H. Shin and I. Tonks, 2005. Opening and closing the market: Evidence from the London stock exchange. *Journal of Financial and Quantitative Analysis* 40, 779-801.

Glosten, Lawrence R., 1994. Is the electronic open limit order book inevitable? *Journal of Finance* 49, 1127–1161.

Glosten, L. R. and P. R. Milgrom, 1985. Bid, ask and transaction prices in specialist market with heterogeneously informed traders. *Journal of Financial Economics* 14, 71–100.

Gonzalo, J. and C. Granger, 1995. Estimation of common long-memory components in cointegrated systems. *Journal of Business and Economic Statistics* 13, 27-35.

Handa, P., R. A. Schwartz and A. Tiwari, 2003. Quote setting and price formation in an order driven market. *Journal of Financial Markets* 6, 461–489.

Harris, L., 1998. Optimal dynamic order submission strategies in some stylized trading problems. *Financial Markets, Institutions, and Instruments* 7, 26-74.

Hasbrouck, J., 1991. Measuring the information content of stock trades. *Journal of Finance* 46, 179-207.

Hasbrouck, J., 1995. One security, many markets: Determining the contributions to price discovery. *Journal of Finance* 50, 1175-1199.

Hasbrouck, J., 2007. Empirical market microstructure. Oxford University Press.

Hawkins, D. M., 1980. Identification of outliers. Chapman and Hall, London.

Hendershott, T. and P.C. Moulton, 2011. Automation, speed, and stock market quality: The NYSE's hybrid. *Journal of Financial Markets* 14, 568-604.

Holden C. and S. Jacobsen, 2014. Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions. *Journal of Finance,* forthcoming.

Huang, R. D., 2002. The quality of ECN and Nasdaq market maker quotes. *Journal of Finance* 57, 1285–1319.

Kalev, P. S. and L. T. Pham, 2009. Intraweek and intraday trade patterns and dynamics. *Pacific-Basin Finance Journal* 17, 574-564.

Kaniel, R. and H. Liu, 2006. So what orders do informed traders use? *Journal of Business* 79, 1867-1914.

Kyle, A. S., 1985. Continuous auctions and insider trading. *Econometrica* 53, 1315-1335.

Lee C. M. C. and M. Ready, 1991. Inferring trade direction from intraday data. *Journal of Finance* 46, 733-747.

Menkhoff, L. and M. Schmeling, 2010. Whose trades convey information? Evidence from a cross-section of trades. *Journal of Financial Markets* 13, 101-128.

Odders-White, E., 2000. On the occurrence and consequences of inaccurate trade classification. *Journal of Financial Markets* 3, 259-286.

O'Hara, M., C. Yao and M. Ye, 2014. What's not there: The odd-lot bias in TAQ data. *Journal of Finance,* forthcoming.

Pardo, A. and R. Pascual, 2012. On the hidden side of liquidity. *European Journal of Finance* 18, 949-967.

Smith, J. W., 1998. The effects of the order handling rules and 16ths on Nasdaq: a Cross-sectional analysis. *NASD Working Paper* 98-02.

Stoll, H.R. 2000. Presidential address: Friction. *Journal of Finance*, 55, 1479-1514.

van Bommel, J., 2011. Measuring price discovery: The variance ratio, the $R^2$, and the weighted price contribution. *Finance Research Letters* 8, 112-119.

Wang, J. and M. Yang, 2010. How well does the weighted price contribution measure price discovery? *SSRN Working Paper*.

**TABLE I**
**Descriptive statistics**

This table contains descriptive statistics on our sample of 55 stocks traded in the SSE from July 2000 to December 2006. Stocks are split into 4 portfolios based on the market capitalization at the beginning of each year. C1 are the largest stocks and C4 are the smallest stocks. We report the daily median and interquartile range (in parenthesis) for each variable across stocks. "Trades" is the number of trades completed; "Volume" is the total amount transacted in million Euros; "Trade size" is the median trade size in shares; "price" is the trading price (average between maximum and minimum); "Volatility" is the daily quote midpoint standard deviation; "Relative spread" is the bid-ask spread divided by the quote midpoint, and "Depth" is the number of shares displayed at the best ask and bid quotes.

