Choice of Order Size and Price Discovery: The Last Digit Puzzle

René Wells

First version: March 15, 2016

This version: January 4, 2017

(Preliminary; Comments are appreciated and will be acknowledged.)

Luxembourg School of Finance, University of Luxembourg

Choice of Order Size and Price Discovery: The Last Digit Puzzle[1]

René Wells

Luxembourg School of Finance[2], University of Luxembourg

[1]The email of the author is rene.wells@uni.lu. Comments and suggestions are appreciated and will be acknowledged. I thank the Fonds National de la Recherche du Luxembourg for its financial grant. I am grateful for the support and guidance provided by Jos Van Bommel, Rajnish Mehra, and Gautam Tripathi. I am also grateful for the comments and suggestions provided by Yoav Benjamini, Kalle Rinne, Robert Algrem, Andreas Chouliaras and the seminar participants at the Luxembourg School of Finance.

[2]Luxembourg School of Finance, 4, rue Albert Borschette, L-1246 Luxembourg

Choice of Order Size and Price Discovery: The Last Digit Puzzle

Abstract

I claim that uninformed traders prefer ending the size of their orders with a zero (e.g. 11<u>0</u> shares) but it is not the case for informed traders, creating an information channel and providing a signal. I propose the Last Digit Hypothesis (LDH): i) some traders exhibit a last digit preference for the digit 0 and other traders do not while ii) the latter are better able to trade on information than the former. The LDH predicts that a trade arising from a marketable order with a size ending with a 0 on average contributes less to price discovery than other trades. My empirical findings support the LDH. However, the LDH is not an equilibrium since informed traders have an incentive to mimic the preferences of uninformed traders to avoid detection and face little constraints or costs to do so. It is puzzling that I find no evidence of such mimicking.

Keywords: Price discovery, digit preferences, order size, informed trading, market microstructure

JEL Classification: G02, G10; G14; G15; G18

1 The Last Digit Hypothesis

Modern financial markets seem to cater to the preferences of investors and regulators by simultaneously providing transparency and anonymity, often in an à-la-carte manner. This said, regulators seem to favor markets with a high degree of pre and post-trade transparency while at the minimum allowing traders the choice to trade anonymously (or not). This gives rise to a curious but challenging situation whereby practitioners and academics alike have vast quantities of pre and post-trade data at their disposal, in real time or on an archived basis, but for which the identities of market participants are fully concealed. The optimality of transparency and anonymity of trading data in term of market design has been researched extensively. My purpose here is mundane by comparison, I am trying to breach the veil of anonymity, at least partially, but in the dimension that matter most.

For investors, trading securities is concerned with decisions regarding prices, quantities, and time required to implement some portfolio decision. When trading securities through an electronic limit order book, orders are submitted with a size, e.g. number of shares expressed as a positive integer. I study the extent to which the decision pertaining to such quantity creates an information channel regarding the characteristics of the trader and the resources used to arrive at such decision and if these characteristics have consequences of interest that can be empirically detected. Of particular interest is to investigate what can be inferred about the trader's information set and skill from such characteristics in an attempt to overcome limitations imposed by the anonymity of trading data commonly available.

More specifically, I investigate if there is variation in the informativeness of the last digit of the size of marketable orders in limit order books and for that purpose I formulate and empirically test a hypothesis that I call the Last Digit Hypothesis (LDH). This hypothesis arises as securities are being bought and sold, whereby i) two types of traders coexist: a first type exhibits a preference for specific digits when submitting orders while a second type exhibit no such preference and ii) the second type is better able than the first type to trade on information (e.g. selling when the market price is higher than the fundamental price) because of a decision process with a higher resource endowment (e.g. skills, information, time, attention, automation). Traders of the first type are hereby referred to as 'uninformed traders' and traders of the second type as 'informed traders'. A general setting for the LDH assumes the existence of a continuous spectrum of traders characterized by the strength of their preference for certain digits and their ability to trade on information as a result of the higher resource intensity afforded to do so, while these two characteristics are negatively correlated, along with a certain mass of traders who exhibit no digit preferences and an above-average ability to trade on information. In practical terms, the degree of uninformativeness displayed by the uninformed traders with regard to the degree of informativeness that can be inferred of the decisions taken by the informed traders is relative rather than absolute.

In the context of traders choosing the last digit of the quantity (size[3]) of their marketable orders[4] when trading securities through an electronic limit order book during continuous trading, a testable prediction of the LDH is that digit preferences of uninformed traders will map into the lack of informativeness of their trades as revealed by the last digit of the size of such trades. Per the findings provided by the relevant empirical literature (reviewed in the next section), I expect the strongest last digit preference of uninformed traders to be for the digit 0. So, the LDH predicts uninformed traders are predominantly choosing to end the size of their marketable orders with the digit zero (e.g. a market order to buy 10<u>0</u> shares), ensuing that trades with a size ending with a zero are on average less informative and contribute less to price discovery than trades arising from other marketable orders.

What has just been exposed in two short paragraphs could appear quite intuitive to the point of being trivial to some or, a contrario, be quite counter-intuitive to others. Why the LDH might not be so trivial after all will be addressed later in this section. For the purpose of clarity, I illustrate graphically in Figure 1 what interplay is foreseen between last digit preferences and information, while assuming, for simplicity, a one--to--one mapping between marketable order sizes and trade sizes — more on that later. If all traders are immune to any form of last digit preference, it would result in an equal proportion of trades for each last digit of trade size, and also to a uniform contribution to price discovery by last digit of trade size, as illustrated in Panel A of Figure 1. In such case, the last digit of size is uninformative (i.e. knowing the last digit of size does not help predict the contribution of the trade to price discovery or its likely permanent price impact relative to other trades). However, consistent with empirical findings across disciplines, Panels B and C depict a more realistic construct: traders 'a' are immune to any digit preference, traders 'b' prefer even digits but not odd digits, traders 'c' prefer digits 0 and 5 equally but not the other digits while traders 'd' only use zeros for the last digit of their orders' size. For illustration purposes, I halve the mass of each trader type successively from 'a' to 'd', but it is nevertheless apparent that the aggregate trader mass at zero is significantly higher than for any other digit. In Panel B, as a result of traders being otherwise homogeneous, having the same information set or being randomly informed, the density of price contribution is uniform by last digit ensuring that each individual trade is still expected to contribute the same to price discovery irrespective of

[3]Size of an order (or a trade) refers to the number of shares comprising the order (or the trade) in the context of cash equity markets.

[4]Marketable orders are orders that immediately trigger a trade upon submission to the limit order book (i.e. demanding liquidity) as opposed to resting orders which rest in the limit order book and becomes part to a trade only when matched with a marketable order (i.e. providing liquidity). Examples of marketable orders are market orders and limit orders which price limit is the same or exceed the other side of the order book.

its last digit of size. For this case, the contribution to price discovery is now non-uniform by last digit category (i.e. last digit 0 will contribute more than last digit 1 since trades ending with a 0 are more numerous than trades ending with a 1). Yet, if one weights the contribution to price discovery by last digit according to the proportion of trades by last digit, the relative price contribution by last digit converges to a constant ex-post (conditional to the number of observations being large enough). Panel C depicts the most interesting case: if, as per the LDH hypothesis, the information content of the marketable orders by trader type is negatively correlated with the strength of last digit preferences, the information density of the trades originated by traders of type 'a' will be high but conversely low for traders of type 'd'. It results in end-digit 0 trades to be, on average, less informative than other trades as illustrated. The empirical literature suggests the number of traders of types 'b' to be very small and 'c' to be small, when compared to the number of traders of types 'a' and 'd'.

The LDH in its simplest form results from the aggregation of two distinct data generation processes. In the first process, the size of the marketable order is not subjected to a specific constraint nor results from an optimization attempt; for example, the trading intent is triggered by a near-arbitrary portfolio decision for which simplifying heuristics play a central role (see the portfolio decision described in the introduction of KAHNEMAN (2011)). For such process, the trade size is quasi random, but the decision maker nevertheless choose to predominantly end the order size with a zero using such simplifying heuristic in the context of an overall complicated decision despite its lack of explicit structure and being resource deprived. So, while the end digit of the trade size is non-random, it nevertheless originates from a decision process devoid of much structured information processing and therefore likely to be uninformed and its output uninformative. The second process is largely the mirror image of the first one: the size of the marketable order is the precise output of an optimization or a decision rule but results in the last digit of size to be random, assuming the average trade size exceeds one digit, which is typically the case for stock markets. In the first case, noise is filtered by a heuristic giving it a deterministic component that can be detected while in the second case the randomness of the last digit is indicative of information having been processed in a structured manner and provided with sufficient resources. Detection is made possible because uninformed traders predominantly use the same heuristic (end trade size with a zero, a single digit amongst ten) while at the same time informed traders predominantly distribute the last digit of their marketable orders at random among ten digits.

To investigate and test the LDH, I use comprehensive stock trading datasets comprising in excess of 90% of all intra-day continuous trading activity having taken place on Euronext Paris, Deutsche Börse Xetra and the London Stock Exchange from 2002 to 2015. The trade data is aggregated carefully to proxy marketable orders with regard to trade size and price impact. The resulting number of trades (or number of shares traded) arising out of marketable orders is compiled by stock-day. The distribution of trades according to the last digit of trade size validates the prior with regard both the number of trades and the number of shares traded (volume) for all three markets and all years (i.e. a net prevalence of digit zero, followed by a small overweight of digit 5 over the other digits, and a very small prevalence of even digits over the odd digits among the remaining eight digits). The only non-negligible deviation is the proportion of volume for the end-digit 7 on the UK market in the pre MIFID1 period (some traders in stocks of large firms appear to seek luck by ending their larger than average marketable orders with the digit 7). It illustrates that preferences can occur simultaneously and superimpose their effects. I also find minimal cross-sectional variation in trade size per last digit. It rules out clustering of small marketable orders with a last digit of size zero and negates confounding effects with trade size stealth trading. Then, using the Weighted Price Contribution, an often-used price contribution estimator since Barclay and Warner (1993), I find that a trade (or a traded share) arising from a marketable order with a last digit 0 size is on average consistently less informative than its non-digit 0 counterparts for each of the three markets (whole samples) and for each year in the period for each market (annual sub-samples).

In a manner similar to the stream of literature initiated by Barclay and Warner (1993), I test the LDH against two classic alternates, the Public Information Hypothesis (PIH) and the Trading Volume hypothesis (TVH) and I find considerable empirical support in favor of the LDH. Succinctly (more later), the PIH (TVH) expects price discovery to map into the number of trades (volume) - it corresponds to Panel B of Figure 1 if last digit preferences affect traders and to Panel A if not. The LDH is never found to be rejected in favor of the PIH or the TVH either when using whole samples, or annual sub-samples or sub-samples by tercile of firm size (per trading volume). This said, when using annual sub-samples, noise is seemingly an issue with the London Stock Exchange dataset, leading to both the LDH and the alternates to be rejected simultaneously in in 18 years out of 28 at the 5% level. The findings are more definite for the two other markets whereby both the LDH and the PIH are simultaneously rejected in 23 years out of 28 at the 1% level but only in 2 years out of 28 for the LDH and the TVH. Said otherwise, in term of informativeness, volume (number of shares) seems to matter while it is not much the case for the number of trades.

In addition, I find that the prevalence of end-digit 0 marketable orders has been steadily decreasing over the time frame, but is still twice as prevalent as any other last digit by the end of the period. I interpret this finding as indicative of a relatively slow technology diffusion process. It appears as if the automation of marketable orders is increasingly prevalent, but at a slower and decreasing pace, to the point that human traders appear to be here to stay. Finally, I test the LDH in the cross-section of firms using samples by terciles of trading volumes in monetary units (Euros or Pounds). The findings are the same across the three markets: the LDH is in evidence in each tercile but its magnitude increases markedly as the firm's trading volume increases in the cross-section. It indicates that the empirical findings in addition to being statistically significant are also economically meaningful (since the largest magnitude of the effect is found for the tercile of the largest firms).

According to the LDH, the non-uniform distribution of trading quantities by their last digit results from traders' digit preferences. However, in the spirit of KYLE (1985), the LDH is not an equilibrium and should be rejected in the data to the extent that informed traders chose to respond strategically to the digit preferences of the uninformed traders and mimic such preferences (in order to make their trading patterns less detectable by other traders). Informed traders have an incentive in that regard as per the stealth trading literature à la BARCLAY and WARNER (1993) inspired by KYLE (1985). They are facing very little constraints to do so, if any, since there is no incremental cost of note to be incurred by an informed trader in mimicking the last digit preferences of uninformed traders. However, I find no evidence in the data to suggest that informed traders indeed mimic the digit preferences of the uninformed traders. On the contrary, if informed traders were to mimic the digit preferences of uninformed traders, one should expect the price informativeness of a trade (or a share traded) to be very similar (uniform) with respect to the last digit of size of its marketable order, which is not what I find at all. In an extreme case, if all uninformed traders end the size of their marketable orders with a zero and informed traders mimic such preference, the LDH would always be rejected in favor of the PIH (or TVH). However, by finding the opposite, that variation of price discovery by last digit of size is systemic across markets, time and trading volume in the cross-section of firms (even once adjusted for number of trades or volume), one shall conclude that little mimicking by informed traders is taking place, if any. Furthermore, for one market over the whole period and for the two others in the post MIFID1 period, price discovery is quite uniform by last digit (around 10% for each digit) irrespective to the variation of the number of trades or volume by last digit, exactly as if informed traders were paying no attention whatsoever to the last digit preferences of uninformed traders.

My empirical findings indicate conclusively the existence of a separating equilibrium for which uninformed traders prefer using marketable orders with size ending with a 0 while informed traders have no such preference. In addition, my empirical setting, contrary to the stealth trading literature initiated by BARCLAY and WARNER (1993), does not requires a prior regarding how to categorize trades (i.e. the last digit of size give rise naturally to ten distinct categories while the extant stealth trading literature requires the choice of size thresholds in order to determine trade size categories). Furthermore, the strategic decision to mimic the last digit 0 default decision of the uninformed traders is trivial in costs and consequences for the informed trader in comparison to a strategic decision to stealth trade with regard to size (e.g. the smaller the trade size the larger the execution costs and the longer the time required to establish a given position). There are several candidate explanations for the existence of the said separating equilibrium. First, the informed traders might be unaware of the last digit preferences of the uninformed and the signal therefore created. In fact, when the information content of the last digit of trades' size is estimated using unaggregated data, it yields results mostly inconclusive since the last digit of trade size is a randomized version of the last digit of marketable order size given that a single marketable order is often matched with more than one resting limit order. Second, any benefit derived by the informed trader from mimicking only arises to the extent that markets makers themselves extract and make use of the signal accordingly. Assuming the markets makers are knowledgeable of the existence of the signal, the signal nevertheless need to be extracted in real time by aggregation on time stamps or preferably extracted from the low level information available in a direct data feed. The cost of doing so in term of computing resources and processing time might exceed the benefits for the market makers, notably because of the existence of other signals having a better cost-benefit profile or the role played by more strategic dimensions like speed drive prioritization of resources in a way that market makers ignore the signal and therefore deprive the informed traders of any benefits out of mimicking the last digit preferences of uninformed traders. Third, a dominant strategy used by informed traders might supersede any need to mimic the last digit preferences of the uninformed traders (e.g. using aggressive kill-or-fill limit orders to capture attractive liquidity as it arises in the limit order book in an ADMATI and PFLEIDERER (1988)–inspired manner). And finally, fourth, hiding from the maker makers might be a second-degree concern with regard to hiding from the 'sniffing algorithms', the purpose of the latter being to back-run the informed meta order once detected. As reported in DOMOWITZ and YEGERMAN (2005, p. 2), lack of randomization results in higher transactions costs for algorithmic trading, presumably from front-running. Therefore, the increasingly deterministic nature of the trade execution processes employed by the informed traders call for some randomization, while the said mimicking would run counter to that.

I comprehensively investigate, for the first time to my knowledge, the information content of the last digit of marketable orders' size and its recent evolution over 14 years in the three major European equity markets. This allows for multiple contributions to the field to be achieved in a comprehensive manner. First, by considering not only theoretical and empirical cognitive physiology, but also findings across several disciplines like demographics, statistics and health sciences, I predict and empirically confirm the distribution characteristics of the choice by traders of the last digit of the size of their marketable orders. Second, I propose, test and validate an hypothesis that links what is known in those fields about the strength of last digit preferences and the likely information set used by traders and their skills and resources in relation to the strength of their last digit preferences. The insights and the findings allow for an understanding that is deeper than, for example, what was proposed regarding 'clustering' like the resolution hypothesis of BALL, TOROUS and TSCHOEGL (1985), the pure attraction hypothesis of GOODHART and CURCIO (1991), the negotiation hypothesis of HARRIS (1991): a non-uniform distribution by last digit arise out of last digit preferences, not out a purposeful decision or a vague attraction to round numbers — at least in modern-age equity trading. Third, a new information channel is therefore identified which provides a trading signal regarding the informativeness of marketable orders. For practitioners, it could, for example, be used to assign a lower (higher) probability to have to reposition the price of an existing limit order subsequent to the arrival of a marketable order with a last digit size of 0 (other than 0). Fourth, the empirical analysis indicates that most if not the vast majority of informed traders do not use the signal to stealth trade while facing apparently no obvious impediment or costs to do so. It suggests that limits to rationality exists, that rationality requires to be empirically tested rather than assumed, that rationality might come in degree and that traders might choose some dimensions to be optimized while other dimensions are either ignored, are chosen not to be optimized or cannot be optimized because of some mutual exclusion. Fifth, the empirical findings also indicates that the speed at which new trading technology is adopted by market participants and how fast the progress toward more structured decision processes is unrelentingly happening but at a rather slow pace for the market as a whole. The final contribution is to confirm the findings of UPSON, JOHNSON and MCINISH (2015) to the effect that theoretical predictions and hypotheses regarding orders shall be empirically testing using orders rather than trades and that can be achieved for marketable orders by aggregating trades having the same timestamp.

The paper is organized as follows. Section 2 explains last digit preferences, Section 3 reviews the literature. Section 4 describes the data. Section 5 presents the methodology used to conduct the empirical analysis. Section 6 presents and discusses the empirical findings. Section 7 presents the robustness checks. Section 8 concludes.

2 Last Digit Preferences

In this section, leveraging existing literature, I outline what is meant by a digit preference, why it is expected to occur, why some traders are expected to be affected more than others, and why it is expected to be negatively correlated with the performance of traders.

Much has been written about human nature and numbers, and MITCHELL (2001, pp. 403-417) presents a practical review and an insightful discussion. Digit preferences could originate from cognitive processes, cognitive limitations, habits, cultural norms, etc. It can be conjectured that the strength of such preferences is: (i) indicative of type 1 decision making (i.e. impulsive or intuitive) rather than type 2 (i.e. reflective and deliberate) — see KAHNEMAN (2011) and EVANS and STANOVICH (2013); and/or (ii) indicative of low decision-making abilities and lack of economic rationality — see CHOI et al. (2014). For the purposes of this research, the exact origin or explanation of last digit preferences is of less relevance than being able to develop and characterize a prior of its manifestation for the last digit of a positive integer number such as the number of shares when submitting an order.

Last digit preferences have been empirically observed in numerous but distinct contexts, notably in circumstances involving no economic decisions as well as no economic outcomes such as gains or losses. The phenomenon is reported as heaping, end-digit preference, or terminal digit preference. It has been shown to affect laboratory work (e.g. OWEN (1968)), age reporting in demographic data (e.g. MYERS (1940, pp. 402-409), SIEGEL, SWANSON and SHRYOCK (2004, pp. 136-141), CAMARDA, EILERS and GAMPE (2008, p. 387)), various health statistics like self-reported time-to-pregnancy (e.g. RIDOUT and MORGAN (1991)), self-reported weight and height (e.g. NIEDHAMMER et al. (2000, p. 1113)), weight measurements (e.g. ROSENBAUM (1954, p. 332)), but also blood pressure measurements taken by trained professionals (e.g. HESSEL (1986), NARGESI et al. (2014)). These preferences manifest themselves as an excessive use of integer multiples of the base and convenient sub-units (e.g. five and even digits would be convenient sub-units for base 10). While most empirical studies are concerned with contexts involving base 10, empirical findings of the same nature have also been found for bases 8, 12, 14, and 20. Smokers over-report 20, 10, and 0 since cigarette packs sold in the US contains 20 cigarettes (e.g. WANG and HEITJAN (2008)). In ROSENBAUM (1954, p. 332), weight is over-reported in multiples of 14 pounds (an English stone is 14 pounds). Time to conception is over-reported in 12 and 6 month-cycles (e.g. RIDOUT and MORGAN (1991)). Unemployment spells is over reported in 1, 2, 3, 6, 12, and 24 months even if the respondent is asked for and did provide an answer in weeks (e.g. SIDER (1985, p. 464), BAKER (1992, p. 118)). Stock prices in eight could be deemed an example of base 8 (e.g. HARRIS (1991)). While the graduation of a measurement device or other extraneous influences can lead the observer to a given preference, empirical studies for which such external influences have been removed still find a preference to manifest itself, although at a lower magnitude (e.g. NARGESI et al. (2014)). Therefore, if traders are affected by last digit preferences when trading securities, it should result in a non-uniform distribution of the last digit of trade sizes or trade prices. Empirical market microstructure research concludes this being indeed the case for quantities (e.g. ALEXANDER and PETERSON (2007), O'HARA, YAO and YE (2014, p. 2212)) as well as for prices (e.g. HARRIS (1991) prior to decimalization and IKENBERRY and WESTON (2007) after decimalization).

The above empirical observations are best explained in ROSCH (1975) who theorizes that the human brain uses 'reference points' when processing data like colors, lines and numbers. In a number system, the obvious candidate for reference point is the base of the system. Other theories by psychologists or neurologists might also explain all or some of these empirical findings.

When trading securities is performed using a base 10 for both prices and quantities as it is now most often the case, my prior, the expected digital signature of the last digit preferences per above empirical and theoretical literature, is a net prevalence of digit 0, followed by a small overweight of digit 5 over the other digits, and a very small prevalence of even digits over the odd digits among the remaining eight digits. Since this prior is indeed confirmed in my trading data, then other characteristics of last digit preferences could be assumed to be carried over. It is therefore of interest to ascertain if these characteristics further rationalize the LDH or not.

The first such characteristic is heterogeneity as it has been documented that last digit preferences are not affecting individuals in a uniform manner. OWEN (1968) observes two laboratory technicians with a strong preference for 'values of 130 and 140' while another has some preference for even values but not specifically for 130 or 140. HESSEL (1986, p. 124) reports 'wide variability' of prevalence of digit preferences among his sample of 12 examining doctors. KUO, LIN and ZHAO (2015, p. 857) identifies 'a large cross-sectional heterogeneity in the submission ratio at round-number prices' for both individual investors and institutional investors.

