Pre-trade trasparency on the Italian Stock Exchange: a trade size model on panel data*

Caterina Lucarelli† Maria Elena Bontempi‡
Camilla Mazzoli§ Anna Grazia Quaranta¶

Abstract

The purpose of this study was to analyze the effects that have been caused by changes in pre-trade transparency upon the behavior of stock traders. We used a trade size model and tested it *before*, *during* and *after* the period when the Italian Stock Exchange introduced a 20-level order book with disaggregated orders. Tick by tick data of the whole set of stocks (up to 277) listed on the Italian Stock Exchange were studied through fixed-effects panel models, within intra day (every 30 minutes and 150 minutes) and daily time frames. Our results indicate that order flows, bid-ask spreads, levels of risk and some information events *differentially* affect trade sizes when investors receive better information prior to negotiation. Both (intra day) informed and uninformed traders operating in a more transparent market became more reticent, with reduced trades sizes and higher orders' cancellations. Moreover, it appears that the higher degree of order book disclosure permits traders to downsize their level of risk aversion; i.e. it reduces the 'uncertainty' that would otherwise result in disrupted trading activity under conditions of information opacity.

JEL Classification: G14; D03; D82; C23; C82.
Key words: pre-trade transparency; trade size; private information; panel data.

*We are grateful to Investnet-IW Bank for providing us with the trading platform used to download the data set. We would like also to thank Roberto Golinelli for his useful comments and advice; Massimiliano Ghedini, Ivan Girotto and CINECA for providing a powerful HPC computer to deal the huge dimension of the dataset.

†Corresponding author, Università Politecnica delle Marche, Department of Economics, Piazzale Martelli 8, I-60121, Italy, tel. +39 071 2207196, fax: +39 071 2207253, *c.lucarelli@univpm.it*.

‡Università degli Studi di Bologna, *mariaelena.bontempi@unibo.it*.

§Università Politecnica delle Marche, *c.mazzoli@univpm.it*.

¶Università degli Studi di Bologna, *annagrazia.quaranta@unbo.it*.

1 Introduction

The way in which private information is impounded in the trading process forms an important aspect of studies into security market microstructures (O'Hara, 1995; Madhavan, 2000; Hasbrouck, 2007). The pre-trade transparency (PTT), inherent to electronic order books, permits the dynamic visibility of prices, quotes and volumes, and increases the level of disclosure in the trading process. As a consequence, it can be argued that when PTT is enhanced, the trading process becomes more informative and the number of market participants able to interpret the trading behavior of informed traders is increased. Furthermore, the higher the level of PTT, the greater the extent that private information held by informed traders can be shared (indirectly, via the trading process) with all the market participants. It is upon this basis that it has been often stated that informed traders resist transparency, while liquidity/noise traders welcome it (among the others Madhavan, 2000).

Since the beginning of electronic trading, the Italian Stock Exchange's (ISE) pre-trade transparency standards have been largely asymmetric between institutional (mainly banks) and retail traders. Whilst institutional traders have had access to transparent order books (unanonymous with disaggregated orders throughout), retail traders have only had access to 5-level anonymous order books containing aggregated orders. The purpose of this study was to reveal how changes in PTT can affect trading behavior. We focused on the effects that an innovative trading device has had upon the market microstructure. The trading device, called 'Book Profondo' (BP), was introduced to the ISE between 1^{st} and 31^{st} July 2007; it was based on a 20-level order book with disaggregated orders and was distributed to all of the market's participants. Even if the asymmetry in anonymity has remained the same, the introduction of the BP represents a relevant improvement in PTT since informed traders are forced to manage the exposure of their orders when facing higher transparency (Bohemer *et al.*, 2005; Harris, 1996). The segment of the market affected by BP is the 'Mercato Telematico Azionario' (MTA) - the ISE traditional equity market. This segment is an electronic order driven market where negotiations are based on a limit order book. The BP innovation increased the amount of information made available during the trading process in terms of enhanced visibility of order flow dynamics, especially in the favor of retail traders. Assuming a higher probability that institutional traders exercise information-based trading (Simaan *et al.*, 2003), we can argue that the BP would have changed the proportion of traders that are able to share private information. This would happen because a larger number of market participants (retail traders) are allowed to tread in the footsteps of informed traders (institutional traders)[1]. Thus, this study considers the behavior of both institutional and retail investors, who have respectively lost and

[1]In Italy, a high level of PTT asymmetry exists between market participants. It is not reasonable that informed, but non-institutional, traders are able to exploit private information using an opaque order book, whilst knowing that their orders are visible to institutional investors. In fact, before the release of BP, some big (informed) retail traders used to trade through the fully transparent platforms of institutional traders (whose names are to be considered confidential, because this procedure was *contralegem*). Thus, in the present study, we have assumed that informed traders always take advantage (either directly or indirectly) of the degree of PTT that is disclosed to institutional investors.

gained advantages following the introduction of BP.