|        | Trades   | Volume (millions €) | Trade Size (shares) | Price  | Volatility | Rel. spread (x100) | Depth (shares) |
|--------|----------|---------------------|---------------------|--------|------------|--------------------|----------------|
| Sample | 397.00   | 6.216               | 6529.2              | 14.86  | 0.057      | 0.200              | 5133.5         |
|        | (150.25) | (4.892)             | (4325.4)            | (3.92) | (0.028)    | (0.140)            | (3073.8)       |
| C1     | 1060.00  | 40.967              | 13634.50            | 16.02  | 0.059      | 0.110              | 11080.5        |
|        | (447.25) | (26.942)            | (4738.0)            | (4.70) | (0.036)    | (0.060)            | (7843.9)       |
| C2     | 395.50   | 7.583               | 6985.7              | 18.64  | 0.075      | 0.190              | 3596.3         |
|        | (184.00) | (5.316)             | (6452.8)            | (7.07) | (0.048)    | (0.130)            | (1788.8)       |
| C3     | 350.00   | 4.717               | 5536.3              | 16.26  | 0.072      | 0.240              | 2091.8         |
|        | (178.00) | (4.677)             | (3765.8)            | (6.75) | (0.045)    | (0.170)            | (3215.5)       |
| C4     | 247.00   | 2.792               | 4572.6              | 10.56  | 0.036      | 0.280              | 5321.3         |
|        | (105.25) | (2.020)             | (2985.1)            | (2.47) | (0.018)    | (0.140)            | (3477.2)       |

## TABLE II
## WPC friction-related bias

This table reports the daily average weighted price contribution (WPC), the daily average friction-free weighted price contribution ($\overline{WPC}$), the proportion of trades (%T), and the proportion of volume in shares (%V) for small-, medium-, and large-size trades in our sample. WPC is computed using trade price changes ($\Delta p$), while $\overline{WPC}$ is computed using the unexpected quote midpoint changes ($\Delta m$), estimated by the residuals of the time series model in eq. [5], with $r = 10$. We consider BW93 trade-size classification: (0 499] small-sized, [500 9,999) medium-size, [10,000 ∞) large-sized. $R$ is the daily (open-to-close) log return.

| Sample | Size category | WPC | $\overline{WPC}$ | %T | %V |
|--------|---------------|-----|-----|-----|-----|
| All days | (0 499] | -1.63 | 37.09 | 58.71 | 15.74 |
| | [500 9,999] | 83.17 | 55.43 | 39.11 | 61.02 |
| | [10,000 ∞) | 18.46 | 7.48 | 2.19 | 23.24 |
| | | | | | |
| $R$>1% | (0 499] | -39.79 | 32.92 | 57.29 | 14.20 |
| | [500 9,999] | 113.88 | 57.10 | 40.32 | 60.85 |
| | [10,000 ∞) | 25.91 | 9.98 | 2.39 | 24.95 |

**TABLE III**
**Friction-related bias: Regression analysis**

This table reports the estimated coefficients of pooled weighted least square regressions. We use two alternative dependent variables: (a) the friction-related bias in the daily price contribution (*PC*), and (b) the absolute value of that bias. We use three trade-size categories: (0 499] small-sized, [500 9,999] medium-sized, and [10,000 ∞) large-sized. For stock (*i*), day (*d*), and trade size category (*s*), we compute $PC_{ids}$ as the sum of the trade price changes divided by the daily cumulative price change. The friction-related bias is the difference between $PC_{ids}$ and the friction-free $PC_{ids}$. The latter is obtained using the unexpected quote midpoint changes given by the residuals of the time series model in eq. [5] ($r = 10$). We only consider days with open-to-close returns larger than 1%. Weights are given by eq. [3]. The regression variables *M* and *L* are dummies for medium-sized and large-sized trades, respectively. The omitted category is small-size trades. The dummy *LCaps* equals 1 when the observation corresponds to a large cap in our sample. Large caps are defined as the stocks in the upper quartile in terms of yearly average market capitalization. Large caps constituents are revised yearly. The model includes dummies for years 2001 to 2006 (2000 is the omitted year), and stock-specific dummies (not reported).

| Coeff. | Bias | |Bias| |
|---|---|---|
| *M* | 0.093 ** | -0.026 ** |
| *L* | 0.066 ** | -0.101 ** |
| *LCaps* | -0.011 | -0.013 * |
| *Y1* | -0.001 | -0.011 ** |
| *Y2* | 0.003 | 0.002 |
| *Y3* | 0.001 | -0.019 ** |
| *Y4* | -0.002 | -0.031 ** |
| *Y5* | -0.002 | -0.036 ** |
| *Y6* | -0.001 | -0.017 ** |
| *Intercept* | -0.048 ** | 0.148 ** |
| *(Stock dummies ommited)* | | |
| *Obs.* | 85740 | |
| *Adj. R$^2$* | 0.046 | 0.128 |