The second characteristic of interest is that the above cross-sectional variation is not induced by randomness but reflects individual characteristics. MYERS (1940, p. 404) finds that U.S. census data is the most (least) affected by terminal digit preferences in states having the highest (lowest) degree of illiteracy, documenting a negative association between such preferences and literacy. KLESGES, DEBON and RAY (1995, p. 1230) confirms this association by estimating that a 5% decrease in digit preferences results from each additional 4-year increment in education.

The third characteristic is that individuals have persistent last digit preferences. According to KUO, LIN and ZHAO (2015, p. 857), digit preferences are persistent one year to the next for both individual investors and institutional investors and HESSEL (1986) reports 11 examining doctors preferring over time the end-digit 0 while a twelfth prefers 2.

A fourth characteristic is that digit preferences are context, process and technology sensitive: MYERS (1940, p. 404) finds that asking at the same time age and date of birth (rather than age only) reduces the occurrence of digit preferences affecting age; NARGESI et al. (2014) shows that partial automation of blood pressure measurement decreases the prevalence of digit preferences but does not eliminate it while according to OSTCHEGA et al. (2003) training, protocol standardization and certification can almost eliminate the preference for last digit 0 for blood pressure measurement, but also potentially lead some examiners to avoid zero and thus displace the last digit preference to another digit.

The fifth and last characteristic is that several studies find a decrease of the prevalence of digit preferences as time goes by (e.g. MYERS (1940, p. 403), ALSANJARI et al. (2012, p. 39)). It could result from an increase in educational levels over time. In addition, since digit preferences are inducing measurement errors hindering the end-purpose of data collection, the sponsors of such data collection efforts might try reducing its manifestation.

This brief characterization of last-digit preferences suggests some plausible explanations why informed traders could be expected be less affected or not at all, therefore providing further rationale in support of using last digit preferences as hereby considered.

That informed traders are expected to exhibit no last digit preference or be less affected could be explained by individual characteristics such as higher rationality in terms of decision-making, higher IQ or advanced educational achievement. It could be the case as a by-product of specific training, of structured institutional environment in term of decision-making process or control. Besides, informed traders have to recover the costs incurred in the acquisition of information and/or in the processing of information, and for doing so are likely to use resources and processes that potentially have the by-product effect of removing partially or completely end-digit preferences from their trading. For example, by splitting meta-orders using computerized algorithms to time distribute trading irrespective of the last digit of the size of the actual individual orders, the 'human element' is partially or completely removed from trading. Furthermore, when departing from classical active management, such investment styles as indexing, factor investing, quantitative investing, smart-beta, ETFs or high frequency trading are all likely to use a higher degree of automation of portfolio rebalancing decisions calling for a commensurate and concomitant automation at the trade implementation level. In these environments, the 'human element' is increasingly being far removed from the choice of the last digit of trading quantities at the individual order level, as suggested by EASLEY, DE PRADO and O'HARA (2016) that the non-uniform distribution of the last digit of trade sizes results from the interplay of 'human traders' and 'silicon traders'.

3 Review of Literature

In the context of capital markets, à la HAYEK (1945), some agents specialize in "... performing eminently useful functions based on special knowledge of circumstances of the fleeting moment not known to others." and two such functions are the provision of liquidity (i.e. market makers asynchronously matching the synchronously mismatched supply and demand of securities), and price discovery (i.e. informed investors arbitraging away significant deviations of the posted and transacted price from an unobservable fundamental price that incorporates any and all information). Investors welcome being able to trade securities at low cost reasonable quantities at fair prices. While liquidity and price discovery are expected to arise endogenously in well-functioning financial markets, their respective dynamics are still not well known. The LDH is of interest in the context of assuming that well-functioning capital markets require an "equilibrium level of disequilibrium", notably in the modeling of agents specializing in the provision of liquidity and/or price discovery who need to earn on average an economic return commensurate with the costs incurred, the capital employed and the risks borne in their endeavors. Testing the LDH is an indirect test of the extent to which the informed traders who are assumed to act in a fully rational manner indeed are indeed acting in such an idealized way. So, I briefly review in this section how such rationality assumption for strategic behavior arises in the theoretical work and how the extant empirical literature of interest went to test it. Of note, in line with the latter, the contribution to price discovery is hereby focused on a given characteristic of marketable orders over a certain macroscopic time scale, and not on specific orders at very fine time scales.

Eugene Fama (and co-authors), notably in FAMA (1970), developed the Efficient Market Hypothesis (i.e. financial markets are on average quite efficient in incorporating quickly and accurately the release of public information into securities‘ prices, and in its strong form that markets are able to do so even for private information). However, this strand of literature is not providing much in term of micro-foundations in support of its empirical evidence and conclusions. This was remedied notably by GROSSMAN and STIGLITZ (1980) and HELLWIG (1980) who introduced 'noisy rational expectations equilibrium' (NREE) models to resolve the conundrum that the fully-revealing nature of a rational expectations equilibrium would deprive agents from any incentive to spend efforts and incur costs to acquire and process information since the output of such activity would be observable at no cost to other agents in such equilibrium. This is known as the Grossmann-Stiglitz paradox. Such NREE class of models assume that the required noise in the price system arise endogenously as a result of the uninformed agents trading for whatever reason to the exception of having relevant information to trade on, basically assuming that such uninformed agents trade on noise (i.e. irrelevant data) as 'noise traders'. Such progress could also be deemed to be problematic since if a rational expectations equilibrium has noise as required component (i.e. the equilibrium degree of disequilibrium), one shall assume that markets participants are aware of the role of such noise in the price system and that such agents have incentives to act strategically and will do so rather than remain simple price takers. This was accounted for thereafter, notably, in KYLE (1985) (i.e. in order to maximize profits from monetizing their information, informed traders chose to split trading volume in smaller trades which are more similar to uninformed trades therefore more difficult to detect), GLOSTEN and MILGROM (1985), and EASLEY and O'HARA (1987), among others. Various empirical evidence supports such strategic trading models and informed traders are sensitive to overall market conditions, specifics, and dynamics.

Various empirical evidence supports strategic trading models and informed traders are sensitive to overall market conditions, specifics, and dynamics. I focus here on one specific strand of empirical research. For the purpose of investigating the predictions of KYLE (1985) from an empirical perspective, BARCLAY and WARNER (1993) formulates three competing hypotheses regarding the dynamics by which information is impounded in the price of traded securities.

The Public Information Hypothesis (PIH) suggests that stock price movements are due to the arrival of public information, implying that any characteristic of a given trade as determined by its corresponding marketable order is irrelevant to its information content. Assuming that the strong EMH holds and that the market is frictionless, according to the PIH the bid and ask prices already contain all information prior to the arrival of any marketable order. Therefore, a trade only reflects the price information already contained in the limit-order book at a specific point in time. Consequently, the likelihood that a price change occurs on a trade of a given category (e.g. size, or last digit of size) increases in the relative frequency of that category. The testable implication of the PIH is that the price contribution of a trade category is proportional to the percentage of trades in that category. Said otherwise, a trade category accounting for a given proportion of total trades shall account in the same proportion of total price discovery. The PIH can also be interpreted as if the marginal price elasticity ($\triangle p / \triangle q$) is spurious. It mechanically induces a 1-to-1 mapping between the frequency of a given trade characteristic and the informativeness assigned to the said trade characteristic. The PIH does not provide much opportunity for markets participants to trade strategically when using marketable orders.

The Trading Volume Hypothesis (TVH) suggests that stock price movements are proportional to trading volume. Assuming that the strong EMH does not hold (e.g. traders are asymmetrically informed) and/or that the market is not frictionless (e.g. risk-adverse liquidity providers are sensitive to inventory risks), according to the TVH large trades move prices more than small trades, so that the informativeness of a price change on a trade increases in its size. Thus, the likelihood that a price change occurs on a trade of a given category (e.g. size, or last digit of size) increases in the relative trading volume of that category. The testable implication of the TVH is that the price contribution of a trade category is proportional to the percentage of trading volume in that category. Said otherwise, a trade category accounting for a given proportion of total volume traded shall correspondingly account in the same proportion of total price discovery. The TVH can also be interpreted as if the marginal price elasticity ($\mid {\triangle p}/{\triangle q} \mid$) is constant. It results in a 1-to-1 mapping between the trading volume of a given trade characteristic and the informativeness assigned to the said trade characteristic. While the TVH does not rule-out markets participants trading strategically, it does not take it into account past first-level effects.

The Stealth Trading Hypothesis (STH), as postulated by BARCLAY and WARNER (1993) "stock price movements are due mainly to private information revealed through trading", suggests that private information held by informed traders is revealed through their trading but, in order to maximize the monetization of their private information, à la KYLE (1985) they deliberately choose to trade using smaller trades that would otherwise be the case in order to slow down information acquisition by uninformed market participants (rather than using large aggressive trades that would be easy to detect). The STH assumes not only that the strong EMH does not hold and that traders are acting strategically, but also that informed traders are subjected to some market frictions (e.g. trading costs, risk-adverse liquidity providers are sensitive to inventory risks, capital constraints) and/or facing other type of constraints (e.g. they are time constrained as a result of an exogenous time-induced decline of the value of their private information) and/or are competing against each other. Otherwise, informed traders would perfectly mimic uninformed traders to the point of duplicating the PIH or the TVH. Thus, if the STH holds, then large (medium) trades would be less (more) informative than it would otherwise be the case. It is also envisioned that informed traders would also avoid (very) small trades as being not optimal because of additional trading costs and the additional uncertainty resulting from the longer duration required to achieve a certain total trading volume out of small trades. The testable implication of the STH is that the price contribution of a trade category will deviate from the predictions of both the PIH and the TVH. The STH can also be interpreted as if the marginal price elasticity ($\mid {\triangle p}/{\triangle q} \mid$) is not a constant, but adopts a certain functional form, likely concave (e.g. STH predicting that middle size trades are more informative than either large or small trades).

The empirical strategy of BARCLAY and WARNER (1993) is essentially to test the STH against the PIH and TVH as contrafactuals. In nearly 25 years since BARCLAY and WARNER (1993), a number of successor studies have been conducted using different datasets reflecting different market structures across time as well as with some minor variations of the initial weighted price contribution (WPC) methodology. This aggregate empirical work is providing evidence that trade size has an influence over price impact (i.e. information content varies across trade sizes), which has been interpreted as evidence to the effect that informed traders are indeed making a deliberate choice of trade size in order to maximize trading profits by 'stealth trading' (i.e. strategically choosing trade sizes in an attempt of making their trading difficult to detect by other market participants).

Of note, BARCLAY and WARNER (1993) suggest additional considerations as potential motivation for informed traders to fragment their trades and therefore use medium size trades executed over a certain period of time (rather than a large trade at once). In addition to the likely price concession resulting from large trades, an informed trader is confronted to uncertainty about the expected timing and magnitude of the stock price movement that is anticipated from the private information held, in a context of general stock market volatility and the possibility of new information arrival. In addition, trading on some of that information could be deemed to carry legal risk. Confronted to such risks and reasonably assuming that informed traders are risk-adverse, have limited wealth, and are constrained regarding borrowing and short selling, all point out toward informed traders managing their positions carefully and this is easier to achieve using multiple medium-size trades than with a few very large trades.

The STH and the LDH are somewhat similar, in the sense that informed traders acting strategically should be sensitive to whatever information channel that could 'give them away'. The STH posits that informed traders are taking into consideration the trade size choices of other traders and adjust their own trade size choices accordingly given the constraints they face. The LDH posits that uninformed traders reveal themselves by choosing to terminate their trade sizes with a specific digit (0) rather than a random digit. If informed traders do not mimic such last-digit choice by uninformed traders, they then also reveal themselves as informed traders from a statistical perspective. It provides ground to adopt the same approach in using the PIH and TVH as alternative hypotheses to the LDH and to implement empirical and econometric approaches that are now well established in this stream of literature with the departure that trades once aggregated are used to proxy marketable orders.

This paper is connected to and has been inspired by several streams of empirical literature, like trade-size and price clustering, stealth trading, trade-size price impact, price discovery, cognitive limitations of traders, etc. The closest relative is KUO, LIN and ZHAO (2015), which shows a negative correlation between a preference for round-number prices when submitting limit orders and investment performance. They hypothesize that such preference is indicative of cognitive limitations explaining the lower performance since cognitive ability is positively correlated with investment performance. These authors use a rich dataset by the virtue of being anonymized but not anonymous, allowing them to use investor identifiers to analyze trading behavior and investment performance at the individual level over time. While obtained using anonymous data and focusing on marketable orders rather than limit orders, I interpret my findings to be coherent with and complement those of KUO, LIN and ZHAO (2015). A second related paper is BHATTACHARYA et al. (2014), which documents a negative correlation between a choice of limit order price driven by superstition and investment performance by Taiwanese investors. I also observe that a choice of last digit driven by superstition (7 on the London Stock Market) is uninformative. I also interpret such findings to be coherent with and complement those of BHATTACHARYA et al. (2014). Another related paper is ALEXANDER and PETERSON (2007), which provides an analysis of the relation between trade-size clustering and stealth trading and finds a certain degree of sub-optimization by informed traders who use rounded trades (rather than cheaper unrounded trades). This paper differs significantly from ALEXANDER and PETERSON (2007), notably with respect to the data used arising from a different institutional setting in a different era, and the fact that I use no priors other than the existence of ten digits. This said, I also find evidence to the effect that informed traders act in an apparent sub-optimal manner.

4 Data

The equity trading data come from Thomson Reuters Tick History database (TRTH) though SIRCA [http://www.sirca.org.au] and is comprised of firms listed and traded primarily on one of the main European exchanges (Paris, Frankfurt, and London). The markets were chosen and the samples constructed with the aim of being highly representative of trading conducted in equity markets over a long period. Given such aim it was found advantageous to focus on European data rather than U.S. data since i) until December 2013, U.S. equity datasets omit trades of less than 100 shares and are therefore truncated, see O'HARA, YAO and YE (2014, p. 2233), constituting a data limitation and ii) several exchanges still impose a minimum trade unit (MTU) of 100 shares while offering a distinct book for orders below 100 shares and per GOZLUKLU et al. (2015, p. 906) this is the case for both NYSE and Nasdaq. The datasets for Euronext Paris and the London Stock Exchange have the full range of trade sizes and it is the case for Deutsche Börse since August 1st, 2002 (i.e. for the German market trades below 100 shares are absent from the dataset for the first seven months of 2002).

The sample period spans 2002 to 2015 (from January 2nd 2002 through December 31, 2015). On European markets (contrary to American markets), a rather limited number of listed firms are actively traded. Given the methodology used, my unit of observation is the stock day. Therefore, it might appears that such characteristic of European markets is disadvantageous in term of econometric power. However, in the context of Panel data it provides for the datasets to be larger in time series (T) than cross section (N), even within a single year, which is prudent. It also allows tackling three non-overlapping distinct markets rather than one. A long period (14 years) allows for a large number of stock days, which is useful to observe how the phenomenon being researched evolves through time (or not). As we will see later, it is possible for European equity markets to capture in excess of 90% of all trading on a given trading venue with less than 200 firms.

The MIFID1 reform became operational on November 1st, 2007 at which date several major institutional changes became effective for all European markets, notably competition between exchanges and trading venues as well as the possibility for broker-dealers to internalize the trades of their clients (i.e. acting as counterparty to the client's order rather than submitting the order to an exchange). Such internalization practice was permissible pre-MIFID1 for the English and German markets but was prohibited for the French market. It does not strictly create a quasi-natural experiment for the French market with regard to internalization since several other institutional changes simultaneously took place, but from a regulatory perspective, it was a key difference between the French market and the two others that was eliminated by MIFID1. The pre-MIFID1 period spans from January 2nd 2002 to October 31, 2007 (70 months) and the MIFID1 period from November 1^{st}, 2007 to December 31, 2015 (98 months).

While it would have been technically possible to use earlier data (1996 onward), 2002 is chosen as beginning of period to allow some time for the various reforms of the late 1990s to settle down (e.g. trading became anonymous on the Paris Bourse in 2001) and taking into consideration that trading is somewhat thin in the early years as this is still apparent in 2002 when compared to later years.

Each of the samples comprised initially all firms that were included in a given index at any point in time throughout the sample period. Such selection is feasible using a functionality for that purpose provided by TRTH ("chaining"). The CAC40, CAC Next 20, and CAC Mid 100 are the indexes used for the Euronext Paris sample while the DAX and the FTSE are used for Deutsche Börse and London Stock Exchange, respectively.

It is not unusual that the universe of firms listed on a given stock exchange comprises a number of firms with minimal trading and low market capitalization (even if such firms were once in an index). While including all such firms in a dataset could be attractive from an econometric power perspective, it could be at the cost of findings of less economic significance and at the risk of incurring excessive noise. Firms are excluded using the following four steps process:

1. For each firm, stock days that are unavailable or incomplete in such a way that they are unusable are identified as missing;

2. For each firm, stock days for which the average trade price is below a certain threshold are excluded and identified as missing;

3. Since I notably intent to bucket trades according to the last digit of trade size on a daily basis, a minimum number of trades per day is warranted. I therefore set arbitrarily a minimum threshold of 25 trades a day below which the stock day is excluded and identified as missing;

4. Firms for which more than 35% of stock days are missing ("too little trading") are excluded from the samples.

The percentage of firms excluded and the related trading volumes are deemed sufficiently low enough to be economically insignificant as Table 1 indicates. Stock days having been identified as missing in the above steps 1, 2, and 3 are excluded from the samples for the firms having been retained in the samples. The stock days so removed contain a very small proportion of the data. Since about or in excess of 90% of all trading activity on each of the three markets is captured, each sample is deemed as having economically relevant characteristics and dynamics indistinguishable from its corresponding market.

[insert Table 1 here]

Since the data is transaction data, filtering operations are carried out to achieve data integrity of scientific quality. However, data is discarded only as a last resort or when there is evidence that the data is erroneous. Data from the open, close, and intraday auctions (if any) is carefully discarded, as auctions constitute a data generating process different from continuous trading. Data identified as arising outside of the continuous trading process occurring through the order limit book is also excluded (e.g. cross-trades, block trades, trades reported for regulatory purposes). A very small proportion of data consisting apparently of data recording errors is subsequently further removed from the stock days remaining in the samples. Details are available by contacting the author.

It is known that price discovery as well as trading volume and intensity exhibits some significant cross-sectional variability across firm size. Typically large firms, the so-called large caps, have large trading volumes, good liquidity, small spreads and are being followed by many security analysts while it is not the case for small firms. To investigate if the LDH is holding in the cross-section of firm size, sub-samples are created by tercile of trading volume (€ or £). It is apparent from Table 2 that the cross-sectional variation by tercile of trading volume is large.

[insert Table 2 here]

Descriptive statistics for each sample and sub-samples are provided in Tables 3, 4 and 5.

[insert Tables 3, 4 and 5 here]

In an electronic limit order book, a trade results from a match between two orders. In a sense, the trade is the child and the orders the parents. While for some it could be natural to think of a trade of 100 shares as being the result of two orders of 100 shares each, in a like-wise fashion of a negotiated face-to-face bilateral trade, in fact it is unlikely to be the case since such clean-cut scenario rarely occurs in an electronic limit order book. Given the price-time priority afforded to limit orders, marketable orders are sequentially matched to the next in line limit order which is likely to be of a different size than the incoming marketable order (especially if the said limit order has already been previously partially matched with one or several marketable orders). Any partial match in quantity between the marketable order triggering the trade and the limit order providing the liquidity will result in either orders (or both) giving rise to more than one trade. This contributes to several empirical properties of trades (e.g. autocorrelation of trade direction) and creates a wedge between the trading intent of market participants as expressed by their orders and the realization of such intent as observable in trades resulting from the said orders. One key theoretical and empirical issue arising from such limit order book matching process is the extent to which and if so how a trade inherits the information properties of the parent orders or that its information properties are determined otherwise (e.g. simply by its size).

The trade fragmentation issue just referred to does not seem to have been adequately addressed in the literature; its relevance kept vague, but arises frequently. It has been acknowledged in FOUCAULT, PAGANO and RÖELL (2013, p. 174) as follows: “This illustrates a problem that often confronts empirical researchers: the data report trades—not the orders that generate them, but the predictions of the theory refer to orders. So the researcher has to infer which sequences of consecutive orders are likely to stem from a single order." A single marketable order, unless very small, results often in a sequence of consecutive trades. A limit order might also result in a sequence of trades but not necessarily consecutive trades. It seems logical to resolve the issue by correcting the observed trade flow in the direction of the unobservable flow of marketable orders. Such approach is not new - see **?**, , HUANG and STOLL (1997, p. 1019). Therefore, trades are aggregated according to the timestamp at the microsecond level, per UPSON, JOHNSON and MCINISH (2015). In the context of this empirical research, I am proxying marketable orders using trade data and the consolidation of trade data as just explained is deemed to result in a proxy which characteristics, notably information content, are closer to what is being proxyied than it would have been otherwise the case.

5 Methodology

The informativeness of trades by their last digit of trade size is estimated using the weighted price contribution (WPC) methodology as per the strand of literature initiated by BARCLAY and WARNER (1993). The aim of the WPC is to estimate how much the trades of a given category contributes to price discovery (i.e. price change or return of a given stock over a given time period). The methodology follows a two-step process. First, for each stock-day, all trades are coded (binned) by their last digit of trade size, the log return of each trade is calculated, and the returns aggregated per bin as a proportion of the daily log return providing for an estimate of the relative contribution of each digit to price discovery for a given stock-day. For the second step, the cross-section of daily means is calculated (weighted by the relative absolute magnitude of the daily return of each stock in the cross section) and then the daily means are averaged providing for an estimate of the relative contribution of each digit to price discovery for a given sample.

Formally, the price contribution (PC) of the trades of a given category (k) for the log return (r) of a given stock (s) trading N times over a certain time interval (t) is defined as:

$$PC_k^{s,t} = \frac{\sum_{n=1}^{N} r_n^{s,t} \delta_{n,k}}{r^{s,t}}, \quad \forall\, r^{s,t} \neq 0, \quad r^{s,t} = \sum_{n=1}^{N} r_n^{s,t} \tag{1}$$

- $r_n^{s,t}$ is the log return of the n^{th} trade for stock s in time period t (a given day), $t = \{1, ..., T\}$;
- $\delta_{n,k}$ is the indicator function of the trade category ($\delta_{n,k}$ = 1 if trade n is of category k, 0 otherwise), $k = \{1, ..., K\}$;
- $s = \{1, ..., S\}$: a stock universe comprising S stocks;
- by construction $\sum_{k=1}^{K} PC_k^{s,t} = 1$.