Our models assume that trade sizes reflect the choices of traders formed on the basis of the information they obtain (Chan, 2000; Simaan *et al.*, 2003; Lin *et al.*, 1995). As a consequence, we have tested trade size models under different conditions of market transparency.

The study addressed different time frames: a very short, intra day time frame (the 30 minute analysis), a medium-short intra day time frame (the 150 minute analysis) and a daily time frame. The rationale was to capture the behavior of all traders, whatever their time-horizon, even though the PTT devices are known to be mainly exploited by short-term traders (the so called 'scalpers' or 'momentum' traders). It was therefore expected that the intra day models would best demonstrate the effect of enhanced PTT.

We have considered tick by tick data on the whole set of stocks listed on the ISE and traded on the MTA equity market, for a maximum of 277 shares, over a six months period running between 15^{th} May 2007 and 15^{th} November 2007. This period was divided into three sub-periods, in accordance with the methodology already used by Bohemer *et al.* (2005):

1. from 15^{th} May 2007 to 30^{th} June 2007: prior to the introduction of BP and characterized by the lowest level of pre-trade transparency (Book 1 or B1);

2. from 1^{st} July 2007 to 30^{th} August 2007: during the introduction of BP and characterized by an improvement in the level of pre-trade transparency (Book 2 or B2);

3. finally, from 1^{st} September 2007 to 15^{th} November 2007: following the introduction of BP and with the highest level of pre-trade transparency ever experienced by ISE (Book 3 or B3).

Fixed-effects panel models were tested over the aforementioned time frames, being either intra day (30 minute and 150 minute) or daily analyses. The estimates were run separately for the three sub-periods (B1, B2 and B3) in order to capture how a set of explanatory variables can differentially affect choices made on trade size when investors receive better information prior to trading. For each sub-period and time frame, different models were run to test various indicators of order flows (order intensity or a set of inter quote durations), levels of risk (the Garman and Klass (1980) indicator or the degree of statistical volatility) and, finally, of the disclosure of information events (time series, such as future/spot interest rates or information dummies). The results of these analyses also contributed therefore to identifying the choice of variables that best fit the models for each of the different time frames observed.

Moreover, our methodology adds *toutcourt* information to the wealth of data on market microstructures. In fact, a very few studies published up to now within this field have used panel data based on tick by tick data as we have done. The panel model permits a full explanation of cross-individual heterogeneity and of time-dimension variability. This introduces a third dimension of analysis compared to the single stock and market wide perspectives (see for all Lo and Wang, 2001).

Regarding the intra day time frames, the results prove our models to be statistically informative, especially when pre-trade transparency is at its best, whilst on the daily time frame models weaken as far as both reliability and clearness of results As a general result, we found that order flows, bid-ask spreads, levels of risk and some information events differentially affect trade sizes when all the investors have access to better information prior to negotiation. The introduction of BP induced traders becoming more cautious, as far the disclosure of their trading choices is concerned, and less averse to risk.

In the following sections of this paper, section 2 reviews the main contributions available in literature regarding pre-trade transparency; section 3 describes the main clearing and filtering rules applied to manage the tick by tick data set and to fit it to the panel framework; section 4 describes the variables and the models, and section 5 summarises the main findings obtained; section 6 presents the conclusions of this study and outlines the future lines of research required to deepen our understanding of the effects of PTT.