* (**) means p-value < 0.05 (0.01)

## TABLE IV
## Sources of bias: bid-ask bounce vs. serial correlation

This table decomposes the WPC friction-related bias into two orthogonal components: a bid-ask bounce driven component and a bid-ask-bounce-free serial correlation driven component. We use BW93's trade-size classification: (0 499] small-sized, [500 9,999] medium-size, [10,000 $\infty$) large-sized. $R$ is the daily (open-to-close) log return.

| | | | Bias decomposition | |
| | Trade-size | | Bid-ask | Serial |
| Days | category | Bias | bounce | correlation |
| --- | --- | --- | --- | --- |
| All | (0 499] | -38.71 | 0.856 | 0.144 |
| | [500 9999] | 27.73 | 0.862 | 0.138 |
| | [10000 $\infty$) | 10.98 | 0.842 | 0.158 |
| | *mean* | | *0.853* | *0.147* |
| | | | | |
| R>1% | (0 499] | -72.71 | 0.892 | 0.108 |
| | [500 9999] | 56.78 | 0.888 | 0.112 |
| | [10000 $\infty$) | 15.93 | 0.908 | 0.092 |
| | *mean* | | *0.896* | *0.104* |

**TABLE V**
**Trade aggressiveness and trading frictions**

In Panel A, we provide some general statistics about the distribution of aggressive and non-aggressive trades among trade-size categories. An aggressive trade walks up or down the book. A non-aggressive trade does not consume all the available depth at the opposite side of the market. In Panel B, we report pooled regression in eq. [7] on the stock-daily-trade-size proportion of aggressive and non-aggressive trades that: [a] contribute to the WPC; [b] contribute to the WPC but with bid-ask bounce driven price change; [c] are wrong trades, and [d] are wrong trades that contribute with a bid-ask bounce driven price change. A trade is wrong if it causes a positive (negative) price change whilst the daily cumulative price change, the denominator in eq. [1], is negative (positive). We use BW93's trade-size classification: (0 499] small-sized, [500 9,999] medium-size, [10,000 ∞) large-sized. The regression variables *M* and *L* are dummies for medium-size and large-size non-aggressive trades, respectively. *AS*, *AM* and *AL* are dummies for small-size, medium-size and large-size aggressive trades, respectively. The omitted category is small-size non-aggressive trades. The dummy *LCaps* equals 1 when the observation corresponds to a large cap in our sample. Large caps are defined as the stocks in the upper quartile in terms of yearly average market capitalization. Large caps constituents are revised yearly. The model includes dummies for years 2001 to 2006 (2000 is the omitted year), and stock-specific dummies (not reported).

Panel A: Statistics on aggressiveness

|            | Small      | Medium     | Large     | Total      |
|------------|------------|------------|-----------|------------|
| *Aggr.*    | 3,347,577  | 6,953,005  | 1,299,606 | 11,600,188 |
|            | 17.20%     | 35.03%     | 56.83%    |            |
| *Non-aggr.*| 20,289,541 | 18,234,297 | 2,581,269 | 41,105,107 |
|            | 82.80%     | 64.97%     | 43.17%    |            |

Panel B: Pooled regressions

| Coeff.    | [a] Contrib. to WPC | [b] Bid-ask bounce driven | [c] Wrong trades | [d] Wrong + bid-ask bounce driv. |
|-----------|---------|---------|---------|---------|
| *M*       | -0.031 * | -0.012 * | -0.013 * | -0.015 * |
| *L*       | 0.018 *  | 0.028 *  | -0.020 * | -0.010 * |
| *AS*      | 0.031 *  | -0.258 * | 0.128 *  | -0.137 * |
| *AM*      | 0.123 *  | -0.239 * | 0.154 *  | -0.127 * |
| *AL*      | 0.293 *  | -0.127 * | 0.161 *  | -0.070 * |
| *LCaps*   | 0.017 *  | -0.007 * | 0.012 *  | -0.004   |
| *Intercept* | 0.506 * | 0.329 * | 0.098 *  | 0.175 *  |
| *(Year dummies and stock dummies ommited)* | | | Obs. | 385547 |
| Adj. R$^2$ | 0.325  | 0.337   | 0.267   | 0.175   |

*means p-value < 0.01.