The weighted price contribution (WPC_k) for the trades of a given category (k) for the sample is given by:

$$WPC_k = \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{S} w^{s,t} PC_k^{s,t} \tag{2}$$

while $w^{s,t}$ is the cross-sectional weight of a given stock (s) for time interval (t):

$$w^{s,t} = \frac{\mid r^{s,t} \mid}{\sum_{s=1}^{S} \mid r^{s,t} \mid} \tag{3}$$

The hypotheses are tested in a manner consistent with the methodology initiated by BARCLAY and WARNER (1993) and often used in the literature (e.g. O'HARA, YAO and YE (2014)). Per equations 4 and 5, it consists in running weighted least squares regressions with the dependent variable being the stock-day price contributions and the independent variables being digit dummies and the proportion of trades (PT) when testing the LDH against the PIH and the proportion of volume (PV) when testing the LDH against the TVH. The Panels are constructed by matching each $PC_k^{s,t}$ observation with its corresponding $PT_k^{s,t}$ ($PV_t^{s,t}$) value while generating a vector of dummies with 1 for the digit category having given rise to the observation and zero otherwise. The weights used are as per equation 3.

$$PC_k^{s,t} = \boldsymbol{\alpha}_{0,...,9}\boldsymbol{D}_{0,...,9}^{s,t} + \beta PT_k^{s,t} + \varepsilon_i^{s,t},\ k \in (0,...,9) \tag{4}$$

$$PC_k^{s,t} = \boldsymbol{\alpha}_{0,...,9}\boldsymbol{D}_{0,...,9}^{s,t} + \beta PV_k^{s,t} + \varepsilon_i^{s,t},\ k \in (0,...,9) \tag{5}$$

$$D_i^{s,t} = 1 \text{ if } i = k,\ 0 \text{ otherwise, for a given } s \text{ and } t$$

For the PIH (TVH), under the null the coefficient of PT (PV) shall be 1 (i.e. $\beta = 1$) while the coefficients of the digit dummies shall all be zero (i.e. $\alpha_k = 0,\ \forall\, k$). The alternate being vice versa (i.e. $\beta \neq 1$ or $\alpha_k \neq 0,$ for some k). The LDH shall be rejected if one of the α coefficients for the non-0 digits is smaller than the α coefficient for the digit 0, therefore the null is $\alpha_0 \geq \alpha_j$, for some $j \in (1, ..., 9)$ while the alternate is the LDH with $\alpha_0 < \alpha_j,\ \forall\, j \in (1, ..., 9)$. Per BENJAMINI and HELLER (2008), we fail to reject the null according to the highest p-value amongst the p-values of each inequality (if we fail to reject the null, the LDH is rejected).

For robustness purposes, per BIAIS, HILLION and SPATT (1999) as per VAN BOMMEL (2011), the adjusted R^2 of a regression of the total price change on the price change attributable to a given subset of trades is an estimate of the informativeness of the said subset:

$$r^{s,t} = \beta_k r_k^{s,t} + \varepsilon_i, \quad r_k^{s,t} = \sum_{n=1}^{N} r_n^{s,t}\delta_{n,k} \quad for\ k \in (0, ..., 9) \tag{6}$$

- $r_n^{s,t}$ is the log return of the n^{th} trade for stock s in time period t (a given day);
- $\delta_{n,k}$ is the indicator function of the trade category ($\delta_{n,k}$ = 1 if trade n is of category k, 0 otherwise), $k = \{1, ..., K\}$;
- $s = \{1, ..., S\}$: a stock universe comprising S stocks;
- by construction $\sum_{k=1}^{K} r_k^{s,t} = r^{s,t}$.

6 Empirical Results

6.1 Are traders exhibiting last digit preferences?

The empirical evidence provided by my samples indicates that traders exhibit last digit preferences across markets, across time and across firm size (as ranked per their respective trading volume). For each stock day, the proportion of trades according to the last digit of size is calculated and then averaged for all stock days. Proportions of volume (number of shares) are also calculated in the same manner. These statistics are reported in the Panel A of Tables 6, 7, and 8. Trades with a trade size ending with a zero are on average about three to six times more numerous than for any of the other digits individually. It indicates that traders have indeed a strong preference for ending the size of marketable orders with a zero. The distributions of volume by last digit are very similar to the distributions of the number of trades across markets. While the statistical significance of the differences in means of the proportion of trades (or volume) across last digits is not reported, it is apparent from how small the standard deviations are that such differences are highly significant. There is not much cross-digit variation in the average size of trades as indicated by the Relative Size metric in the tables. The only exception is for digit 7 on the London Stock Exchange as if some traders choose that 'lucky' end-digit when submitting larger than average trades for large firms on this market, as per Panel D of Table 8.

[insert Tables 6, 7, and 8 here]

6.2 Are traders' last digit preferences of the common type?

As shown in Table 12, the distribution of the number of trades by last digit is as per the expected order ranking and relative magnitude of end-digit prevalence for each of the three markets. It is also the case for volumes all across. This confirms that the observed deviation from a uniform digit distribution in my datasets is indeed the by-product of the end digit preference phenomenon that has been observed and documented in many distinct contexts as reported in Section 2 above.

[insert Table 12 here]

6.3 Are firms' size and trading volumes influencing traders' last digit preferences?

It is known that the size of the firm in term of its market capitalization has a significant influence over its trading volumes as well as other trading characteristics. There is considerable variation in trading volume across my cross-section sub-samples as the average trade volume increases by a factor between 8 and 30 from the small volume samples to the large volume samples, as shown in Tables 3, 4, and 5. Likewise, trading statistics such as the size of trades, the depth of the limit order book, the number of quote revisions at the top of the book as well as spreads all see considerable cross-sectional variation in each of the three markets. Interestingly enough, trading volumes in the cross section of firms appear to have an almost negligible influence over the digit preferences of traders. It could be seen in Table 12 that the proportion of trades by last digit is almost insensitive to trading volumes in each of the three markets. Meanwhile the distribution of volumes by last digit is getting slightly more uniform as firms' trading volume increases but still remains as uneven or more than the corresponding distribution of trades all across. This evidence indicates that variation of the trading volumes in the cross-section has no material influence over the last digit preferences of traders. This has to be expected if last digit preferences are indeed a persistent characteristic of individuals.

6.4 Is the informativeness of end-digit 0 trades different from other trades?

In my data, the distribution of trades by last digit of trade size differs significantly from a uniform distribution, and it results from traders having last digit preferences as evidenced by i) the distribution by last digit being a match to the distribution having been previously observed in other contexts by numerous studies and ii) variation of the trading volumes in the cross-section having no material influence over the said last digit preferences of traders, confirming the nature of such preferences (e.g. a persistent characteristic of individuals). Thus, our attention turns to the extent to which the distribution of contribution to price discovery by last digit of trade size differs, or not, from the corresponding distribution of trades (volume) by last digit of trade size. The LDH predicts that the informativeness of end-digit 0 trades to be lower on average than other trades and this is what we find in the data across markets (and, as we will see later, across time and across firm size).

For the three samples, as shown in Panel A of Tables 6, 7, and 8, the estimated weighted price contribution (WPC) attributable to trades with a trade size ending with a zero is smaller than its corresponding proportion of trades, while it is the reverse for trades with a trade size not ending with a zero. I test these differences of means for statistical significance, reported in Panel A of Tables 9, 10 and 11, and find all differences significant at the 0.1% level, except for digit 5 of the Euronext Paris sample. These findings indicate that trades with a size ending with a 0 are average less informative than trades not ending with a zero. Regarding volumes, I find all differences of means to be statistically significant at the 0.1% level, which indicates that a share traded with a trade size ending with a 0 is also less informative on average than a share traded with a trade size not ending with a 0. These findings confirm that traders who have no information tend to choose to end their order size with a zero and vice versa for those having information. Since being informed or uninformed is an endogenous choice (i.e. not a random variable), such findings indicate that uninformed traders have a preference for ending marketable orders with the digit 0 while it is not (less) the case for informed traders.

[insert Tables 9, 10, and 11 here]

The deviation from unity of the relative size statistic for a given digit category indicates the extent to which the average trade size of that category deviates from the sample average. Panel A of Tables 6, 7 and 8 indicates that such deviations are quite minor in the data: the average trade size ending with a zero being slightly bigger, smaller when ending with a 5 and quite close to 1 when ending with other digits. Therefore, trade size cross-digit variations can be assumed not playing a significant role since the variations are small when compared to previously reported variations in price discovery which resulted from large trade sizes variations (e.g. one order of magnitude and above). It indicates that the Stealth Trading Hypothesis and the Last Digit Hypothesis are unrelated from an empirical perspective in my samples on an aggregate basis.

The only small deviation of note is the peculiarity of digit 7 for the London Stock Market: its relative size is the largest of the non-zero digits but its contribution to price discovery is lower than the other non-zero digits. As indicated in Panel D of Table 8, it results from traders who on average are submitting fewer but larger marketable orders ending with a 7 when trading large firms while such trades are less informative, if at all. It indicates, in a manner somewhat similar to BHATTACHARYA et al. (2014), that superstitious traders are uninformed.

6.5 Is firm's size (by trading volume) influencing contribution to price discovery by last digit?

It has been shown above that firms' size as inferred by their trading volumes has no influence of note over the traders' last digit preferences. However, it is quite the opposite regarding the contribution to price discovery by last digit. As indicated in Panels B, C, and D of Tables 6, 7, and 8, the estimated weighted price contribution attributable to trades with a trade size ending with a zero is consistently smaller across trade volume terciles than its corresponding proportion of trades or volume, while it is the reverse for trades with a trade size not ending with a zero. This is the case for all three terciles of trading volumes across the markets, and supports the LDH. However, the magnitude of the discrepancy significantly increases as trading volumes increase in the cross-section.

For firms with small trade volumes (see Panels B of Tables 9, 10, and 11), the price discovery contribution attributable of last-digit 0 trades is less than its corresponding trade metric, a difference of 1% to 6%, across markets and slightly larger, between 5 and 12% against volumes (since zero-digit trade size is higher for small firms). For such small firms, these differences of means for the digit zero are statistically significant at the 0.1%, but statistically insignificant in the case of the Deutsche Börse sub-sample for the difference between WPC and trades. For middle-volume firms, per Panels C, these differences increase to between 5 to 17% and 7% to 21% respectively and are all statistically significant at the 0.1% level. Then, as the price contribution for end-digit zero trades becomes surprisingly negative for firms with large trade volumes, these differences further increase to between 35 to 41% and 34% to 45% respectively and are all statistically significant at the 0.1% level. Since the WPC is a signed metric, a negative WPC indicates that the price impact of the trade category was on average in the opposite direction to the market trend. Thus, for such large firms, the price contribution of last digit 0 trades is not only lower than its corresponding proportion of trades or volume, but its average is consistently negative across markets.

The influence of firms' size (as inferred by their trading volumes) over the contribution to price discovery by last digit of trade size when using sub-samples by tercile of trading volume provides further support to the LDH and expands our understanding of such dynamics. The large influence of the variation of trading volumes in the cross-section over contribution to price discovery by last digit of trade size requires further research to be better understood (outside the scope of this paper).

6.6 Are formal hypothesis tests supporting the LDH against the PIH and the TVH?

Having found significant support for the LDH in the data, it is customary in that line of empirical research, as envisioned above, to formally test the LDH against two alternates, the PIH and the TVH. Our focus is on the two leftmost 'All Firms' columns of Tables 14, 15, and 16, which is summarized in the leftmost column of Table 13 at a 5% level of significance.

[insert Table 13 here]

For the Euronext Paris sample, the PIH and the TVH are simultaneously rejected (since $\beta \neq 1$ and $\alpha_k \neq 0$ for all k for both). For the PIH regression, β_{PIH} is close to zero and statistically insignificant leading the α to be very similar to the average WPC provided in Panel A of Table 6. It explains why the LDH is rejected in this PIH regression as a consequence of the null not being rejected (α_0 is larger than some α_j; the largest p-value being 0.981). For the TVH regression, the LDH is supported as the null is rejected (α_0 is found not being larger or equal to any of the α_j; the largest p-value being

0.000). With the exception of β_{PIH}, all t-values are large (in excess of 5) and the p-values of the tests either very small (0.000) or large (0.981).

[insert Table 14 here]

For the Deutsche Börse sample, the PIH is rejected (since $\beta \neq 1$ and $\alpha_k \neq 0$) and the TVH is also rejected (since $\alpha_k \neq 0$, but $\beta = 1$ with a p-value of 0.719, possibly as a result of a very large α_0, given that the coefficients sum together to 1). For both the PIH and the TVH regressions, the LDH is supported as the null is rejected (since α_0 is found not being larger or equal to any of the α_j). All t-values are large (in excess of 13) and the p-values of the tests either very small (0.000) or large (0.719).

[insert Table 15 here]

For the London Stock Exchange sample, the PIH and the TVH are rejected (since $\beta \neq 1$ and $\alpha_k \neq 0$). For the PIH regression, the LDH is supported as the null is rejected (since α_0 is found not being larger or equal to any of the α_j). For the TVH regression, the LDH is rejected at the 1% level of significance but not at the 5% level (since α_0 is larger than some α_j at the 1% level, but not at the 5% level). The t-value for β_{TVH} is the lowest at 3.17 while all other t-values are large (in excess of 5), and the p-values of the tests are small (0.000 or 0.018).

[insert Table 16 here]

The Deutsche Börse dataset provides the strongest support to the LDH given the large negative magnitude of the coefficients for the end-digit 0 along with large t-values. Given that the coefficients for both PT and PV and close to unity along with large t-values it raises the question to which extent the LDH and the two competing PIH and TVH hypotheses are mutually exclusive when tested. For the Euronext dataset we can reach the same strong LDH support conclusion for the TVH regression, but it is somewhat surprising and counter-intuitive that the PT coefficient is indistinguishable from zero in the PIH regression. For the London Stock Exchange dataset, the estimated coefficients for the PIH regression are very similar to its Deutsche Börse counterpart, so we reach the same strong LDH support conclusions but the TVH regression provides a slightly weaker support for the LDH at a 2% significance level.

Full sample testing rejects both the PIH and the TVH for the three datasets at all levels of significance while the LDH is not rejected four times out of six at the 1% level and five times out of six at the 2% level. Overall, with the exception of the PIH regression for Euronext Paris, the full sample empirical results are congruent across markets. I interpret such findings to constitute a strong empirical support in favor of the LDH.

I also test the LDH against the two alternates, the PIH and the TVH, using the sub-samples per tercile of trading volume for each of the three markets (refer to the six rightmost columns of Tables 14, 15, and 16 - summarized in Table 17 at a 5% level of significance). For The Euronext Paris sub-samples, the test results confirms the whole sample results, with the caveat that since the weighted price contribution is a signed metric if might erroneously induce the β_{PIH} and β_{TVH} coefficients to be negative which is apparent with the large trade volume regressions. For the Deutsche Börse sub-samples, the test results confirms the whole sample results, with the exception of the PIH regression for small trade volume, which is not too surprising since the difference between the price contribution and the trades proportion is only of 1% for that sub-sample as per Panel B of Table 7. For the London Stock Exchange sub-samples, the test results confirms the whole sample results, except for the TVH regressions for the small and medium trade volumes for which the LDH is rejected (allowing to assign the source of the p-value of 0.018 against the LDH in the full sample results to comes solely from the small and medium-size terciles). I interpret such cross-section findings to also constitute a strong empirical support in favor of the LDH.

As we will see later the above empirical findings are consistent with the results of the hypothesis tests conducted using annual sub-samples.

6.7 Are digit preferences evolving over time?

For ease of understanding and ready comparison between the three datasets, the empirical results for the annual sub-samples are presented in graphical form (except for hypothesis testing). As shown below the magnitude of the manifestation of digit preferences has been decreasing over time throughout the sample period. Of note, the ranking of the proportions of trades and volumes by last digit of trade size as previously observed in whole samples is confirmed for all annual samples for each of the three markets.

6.7.1 Number of trades and Volume (number of shares)

As Figures 2 and 3 show, as years went by over the period the digit-0 trades and volumes saw their proportion decline progressively form 40% to 80% toward 20% in a similar manner in each market. These distributions are still far from uniform since a digit-0 proportion of about 20% in 2015 is still twice as much as an equal weight of 10% a uniform distribution implies. The other digits' trades and volumes increased about equally toward 10% concomitantly with the digit 0 decrease. Even digits seem marginally more numerous than odd digits over the whole period. Overall, starting from very uneven distributions, the trend is slowly moving toward more uniform distributions. Since this trend has been relatively stable over the whole period across markets, there is no reason to suspect MIFID1 played a significant role in such evolution.

[insert Figures 2 and 3 here]

6.7.2 Informativeness as per the Weighted Price Contribution measure

As indicated in Figure 4, while apparently exhibiting some noise, the price discovery metric for all non-zero digits is uniformly close to or slightly in excess of 10% for all three markets for the whole period. For Euronext Paris, the WPC of last-digit 0 experienced a sort of V shape, but except for 2008 it stayed within a +/- 5% corridor around the 10% level. For Deutsche Börse, the WPC of last-digit 0 declined from slightly above 60% in 2002 into about the same to +/- 5% corridor around the 10% level in the MIFID1 period (2008 onward). For the LSE, the WPC of last-digit 0 declined from about 40% in 2002 to about -10% in 2007 and then after a temporary increase in 2008-2009 settled around the 10% level.

If all informed traders were immune to any end-digit preference but did not act strategically (i.e. not mimicking the digit preferences of the uninformed traders), one should expect to find a uniform distribution of WPC per end-digit. While it was clearly not the situation at the beginning of the period for both Deutsche Börse and the London Stock Exchange, it looks to be the case over the last five years of the period for these two markets if one takes into account a 95% confidence interval (see Figure 5). If a 95% confidence interval is also taken into account a uniform distribution for Euronext Paris over the whole period cannot be ruled out meaningfully, except for 2008. The preponderance of evidence is that informed traders have removed any end-digit preference from their trade execution over the period and are not acting strategically with regard to the digit preferences of uninformed investors.

[insert Figure 4 here]

6.7.3 Last Digit Zero

Since the last digit zero category is the only one seeing a significant evolution over time, for ease of comparison Figure 5 shows all three metrics on the same graph along with a 95% confidence interval for the price contribution estimate (the two other metrics would have a very narrow confidence interval, so confidence intervals are done without).

For the Euronext Paris sample, as shown in Panel A of Figure 5, last digit 0 trades have seen their proportion of trades and volume decline while their price contribution as first declined in parallel to trades and volume proportion up to 2008, but for 2009 onward it has been converging toward the level of trades and volume proportions. The proportions of trades and volume remain closely grouped together over the period, indicating that there has not been a significant change in the relative trade size for end-digit 0 trades.

For the Deutsche Börse sample, as shown in Panel B of Figure 5, last digit 0 trades have seen their proportion of trades and volume decline while their price contribution declined in parallel throughout the period and is not converging much toward the level of trades and volume proportions, if at all. The proportions of trades and volume remain closely grouped together over the period, indicating that there has not been a significant change in the relative trade size for end-digit 0 trades.

For the London Stock Exchange sample, as shown in Panel C of Figure 5, trades with last digit 0 have seen their proportion of trades and volume decline while their price contribution declined in parallel at first, then converge strongly toward the level of trades and volume only to finally diverge again. The proportions of trades and volume were apart initially, but suddenly converged upon MIFID1 to stay neatly grouped together over the MIFID1 period, indicating that there hasn't been a significant change in the relative trade size for end-digit 0 trades, except for apparent discontinuity upon the MIFID1 implementation.

[insert Figure 5 here]

Overall, for each of the fourteen years across the three markets, the proportion of price discovery attributed to last digit 0 trades, as estimated using the WPC, is always less that the corresponding proportion of either trades or volumes, even when taking into consideration a 95% confidence interval for the WPC. It provides further empirical support for the LDH.

6.7.4 Does hypothesis testing using annual sub-samples confirm already obtained findings?

The empirical results of the hypothesis testing by yearly subsamples are presented in Tables 18, 19, 20, 21, 22, 23, and summarized in Table 17 at a 5% level of significance. Again, the LDH is tested against two alternates, the PIH and the TVH, but using annual sub-samples for each of the three markets. Our purpose is to confirm (or infirm) the already obtained empirical findings. The findings for the Euronext Paris and for Deutsche Börse markets are quite conclusive and convincing in confirming already obtained empirical findings, but it is less the case for the London Stock Exchange market. As shown in Table 17, the findings of the whole samples tests (left column) are broadly supported by the annual samples tests (right column), despite using annual sub-samples with reduced econometric power (as annual sub-samples are significantly smaller than the corresponding full samples for data that is inherently quite noisy).

[insert Table 17 here]

As per Table 18, when using a PIH regression with each of the Euronext Paris annual samples, the PIH and the LDH are both rejected at the 1% in 14 years out of 14. At the 5% level, the LDH is marginally supported in two years out of 14. B_{PIH} being equal to 1 is rejected with a p-value of 0.000 in 14 years out of 14. All this constitutes a very strong endorsement of the above findings for the whole sample. The proportion of trades as a control variable in these regressions for this sample seems irrelevant.

As per Table 19, when using a TVH regression with each of the Euronext Paris annual samples, the TVH is rejected in 14 years out of 14 at the 1% level and the LDH is not rejected in 12 years out of 14 at the 1% level. B_{TVH} being equal to 1 is rejected with a p-value of 0.000 in 11 years out of 14 while the p-values for the three remaining years are relatively small (~5%). All this constitutes again a very strong endorsement of the findings for the whole sample. The proportion of volume as a control variable in these regressions for this sample is working well. The fact that the p-values for the LDH test are 0.000 from 2002 to 2012, then increases to 0.096 for 2013 and 0.675 for 2014 while reverting to 0.001 in 2015, along with the coefficient for digit 0 being not statistically significant for 2015 is interpreted to confirm that the support for the LDH is weakening significantly toward the end of the sample period.

As per Table 20, when using a PIH regression with each of the Deutsche Börse annual samples, the PIH is rejected in 14 years out of 14 at the 1% level and the LDH is not rejected in only 5 years out of 14 at the 1% level and in 6 out of 14 at the 5% level. This constitutes a mixed endorsement of the above findings for the whole sample. However, the cumulative p-values for the LDH test are only 0.112 for the first 6 years of the sample (from 2002 to 2007), but nearly 4 (3.806) for the last four years (2012 to 2015). This is interpreted to indicate that the support for the LDH is disappearing toward the end of the sample period. I conclude that the proportion of trades as a control variable in these regressions for this sample is working, especially in the pre-MIFID1 period when B_{PIH} being equal to 1 is not rejected at a 1% level in 5 years out of 6.

As per Table 21, when using a TVH regression with the each of Deutsche Börse annual samples, the TVH is rejected in 14 years out of 14 at the 1% level while the LDH is not rejected in each year all with p-values equal to 0.000 or 0.001. B_{TVH} being equal to 1 is rejected with a p-value less than 0.008 in 9 years out of 14 while the p-values are in excess of ~5% for only three of the remaining five years. All this constitutes a strong endorsement of the above findings for the whole sample. However, while the annual sample testing cannot reject B_{TVH} being equal to 1, the magnitude of this coefficient in the annual sub-samples has been decreasing steadily over time and was either small or not statistically significant in the last four of the sub-sample. The coefficient for digit 0 is negative and highly significant for the first 10 years of the sample, but indifferent from 0 in the last four years of the sample (2012 to 2015). As this confirm the above earlier findings for the annual PIH regressions for the Deutsche Börse sample, this is also interpreted to indicate that the support for the LDH is disappearing toward the end of the sample period. I conclude that the proportion of volume as a control variable in these regressions for this sample is working well.