2 Literature Review

Pre-trade transparency refers to the entitlement of market participants to observe the pending trading interests of others; or, in other words, the exposure of the limit order books' contents (Bohemer, Saar and Yu, 2005). It is often cited in literature that the level of PTT is able to influence the trading behaviour. Previous studies on this topic have mainly focused on the information disclosed to investors in quote driven markets. As a consequence, such studies have addressed whether a change in the PTT is able to influence the behaviour of liquidity providers who play a crucial role in quote driven markets. The main assumption of these studies is that a higher level of transparency (very often associated with the removal of anonymity) allows market makers to discriminate between informed traders and uninformed traders; as a consequence, the cost of trading decreases in terms of the bid-ask spread in accordance with the inventory and with the adverse selection paradigms (Stoll, 1976, 1978; Amihud and Mendelson, 1980). Chowdhry and Nanda (1991) and Forster and George (1992) have both proposed models where a higher level of transparency is provided in order to attract more uninformed traders and discourage insider trading; in these models, market makers charge lower bid-ask spreads because they face lower adverse selection costs. Harris and Schultz (1997) found that market makers fix wider bid-ask spreads when they trade on the anonymous Small Order Execution System (SOES) of the Nasdaq compared to when they trade on the non-anonymous dealer market. Theissen (2003) found similar results for the Frankfurt Stock Exchange showing that when market makers trade on the non-anonymous, floor-based trading system, they can identify uninformed traders and consequently bear less risk. Huang and Stoll (1996) have compared two markets: the NYSE, where market makers can see the limit order book, and the Nasdaq, where they cannot; they found that spreads were generally higher on the Nasdaq.

Nowadays, a large number of stock exchanges worldwide are characterised by order driven markets (or hybrid markets), rather than by quote driven markets. Focault, Moinas and Theissen (2007) analyzed the changes in the Euronext limit order market after its transition to an anonymous order book on 23^{rd} April 2001. They found evidence that after this transition the bid-ask spread significantly decreased. They maintain that the information content of the order book reduces when trading is anonymous. Informed traders react by posting aggressive offers more frequently when their identities are hidden. Bohemer, Saar and Yu (2005) investigated the consequences of increased levels of pre-trade transparency in the NYSE stock exchange upon the trading strategies of both traders and liquidity providers. They found that informed traders react to an increase in the level of transparency by trading more actively; they tend to break up their orders, submitting limited orders of smaller sizes, and they increase cancellation rates in order to avoid any front running of their orders by others.

In contrast with the previous literature, we have developed a trade size model instead of a bid-ask spread model. Many authors use bid-ask spread models to analyze the effects of a change in the anonymity of the order book. In this case, the ISE's PTT anonymity standards were not affected by the introduction of the BP: institutional traders still rely on a non-anonymous order book, while retail traders deal with an anonymous one. Moreover, our main interest is related to the behaviour of traders in relation to the information they hold; assuming that trading at large volume is more likely to be associated with informed trading (Chan, 2000; Simaan *et al.*, 2003; Lin *et al.*, 1995), a trade size model seems to be more appropriate.

3 The database management

The whole data set is stored in a data warehouse and configured according to the following structure:

Table 1: The original data set layout

Stock_j	Day_j	Time_t	P_{it}	V_{it}	TT_{it}
PG.MIL	15/05/2007	9.03.01	0.462	4414019	0
PG.MIL	15/05/2007	9.05.05	0.462	114256	0
PG.MIL	15/05/2007	9.05.05	0.462	2204	0
PG.MIL	15/05/2007	9.05.10	0.474	32700	2
PG.MIL	15/05/2007	9.05.13	0.460	23150	1
PG.MIL	15/05/2007	9.05.27	0.475	12500	2

P_{it} and V_{it} are the price and volume associated either to a trade or to an order (bid or ask) that refer to stock i at time t. TT_{it} indicates whether the operation is a trade or an order: $TT_{it} = 0$ is for a trade; $TT_{it} = 1$ is for a bid order; $TT_{it} = 2$ is for an ask order.

We used the following filtering rules in order to clean the data set and produce a tick by tick structure coherent with a panel framework:

1. data gaps or server errors were excluded when price (P_{it}) or volume (V_{it}) were equal to 0;

2. the overall data set, based on a tick by tick framework, could have repeated observations of trade/quote prices and volumes across data points, for each second. Table 2 and Table 3 in Appendix 1 describe typical repeated observations and the rules we followed to manage them.

3. observations were limited to when the stock exchange was open, excluding the first and the last hour of negotiations in order to omit deals that might have been influenced by market makers' inventory strategies. Thus, transactions performed outside the period 9.31.00AM-5.00.59PM were discarded.