## TABLE VI
## Trade aggressiveness and the WPC bias

This table reports the daily average weighted price contribution (WPC), the daily average friction-free weighted price contribution ( $\overline{WPC}$ ), the friction related bias in WPC, the proportion of trades (%T), and the proportion of volume in shares (%V) for small-, medium-, and large-size aggressive and non-aggressive trades in our sample. A trade is aggressive if it consumes at least the available depth (both displayed and non-displayed) at the best quote on the opposite side of the market. We consider BW93's trade-size classification: (0 499] small-sized, [500 9,999] medium-sized, [10,000 ∞) large-sized. $R$ is the daily (open-to-close) log return.

| Sample | Size category | Aggr. | WPC | $\overline{WPC}$ | Bias | %T | %V |
|---|---|---|---|---|---|---|---|
| All days | (0 499] | *No* | -27.72 | 21.37 | -49.09 | 44.95 | 10.57 |
| | | *Yes* | 26.92 | 15.86 | 11.05 | 11.07 | 3.69 |
| | [500 9,999] | *No* | 21.09 | 11.98 | 9.11 | 26.89 | 31.82 |
| | | *Yes* | 61.68 | 43.27 | 18.42 | 13.96 | 26.76 |
| | [10,000 ∞) | *No* | 6.29 | 1.13 | 5.16 | 1.88 | 11.02 |
| | | *Yes* | 11.73 | 6.39 | 5.35 | 1.24 | 16.15 |
| $R$ >1% | (0 499] | *No* | -57.85 | 12.58 | -70.43 | 45.09 | 10.55 |
| | | *Yes* | 32.08 | 22.70 | 9.38 | 11.86 | 4.12 |
| | [500 9,999] | *No* | 36.48 | 5.70 | 30.77 | 25.55 | 31.06 |
| | | *Yes* | 69.03 | 50.55 | 18.48 | 14.38 | 27.52 |
| | [10,000 ∞) | *No* | 7.19 | 0.64 | 6.55 | 1.81 | 10.41 |
| | | *Yes* | 13.08 | 7.82 | 5.26 | 1.32 | 16.33 |

# TABLE VII
## Friction-related bias: regression analysis

This table reports the estimated coefficients of pooled weighted least square regressions. We use two alternative dependent variables: (a) the friction related bias in the daily price contribution (PC), and (b) the absolute value of that bias. We use three trade-size categories: (0 499] small-sized, [500 9,999] medium-sized, and [10,000 ∞) large-sized. We also separate aggressive from non-aggressive trades in each trade-size category. An aggressive trade consumes at least the available depth (both displayed and non-displayed) at the best quote on the opposite side of the market. For each resulting trade size-aggressiveness (*sa*) category, for each stock *i*, and each day *d*, we compute $PC_{idsa}$ as the sum of the trade price changes divided by the daily cumulative price change. The friction related bias is the difference between $PC_{idsa}$ and the friction-free $PC_{idsa}$. The latter is obtained using the unexpected quote midpoint changes given by the residuals of the time series model in eq. [5] ($r = 10$). We only consider days with open-to-close returns larger than 1%. Weights are given by eq. [3]. The regression variables *AS*, *M*, *AM*, *L*, *AL* are dummies for the following categories of trades: small-sized and aggressive, medium-sized and non-aggressive, medium-sized and aggressive, large-sized and non-aggressive, and large-sized and aggressive, respectively. The omitted category is small-size non-aggressive trades. *N* is the log number of daily trades and the dummy *LCaps* equals 1 when the observation corresponds to a large cap in our sample. Large caps are defined as the stocks in the upper quartile in terms of yearly average market capitalization. Large caps constituents are revised yearly. The model includes dummies for years 2001 to 2006 (2000 is the omitted year), and stock-specific dummies (not reported).