As per Table 22, when using a PIH regression with each of the London Stock Exchange annual samples, the TVH is rejected in 14 years out of 14 at the 1% level while the LDH is not rejected in 7 out of 14 years at a 1% level. This constitutes a weak endorsement of the above findings for the whole sample. There is no discernable trend. The proportion of trades as a control variable in these regressions for this sample does not appear to work too well. Interestingly enough, B_{PIH} being equal to 1 is not rejected at a 1% level in the first 8 years and then rejected with p-values = 0.000 in the remaining 6 years, but for 2003 with 0.030. Of note, the adjusted R^2 and the F-values are remarkably lower in the pre-MIFID1 period than afterward.

As per Table 23, when using a TVH regression with the each of London Stock Exchange annual samples, the TVH is rejected in 14 years out of 14 at the 1% level while the LDH is rejected in 13 out of 14 years at a 1% level and 11 out of 14 at the 5% level. This constitutes an endorsement of the above findings for the whole sample (i.e. marginal rejection of the LDH). There is no discernable trend. The proportion of volume as a control variable in these regressions for this sample does not appear to work too well. B_{TVH} being equal to 1 is rejected in each year with a p-value of 0.000 except for 2005. Of note, the adjusted R^2 and the F-values are also remarkably lower in the pre-MIFID1 period than afterward.

[insert Tables 18, 19, 20, 21, 22, 23 here]

Noise is seemingly an issue for the London Stock Exchange (LSE) dataset, leading to both the LDH and the classic alternate to be rejected simultaneously in in 18 years out of 28 at the 5% level. The findings are more definite for the two other markets whereby both the LDH and the PIH are simultaneously rejected in 23 years out of 28 at the 1% level but only in 2 years out of 28 for the LDH and the TVH. Said otherwise, in term of informativeness, volume (number of shares) seems to matter while it is not much the case for the number of trades. Interestingly enough, hypothesis-testing using annual sub-samples indicates a definite weakening the support of the LDH toward the end of the period for both the Euronext Paris and the Deutsche Börse markets.

6.8 What a negative WPC is telling us?

For firms with large trade volumes, the estimated WPC is negative for the last digit 0 trade category in each of the three markets (see Panel D in Tables 6, 7, and 8). How should a systematic negative contribution to price discovery be understood and interpreted?

As O'HARA, YAO and YE (2014, p. 2223) mentions it, the WPC is a signed measure since the price contribution for individual trades, as well as for a given trade category on a given day, can either be positive or negative. When the returns of the trades of a given category, once aggregated, are of the opposite sign to the day's return, the contribution to price discovery of that category is negative. It means that these trades together moved the price in the opposite direction to the price movement resulting from all the trades altogether. It does not mean that all trades of the said category were systematically in the opposite direction to the day's price movement, but only that for that category the overall balance between the trades in the same direction of the day's return and the trades in the opposite direction was found to be in the opposite direction on that day and therefore resulted in a negative price contribution. So, in itself a negative WPC is not mysterious and could simply indicates an overall balance in favor of aggressive contrarian trading taking place in a given trade category (e.g. selling in an up-trend and vice-versa).

In our samples, it is apparent, per Tables 6, 7, and 8, that the WPC for the last digit 0 category is negatively correlated with the firm trade volumes in each of the three markets (and positively correlated for the other digits). The WPC estimates, when compared to the proportions of trades (or volume), indicate that the difference in information density across last digits is large for large cap firms and small for small caps in each of the three markets. This could parallel the level of activity of algorithmic traders, which is known to be positively correlated with trade volumes. Additional empirical investigation is required to confirm or infirm such explanation, but it is intuitively appealing.

As per Figure 4, negative whole-sample WPC to be found statistically and economically significant for any last digit category in a given year is rare. It only occurred for the last digit 0 category for the LSE in 2006 and 2007 (in the pre-MIFID1 period). In the difference of means and in the regression tests, a negative figure indicates how much the given category under contributes to price discovery relative to its proportion of trades (or volume). It has to be interpreted as the extent to which a given trade (or share) of that category contributes to price discovery relative to a mean contribution which would have a zero difference of means (or a coefficient of zero in the said regressions).

The above said, in the formal tests of the LDH against the PIH and the TVH for the Euronext Paris sample as reported in Table 14, the PT coefficient for the medium trade volume tercile and both the PT and PV coefficients for the large volume tercile are negative and statistically significant. This is counter intuitive to the point of being non-sensical since more numerous or larger trades are unlikely to have increasingly smaller price impacts. Once the PT (PV) coefficient is constrained to be positive (not shown) the LDH is no longer rejected in the PIH regression for large trade volume but still is in the PIH regression for medium trade volume. Therefore, the formal testing methodology commonly used is not fail-proof when the WPC is significantly lower than its corresponding proportion of trades or volume for the trade category of interest. However, it could only lead to a spurious rejection of the hypothesis tested against the PIH/TVH, the LDH in the case of this study.

7 Robustness Tests

7.1 Relative trade size

An implicit assumption of the LDH is that there is not much systematic cross-digit variation in trade size. Otherwise the LDH could piggyback on the STH or be deemed to result from the instantaneous price impact differential mechanically induced by a systematic and large cross-digit variation in trade size. As per Figure 6, the relative trade size by last digit of trade size stays within relatively narrow bands, except for the LSE in the pre-MIFID1 period. In addition, the said narrow bands have been narrowing somewhat toward an increased uniformity of trade size. Given the relatively narrow bands (i.e. significantly less than an order of magnitude), there is no reason to expect the difference in trade size to be a significant factor in inducing a difference in price contribution.

[insert Figure 6 here]

7.2 Unbiasedness regressions

Price discovery measures are not without some controversies. In order to alleviate the pitfalls of relying on a single price discovery estimator, I use a second estimator that has previously been used in the literature as described in section 5.

As per Panel A of Figure 7, for Euronext Paris, the price contribution as indicated by the adjusted R^2 for the end-digit 0 category has been consistently below the price contribution of the other digits, despite a spike in 2002. It might be explained by the low proportion of end-digit 0 trades with information as indicated by the WPC estimator (see Panel A of Figure 5).

As per Panel B of Figure 7, for Deutsche Börse, the price contribution as indicated by the adjusted R^2 for the end-digit 0 category saw a drastic fall from about 0.3 in 2002 to 0 in 2008 and staying at such level until the end of the period. It might be explained by an initial high proportion of end-digit 0 trades with information in 2002 (.7/.8, ~88%) continually decreasing until reaching a low proportion in 2008 (.1/.3, ~33%) and remaining at such level until the end of the period as indicated by the WPC estimator (see Panel B of Figure 5).

As per Panel C of Figure 7, for the London Stock Exchange, the adjusted R^2 estimator is extremely noisy in the pre-MIFID1 period but in the MIFID1 period it parallels the ratio of WPC to trades (see Panel C of Figure 5).

Altogether, the adjusted R^2 price discovery estimates indicate that the contribution to price discovery of last-digit 0 is different from the other digits that exhibit similar contribution (except for the LSE pre-MIFID1). The informativeness of trades with a trade size ending with a zero as estimated using the adjusted R^2 price discovery measure is systematically lower for the full sample for Euronext Paris and in post MIFID1 for both Deutsche Börse and the London Stock Exchange. It is possible to reconcile the informativeness measurements as provided by the two estimators to the extent that the ratio of WPC proportion to the proportion of trades (volume) correlates with the adjusted R^2 (i.e. a low ratio of WPC to trades correlates with a low adjusted R^2).

[insert Figure 7 here]

8 Conclusion

I investigate if the information content of marketable orders varies according to the last digit of size using aggregated trades as a proxy. By information content, I mean contribution to price discovery. I find using large datasets that such variation exists and that marketable orders with sizes ending with a zero are on average less informative than when ending with other digits. The Last Digit Hypothesis I propose explains such variation: the strength of last digit preference is negatively correlated with the trader being informed. Uninformed traders exhibit a preference to end the size of their orders with a zero and it is not the case for informed traders. My empirical evidence supports the LDH across three markets, over 14 years and in the cross section of firm size when ranked according to trading volume. Of note, the difference in informativeness between last digit zero and the other digits is positively correlated with trading volumes in the cross-section of firms while it is known that the prevalence of computer driven trading is also positively correlated with trading volumes.

This said, the Last Digit Hypothesis is not an equilibrium since informed traders are expected to act strategically and mimic the last digit preferences of uninformed traders. However, it is a puzzle that I find no evidence of such strategic response and therefore no stealth trading appears to take place in this dimension. More research is warranted to investigate fully this puzzle, but at this stage, it suggests that traders are optimizing along a few key dimensions while disregarding dimensions of perceived less importance. A more prosaic but less plausible explanation is that informed traders are unaware of the digit preferences of uninformed traders and that no strategy to exploit such signal has been developed, a case of collective blissful ignorance. A more plausible explanation is that traders have learned that the execution cost of a meta order can be lowered when child orders are slightly randomized in time and quantity. This would be mutually exclusive with mimicking the last digit preference of uninformed traders for ending the size of their orders with a zero and therefore explain why no mimicking has been observed in the data.

My study is focusing on the size of marketable orders by proxying these using trades, but it is not possible to do so for resting limit orders. However, the LDH applies as well to resting limit orders. To the extent that data availability could be overcome, researching how the informativeness of resting limit orders by their last digit of size as submitted to the limit order book could be quite valuable. A logical follow-up study consists in researching the interaction between marketable and resting orders by their last digit of size and how such dynamics is influencing the information content of the resulting trades. Furthermore, limit orders entail the simultaneous choice of quantity and price that would allow for a finer ranking of limit orders in terms of strength of last digit preferences by analyzing simultaneously quantity and price. All together this could provide for interesting trading signals in a rich setting to research empirically.

References

Admati, Anat R. and Paul Pfleiderer, 1988:. “A theory of intraday patterns: Volume and price variability”, *Review of Financial Studies*, 1, no. 1, pp. 3–40.

Alexander, Gordon J. and Mark A. Peterson, 2007:. “An analysis of trade-size clustering and its relation to stealth trading”, *Journal of Financial Economics*, 84, no. 2, pp. 435–471.

Alsanjari, O.N. et al., 2012:. “Trends and transient change in end-digit preference in blood pressure recording: studies of sequential and longitudinal collected primary care data”, *International journal of clinical practice*, 66, no. 1, pp. 37–43.

Baker, Michael, 1992:. “Digit preference in CPS unemployment data”, *Economics Letters*, 39, no. 1, pp. 117–121.

Ball, Clifford A., Walter N. Torous and Adrian E. Tschoegl, 1985:. “The degree of price resolution: The case of the gold market”, *Journal of Futures Markets*, 5, no. 1, pp. 29–43.

Barclay, Michael J. and Jerold B. Warner, 1993:. “Stealth trading and volatility: Which trades move prices?” *Journal of Financial Economics*, 34, no. 3, pp. 281–305.

Benjamini, Yoav and Ruth Heller, 2008:. “Screening for partial conjunction hypotheses”, *Biometrics*, 64, no. 4, pp. 1215–1222.

Bhattacharya, Utpal et al., 2014:. “Do superstitious traders lose money?” *Available at SSRN 2478124*.

Biais, Bruno, Pierre Hillion and Chester Spatt, 1999:. “Price discovery and learning during the pre-opening period in the Paris Bourse”, *Journal of Political Economy*, 107, no. 6, pp. 1218–1248.

Camarda, Carlo G, Paul HC Eilers and Jutta Gampe, 2008:. “Modelling general patterns of digit preference”, *Statistical Modelling*, 8, no. 4, pp. 385–401.

Choi, Syngjoo et al., 2014:. “Who is (more) rational?” *American Economic Review*, 104, no. 6, pp. 1518–1550.

Domowitz, Ian and Henry Yegerman, 2005:. “The cost of algorithmic trading: A first look at comparative performance”, *Trading*, 2005, no. 1, pp. 30–40.

Easley, David and Maureen O'Hara, 1987:. "Price, trade size, and information in securities markets", *Journal of Financial Economics*, 19, no. 1, pp. 69–90.

Easley, David, Marcos López de Prado and Maureen O'Hara, 2016:. "Discerning information from trade data", *Journal of Financial Economics*.

Evans, Jonathan St BT and Keith E Stanovich, 2013:. "Dual-process theories of higher cognition advancing the debate", *Perspectives on psychological science*, 8, no. 3, pp. 223–241.

Fama, Eugene F, 1970:. "Efficient capital markets: A review of theory and empirical work*", *Journal of Finance*, 25, no. 2, pp. 383–417.

Foucault, Thierry, Marco Pagano and Ailsa Röell, 2013:. *Market liquidity: theory, evidence, and policy*, (Oxford University Press).

Glosten, Lawrence R and Paul R Milgrom, 1985:. "Bid, ask and transaction prices in a specialist market with heterogeneously informed traders", *Journal of Financial Economics*, 14, no. 1, pp. 71–100.

Goodhart, Charles and Riccardo Curcio, 1991:. "The clustering of bid/ask prices and the spread in the foreign exchange market", *LSE Working Paper* Discussion Paper No 110.

Gozluklu, Arie E et al., 2015:. "Lot size constraints and market quality: evidence from the Borsa Italiana", *Financial Management*, 44, no. 4, pp. 905–945.

Grossman, Sanford J and Joseph E Stiglitz, 1980:. "On the impossibility of informationally efficient markets", *American Economic Review*,, pp. 393–408.

Harris, Laurence E., 1991:. "Stock price clustering and discretness", *Review of Financial Studies*, 4, no. 3, pp. 389–415.

Hayek, Friedrich A., 1945:. "The use of knowledge in society", *American Economic Review*, 35, no. 4, pp. 519–30.

Hellwig, Martin F, 1980:. "On the aggregation of information in competitive markets", *Journal of Economic Theory*, 22, no. 3, pp. 477–498.

Hessel, Patrick A, 1986:. "Terminal digit preference in blood pressure measurements: effects on epidemiological associations", *International Journal of Epidemiology*, 15, no. 1, pp. 122–125.

Huang, Roger D and Hans R Stoll, 1997:. "The components of the bid-ask spread: A general approach", *Review of Financial Studies*, 10, no. 4, pp. 995–1034.

Ikenberry, David L. and James P. Weston, 2007:. "Clustering in US stock prices after decimation", *European Financial Management*, 14, pp. 30–54.

Kahneman, Daniel, 2011:. *Thinking, fast and slow*, (Macmillan).

Klesges, Robert C, Margaret Debon and Joanne White Ray, 1995:. "Are self-reports of smoking rate biased? Evidence from the Second National Health and Nutrition Examination Survey", *Journal of Clinical Epidemiology*, 48, no. 10, pp. 1225–1233.

Kuo, Wei-Yu, Tse-Chun Lin and Jing Zhao, 2015:. "Cognitive limitation and investment performance: Evidence from limit order clustering", *Review of Financial Studies*, 28, no. 3, pp. 838–875.

Kyle, Albert S., 1985:. "Continuous auctions and insider trading", *Econometrica*, 53, no. 6, pp. 1315–1335.

Mitchell, Jason, 2001:. "Clustering and psychological barriers: The importance of numbers", *Journal of Futures Markets*, 21, no. 5, pp. 395–428.

Myers, Robert J., 1940:. "Errors and bias in the reporting of ages in census data." *Transaction of the Actuarial Society of America*, 41, pp. 395–415.

Nargesi, Arash A et al., 2014:. "Manual or automated sphygmomanometer? A historical cohort to quantify measurement bias in blood pressure recording", *The Journal of Clinical Hypertension*, 16, no. 10, pp. 716–721.

Niedhammer, I et al., 2000:. "Validity of self-reported weight and height in the French GAZEL cohort." *International journal of obesity and related metabolic disorders: Journal of the International Association for the Study of Obesity*, 24, no. 9, pp. 1111–1118.

O'Hara, Maureen, Chen Yao and Mao Ye, 2014:. "What's not there: Odd lots and market data", *Journal of Finance*, 69, no. 5, pp. 2199–2236.

Ostchega, Yechiam et al., 2003:. "National Health and Nutrition Examination Survey 1999-2000: effect of observer training and protocol standardization on reducing blood pressure measurement error", *Journal of Clinical Epidemiology*, 56, no. 8, pp. 768–774.

Owen, J. A., 1968:. “Terminal digit preference in the laboratory”, *Proceedings of the Association of Clinical Biochemists*, 5, no. 2, pp. 62–62.

Ridout, Martin S and Byron JT Morgan, 1991:. “Modelling digit preference in fecundability studies”, *Biometrics*,, pp. 1423–1433.

Rosch, Eleanor, 1975:. “Cognitive reference points”, *Cognitive Psychology*, 7, no. 4, pp. 532–547.

Rosenbaum, S, 1954:. “Heights and weights of the army intake, 1951”, *Journal of the Royal Statistical Society. Series A (General)*, 117, no. 3, pp. 331–347.

Sider, Hal, 1985:. “Unemployment duration and incidence: 1968-82”, *American Economic Review*, 75, no. 3, pp. 461–472.

Siegel, Jacob S, David Arthur Swanson and Henry S Shryock, 2004:. *The methods and materials of demography*, (Elsevier/Academic Press).

Upson, James, Hardy Johnson and Thomas H McInish, 2015:. “Orders verses trades on the consolidated tape”, *Available at SSRN 2625401*.

Van Bommel, Jos, 2011:. “Measuring price discovery: The variance ratio, the R^2, and the weighted price contribution”, *Finance Research Letters*, 8, no. 3, pp. 112–119.

Wang, Hao and Daniel F Heitjan, 2008:. “Modeling heaping in self-reported cigarette counts”, *Statistics in medicine*, 27, no. 19, p. 3789.

Figure 1: Last Digit Distributions

The figure plots the relative proportion of trades by last digit of trade size. The color scheme depicts informativeness (a tilt to the red being indicative of increased informativeness, and a tilt to the blue the opposite, increased uninformativeness).

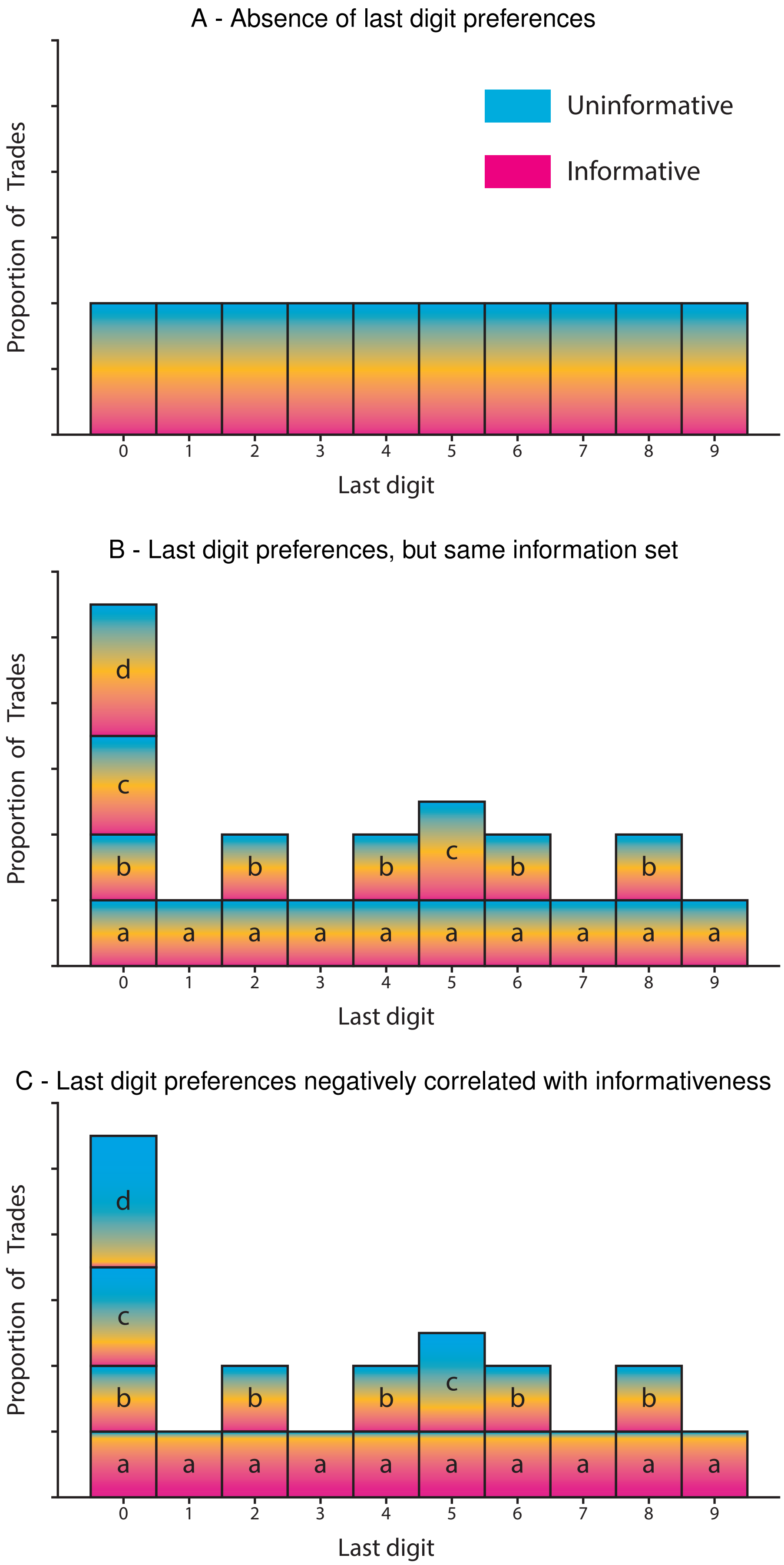

Table 1: Datasets

Market	Trading Venue	Initial Sample		Excluded		Final Sample		
		Stocks	Data	Stocks	Data	Stocks	Data	Stock-days
Paris	Euronext	178	100%	47	3.2%	131	96.8%	401,507
Frankfurt	XTRA	133	100%	30	2.2%	103	97.8%	299,988
London	LSE	202	100%	20	1.3%	182	98.7%	505,365

Table 2: Sub-samples by terciles of trade volume (€ or £)

Market	Trading Venue	Small trade volume		Medium trade volume		Large trade volume	
		Stocks	Volume	Stocks	Volume	Stocks	Volume
Paris	Euronext	44	2.3%	44	11.5%	43	86.2%
Frankfurt	XTRA	35	1.7%	34	8.2%	34	90.1%
London	LSE	61	7.0%	61	16.7%	60	76.3%

Table 3: Descriptive Statistics - Euronext Paris

This table reports descriptive statistics for the intraday continuous trading period for my sample of 131 stocks traded on Euronext Paris from January 2, 2002 to December 31, 2015. “Trades” is the average number of trades per stock day. “Volume” is the average number of shares traded per stock day. “Size” is the average number of shares traded per trade. “Depth B” (“Depth A”) is the average number of shares displayed at the best bid (ask) price at the time of a trade. “Quotes” is the average number of quote revisions of the top of the book per stock day. “Spread T” (“Spread Q”) is, in basis points, the average bis-ask spread divided by the quote midpoint existing prior to a trade (at the time of a quote message).