4. extreme outliers that characterized relative-bid-ask spread, levels of risk and order imbalance indicators were also dropped (see section 4). A total of seven stocks were eliminated whose time-dimensions were not adequately long enough for our analysis.

The data management and cleaning procedures, the identification of the variables, the exploratory data analysis and empirical estimates were all carried out using the Stata/MP 10 Parallel Edition for Linux64 and a powerful HPC computer (IBM-BCX) that was kindly made available to academic users by CINECA (at the Italian Ministry of Research). IBM-BCX is a cluster of 2560 dual-core AMD opteron processors (5120 cores in total) dedicated to massive parallel applications and special High-End projects. Following the initial cleaning procedures, the tick by tick data sets of bid and ask quotes (in Stata format) each reached 2 gigabytes in file dimensions (more than 50 million panel observations, 277 stocks with an average of about 181,000 temporal observations); the data set of trades became a (in Stata format) 0.5 gigabyte file (more than 10 millions observations, 277 stocks with an average of about 38,000 temporal observations).

4 The model

We developed a trade size model and tested under different conditions of market transparency, according to the methodology already used by Bohemer *et al.* (2005) (see section 1). Dynamics are important in our model due to the persistent nature of trading behavior. For example, if a share is traded according to a specific trade size in the period t-1, it may also be reasonably traded with a similar size in the period t. A significant change in the trade size may be interpreted by the market participants as a signal of private information; that is why such activity is prudently avoided by informed traders, especially when the market is transparent. Moreover, traders are used to being attracted to 'glittering' or 'commonly-used' stocks (Barber and Odean, 2008) that intensify the persistence of trades. Our model also aims to identify the other factors that affect stock trade sizes: order flows, transaction costs (the relative bid-ask spread), levels of risk of stocks, the contingent situation of the market (order imbalance) and the disclosure of information events. The greater the intensity of the order flows, the more 'emotional' the trading process becomes; in turn, this may spread about a sort of 'trading pressure' among the market participants and affect their trade sizes. The lower the (fixed) transaction cost of a share, the more liquid it becomes, thus inducing traders to deal larger trade sizes, as already proposed by Brennan and Subrahmanyam

(1998). Moreover, the more liquid a stock is (with a low bid-ask spread), the more comfortable a trader will feel with higher trade sizes. This is to say that more liquid stocks attract informed traders willing to deal with larger trade sizes. By incorporating risk measures into our model we aim to shed light on the well known debate of volume-volatility interconnections. In general, results on the US markets, especially on the daily time frame, have revealed a positive link (Brennan and Subrahmanyam, 1998). Here we look for signs of such a relationship, whilst referring to a different stock exchange (the ISE) and over different time frames. The variable 'order imbalance' might also be able to capture the contingent situation of the market; more precisely, whether it is bearish (with considerable selling pressure) or bullish (considerable bid pressure). In the study of a US market, the order imbalance generally positively affects trading volumes (Chordia *et al.*, 2001).

Finally, our model was applied across three different time frames: 30 minutes, 150 minutes (intra day time frames) and daily. The main reasons to run this experiment were:

- to identify differences between traders' behaviors (whatever their time-horizon specialization) and to compare the results between the intra day and daily time frame. We expected that a change in the PTT would be more significant within the intra day perspectives;
- to test (across the different time frames) the various indicators of order flows (the order intensity or a set of inter quote durations) and risk (the Garman-Klass indicator or the statistical volatility) and assess the effects of the disclosure of information events (information time series such as future/spot interest rates, or news dummies);
- to manage the trade-off between running a very short term analysis and the need to avoid the micro-structure noise that has been proved to bias high frequency data (Hansen and Lunde, 2006; Bandi and Russell, 2006).

4.1 The variables

Our empirical model is based on the variables listed below. Variable i represents the stocks investigated, where $i = 1, \ldots, 277$; variable j stands for the day, where $j = 15/05/2007, \ldots, 15/11/2007$; and variable k represents the temporal interval, where $k : 1, 2, \ldots, K$. Note, the latter variable has a breadth (z) that can be 30 minutes, 150 minutes or 450 minutes.