| Coeff. | Bias | |Bias| |
|---|---|---|
| *AS* | 0.132 ** | -0.120 ** |
| *M* | 0.153 ** | -0.089 ** |
| *AM* | 0.166 ** | -0.083 ** |
| *L* | 0.133 ** | -0.165 ** |
| *AL* | 0.132 ** | -0.157 ** |
| *LCaps* | -0.011 | -0.013 * |
| *Y1* | -0.001 | -0.011 ** |
| *Y2* | 0.003 | 0.002 |
| *Y3* | 0.001 | -0.020 ** |
| *Y4* | -0.002 | -0.031 ** |
| *Y5* | -0.002 | -0.037 ** |
| *Y6* | -0.001 | -0.017 ** |
| *Intercept* | -0.115 ** | 0.209 ** |
| *(Stock dummies ommited)* | | |
| *Obs.* | 85740 | |
| *Adj. $R^2$* | 0.089 | 0.173 |

\* (\*\*) means p-value $< 0.05$ (0.01)

## TABLE VIII
## The friction-related bias and formal tests of the STH

In Panel A, we report the estimated pooled weighted least square regression in eq. [4]. The dependent variable is either the price contribution (PC) or the friction-free price contribution ($\overline{PC}$) of the trade size category $k$, for stock $s$, and day $d$. We consider three trade size categories: (0 499] small-sized, [500 9,999] medium-size, [10,000 ∞) large-sized. The weights are given by eq. [3]. PIH stands for public information hypothesis and TVH for trading volume hypothesis. Under the PIH (TVH), the null hypothesis is that the three trade-size coefficients equal zero and the PT (PV) coefficient equals one. Wald tests are provided. $R$ is the daily cumulative price change. In Panel B, we extend eq. [4] to consider six trade-size categories: aggressive and non-aggressive small-, medium-, and large-size trades. We focus on stock-days with open-to-close returns (R)>1%.

| Coeff. | | PC | | $\overline{PC}$ | |
|---|---|---|---|---|---|
| | | PIH | TVH | PIH | TVH |
| **Panel A: Ignoring trade aggressivness** | | | | | |
| S | | -1.287 ** | -0.522 ** | -0.069 ** | 0.273 ** |
| M | | 0.452 ** | 0.242 ** | 0.286 ** | 0.277 ** |
| L | | 0.178 ** | -0.118 ** | 0.064 ** | -0.051 ** |
| PT | | 1.656 †† | | 0.719 †† | |
| PV | | | 1.395 †† | | 0.501 †† |
| Adj. R$^2$ | | 0.189 | 0.189 | 0.237 | 0.232 |
| Obs. | | 44059 | 44059 | 43623 | 43623 |
| **Panel B: Controlling for trade aggressivness** | | | | | |
| S | *non-aggr.* | -0.861 ** | -0.760 ** | -0.169 ** | 0.073 ** |
| | *aggr.* | 0.243 ** | 0.252 ** | 0.147 ** | 0.208 ** |
| M | *non-aggr.* | 0.250 ** | -0.028 * | -0.101 ** | -0.074 ** |
| | *aggr.* | 0.639 ** | 0.302 ** | 0.417 ** | 0.386 ** |
| L | *non-aggr.* | 0.072 ** | -0.024 ** | -0.005 ** | -0.037 ** |
| | *aggr.* | 0.131 ** | -0.084 ** | 0.072 ** | 0.005 |
| PT | | 0.539 †† | | 0.639 †† | |
| PV | | | 1.353 †† | | 0.438 †† |
| Adj. R$^2$ | | 0.134 | 0.149 | 0.117 | 0.115 |
| Obs. | | 88132 | 88132 | 87249 | 87249 |

*, ** means statistically different from zero at the 5% (1%) level.

† (††) means statistically different from one at the 5% (1%) level (Wald test)

## Table IX
## Monthly WPC

For small-, medium-, and large-size aggressive and non-aggressive trades in our sample, this table reports the monthly average weighted price contribution (WPC); the monthly average friction-free weighted price contribution ($\overline{WPC}$); the friction related bias in WPC, the proportion of trades (%T), and the proportion of volume in shares (%V). A trade is aggressive if it consumes at least the available depth (both displayed and non-displayed) at the best quote on the opposite side of the market. We consider BW93's trade-size classification: (0 499] small-sized, [500 9,999] medium-sized, [10,000 ∞) large-sized. $R$ is the daily (open-to-close) log return. We report trades for stock-months with $R > 2.5\%$ to increase the likelihood of observing stealth trading.