Panel A - All Firms

	Trades	Volume	Size	Depth B	Depth A	Quotes	Spread T	Spread Q
2002	1,082	760,930	703	2,571	2,655	2,135	17	24
2003	1,359	917,698	675	3,086	3,210	2,451	13	18
2004	1,317	1,002,113	761	7,455	7,715	2,373	11	14
2005	1,449	986,679	681	10,309	10,352	2,569	9	12
2006	1,893	1,218,797	644	12,045	12,508	3,454	10	13
2007	2,670	1,119,398	419	3,685	3,917	6,000	7	12
2008	3,528	1,460,400	414	2,213	2,177	12,706	10	18
2009	3,027	1,217,125	402	1,609	1,638	17,767	9	18
2010	3,364	1,098,103	326	1,674	1,688	25,282	6	11
2011	3,979	1,191,269	299	1,610	1,603	41,585	7	11
2012	3,355	1,193,324	356	2,490	2,486	45,335	7	9
2013	3,036	1,012,796	334	2,372	2,470	40,427	6	7
2014	3,264	986,856	302	3,036	2,728	40,672	6	7
2015	4,118	1,074,659	261	1,896	1,883	65,672	6	6
All	2,769	1,101,780	398	3,181	3,214	23,254	8	10
Panel B - Firms with small trade volume								
All	539	75,912	141	686	438	4,325	18	34
Panel C - Firms with medium trade volume								
All	1,728	610,736	353	5,511	5,577	14,229	11	13
Panel D - Firms with large trade volume								
All	5,928	2,567,241	433	2,732	2,780	50,260	6	7

Table 4: Descriptive Statistics - Deutsche Börse Xetra

This table reports descriptive statistics for the intraday continuous trading period for my sample of 103 stocks traded on Deutsche Börse Xetra from January 2, 2002 to December 31, 2015. “Trades” is the average number of trades per stock day. “Volume” is the average number of shares traded per stock day. “Size” is the average number of shares traded per trade. “Depth B” (“Depth A”) is the average number of shares displayed at the best bid (ask) price at the time of a trade. “Quotes” is the average number of quote revisions of the top of the book per stock day. “Spread T” (“Spread Q”) is, in basis points, the average bis-ask spread divided by the quote midpoint existing prior to a trade (at the time of a quote message).

Panel A - All Firms

	Trades	Volume	Size	Depth B	Depth A	Quotes	Spread T	Spread Q
2002	1,148	1,350,552	1,176	2,649	2,628	3,131	15	25
2003	1,277	1,528,516	1,197	3,503	3,629	3,654	11	19
2004	992	1,309,715	1,321	4,953	4,812	2,721	8	12
2005	984	1,245,129	1,265	6,219	6,209	2,384	7	11
2006	1,243	1,367,784	1,100	5,833	5,700	2,865	7	13
2007	1,896	1,597,448	842	3,384	3,203	7,559	9	14
2008	2,611	1,819,460	697	2,248	2,403	17,006	12	19
2009	1,911	1,337,511	700	3,496	2,817	16,875	12	17
2010	1,982	1,192,654	602	1,665	1,683	9,336	7	13
2011	2,651	1,590,044	600	1,894	1,865	9,444	8	13
2012	2,180	1,473,826	676	2,429	2,437	8,833	7	12
2013	1,930	1,037,216	537	1,645	1,648	7,747	6	10
2014	1,986	881,607	444	1,428	1,424	21,164	7	11
2015	2,353	943,132	401	1,315	1,319	35,864	7	9
All	1,856	1,333,633	719	2,623	2,563	11,190	9	13

Panel B - Firms with small trade volume

All	370	121,467	328	680	697	3,759	23	37

Panel C - Firms with medium trade volume

All	1,036	383,987	371	847	839	6,383	13	19

Panel D - Firms with large trade volume

All	3,836	3,202,342	835	3,211	3,133	21,732	6	8

Table 5: Descriptive Statistics - London Stock Exchange

This table reports descriptive statistics for the intraday continuous trading period for my sample of 182 stocks traded on the London Stock Exchange from January 2, 2002 to December 31, 2015. “Trades” is the average number of trades per stock day. “Volume” is the average number of shares traded per stock day. “Size” is the average number of shares traded per trade. “Depth B” (“Depth A”) is the average number of shares displayed at the best bid (ask) price at the time of a trade. “Quotes” is the average number of quote revisions of the top of the book per stock day. “Spread T” (“Spread Q”) is, in basis points, the average bis-ask spread divided by the quote midpoint existing prior to a trade (at the time of a quote message). Depth is unavailable until April 20, 2009.

Panel A - All Firms

	Trades	Volume	Size	Depth B	Depth A	Quotes	Spread T	Spread Q
2002	860	11,265,130	13,102			1,310	23	37
2003	1,042	11,246,828	10,793			2,081	18	28
2004	1,040	10,211,298	9,823			2,359	14	20
2005	1,163	9,647,784	8,293			2,622	12	17
2006	1,591	10,043,290	6,313			3,862	12	16
2007	2,631	10,696,200	4,065			6,865	11	14
2008	3,806	9,205,935	2,418			17,240	15	19
2009	3,192	7,022,840	2,200	11,719	11,883	18,779	12	15
2010	2,894	5,863,630	2,026	10,101	10,139	26,029	7	11
2011	3,123	4,938,393	1,581	8,213	8,069	32,813	8	12
2012	2,968	4,349,254	1,465	7,872	7,481	29,846	7	10
2013	2,814	3,653,581	1,298	6,365	6,415	30,831	7	9
2014	3,405	3,778,362	1,110	5,686	5,641	30,659	6	8
2015	4,231	4,109,133	971	4,826	4,795	40,346	6	7
All	3,195	4,679,363	1,465	7,528	7,468	30,182	10	11

Panel B - Firms with small trade volume

	Trades	Volume	Size	Depth B	Depth A	Quotes	Spread T	Spread Q
All	1,379	1,229,453	892	3,693	3,639	12,550	15	20

Panel C - Firms with medium trade volume

	Trades	Volume	Size	Depth B	Depth A	Quotes	Spread T	Spread Q
All	2,277	2,781,767	1,222	6,423	6,216	21,408	13	15

Panel D - Firms with large trade volume

	Trades	Volume	Size	Depth B	Depth A	Quotes	Spread T	Spread Q
All	6,036	10,222,409	1,694	8,874	8,865	57,629	7	8

Table 6: Price Discovery Statistics - Euronext Paris

This table reports price discovery statistics in percentages for each digit category (0, ..., 9). All trades are grouped per stock day by their last digit of trade size (number of shares transacted for each individual trade), for the intraday continuous trading period for my sample of 131 stocks traded on Euronext Paris from January 2, 2002 to December 31, 2015. "WPC" is the weighted price contribution. "Trades" is the average proportion of trades. "Volume" is the average proportion of shares traded. "Relative Size" is the average trade size of the digit category relative to the average. The price discovery estimator is cross section weighted then time series averaged while the other estimates are equally weighted stock day averages. The number of observations (N) is as indicated for each estimate. The standard deviation of the estimate is provided in parentheses.

Panel A - All Firms

	0	1	2	3	4	5	6	7	8	9
WPC	10.44	10.37	9.99	10.00	9.88	10.10	9.89	9.85	9.89	9.59
N = 3,574	(0.52)	(0.15)	(0.14)	(0.12)	(0.13)	(0.15)	(0.12)	(0.11)	(0.12)	(0.11)
Trades	29.60	7.96	8.07	7.47	7.64	10.10	7.52	7.17	7.49	6.98
N = 3,574	(0.13)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Volume	30.97	7.35	7.70	7.37	7.62	8.95	7.62	7.38	7.68	7.36
N = 3,574	(0.15)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Relative Size	1.04	0.92	0.95	0.99	1.00	0.89	1.01	1.03	1.02	1.06
N = 3,574	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Panel B - Firms with small trade volume

	0	1	2	3	4	5	6	7	8	9
WPC	25.63	8.43	7.72	8.09	7.70	10.26	8.34	8.00	8.15	7.68
N = 3,572	(0.47)	(0.21)	(0.20)	(0.18)	(0.21)	(0.27)	(0.18)	(0.17)	(0.20)	(0.17)
Trades	30.80	7.81	8.02	7.27	7.46	11.01	7.16	6.81	7.14	6.51
N = 3,572	(0.17)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Volume	34.66	6.75	7.18	6.84	7.11	9.48	7.08	6.86	7.20	6.83
N = 3,572	(0.18)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)	(0.03)
Relative Size	1.13	0.87	0.90	0.94	0.95	0.87	0.99	1.01	1.01	1.05
N = 3,572	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Panel C - Firms with medium trade volume

	0	1	2	3	4	5	6	7	8	9
WPC	12.67	10.79	10.05	9.67	9.80	9.86	9.10	9.50	9.27	9.29
N = 3,573	(0.72)	(0.27)	(0.21)	(0.19)	(0.20)	(0.25)	(0.19)	(0.19)	(0.20)	(0.18)
Trades	29.82	7.97	8.01	7.48	7.60	9.83	7.55	7.21	7.51	7.04
N = 3,573	(0.14)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Volume	30.74	7.41	7.75	7.41	7.65	8.79	7.67	7.44	7.73	7.42
N = 3,573	(0.16)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Relative Size	1.02	0.93	0.97	0.99	1.01	0.90	1.01	1.03	1.03	1.05
N = 3,573	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Panel D - Firms with large trade volume

	0	1	2	3	4	5	6	7	8	9
WPC	-6.35	11.92	12.23	12.21	12.05	10.14	12.06	11.91	12.09	11.75
N = 3,574	(1.06)	(0.29)	(0.26)	(0.23)	(0.24)	(0.28)	(0.24)	(0.23)	(0.24)	(0.23)
Trades	28.80	7.97	8.10	7.57	7.79	9.58	7.76	7.39	7.74	7.30
N = 3,574	(0.10)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Volume	28.09	7.77	8.09	7.76	8.03	8.65	8.03	7.76	8.05	7.76
N = 3,574	(0.13)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Relative Size	0.96	0.97	1.00	1.02	1.03	0.91	1.03	1.05	1.04	1.06
N = 3,574	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Table 7: Price Discovery Statistics - Deutsche Börse

This table reports price discovery statistics in percentages for each digit category (0, ..., 9). All trades are grouped per stock day by their last digit of trade size (number of shares transacted for each individual trade), for the intraday continuous trading period for my sample of 103 stocks traded on Deutsche Börse from January 2, 2002 to December 30, 2015."WPC" is the weighted price contribution. "Trades" is the average proportion of trades. "Volume" is the average proportion of shares traded. "Relative Size" is the average trade size of the digit category relative to the average of all trades. The price discovery estimator is cross section weighted then time series averaged while the other estimates are equally weighted stock day averages. The number of observations (N) is as indicated for each estimate. The standard deviation of the estimate is provided in parentheses.

Panel A - All Firms

	0	1	2	3	4	5	6	7	8	9
WPC	23.68	8.61	8.60	8.04	8.53	9.33	8.47	8.20	8.53	8.01
N = 3,551	(0.59)	(0.12)	(0.13)	(0.12)	(0.12)	(0.13)	(0.11)	(0.12)	(0.12)	(0.11)
Trades	37.65	7.02	7.13	6.76	6.88	8.03	6.79	6.52	6.78	6.43
N = 3,551	(0.28)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Volume	39.91	6.47	6.75	6.45	6.70	7.37	6.69	6.46	6.75	6.46
N = 3,551	(0.27)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Relative Size	1.08	0.92	0.95	0.95	0.97	0.91	0.98	0.99	0.99	1.00
N = 3,551	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Panel B - Firms with small trade volume

	0	1	2	3	4	5	6	7	8	9
WPC	38.56	6.76	6.79	6.37	6.87	8.93	6.55	6.40	6.55	6.21
N = 3,550	(0.60)	(0.20)	(0.20)	(0.19)	(0.19)	(0.25)	(0.19)	(0.18)	(0.18)	(0.18)
Trades	39.48	6.87	6.97	6.61	6.62	8.25	6.46	6.20	6.48	6.06
N = 3,550	(0.30)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
Volume	43.84	5.95	6.27	5.98	6.21	7.31	6.19	5.99	6.28	5.97
N = 3,550	(0.28)	(0.04)	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)
Relative Size	1.15	0.88	0.91	0.91	0.94	0.88	0.96	0.97	0.97	0.99
N = 3,550	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Panel C - Firms with medium trade volume

	0	1	2	3	4	5	6	7	8	9
WPC	31.07	7.85	7.74	7.24	7.67	8.92	7.52	7.41	7.56	7.03
N = 3,510	(0.61)	(0.19)	(0.19)	(0.19)	(0.19)	(0.21)	(0.17)	(0.18)	(0.18)	(0.17)
Trades	35.99	7.23	7.33	6.93	7.07	8.12	6.99	6.71	6.97	6.65
N = 3,510	(0.27)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
Volume	38.50	6.62	6.91	6.59	6.86	7.51	6.86	6.61	6.92	6.63
N = 3,510	(0.27)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Relative Size	1.09	0.92	0.94	0.95	0.97	0.92	0.98	0.98	0.99	1.00
N = 3,510	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Panel D - Firms with large trade volume

	0	1	2	3	4	5	6	7	8	9
WPC	-1.05	11.48	11.57	10.67	11.42	10.45	11.54	11.04	11.83	11.06
N = 3,511	(1.24)	(0.25)	(0.25)	(0.24)	(0.23)	(0.24)	(0.22)	(0.23)	(0.23)	(0.23)
Trades	36.66	7.04	7.18	6.82	7.03	7.96	6.98	6.72	6.97	6.64
N = 3,511	(0.25)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Volume	36.83	6.85	7.12	6.81	7.09	7.48	7.07	6.81	7.10	6.82
N = 3,511	(0.25)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Relative Size	1.00	0.97	0.99	1.00	1.00	0.93	1.01	1.01	1.01	1.02
N = 3,511	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Table 8: Price Discovery Statistics - London Stock Exchange

This table reports price discovery statistics in percentages for each digit category (0, ..., 9). All trades are grouped per stock day by their last digit of trade size (number of shares transacted for each individual trade), for the intraday continuous trading period for my sample of 182 stocks traded on the LSE from January 2, 2002 to December 31, 2015. "WPC" is the weighted price contribution. "Trades" is the average proportion of trades. "Volume" is the average proportion of shares traded. "Relative Size" is the average trade size of the digit category relative to the average of all trades. The price discovery estimator is cross section weighted then time series averaged while the other estimates are equally weighted stock day averages. The number of observations (N) is as indicated for each estimate. The standard deviation of the estimate is provided in parentheses.

Panel A - All Firms

	0	1	2	3	4	5	6	7	8	9
WPC	11.33	10.01	10.03	9.80	9.87	9.73	9.88	9.21	10.43	9.71
N = 3528	(0.80)	(0.23)	(0.27)	(0.35)	(0.28)	(0.46)	(0.29)	(0.26)	(0.30)	(0.32)
Trades	31.80	7.57	7.71	7.38	7.61	8.19	7.56	7.33	7.55	7.30
N = 3528	(0.15)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Volume	36.79	6.91	7.09	6.82	7.07	7.31	7.03	7.11	7.05	6.82
N = 3528	(0.25)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Relative Size	1.12	0.90	0.91	0.91	0.91	0.88	0.92	0.96	0.92	0.92
N = 3528	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Panel B - Firms with small trade volume

	0	1	2	3	4	5	6	7	8	9
WPC	25.47	8.60	8.80	7.65	8.11	8.36	8.20	7.58	8.58	8.64
N = 3528	(0.69)	(0.40)	(0.47)	(0.64)	(0.20)	(0.22)	(0.56)	(0.42)	(0.30)	(0.75)
Trades	31.10	7.69	7.81	7.49	7.68	8.21	7.61	7.42	7.61	7.39
N = 3528	(0.17)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Volume	37.53	6.83	7.04	6.76	7.01	7.32	6.95	6.84	6.95	6.76
N = 3528	(0.26)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Relative Size	1.18	0.88	0.89	0.89	0.90	0.88	0.90	0.91	0.90	0.90
N = 3528	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Panel C - Firms with medium trade volume

	0	1	2	3	4	5	6	7	8	9
WPC	15.58	9.19	9.77	9.51	9.49	8.71	9.35	9.68	9.42	9.30
N = 3528	(0.91)	(0.26)	(0.38)	(0.28)	(0.24)	(0.56)	(0.43)	(0.46)	(0.31)	(0.24)
Trades	31.68	7.58	7.73	7.41	7.64	8.15	7.59	7.35	7.56	7.31
N = 3528	(0.15)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Volume	36.53	6.96	7.13	6.87	7.11	7.32	7.08	7.02	7.09	6.87
N = 3528	(0.26)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Relative Size	1.11	0.91	0.91	0.91	0.92	0.89	0.92	0.94	0.92	0.93
N = 3528	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Panel D - Firms with large trade volume

	0	1	2	3	4	5	6	7	8	9
WPC	-8.59	12.52	11.52	12.42	12.07	12.22	12.20	10.66	13.42	11.54
N = 3528	(2.06)	(0.54)	(0.43)	(0.71)	(0.78)	(1.40)	(0.42)	(0.59)	(0.82)	(0.56)
Trades	32.84	7.41	7.58	7.23	7.49	8.19	7.45	7.18	7.45	7.19
N = 3528	(0.13)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Volume	36.45	6.92	7.10	6.83	7.07	7.30	7.05	7.39	7.07	6.83
N = 3528	(0.24)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Relative Size	1.08	0.92	0.92	0.93	0.93	0.88	0.93	1.02	0.94	0.94
N = 3528	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Table 9: Difference of Means - Euronext Paris

This table reports tests of statistical significance of the difference of means between the WPC and the proportion of trades (volume) for each digit category (0, ..., 9). All trades are grouped per stock day by their last digit of trade size (number of shares transacted for each individual trade), for the intraday continuous trading period for my sample of 131 stocks traded on Euronext Paris from January 2, 2002 to December 31, 2015. "WPC-Trades" is the difference between the weighted price contribution and the average proportion of trades. "WPC-Volume" is difference between the weighted price contribution and the average proportion of shares traded. The price discovery estimator is cross section weighted then time series averaged while the other estimates are equally weighted stock day averages. The number of observations (N) is as indicated for each estimate. The t-statistic of the estimate is provided in brackets.

Panel A - All Firms

	0	1	2	3	4	5	6	7	8	9
WPC-Trades	-19.15	2.41	1.92	2.53	2.23	-0.01	2.37	2.69	2.40	2.62
N = 3,574	[-36.01]	[15.55]	[14.09]	[21.14]	[17.52]	[-0.04]	[19.65]	[24.08]	[19.45]	[22.74]
p-$value$	0.000	0.000	0.000	0.000	0.000	0.970	0.000	0.000	0.000	0.000
WPC-Volume	-20.53	3.03	2.29	2.63	2.26	1.15	2.26	2.47	2.20	2.23
N = 3,574	[-38.17]	[19.54]	[16.80]	[22.00]	[17.68]	[7.72]	[18.74]	[22.10]	[17.82]	[19.32]
p-$value$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Panel B - Firms with small trade volume

WPC-Trades	-5.17	0.62	-0.30	0.81	0.24	-0.75	1.18	1.19	1.01	1.17
N = 3,572	[-10.31]	[2.92]	[-1.52]	[4.46]	[1.17]	[-2.77]	[6.44]	[6.85]	[5.09]	[6.94]
p-$value$	0.000	0.003	0.129	0.000	0.242	0.006	0.000	0.000	0.000	0.000
WPC-Volume	-9.03	1.68	0.53	1.24	0.59	0.79	1.26	1.14	0.95	0.85
N = 3,572	[-17.86]	[7.95]	[2.68]	[6.80]	[2.85]	[2.90]	[6.84]	[6.55]	[4.77]	[5.02]
p-$value$	0.000	0.000	0.007	0.000	0.004	0.004	0.000	0.000	0.000	0.000

Panel C - Firms with medium trade volume

WPC-Trades	-17.15	2.82	2.04	2.20	2.20	0.04	1.55	2.29	1.76	2.25
N = 3,573	[-23.43]	[10.49]	[9.58]	[11.23]	[11.20]	[0.15]	[8.15]	[11.96]	[9.00]	[12.36]
p-$value$	0.000	0.000	0.000	0.000	0.000	0.878	0.000	0.000	0.000	0.000
WPC-Volume	-18.07	3.37	2.30	2.26	2.15	1.08	1.43	2.07	1.55	1.87
N = 3,573	[-24.51]	[12.56]	[10.78]	[11.54]	[10.91]	[4.25]	[7.54]	[10.75]	[7.89]	[10.26]
p-$value$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Panel D - Firms with large trade volume

WPC-Trades	-35.15	3.94	4.14	4.64	4.26	0.56	4.30	4.52	4.35	4.46
N = 3,574	[-32.94]	[13.40]	[15.65]	[19.72]	[17.89]	[1.99]	[17.89]	[19.99]	[18.06]	[19.38]
p-$value$	0.000	0.000	0.000	0.000	0.000	0.047	0.000	0.000	0.000	0.000
WPC-Volume	-34.45	4.14	4.14	4.46	4.02	1.49	4.02	4.15	4.04	3.99
N = 3,574	[-32.17]	[14.07]	[15.66]	[18.93]	[16.89]	[5.27]	[16.72]	[18.36]	[16.77]	[17.35]
p-$value$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 10: Difference of Means - Deutsche Börse

This table reports tests of statistical significance of the difference of means between the WPC and the proportion of trades (volume) for each digit category (0, ..., 9). All trades are grouped per stock day by their last digit of trade size (number of shares transacted for each individual trade), for the intraday continuous trading period for my sample of 103 stocks traded on Deutsche Börse from January 2, 2002 to December 30, 2015. “WPC-Trades” is the difference between the weighted price contribution and the average proportion of trades. “WPC-Volume” is difference between the weighted price contribution and the average proportion of shares traded. The price discovery estimator is cross section weighted then time series averaged while the other estimates are equally weighted stock day averages. The number of observations (N) is as indicated for each estimate. The t-statistic of the estimate is provided in brackets.

Panel A - All Firms

	0	1	2	3	4	5	6	7	8	9
WPC-Trades	-13.97	1.59	1.47	1.28	1.65	1.30	1.68	1.68	1.75	1.57
N = 3,551	[-21.48]	[12.26]	[11.19]	[10.16]	[13.76]	[9.93]	[14.34]	[13.87]	[14.53]	[13.52]
p-$value$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
WPC-Volume	-16.22	2.14	1.85	1.59	1.82	1.96	1.78	1.75	1.78	1.55
N = 3,551	[-25.07]	[16.68]	[14.21]	[12.66]	[15.27]	[14.97]	[15.14]	[14.42]	[14.74]	[13.27]
p-$value$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Panel B - Firms with small trade volume

WPC-Trades	-0.92	-0.10	-0.19	-0.24	0.25	0.68	0.09	0.20	0.07	0.15
N = 3,550	[-1.36]	[-0.52]	[-0.91]	[-1.21]	[1.35]	[2.75]	[0.48]	[1.06]	[0.38]	[0.84]
p-$value$	0.175	0.604	0.363	0.225	0.178	0.006	0.629	0.290	0.702	0.399
WPC-Volume	-5.28	0.81	0.52	0.39	0.67	1.62	0.36	0.41	0.27	0.24
N = 3,550	[-7.91]	[4.06]	[2.55]	[1.96]	[3.54]	[6.54]	[1.91]	[2.18]	[1.43]	[1.35]
p-$value$	0.000	0.000	0.011	0.050	0.000	0.000	0.057	0.029	0.154	0.178

Panel C - Firms with medium trade volume

WPC-Trades	-4.86	0.61	0.40	0.31	0.59	0.79	0.52	0.69	0.58	0.37
N = 3,550	[-7.13]	[3.13]	[2.06]	[1.63]	[3.14]	[3.85]	[2.93]	[3.88]	[3.23]	[2.14]
p-$value$	0.000	0.002	0.039	0.104	0.002	0.000	0.003	0.000	0.001	0.032
WPC-Volume	-7.34	1.21	0.82	0.64	0.79	1.40	0.65	0.79	0.63	0.40
N = 3,550	[-10.79]	[6.22]	[4.23]	[3.40]	[4.25]	[6.78]	[3.69]	[4.44]	[3.49]	[2.28]
p-$value$	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.023

Panel D - Firms with large trade volume

WPC-Trades	-37.28	4.39	4.34	3.80	4.33	2.46	4.52	4.27	4.81	4.37
N = 3,551	[-29.47]	[17.77]	[17.48]	[16.04]	[18.99]	[10.30]	[20.48]	[18.55]	[20.77]	[18.63]
p-$value$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
WPC-Volume	-37.45	4.57	4.40	3.81	4.28	2.93	4.42	4.18	4.68	4.19
N = 3,551	[-29.58]	[18.52]	[17.72]	[16.09]	[18.74]	[12.29]	[20.03]	[18.12]	[20.19]	[17.86]
p-$value$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 11: Difference of Means - London Stock Exchange

This table reports tests of statistical significance of the difference of means between the WPC and the proportion of trades (volume) for each digit category (0, ..., 9). All trades are grouped per stock day by their last digit of trade size (number of shares transacted for each individual trade), for the intraday continuous trading period for my sample of 182 stocks traded on the LSE from January 2, 2002 to December 31, 2015. "WPC-Trades" is the difference between the weighted price contribution and the average proportion of trades. "WPC-Volume" is difference between the weighted price contribution and the average proportion of shares traded. The price discovery estimator is cross section weighted then time series averaged while the other estimates are equally weighted stock day averages. The number of observations (N) is as indicated for each estimate. The t-statistic of the estimate is provided in brackets.

Panel A - All Firms

	0	1	2	3	4	5	6	7	8	9
WPC-Trades	-20.44	2.44	2.31	2.42	2.26	1.54	2.31	1.87	2.84	2.44
N = 3,529	[-25.07]	[10.69]	[8.70]	[6.88]	[7.98]	[3.38]	[8.10]	[7.26]	[9.43]	[7.61]
p-$value$	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000
WPC-Volume	-25.43	3.10	2.93	2.98	2.81	2.42	2.84	2.09	3.34	2.93
N = 3,529	[-30.26]	[13.49]	[10.98]	[8.46]	[9.87]	[5.29]	[9.92]	[8.08]	[11.07]	[9.09]
p-$value$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Panel B - Firms with small trade volume

WPC-Trades	-5.63	0.91	1.00	0.17	0.42	0.16	0.59	0.16	0.97	1.25
N = 3,528	[-7.93]	[2.26]	[2.10]	[0.26]	[2.08]	[0.70]	[1.07]	[0.37]	[3.19]	[1.67]
p-$value$	0.000	0.024	0.036	0.794	0.037	0.481	0.287	0.709	0.001	0.095
WPC-Volume	-12.06	1.77	1.76	0.89	1.10	1.04	1.25	0.74	1.62	1.89
N = 3,528	[-16.39]	[4.37]	[3.70]	[1.39]	[5.39]	[4.62]	[2.25]	[1.73]	[5.34]	[2.52]
p-$value$	0.000	0.000	0.000	0.166	0.000	0.000	0.025	0.083	0.000	0.012

Panel C - Firms with medium trade volume

WPC-Trades	-16.10	1.61	2.04	2.10	1.86	0.57	1.75	2.33	1.86	1.99
N = 3,528	[-17.48]	[6.11]	[5.42]	[7.43]	[7.86]	[1.01]	[4.10]	[5.06]	[5.92]	[8.25]
p-$value$	0.000	0.000	0.000	0.000	0.000	0.314	0.000	0.000	0.000	0.000
WPC-Volume	-20.95	2.22	2.64	2.64	2.38	1.39	2.26	2.66	2.33	2.43
N = 3,528	[-22.17]	[8.41]	[6.99]	[9.32]	[10.02]	[2.47]	[5.28]	[5.76]	[7.40]	[10.04]
p-$value$	0.000	0.000	0.000	0.000	0.000	0.014	0.000	0.000	0.000	0.000

Panel D - Firms with large trade volume

WPC-Trades	-41.39	5.11	3.94	5.20	4.58	4.03	4.74	3.46	5.93	4.40
N = 3,529	[-20.05]	[9.53]	[9.20]	[7.30]	[5.84]	[2.87]	[11.42]	[5.85]	[7.21]	[7.83]
p-$value$	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.000
WPC-Volume	-45.00	5.60	4.42	5.60	5.00	4.92	5.15	3.25	6.30	4.76
N = 3,529	[-21.69]	[10.43]	[10.30]	[7.86]	[6.38]	[3.51]	[12.37]	[5.49]	[7.66]	[8.46]
p-$value$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 12: Digital Signature

This table summarizes selected proportion data of Tables 6, 7, and 8 per the expected ranking in decreasing order of the proportion of trades and volume by digit (in percentage) from left to right. The digits 2, 4, 6, and 8 are grouped together under the heading 'Even' while digits 1, 3, 7, and 9 are grouped together under the heading 'Odd'.

		Proportion of Trades				Proportion of Volume			
		0	5	Even	Odd	0	5	Even	Odd
Euronext Paris	All	29.60	10.10	30.72	29.58	30.97	8.95	30.62	29.46
	Small	30.80	11.01	29.78	28.40	34.66	9.48	28.57	27.28
	Medium	29.82	9.83	30.67	29.70	30.74	8.79	30.80	29.68
	Large	28.80	9.58	31.39	30.23	28.09	8.65	32.20	31.05
Deutsche Börse Xetra	All	37.65	8.03	27.58	26.73	39.91	7.37	26.89	25.84
	Small	39.48	8.25	26.53	25.74	43.84	7.31	24.95	23.89
	Medium	35.99	8.12	28.36	27.52	38.50	7.51	27.55	26.45
	Large	36.66	7.96	28.16	27.22	36.83	7.48	28.38	27.29
London Stock Exchange	All	31.80	8.19	30.43	29.58	36.79	7.31	28.24	27.66
	Small	31.10	8.21	30.71	29.99	37.53	7.32	27.95	27.19
	Medium	31.68	8.15	30.52	29.65	36.53	7.32	28.41	27.72
	Large	32.84	8.19	29.97	29.01	36.45	7.30	28.29	27.97

Table 13: Summary of Formal Testing using Whole Samples

This table summarizes the results of formally testing the LDH against the PIH and the TVH, as per findings detailed in Tables 14, 15, and 16 (for a 5% level of statistical significance).

	All firms	Per tercile of trading volume		
		Small	Medium	Large
Euronext Paris				
PIH vs LDH	Reject both	Reject both	Reject both	Reject both
TVH vs LDH	Reject TVH Not reject LDH	Reject TVH Not reject LDH	Reject both	Reject TVH Not reject LDH
Deutsche Börse Xetra				
PIH vs LDH	Reject PIH Not reject LDH	Reject both	Reject TVH Not reject LDH	Reject TVH Not reject LDH
TVH vs LDH	Reject TVH Not reject LDH	Reject TVH Not reject LDH	Reject TVH Not reject LDH	Reject TVH Not reject LDH
London Stock Exchange				
PIH vs LDH	Reject PIH Not reject LDH	Reject PIH Not reject LDH	Reject PIH Not reject LDH	Reject PIH Not reject LDH
TVH vs LDH	Reject TVH Not reject LDH	Reject both	Reject both	Reject TVH Not reject LDH

Table 14: Test of Hypotheses - Euronext Paris

This table reports the estimated parameters of the following regressions:

$$PC_k^{s,t} = \boldsymbol{\alpha}_{0,\ldots,9}\boldsymbol{D}_{0,\ldots,9}^{s,t} + \beta PT_k^{s,t} + \varepsilon_i^{s,t},\ k \in (0, ..., 9)$$
$$PC_k^{s,t} = \boldsymbol{\alpha}_{0,\ldots,9}\boldsymbol{D}_{0,\ldots,9}^{s,t} + \beta PV_k^{s,t} + \varepsilon_i^{s,t},\ k \in (0, ..., 9)$$

$PC_k^{s,t}$ is the proportion of the cumulative return of all trades for a given stock s for the time period t attributable to a given category k. $D_k^{s,t}$ is a vector dummy variable equal to 1 when its k corresponds to the k of the other variables. $PT_k^{s,t}$ is the proportion of the number of trades for a given stock s for the time period t of a given category k. $PV_k^{s,t}$ is the proportion of shares traded for a given stock s for the time period t for a given category k. Trades are categorized by their last digit of trade size (number of shares transacted for each individual trade), pertaining to the intraday continuous trading period for my sample of 131 stocks traded on Euronext Paris from January 2, 2002 to December 31, 2015. The regressions are pooled weighted according to the relative importance of the absolute value of the cumulative return of all trades for a given stock s for the time period t. The number of observations (N) is as indicated.

	All Firms		Firms with small trade volume		Firms with medium trade volume		Firms with large trade volume	
	PIH	TVH	PIH	TVH	PIH	TVH	PIH	TVH
D_0	0.11*** [12.71]	−0.03*** [-5.85]	0.03*** [3.16]	0.01 [0.91]	0.25*** [14.12]	0.10*** [9.38]	0.38*** [13.10]	0.06*** [3.00]
D_1	0.11*** [36.81]	0.07*** [35.88]	0.03*** [10.13]	0.04*** [15.43]	0.14*** [23.75]	0.10*** [24.53]	0.24*** [27.73]	0.15*** [23.84]
D_2	0.10*** [35.16]	0.07*** [34.58]	0.02*** [7.38]	0.03*** [11.70]	0.13*** [22.61]	0.10*** [23.93]	0.24*** [27.76]	0.16*** [24.38]
D_3	0.10*** [37.88]	0.07*** [37.54]	0.03*** [11.47]	0.03*** [15.50]	0.13*** [23.11]	0.09*** [24.23]	0.23*** [28.74]	0.15*** [25.48]
D_4	0.10*** [36.79]	0.07*** [35.20]	0.03*** [9.19]	0.03*** [11.77]	0.13*** [23.09]	0.09*** [24.02]	0.24*** [28.19]	0.15*** [24.59]
D_5	0.10*** [28.76]	0.06*** [27.53]	0.02*** [6.67]	0.04*** [12.05]	0.14*** [19.49]	0.09*** [20.45]	0.24*** [23.49]	0.14*** [19.86]
D_6	0.10*** [37.48]	0.07*** [35.95]	0.03*** [13.08]	0.03*** [15.97]	0.12*** [22.08]	0.09*** [22.38]	0.23*** [28.21]	0.15*** [24.60]
D_7	0.10*** [38.88]	0.07*** [37.47]	0.03*** [12.80]	0.03*** [15.32]	0.12*** [23.44]	0.09*** [24.04]	0.23*** [28.79]	0.15*** [25.19]
D_8	0.10*** [37.39]	0.07*** [35.03]	0.03*** [11.85]	0.03*** [13.67]	0.12*** [22.44]	0.09*** [22.65]	0.23*** [28.30]	0.15*** [24.45]
D_9	0.10*** [38.78]	0.06*** [36.17]	0.03*** [12.93]	0.03*** [14.24]	0.12*** [23.48]	0.09*** [23.74]	0.22*** [28.70]	0.15*** [24.74]
PT	−0.03 [-1.01]		0.72*** [26.24]		−0.41*** [-5.89]		−1.49*** [-14.10]	
PV		0.43*** [21.50]		0.71*** [38.70]		0.08* [1.71]		−0.42*** [-5.68]
N	3,940,869	3,940,869	1,278,821	1,278,821	1,313,990	1,313,990	1,348,058	1,348,058
adj. R^2	0.005	0.005	0.009	0.009	0.005	0.005	0.005	0.004
F-value	6,229.399	6,080.963	1,902.145	1,902.493	2,109.104	2,024.612	2,246.426	2,178.863
$PT(PV)$ = 1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$D_{0\ldots9}$ = 0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\max_j (D_0 \geq D_j)$	0.981	0.000	0.734	0.000	1.000	0.950	1.000	0.000

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ *(robust se), t statistics in brackets*

Table 15: Test of Hypotheses - Deutsche Börse

This table reports the estimated parameters of the following regressions:

$$PC_k^{s,t} = \boldsymbol{\alpha}_{0,\ldots,9}\boldsymbol{D}_{0,\ldots,9}^{s,t} + \beta PT_k^{s,t} + \varepsilon_i^{s,t},\ k \in (0,...,9)$$
$$PC_k^{s,t} = \boldsymbol{\alpha}_{0,\ldots,9}\boldsymbol{D}_{0,\ldots,9}^{s,t} + \beta PV_k^{s,t} + \varepsilon_i^{s,t},\ k \in (0,...,9)$$

$PC_k^{s,t}$ is the proportion of the cumulative return of all trades for a given stock s for the time period t attributable to a given category k. $D_k^{s,t}$ is a vector dummy variable equal to 1 when its k corresponds to the k of the other variables. $PT_k^{s,t}$ is the proportion of the number of trades for a given stock s for the time period t of a given category k. $PV_k^{s,t}$ is the proportion of shares traded for a given stock s for the time period t for a given category k. Trades are categorized by their last digit of trade size (number of shares transacted for each individual trade), pertaining to the intraday continuous trading period for my sample of 103 stocks traded on Deutsche Börse from January 2, 2002 to December 30, 2015. The regressions are pooled weighted according to the relative importance of the absolute value of the cumulative return of all trades for a given stock s for the time period t. The number of observations (N) is as indicated.

	All Firms		Firms with small trade volume		Firms with medium trade volume		Firms with large trade volume	
	PIH	TVH	PIH	TVH	PIH	TVH	PIH	TVH
D_0	–0.09*** [-16.60]	–0.17*** [-32.18]	0.02** [2.45]	–0.01* [-1.93]	–0.05*** [-5.90]	–0.08*** [-9.77]	–0.29*** [-20.09]	–0.41*** [-32.00]
D_1	0.03*** [19.51]	0.02*** [17.34]	0.01*** [3.17]	0.02*** [7.47]	0.01*** [5.47]	0.02*** [7.27]	0.06*** [17.58]	0.04*** [13.41]
D_2	0.03*** [18.31]	0.02*** [14.80]	0.01*** [2.59]	0.01*** [5.92]	0.01*** [3.98]	0.01*** [5.07]	0.06*** [17.27]	0.04*** [12.54]
D_3	0.03*** [17.56]	0.02*** [13.43]	0.01*** [2.64]	0.01*** [5.61]	0.01*** [3.63]	0.01*** [4.34]	0.06*** [16.29]	0.03*** [11.24]
D_4	0.03*** [20.13]	0.02*** [15.31]	0.01*** [4.18]	0.01*** [6.55]	0.01*** [4.95]	0.01*** [5.15]	0.06*** [17.72]	0.04*** [12.60]
D_5	0.03*** [15.80]	0.02*** [13.49]	0.01*** [5.31]	0.02*** [9.18]	0.01*** [5.07]	0.02*** [6.83]	0.04*** [11.40]	0.02*** [7.25]
D_6	0.03*** [20.51]	0.02*** [14.99]	0.01*** [3.93]	0.01*** [5.59]	0.01*** [4.61]	0.01*** [4.43]	0.06*** [18.45]	0.04*** [13.27]
D_7	0.03*** [20.64]	0.02*** [15.08]	0.01*** [4.27]	0.01*** [5.76]	0.01*** [5.65]	0.01*** [5.44]	0.06*** [17.69]	0.04*** [12.43]
D_8	0.03*** [20.75]	0.02*** [14.72]	0.01*** [3.60]	0.01*** [4.77]	0.01*** [4.97]	0.01*** [4.45]	0.07*** [19.09]	0.04*** [13.87]
D_9	0.03*** [19.86]	0.02*** [13.34]	0.01*** [4.06]	0.01*** [4.93]	0.01*** [4.09]	0.01*** [3.38]	0.06*** [18.27]	0.04*** [12.66]
PT	0.83*** [53.94]		0.90*** [49.08]		0.95*** [46.66]		0.75*** [18.48]	
PV		1.00*** [84.69]		0.89*** [56.56]		0.97*** [53.74]		1.09*** [35.11]
N	2,956,957	2,956,957	897,797	897,797	974,790	974,790	1,084,370	1,084,370
adj. R^2	0.008	0.009	0.016	0.017	0.012	0.013	0.004	0.005
F-value	5,533.380	5,866.134	1,728.160	1,805.624	1,848.710	1,935.227	2,302.308	2,380.613
$PT(PV)$ = 1	0.000	0.719	0.000	0.000	0.011	0.165	0.000	0.003
$D_{0\ldots9}$ = 0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\max_j(D_0 \geq D_j)$	0.000	0.000	0.966	0.000	0.000	0.000	0.000	0.000

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ *(robust se), t statistics in brackets*

Table 16: Test of Hypotheses - London Stock Exchange

This table reports the estimated parameters of the following regressions:

$$PC_k^{s,t} = \boldsymbol{\alpha}_{0,\dots,9}\boldsymbol{D}_{0,\dots,9}^{s,t} + \beta PT_k^{s,t} + \varepsilon_i^{s,t},\ k \in (0,\dots,9)$$
$$PC_k^{s,t} = \boldsymbol{\alpha}_{0,\dots,9}\boldsymbol{D}_{0,\dots,9}^{s,t} + \beta PV_k^{s,t} + \varepsilon_i^{s,t},\ k \in (0,\dots,9)$$

$PC_k^{s,t}$ is the proportion of the cumulative return of all trades for a given stock s for the time period t attributable to a given category k. $D_k^{s,t}$ is a vector dummy variable equal to 1 when its k corresponds to the k of the other variables. $PT_k^{s,t}$ is the proportion of the number of trades for a given stock s for the time period t of a given category k. $PV_k^{s,t}$ is the proportion of shares traded for a given stock s for the time period t for a given category k. Trades are categorized by their last digit of trade size (number of shares transacted for each individual trade), pertaining to the intraday continuous trading period for my sample of 182 stocks traded on London Stock Exchange from January 2, 2002 to December 31, 2015. The regressions are pooled weighted according to the relative importance of the absolute value of the cumulative return of all trades for a given stock s for the time period t. The number of observations (N) is as indicated. The t-values are provided in brackets for the corresponding estimated parameter.

	All Firms		Firms with small trade volume		Firms with medium trade volume		Firms with large trade volume	
	PIH	TVH	PIH	TVH	PIH	TVH	PIH	TVH
D_0	–0.10*** [-6.77]	0.06*** [5.73]	–0.05*** [-3.81]	0.12*** [11.01]	–0.08*** [-4.08]	0.12*** [9.74]	–0.23*** [-3.99]	–0.05 [-1.39]
D_1	0.05*** [13.43]	0.09*** [28.00]	0.02*** [3.30]	0.07*** [16.53]	0.04*** [6.43]	0.09*** [21.24]	0.10*** [8.19]	0.13*** [14.18]
D_2	0.05*** [12.47]	0.09*** [25.69]	0.02*** [3.20]	0.07*** [14.25]	0.04*** [6.30]	0.09*** [17.66]	0.09*** [7.63]	0.12*** [14.09]
D_3	0.05*** [11.68]	0.09*** [23.18]	0.01 [1.53]	0.06*** [11.16]	0.05*** [7.11]	0.09*** [21.04]	0.10*** [7.77]	0.13*** [12.66]
D_4	0.05*** [12.10]	0.09*** [30.40]	0.01*** [3.33]	0.06*** [22.15]	0.04*** [6.92]	0.09*** [22.42]	0.09*** [7.04]	0.13*** [15.58]
D_5	0.05*** [8.20]	0.09*** [16.07]	0.01*** [2.69]	0.06*** [21.35]	0.03*** [3.44]	0.08*** [11.01]	0.09*** [5.13]	0.13*** [7.48]
D_6	0.05*** [12.40]	0.09*** [24.83]	0.01** [2.42]	0.06*** [11.48]	0.04*** [5.54]	0.09*** [15.32]	0.09*** [8.40]	0.13*** [15.26]
D_7	0.05*** [10.15]	0.08*** [19.92]	0.01 [1.36]	0.06*** [8.68]	0.05*** [6.31]	0.09*** [15.30]	0.08*** [6.58]	0.11*** [11.29]
D_8	0.06*** [13.12]	0.10*** [24.63]	0.02*** [4.25]	0.07*** [18.48]	0.04*** [6.53]	0.09*** [19.54]	0.11*** [7.94]	0.14*** [12.62]
D_9	0.05*** [11.99]	0.09*** [22.35]	0.02*** [2.80]	0.07*** [9.10]	0.04*** [7.23]	0.09*** [22.77]	0.09*** [7.34]	0.12*** [12.65]
PT	0.63*** [14.55]		0.92*** [21.81]		0.70*** [9.18]		0.40*** [2.86]	
PV		0.11*** [3.17]		0.32*** [11.59]		0.07 [1.59]		–0.11 [-1.01]
N	4,952,698	4,952,698	1,577,909	1,577,909	1,626,729	1,626,729	1,748,060	1,748,060
adj. R^2	0.000	0.000	0.000	0.000	0.002	0.002	0.000	0.000
F-value	922.281	985.466	680.710	688.422	720.973	725.649	311.808	314.892
$PT(PV)$ = 1	0.000	0.000	0.052	0.000	0.000	0.000	0.000	0.000
$D_{0\dots9}$ = 0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\max_j (D_0 \geq D_j)$	0.000	0.018	0.000	1.000	0.000	1.000	0.000	0.000

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ *(robust se), t statistics in brackets*

Figure 2: Trades per last digit of trade size

The figure plots in percentage the proportion of trades as categorized by the last digit of trade size (0, ..., 9) over time. Even (odd) digits other than 0 (5) are each plotted separately using the same color.