| Sample | Trade size category | Aggr. | $\overline{WPC}$ | Bias | %T | %V |
|--------|---------------------|-------|-------|---------|-------|-------|
| $R>2.5\%$ | (0 499] | *No* | 30.03 | -358.55 | 43.69 | 7.84 |
| | | *Yes* | 12.27 | 79.67 | 10.16 | 2.87 |
| | [500 9,999] | *No* | 15.76 | 99.48 | 27.98 | 30.03 |
| | | *Yes* | 35.23 | 137.31 | 13.98 | 24.38 |
| | [10,000 ∞) | *No* | 1.71 | 20.96 | 2.59 | 14.71 |
| | | *Yes* | 5.00 | 21.13 | 1.61 | 20.17 |

**TABLE X**
**US markets**

In this table, we provide the weighted price contribution (WPC) analysis for two samples of 150 US common stocks listed in the NYSE in 1995 (Panel A) and 2005 (Panel B), respectively. We use data from the NYSE Monthly TAQ. We consider BW93's trade-size classification: (0 499] small-sized, [500 9,999] medium-sized, [10,000 ∞) large-sized. For each category, we report the friction-related bias in WPC, the proportion of trades (%T), and the proportion of volume in shares (%V). The bias is computed as the difference between the WPC computed using trade-price changes and the friction-free WPC, computed using friction-unrelated changes in the quote midpoint. We report results for two different subsets of trades and quotes: 'N/P/T/Q NBBO' means consolidated data from the NYSE, Pacific Exchange/Arca and NASDAQ; 'NYSE BBO' means NYSE data only. Finally, we report the estimated $\alpha$ (small-size trades), $\beta_1$ (medium-size trades) and $\beta_2$ (large-size trades) coefficients of the eq. [7].

Panel A: 1995 data

| Sample | Size category | N/P/T/Q NBBO | | | | | NYSE BBO | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | WPC | Bias[‡] | Coef. | %T | %V | NYSE | Bias[‡] | Coef. | %T | %V |
| All days | (0 499] | 11.54 | -18.51 | -0.351 ** | 43.43 | 7.16 | 19.18 | -5.38 | -0.069 * | 38.10 | 6.21 |
| | [500 9,999] | 70.16 | 13.71 | 0.295 ** | 51.81 | 55.98 | 63.40 | 5.11 | 0.078 ** | 56.05 | 56.07 |
| | [10,000 ∞) | 18.30 | 4.81 | 0.303 ** | 4.77 | 36.85 | 17.42 | 0.27 | 0.103 ** | 5.84 | 37.73 |
| R>1% | (0 499] | 14.63 | -12.35 | -0.245 ** | 41.59 | 6.50 | 18.21 | -4.80 | -0.024 | 36.51 | 5.64 |
| | [500 9,999] | 66.98 | 7.16 | 0.230 ** | 53.49 | 55.99 | 64.66 | 4.13 | 0.061 ** | 57.45 | 55.95 |
| | [10,000 ∞) | 18.39 | 5.19 | 0.219 ** | 4.93 | 37.50 | 17.13 | 0.66 | 0.039 ** | 6.04 | 38.42 |

Panel B: 2005 data

| Sample | Size category | N/P/T/Q NBBO | | | | | NYSE BBO | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | WPC | Bias[‡] | Coef. | %T | %V | NYSE | Bias[‡] | Coef. | %T | %V |
| All days | (0 499] | 60.23 | -11.08 | -0.166 | 77.98 | 38.96 | 57.65 | -7.77 | -0.094 | 75.63 | 38.31 |
| | [500 9,999] | 35.50 | 9.46 | 0.195 ** | 21.63 | 49.16 | 37.42 | 9.36 | 0.083 ** | 23.84 | 51.20 |
| | [10,000 ∞) | 4.27 | 1.62 | 0.115 ** | 0.39 | 11.88 | 4.93 | -1.59 | 0.032 * | 0.53 | 10.49 |
| R>1% | (0 499] | 48.53 | -16.78 | -0.153 | 76.32 | 35.94 | 32.22 | -12.83 | -0.460 ** | 73.76 | 35.29 |
| | [500 9,999] | 44.66 | 14.96 | 0.492 ** | 23.18 | 50.17 | 57.74 | 15.06 | 0.767 ** | 25.54 | 52.09 |
| | [10,000 ∞) | 6.81 | 1.82 | 0.310 ** | 0.50 | 13.89 | 10.04 | -2.23 | 0.437 ** | 0.70 | 12.62 |

‡ Using Lee and Ready (1991)

* (**) means p-value < 0.05 (0.01)