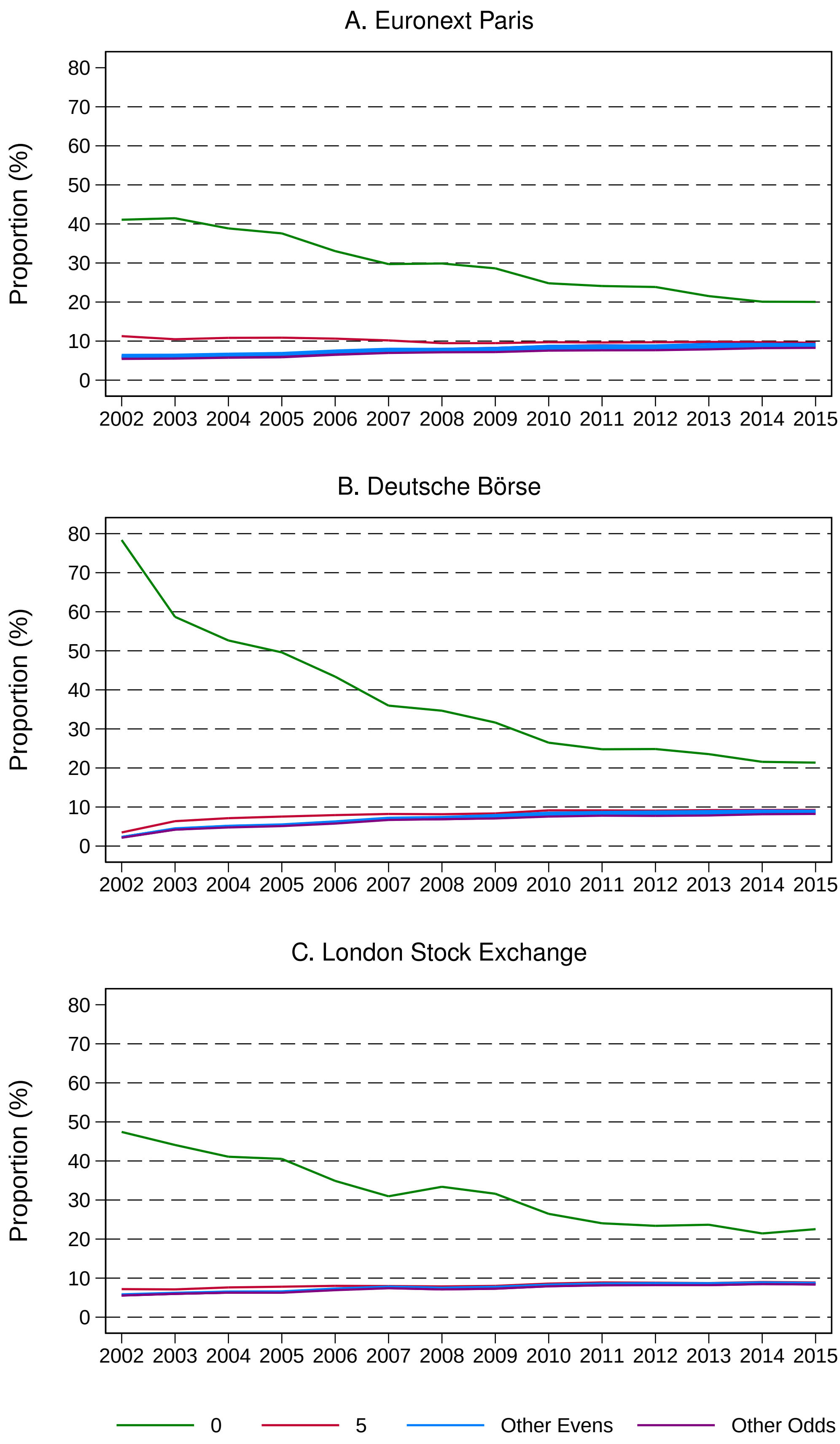

Figure 3: Volume per last digit of trade size

The figure plots in percentage the proportion of volume (number of shares) as categorized by the last digit of trade size (0, ..., 9) over time. Even (odd) digits other than 0 (5) are each plotted separately using the same color.

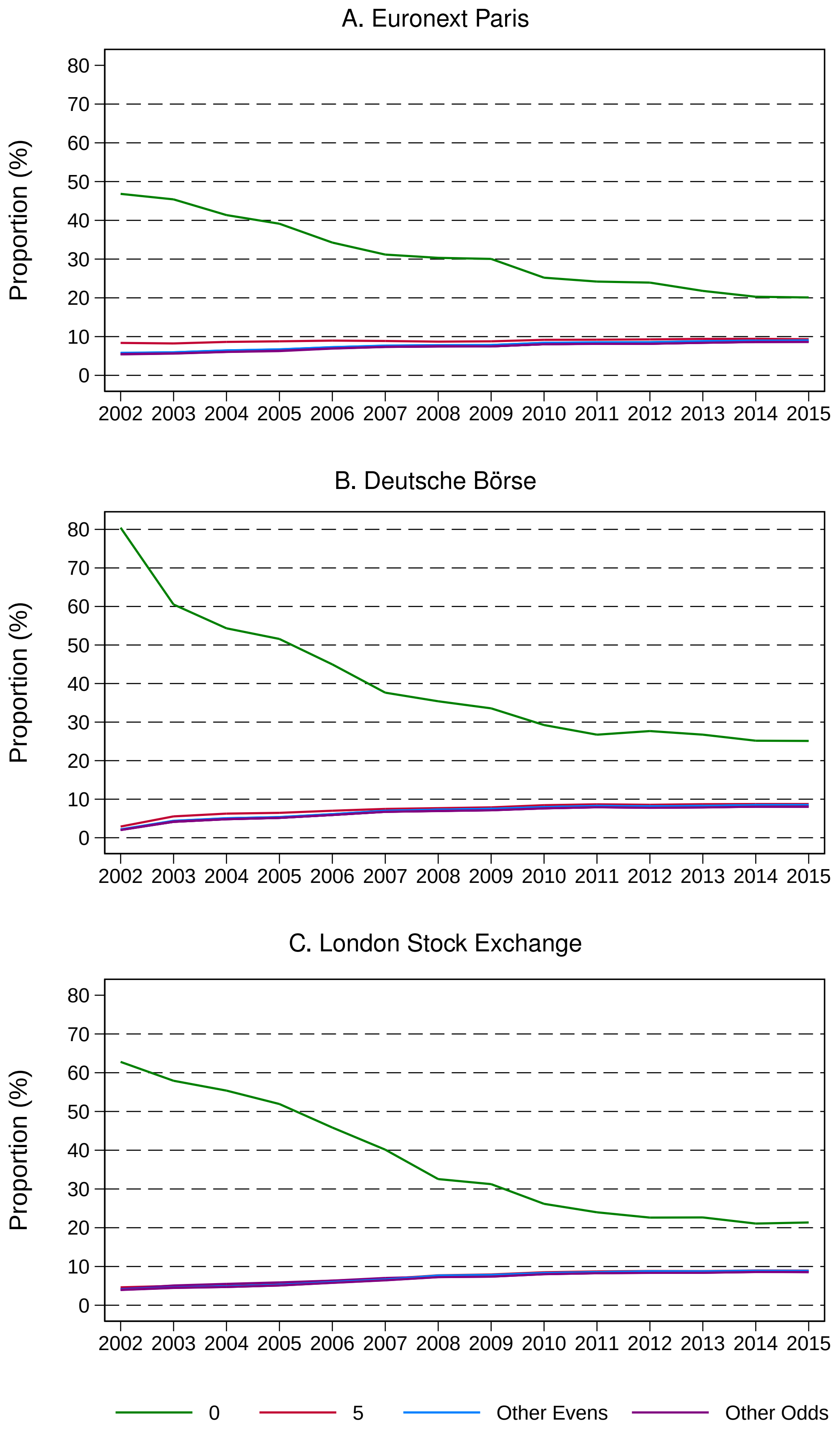

Figure 4: WPC per last digit of trade size

The figure plots in percentage the weighted price contribution by category of last digit of trade size (0, ..., 9) over time. Even (odd) digits other than 0 (5) are each plotted separately using the same color.

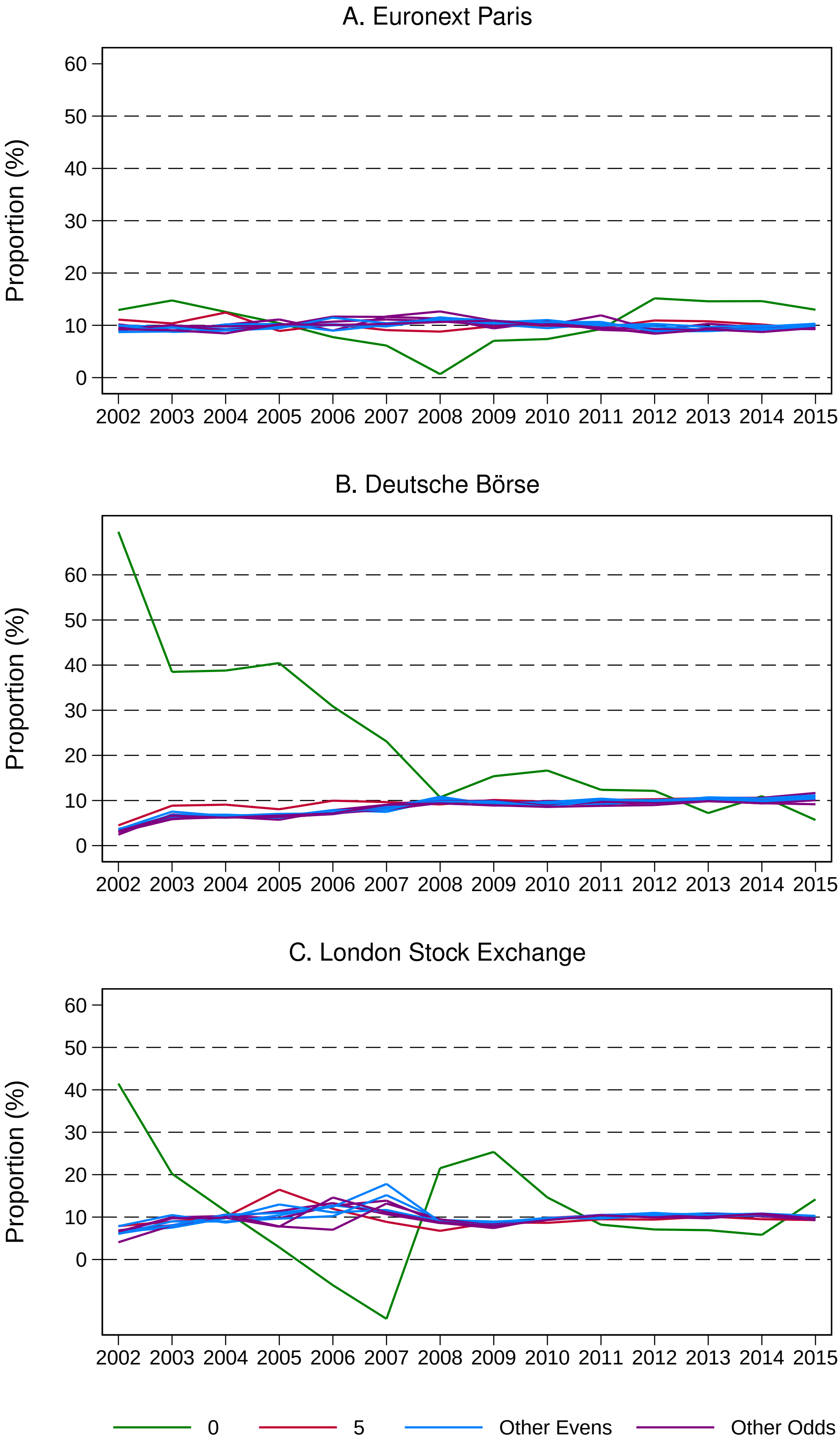

Figure 5: Last digit 0 of trade size

The figure plots in percentage the proportion of trades, of volume, and the weighted price contribution for the last digit of trade size 0 category, over time. A 95% confidence interval is presented for the WPC and same for the proportion of trades or volume would be about the width of the line used.

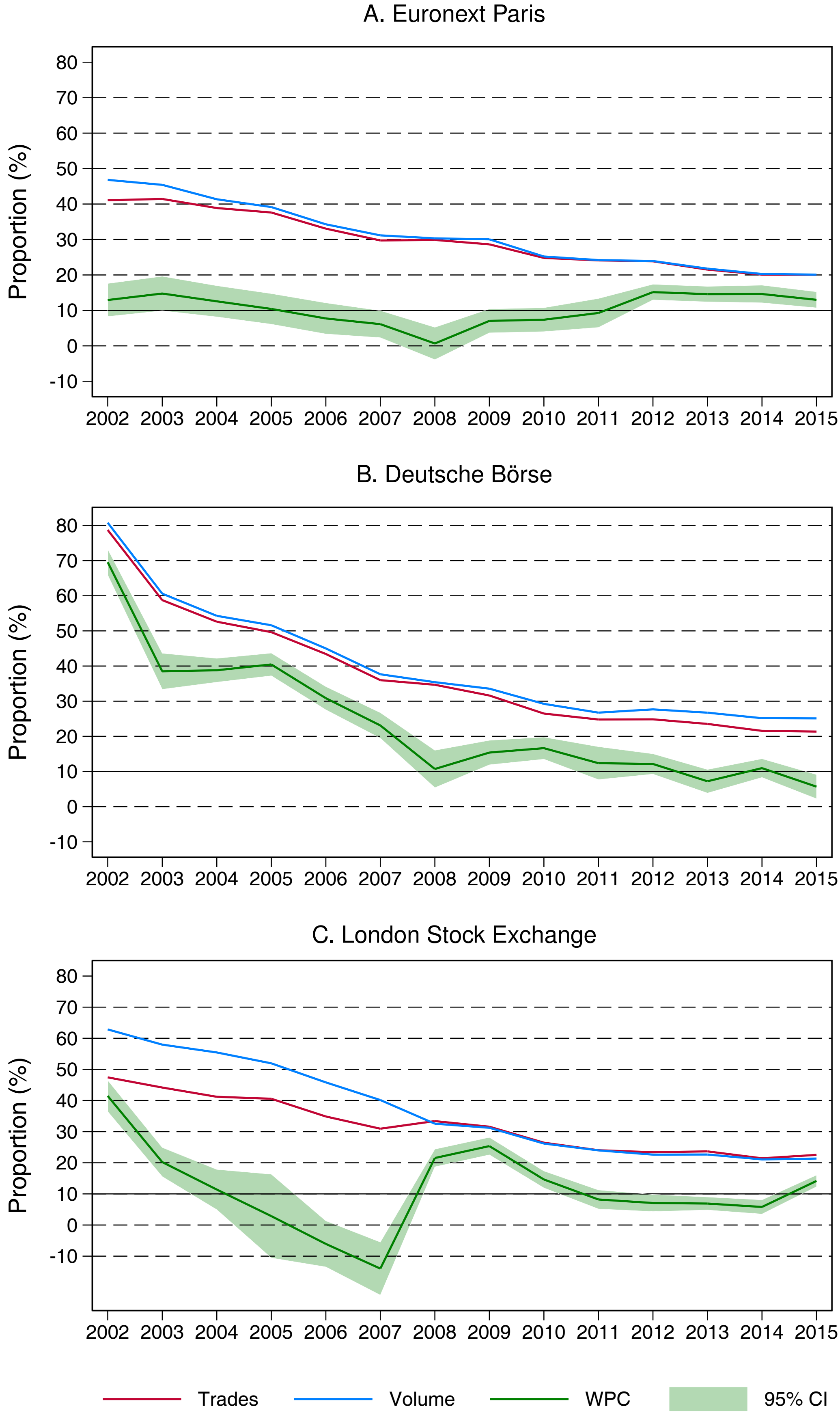

Table 17: Summary of Formal Testing using Annual Samples

This table summarizes the outcome of formally testing the LDH against the PIH and the TVH using annual samples, as per findings detailed in Tables 18, 19, 20, 21, 22 and 23 (for a 5% level of statistical significance). For easy comparison, the results of testing using whole samples are provided in the first column of the table.

	All firms	All firms
	Whole samples	Annual Samples
Euronext Paris		
PIH vs LDH	Reject both	Reject both 14 years out of 14
TVH vs LDH	Reject TVH Not reject LDH	Reject TVH 14 years out of 14 Not reject LDH 12 years out of 14
Deutsche Börse Xetra		
PIH vs LDH	Reject PIH Not reject LDH	Reject PIH 14 years out of 14 Not reject LDH 6 years out of 14
TVH vs LDH	Reject TVH Not reject LDH	Reject TVH 14 years out of 14 Not reject LDH 14 years out of 14
London Stock Exchange		
PIH vs LDH	Reject PIH Not reject LDH	Reject PIH 14 years out of 14 Not reject LDH 7 years out of 14
TVH vs LDH	Reject TVH Not reject LDH	Reject TVH 14 years out of 14 Not reject LDH 3 years out of 14

Table 18: Test of PIH Hypothesis - Euronext Paris

This table reports the estimated parameters of the following regression for each sample-year:

$$PC_k^{s,t} = \boldsymbol{\alpha}_{0,\ldots,9}\boldsymbol{D}_{0,\ldots,9}^{s,t} + \beta PT_k^{s,t} + \varepsilon_i^{s,t},\ k \in (0,\ldots,9)$$

$PC_k^{s,t}$ is the proportion of the cumulative return of all trades for a given stock s for the time period t attributable to a given category k. $D_k^{s,t}$ is a vector dummy variable equal to 1 when its k corresponds to the k of the other variables. $PT_k^{s,t}$ is the proportion of the number of trades for a given stock s for the time period t of a given category k. Trades are categorized by their last digit of trade size (number of shares transacted for each individual trade), pertaining to the intraday continuous trading period for my sample of 131 stocks traded on Euronext Paris from January 2, 2002 to December 31, 2015. The regressions are pooled weighted according to the relative importance of the absolute value of the cumulative return of all trades for a given stock s for the time period t. The number of observations (N) is as indicated.

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
D_0	0.08	0.00	0.09	–0.01	0.29***	0.19***	0.26***	0.06	0.14***	0.11***	0.02	0.12***	0.18***	0.09***
D_1	0.09***	0.07***	0.10***	0.09***	0.13***	0.15***	0.19***	0.11***	0.12***	0.13***	0.05***	0.09***	0.11***	0.08***
D_2	0.08***	0.07***	0.09***	0.08***	0.14***	0.14***	0.18***	0.10***	0.13***	0.11***	0.06***	0.09***	0.11***	0.08***
D_3	0.09***	0.08***	0.09***	0.08***	0.16***	0.15***	0.17***	0.09***	0.13***	0.10***	0.05***	0.08***	0.11***	0.08***
D_4	0.09***	0.08***	0.09***	0.08***	0.15***	0.13***	0.17***	0.10***	0.12***	0.11***	0.05***	0.08***	0.11***	0.09***
D_5	0.10***	0.07***	0.12***	0.06***	0.17***	0.14***	0.17***	0.10***	0.12***	0.10***	0.06***	0.09***	0.12***	0.07***
D_6	0.08***	0.08***	0.09***	0.08***	0.14***	0.13***	0.18***	0.11***	0.13***	0.11***	0.06***	0.08***	0.10***	0.08***
D_7	0.09***	0.08***	0.08***	0.08***	0.15***	0.14***	0.17***	0.10***	0.12***	0.10***	0.05***	0.08***	0.11***	0.08***
D_8	0.09***	0.07***	0.08***	0.08***	0.16***	0.14***	0.17***	0.10***	0.13***	0.11***	0.05***	0.08***	0.11***	0.08***
D_9	0.09***	0.07***	0.08***	0.09***	0.14***	0.13***	0.16***	0.11***	0.12***	0.10***	0.05***	0.08***	0.10***	0.08***
PT	0.12	0.34***	0.08	0.30**	–0.62***	–0.43***	–0.82***	0.04	–0.25	–0.08	0.50***	0.13	–0.16	0.20
N	212,690	222,298	237,210	259,818	280,140	297,618	306,900	306,590	314,550	313,380	308,466	300,264	294,720	286,225
adj. R^2	0.004	0.005	0.004	0.004	0.003	0.003	0.004	0.005	0.005	0.004	0.007	0.008	0.009	0.009
F-value	334.074	338.054	304.339	346.169	333.885	411.494	555.075	577.177	542.331	445.899	572.756	609.642	625.636	759.831
$PT(PV)$ = 1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$D_{0\ldots9}$ = 0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\max_j (D_0 \geq D_j)$	0.474	0.051	0.551	0.046	0.999	0.977	0.994	0.153	0.804	0.734	0.095	0.989	0.997	0.820

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ (robust se)

Table 19: Test of TVH Hypothesis - Euronext Paris

This table reports the estimated parameters of the following regressionfor each sample-year:

$$PC_k^{s,t} = \boldsymbol{\alpha}_{0,\dots,9}\boldsymbol{D}_{0,\dots,9}^{s,t} + \beta PV_k^{s,t} + \varepsilon_i^{s,t},\ k \in (0, ..., 9)$$

$PC_k^{s,t}$ is the proportion of the cumulative return of all trades for a given stock s for the time period t attributable to a given category k. $D_k^{s,t}$ is a vector dummy variable equal to 1 when its k corresponds to the k of the other variables. $PV_k^{s,t}$ is the proportion of shares traded for a given stock s for the time period t of a given category k. Trades are categorized by their last digit of trade size (number of shares transacted for each individual trade), pertaining to the intraday continuous trading period for my sample of 131 stocks traded on Euronext Paris from January 2, 2002 to December 31, 2015. The regressions are pooled weighted according to the relative importance of the absolute value of the cumulative return of all trades for a given stock s for the time period t. The number of observations (N) is as indicated.

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
D_0	–0.18***	–0.26***	–0.09***	–0.20***	–0.13***	–0.16***	–0.14***	–0.11***	–0.04	–0.10***	–0.06***	0.03*	0.06**	0.00
D_1	0.07***	0.04***	0.07***	0.06***	0.05***	0.07***	0.09***	0.07***	0.07***	0.06***	0.02***	0.06***	0.06***	0.05***
D_2	0.05***	0.04***	0.07***	0.05***	0.05***	0.05***	0.08***	0.06***	0.08***	0.04***	0.03***	0.05***	0.06***	0.05***
D_3	0.06***	0.05***	0.07***	0.05***	0.08***	0.07***	0.08***	0.05***	0.07***	0.03***	0.02***	0.05***	0.06***	0.04***
D_4	0.06***	0.05***	0.07***	0.05***	0.06***	0.05***	0.07***	0.06***	0.06***	0.04***	0.02***	0.05***	0.06***	0.05***
D_5	0.06***	0.03***	0.08***	0.02**	0.05***	0.03***	0.05***	0.05***	0.06***	0.03**	0.03***	0.06***	0.06***	0.04***
D_6	0.06***	0.05***	0.06***	0.05***	0.06***	0.05***	0.08***	0.07***	0.07***	0.04***	0.03***	0.05***	0.06***	0.05***
D_7	0.06***	0.05***	0.06***	0.05***	0.07***	0.06***	0.07***	0.06***	0.07***	0.03***	0.03***	0.05***	0.06***	0.04***
D_8	0.06***	0.04***	0.06***	0.05***	0.07***	0.05***	0.07***	0.06***	0.07***	0.04***	0.02**	0.05***	0.06***	0.04***
D_9	0.06***	0.04***	0.06***	0.06***	0.06***	0.05***	0.07***	0.07***	0.07***	0.03***	0.02**	0.05***	0.05***	0.04***
PV	0.64***	0.87***	0.51***	0.75***	0.60***	0.70***	0.47***	0.57***	0.42***	0.76***	0.84***	0.50***	0.40***	0.60***
N	212,690	222,298	237,210	259,818	280,140	297,618	306,900	306,590	314,550	313,380	308,466	300,264	294,720	286,225
adj. R^2	0.004	0.005	0.004	0.004	0.003	0.003	0.004	0.005	0.005	0.004	0.008	0.008	0.010	0.009
F-value	332.499	342.434	296.538	348.179	323.408	402.168	550.372	571.831	528.138	422.819	552.902	580.927	600.654	747.841
$PT(PV) = 1$	0.000	0.043	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.044	0.060	0.000	0.000	0.000
$D_{0...9} = 0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\max_j (D_0 \geq D_j)$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.096	0.675	0.001

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ (robust se)

Table 20: Test of PIH Hypothesis - Deutsche Börse

This table reports the estimated parameters of the following regression for each sample-year:

$$PC_k^{s,t} = \boldsymbol{\alpha}_{0,\ldots,9}\boldsymbol{D}_{0,\ldots,9}^{s,t} + \beta PT_k^{s,t} + \varepsilon_i^{s,t},\ k \in (0,\ldots,9)$$

$PC_k^{s,t}$ is the proportion of the cumulative return of all trades for a given stock s for the time period t attributable to a given category k. $D_k^{s,t}$ is a vector dummy variable equal to 1 when its k corresponds to the k of the other variables. $PT_k^{s,t}$ is the proportion of the number of trades for a given stock s for the time period t of a given category k. Trades are categorized by their last digit of trade size (number of shares transacted for each individual trade), pertaining to the intraday continuous trading period for my sample of 103 stocks traded on Deutsche Börse from January 2, 2002 to December 30, 2015. The regressions are pooled weighted according to the relative importance of the absolute value of the cumulative return of all trades for a given stock s for the time period t. The number of observations (N) is as indicated.

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
D_0	–0.33***	–0.29***	–0.14***	–0.04	–0.17***	–0.03	0.09	0.13***	–0.04	0.09	0.20***	0.24***	0.15***	0.15***
D_1	0.00	0.02***	0.02***	0.01***	0.02**	0.04***	0.10***	0.09***	0.04***	0.09***	0.12***	0.16***	0.12***	0.15***
D_2	0.00*	0.02***	0.02***	0.02***	0.01**	0.02***	0.10***	0.09***	0.03***	0.09***	0.12***	0.17***	0.12***	0.15***
D_3	0.00	0.01***	0.02***	0.02***	0.01	0.03***	0.09***	0.08***	0.03***	0.08***	0.12***	0.16***	0.11***	0.14***
D_4	0.01***	0.02***	0.02***	0.02***	0.02**	0.03***	0.10***	0.09***	0.03***	0.09***	0.12***	0.16***	0.11***	0.15***
D_5	0.00	0.02***	0.02***	0.02**	0.02**	0.04***	0.09***	0.10***	0.03**	0.09***	0.13***	0.17***	0.12***	0.14***
D_6	0.01***	0.02***	0.02***	0.02***	0.01*	0.03***	0.11***	0.09***	0.04***	0.08***	0.12***	0.16***	0.11***	0.14***
D_7	0.00	0.01***	0.02***	0.02***	0.01	0.04***	0.10***	0.10***	0.03***	0.09***	0.12***	0.15***	0.11***	0.14***
D_8	0.01***	0.03***	0.02***	0.03***	0.01	0.04***	0.09***	0.09***	0.03**	0.09***	0.12***	0.16***	0.11***	0.14***
D_9	0.01**	0.02***	0.02***	0.03***	0.01**	0.04***	0.09***	0.09***	0.03***	0.08***	0.11***	0.15***	0.11***	0.13***
PT	1.30***	1.13***	0.96***	0.85***	1.06***	0.71***	0.04	0.06	0.76***	0.14	–0.29	–0.67**	–0.16	–0.42*
N	142,610	154,050	170,770	199,830	214,190	230,680	237,160	229,250	235,070	237,760	230,817	228,398	223,673	222,699
adj. R^2	0.068	0.015	0.017	0.014	0.009	0.005	0.004	0.007	0.007	0.004	0.006	0.006	0.007	0.007
F-value	1,654.954	301.638	350.487	336.776	361.574	345.512	475.911	492.277	479.163	395.559	491.066	510.387	562.624	597.401
$PT(PV) = 1$	0.000	0.079	0.489	0.029	0.524	0.012	0.000	0.000	0.057	0.002	0.000	0.000	0.000	0.000
$D_{0\ldots9} = 0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\max_j (D_0 \geq D_j)$	0.000	0.000	0.000	0.067	0.000	0.045	0.529	0.912	0.001	0.590	0.998	0.981	0.982	0.845

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ (robust se)

Table 21: Test of TVH Hypothesis - Deutsche Börse

This table reports the estimated parameters of the following regression for each sample-year:

$$PC_k^{s,t} = \boldsymbol{\alpha}_{0,\ldots,9}\boldsymbol{D}_{0,\ldots,9}^{s,t} + \beta PV_k^{s,t} + \varepsilon_i^{s,t},\ k \in (0,\ldots,9)$$

$PC_k^{s,t}$ is the proportion of the cumulative return of all trades for a given stock s for the time period t attributable to a given category k. $D_k^{s,t}$ is a vector dummy variable equal to 1 when its k corresponds to the k of the other variables. $PV_k^{s,t}$ is the proportion of shares traded for a given stock s for the time period t of a given category k. Trades are categorized by their last digit of trade size (number of shares transacted for each individual trade), pertaining to the intraday continuous trading period for my sample of 103 stocks traded on Deutsche Börse from January 2, 2002 to December 30, 2015. The regressions are pooled weighted according to the relative importance of the absolute value of the cumulative return of all trades for a given stock s for the time period t. The number of observations (N) is as indicated.

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
D_0	-0.46^{***}	-0.49^{***}	-0.32^{***}	-0.20^{***}	-0.24^{***}	-0.14^{***}	-0.25^{***}	-0.10^{***}	-0.13^{***}	-0.09^{***}	0.01	0.02	0.01	0.01
D_1	0.00	0.01^{**}	0.01^{*}	0.00	0.01^{**}	0.03^{***}	0.04^{***}	0.04^{***}	0.02^{***}	0.04^{***}	0.07^{***}	0.09^{***}	0.07^{***}	0.10^{***}
D_2	0.00	0.01	0.01^{**}	0.01	0.01	0.01	0.03^{***}	0.04^{***}	0.02^{***}	0.04^{***}	0.06^{***}	0.09^{***}	0.07^{***}	0.10^{***}
D_3	0.00	0.00	0.01^{*}	0.01	0.00	0.02^{**}	0.03^{**}	0.04^{***}	0.02^{**}	0.03^{***}	0.06^{***}	0.08^{***}	0.06^{***}	0.09^{***}
D_4	0.01^{***}	0.00	0.01^{*}	0.00	0.01^{*}	0.01^{**}	0.04^{***}	0.04^{***}	0.02^{**}	0.04^{***}	0.07^{***}	0.09^{***}	0.07^{***}	0.09^{***}
D_5	0.00	0.01	0.01^{**}	0.01	0.02^{***}	0.03^{***}	0.02	0.04^{***}	0.01^{*}	0.04^{***}	0.07^{***}	0.09^{***}	0.07^{***}	0.09^{***}
D_6	0.01^{**}	0.01	0.01	0.01^{***}	0.00	0.02^{**}	0.04^{***}	0.04^{***}	0.02^{***}	0.03^{***}	0.07^{***}	0.09^{***}	0.07^{***}	0.09^{***}
D_7	0.00	0.00	0.00	0.01^{*}	0.00	0.02^{***}	0.03^{***}	0.05^{***}	0.01^{**}	0.04^{***}	0.06^{***}	0.09^{***}	0.06^{***}	0.09^{***}
D_8	0.01^{**}	0.01^{***}	0.00	0.01^{***}	0.00	0.02^{***}	0.03^{**}	0.04^{***}	0.01	0.04^{***}	0.07^{***}	0.09^{***}	0.07^{***}	0.09^{***}
D_9	0.00	0.01^{*}	0.00	0.01^{**}	0.00	0.03^{***}	0.03^{**}	0.04^{***}	0.01^{*}	0.03^{***}	0.06^{***}	0.08^{***}	0.06^{***}	0.08^{***}
PV	1.43^{***}	1.43^{***}	1.25^{***}	1.12^{***}	1.18^{***}	0.96^{***}	0.98^{***}	0.73^{***}	0.99^{***}	0.76^{***}	0.41^{***}	0.20	0.40^{***}	0.18
N	142,610	154,050	170,770	199,830	214,190	230,680	237,160	229,250	235,070	237,760	230,817	228,398	223,673	222,699
adj. R^2	0.069	0.016	0.019	0.015	0.009	0.005	0.004	0.007	0.007	0.005	0.006	0.006	0.007	0.007
F-value	2,176.534	357.373	393.286	362.421	372.097	348.481	474.101	481.529	470.946	388.225	468.096	499.101	534.581	586.732
$PT(PV) = 1$	0.000	0.000	0.000	0.026	0.008	0.639	0.879	0.004	0.861	0.056	0.000	0.000	0.000	0.000
$D_{0\ldots9} = 0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\max_j (D_0 \geq D_j)$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.000

$^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$ (robust se)

Table 22: Test of PIH Hypothesis - London Stock Exchange

This table reports the estimated parameters of the following regression for each sample-year:

$$PC_k^{s,t} = \boldsymbol{\alpha}_{0,\ldots,9}\boldsymbol{D}_{0,\ldots,9}^{s,t} + \beta PT_k^{s,t} + \varepsilon_i^{s,t},\ k \in (0,...,9)$$

$PC_k^{s,t}$ is the proportion of the cumulative return of all trades for a given stock s for the time period t attributable to a given category k. $D_k^{s,t}$ is a vector dummy variable equal to 1 when its k corresponds to the k of the other variables. $PT_k^{s,t}$ is the proportion of the number of trades for a given stock s for the time period t of a given category k. Trades are categorized by their last digit of trade size (number of shares transacted for each individual trade), pertaining to the intraday continuous trading period for my sample of 182 stocks traded on the London Stock Exchange from January 2, 2002 to December 31, 2015. The regressions are pooled weighted according to the relative importance of the absolute value of the cumulative return of all trades for a given stock s for the time period t. The number of observations (N) is as indicated.

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
D_0	0.08	–0.11*	–0.38***	–0.91***	–0.30***	–0.34**	–0.02	–0.02	0.20***	0.17***	–0.06	0.03	0.11*	0.08**
D_1	0.03***	0.06***	0.03***	–0.04	0.08***	0.09**	0.03***	0.01	0.11***	0.14***	0.06***	0.09***	0.12***	0.07***
D_2	0.04***	0.06***	0.02	–0.04	0.08***	0.07	0.04***	0.02*	0.11***	0.13***	0.06***	0.09***	0.13***	0.07***
D_3	0.03***	0.04***	0.03	–0.02	0.09***	0.06	0.04***	0.02**	0.11***	0.14***	0.06***	0.08***	0.13***	0.07***
D_4	0.03***	0.04**	0.03**	–0.01	0.06***	0.07	0.04***	0.02**	0.11***	0.13***	0.06***	0.09***	0.13***	0.07***
D_5	0.03***	0.04***	0.02	0.00	0.07***	0.04	0.01	0.02*	0.10***	0.13***	0.05**	0.09***	0.12***	0.07***
D_6	0.02*	0.05***	0.02	–0.04	0.06**	0.11**	0.04***	0.03**	0.11***	0.13***	0.06***	0.09***	0.13***	0.07***
D_7	0.00	0.04***	0.04***	–0.06	0.02	0.09*	0.04***	0.02*	0.11***	0.14***	0.06***	0.09***	0.13***	0.07***
D_8	0.02**	0.04***	0.04**	–0.03	0.08***	0.13***	0.04***	0.02	0.11***	0.14***	0.06***	0.09***	0.13***	0.08***
D_9	0.03***	0.06***	0.03**	–0.06	0.10***	0.07	0.04***	0.02	0.11***	0.13***	0.06***	0.09***	0.13***	0.08***
PT	0.67***	0.67***	1.11***	2.20***	0.66**	0.61	0.71***	0.84***	–0.20	–0.38*	0.54**	0.17	–0.25	0.25
N	295,750	321,720	363,320	373,450	380,530	381,330	374,940	363,609	364,930	358,280	354,760	352,080	342,230	325,769
adj. R^2	0.002	0.001	0.001	0.000	0.001	0.000	0.003	0.005	0.005	0.003	0.006	0.007	0.007	0.009
F-value	99.775	148.979	111.698	104.874	120.254	33.507	465.083	431.837	555.517	580.438	513.008	694.155	748.998	641.581
$PT(PV)$ = 1	0.023	0.034	0.513	0.037	0.219	0.460	0.087	0.267	0.000	0.000	0.030	0.000	0.000	0.000
$D_{0...9}$ = 0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\max_j (D_0 \geq D_j)$	0.876	0.008	0.000	0.001	0.000	0.001	0.189	0.197	1.000	0.950	0.000	0.002	0.445	0.689

$^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$ (robust se)

Table 23: Test of TVH Hypothesis - London Stock Exchange

This table reports the estimated parameters of the following regression for each sample-year:

$$PC_k^{s,t} = \boldsymbol{\alpha}_{0,\ldots,9}\boldsymbol{D}_{0,\ldots,9}^{s,t} + \beta PV_k^{s,t} + \varepsilon_i^{s,t},\ k \in (0, ..., 9)$$

$PC_k^{s,t}$ is the proportion of the cumulative of all trades for a given stock s for the time period t attributable to a given category k. $D_k^{s,t}$ is a vector dummy variable equal to 1 when its k corresponds to the k of the other variables. $PV_k^{s,t}$ is the proportion of shares traded for a given stock s for the time period t of a given category k. Trades are categorized by their last digit of trade size (number of shares transacted for each individual trade), pertaining to the intraday continuous trading period for my sample of 182 stocks traded on the London Stock Exchange from January 2, 2002 to December 31, 2015. The regressions are pooled weighted according to the relative importance of the absolute value of the cumulative return of all trades for a given stock s for the time period t. The number of observations (N) is as indicated.

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
D_0	0.36***	0.22***	0.02	–0.39*	–0.13	0.12	0.09***	0.20***	0.20***	0.10***	0.06***	0.10***	0.13***	0.13***
D_1	0.06***	0.10***	0.10***	0.06***	0.12***	0.18***	0.06***	0.06***	0.11***	0.11***	0.10***	0.12***	0.13***	0.09***
D_2	0.08***	0.11***	0.08***	0.06***	0.12***	0.16***	0.06***	0.07***	0.11***	0.11***	0.10***	0.12***	0.14***	0.09***
D_3	0.07***	0.08***	0.09***	0.08***	0.12***	0.15***	0.06***	0.07***	0.11***	0.11***	0.10***	0.11***	0.14***	0.09***
D_4	0.06***	0.08***	0.09***	0.09***	0.10***	0.16***	0.06***	0.08***	0.11***	0.10***	0.10***	0.12***	0.14***	0.09***
D_5	0.08***	0.09***	0.09***	0.12**	0.11***	0.13***	0.04***	0.08***	0.10***	0.10***	0.09***	0.11***	0.13***	0.09***
D_6	0.06***	0.09***	0.08***	0.06***	0.09***	0.19***	0.06***	0.08***	0.11***	0.11***	0.11***	0.11***	0.14***	0.09***
D_7	0.04***	0.08***	0.10***	0.03	0.06**	0.18***	0.07***	0.07***	0.11***	0.11***	0.10***	0.12***	0.14***	0.09***
D_8	0.06***	0.08***	0.11***	0.07***	0.12***	0.22***	0.06***	0.07***	0.11***	0.11***	0.10***	0.12***	0.14***	0.10***
D_9	0.06***	0.10***	0.09***	0.04	0.14***	0.15***	0.06***	0.07***	0.11***	0.11***	0.10***	0.11***	0.14***	0.09***
PV	0.08	–0.03	0.15	0.79**	0.14	–0.64**	0.38***	0.16*	–0.19**	–0.07	0.05	–0.14**	–0.35	0.06
N	295,750	321,720	363,320	373,450	380,530	381,330	374,940	363,609	364,930	358,280	354,760	352,080	342,230	325,769
adj. R^2	0.002	0.001	0.001	0.000	0.001	0.000	0.003	0.005	0.005	0.003	0.006	0.007	0.007	0.009
F-value	99.645	149.838	108.453	104.060	121.032	34.397	458.079	430.379	549.753	569.712	499.526	689.957	741.748	637.989
$PT(PV)$ = 1	0.000	0.000	0.000	0.569	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$D_{0\ldots9}$ = 0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\max_j (D_0 \geq D_j)$	1.000	0.999	0.164	0.018	0.010	0.464	0.984	1.000	1.000	0.473	0.017	0.281	0.622	1.000

$^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$ (robust se)

Figure 6: Relative trade size per last digit of trade size

The figure plots the relative trade size of each last digit of trade size category (0, ..., 9) over time. The relative trade size is the average trade size of a given category divided by the average trade size of the sample. Even (odd) digits other than 0 (5) are each plotted separately using the same color.

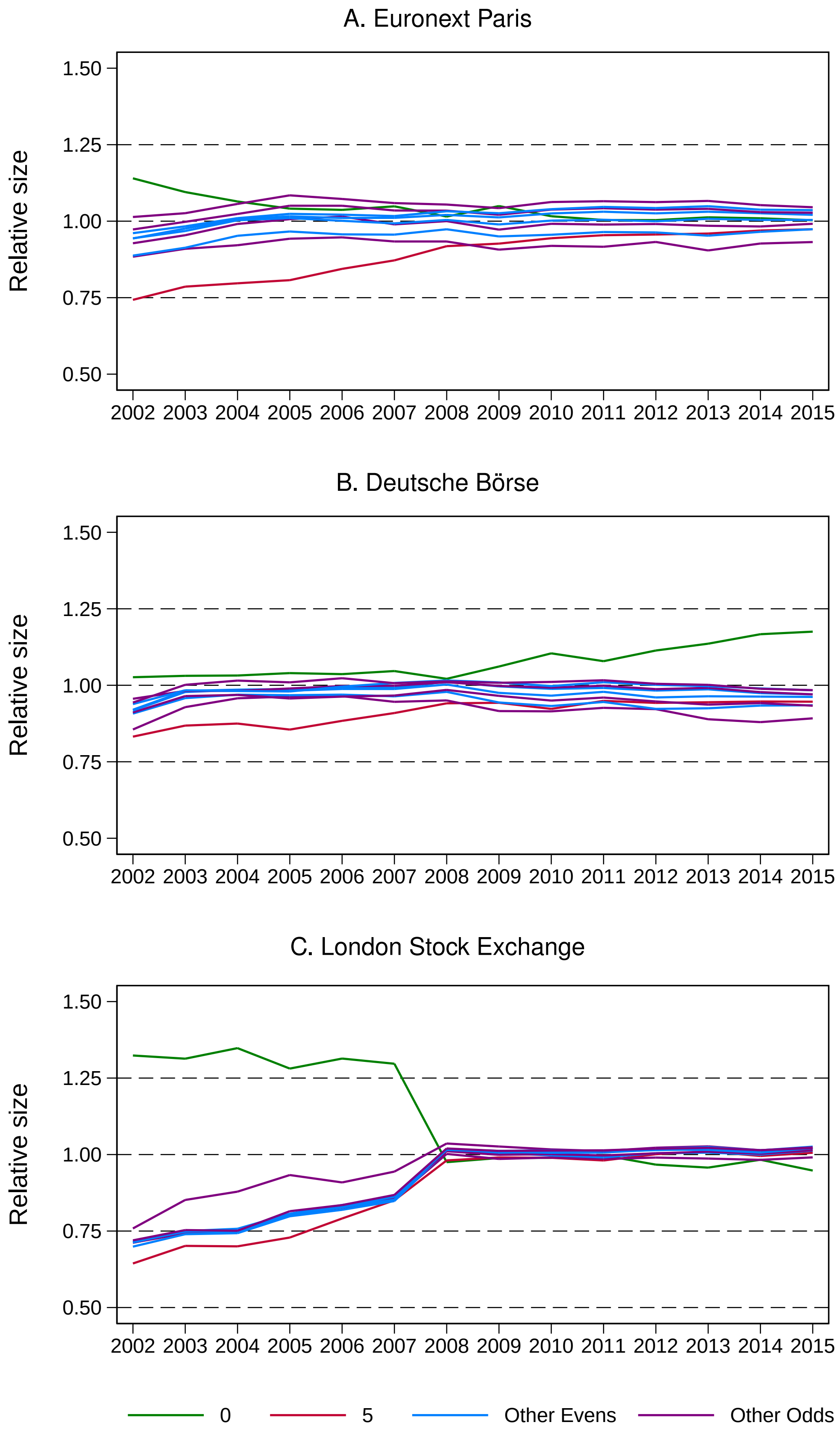

Figure 7: Adjusted R^2 per last digit of trade size

The figure plots the adjusted R^2 of each last digit of trade size category (0, ..., 9) over time. Even (odd) digits other than 0 (5) are each plotted separately using the same color.