Filtering rule number 3 refers to a trading day that runs from 9.31.00AM to 5.00.59PM (i.e. 450 minutes). Thus, three large intra day phases were created, each made up of 150 minutes (2.5 hours): the first phase runs from 9.31.00AM to 12.00.59PM (morning), the second runs from 12.01.00PM to 2.30.59PM (lunch) and the third from 2.31.00PM to 5.00.59PM (evening). This allowed us to catch the main effects of intra day variability. In the same way, we created fifteen intra day phases made up of 30 minutes each and only one phase for the 450 minutes time frame. Within each of the different time frames, the tick by tick observations TT_{it} were aggregated and ordered according to the Time t column.

In our model, when $TT_{it} = 0$, it is associated to its transaction price (P_{it}) and to its trading volume (TV_{it}); when $TT_{it} = 1$, it is associated to its bid price (BP_{it}) and to its bid volume (BV_{it}); when $TT_{it} = 2$, it is associated to its ask price (AP_{it}) and to its ask volume (AV_{it}). Moreover, in the following equations the subscripts i, k and j stand for stock, interval and day respectively.

The model is based on the following variables:

1. **Trade size indicator** ($ts_{i,k,j}$)

   The trade size is the average trading volume within every k interval. It is obtained as:

$$ts_{i,k,j} = \frac{1}{n_{i,k,j}} \sum_{t=1}^{n_{i,k,j}} TV_{i,t,j} \tag{1}$$

   where

   - $ts_{i,k,j}$ = trade size;
   - $TV_{i,t,j}$ = trading volume;
   - $n_{i,k,j}$ total number of ticks (trades) that the stock i experienced within the k interval, of day j.

2. **Order Flows Indicators** ($OF_{i,k,j}$)

   We consider the following alternative proxies of the order flows:

   - *Order intensity indicator* ($int_{i,k,j}$).
     This indicator measures the frequency of bid/ask order within our k intervals. It is:

$$int_{i,k,j} = \frac{bn_{i,k,j} + an_{i,k,j}}{2} \tag{2}$$

     where $bn_{i,k,j}$ is the number of ticks for bid orders and $an_{i,k,j}$ is the number of ticks for ask orders, that the stock i experienced within the k interval, of day j.

   - *Interquote duration indicator* ($iqd_{i,k,j}$).
     This indicator, as the following ones for the order flows, is based on the assumption that our time series are point processes (Hasbrouck, 2007). Bid and ask orders happen in discrete irregular time intervals. Our interquote indicators are based on the bid distance $BD_{i,t,j}$, in seconds, between the bid $TT_{i,t,j}$ and the bid $TT_{i,t-1,j}$ and the ask distance $AD_{i,t,j}$, in seconds, between the ask $TT_{i,t,j}$ and the ask $TT_{i,t-1,j}$. They are inspired to the idea of irregularly spaced data of Engel and Russel (2002). In particular, the interquote duration is:

$$iqd_{i,k,j} = \frac{IQDB_{i,k,j} + IQDA_{i,k,j}}{2} \tag{3}$$

where
- $IQDB_{i,k,j} = \frac{\sum_{t=1}^{bn_{i,k,j}} BD_{i,t,j}}{bn_{i,k,j}}$ is the average interquote duration indicator on the bid side.
- $IQDA_{i,k,j} = \frac{\sum_{t=1}^{an_{i,k,j}} AD_{i,t,j}}{an_{i,k,j}}$ is the average interquote duration indicator on the ask side.

- *Median Interquote duration indicator* ($iqdmed_{i,k,j}$).
  This indicator is obtained by replacing in the previous expression (3) the quantities $IQDB_{i,k,j}$ and $IQDA_{i,k,j}$ respectively with the median of the values $BD_{i,t,j}$ and $AD_{i,t,j}$.
- *Minimum Interquote duration indicator* ($iqdmin_{i,k,j}$).
  This indicator is obtained by replacing in the previous expression (3) the quantities $IQDB_{i,k,j}$ and $IQDA_{i,k,j}$ with the minimum of the values $BD_{i,t,j}$ and $AD_{i,t,j}$.
- *Weighted Interquote duration indicator* ($wiqd_{i,k,j}$).
  This indicator differs from $iqd_{i,k,j}$ because the time distances are weighted for their relative volumes. It is:

$$wiqd_{i,k,j} = \frac{WIQDB_{i,k,j} + WIQDA_{i,k,j}}{2} \tag{4}$$

  where
  - $WIQDB_{i,k,j} = \frac{\sum_{t=1}^{bn_{i,k,j}} BD_{i,t,j} BV_{i,t,j}}{\sum_{t=1}^{bn_{i,k,j}} BV_{i,t,j}}$ is the weighted average interquote duration indicator on the bid side.
  - $WIQDA_{i,k,j} = \frac{\sum_{t=1}^{an_{i,k,j}} AD_{i,t,j} AV_{i,t,j}}{\sum_{t=1}^{an_{i,k,j}} AV_{i,t,j}}$ is the average weighted interquote duration indicator on the ask side.

3. **Relative bid-ask spread indicator** ($rbas_{i,k,j}$).
   It is a proxy of the trading costs and is calculated as

$$rbas_{i,k,j} = \frac{(AAP_{i,k,j} - ABP_{i,k,j})}{\frac{1}{2}(AAP_{i,k,j} + ABP_{i,k,j})} \tag{5}$$

   where
   - $AAP_{i,k,j} = \frac{1}{an_{i,k,j}} \sum_{t=1}^{an_{i,k,j}} AP_{i,t,j}$ is the average ask price within the k interval.
   - $ABP_{i,k,j} = \frac{1}{bn_{i,k,j}} \sum_{t=1}^{bn_{i,k,j}} BP_{i,t,j}$ is the average bid price within the k interval.

4. **Riskiness Indicators** ($R_{i,k,j}$).

   We consider the following different proxies of risk:

   - *Statistical Volatility indicator* ($vol_{i,k,j}$).
     As it is well known, we can easily calculate the statistical volatility as follows

$$vol_{i,k,j} = \sqrt{\frac{1}{\frac{z}{z'} - 1} \sum_{k=1}^{\frac{z}{z'}} (r_{i,k,j} - \overline{r}_{i,k,j})^2} \tag{6}$$

where
- $r_{i,k,j} = ln(\frac{P_{i,k,j}}{P_{i,k-1,j}})$ is the log-return on a fifteen minutes ($z' = 15$) interval basis (k) for stock i on day j. In this formulation, $P_{i,k,j}$ is the last trade price within the k period and $P_{i,k-1,j}$ is the last trade price within the *k-1* period. Note that the frequency $z' = 15$ is chosen to avoid the microstructure noise.
- $\overline{r}_{i,k,j}$ is the average return within the original k periods whose breadth (z) can be alternatively 30 minutes, 150 minutes or 450 minutes.

- *Garman-Klass Range* (Garman and Klass, 1980)($gkr_{i,k,j}$).
After filtering for $TT_{it} = 0$, depending on the selected time frame, the high-frequency/intraday/daily indicator is

$$gkr = \frac{0,511}{K}\sum_{k=1}^{K}(u_k - d_k)^2 - \frac{0,019}{K}\sum_{k=1}^{K}[c_k(u_k + d_k) - 2u_k d_k] - \frac{0,383}{K}\sum_{k=1}^{K} c_k^2 \tag{7}$$

where
- C_{k-1} closing price of the stock for each high-frequency/intraday/daily $k-1$ interval;
- O_k opening price of the stock for each high-frequency/intraday/daily k interval;
- H_k highest price of the stock for each high-frequency/intraday/daily k interval;
- L_k lowest price of the stock for each high-frequency/intraday/daily k interval;
- C_k closing price of the stock for each high-frequency/intraday/daily k interval;
- $c_k = lnC_k - lnO_k = ln\frac{C_k}{O_k}$ is the normalized closing price;
- $o_k = lnO_k - lnC_{k-1} = ln\frac{O_k}{C_{k-1}}$ is the normalized opening price;
- $u_k = lnH_k - lnO_k = ln\frac{H_k}{O_k}$ is the normalized highest price;
- $d_k = lnL_k - lnO_k = ln\frac{L_k}{O_k}$ is the normalized lowest price".

5. **Order Imbalance Indicator** ($oi_{i,k,j}$).

We calculate the order imbalance as

$$oi_{i,k,j} = \frac{APR_{i,k,j}}{BPR_{i,k,j}} \tag{8}$$

where
- $APR_{i,k,j} = \frac{1}{an_{i,k,j}}\sum_{t=1}^{an_{i,k,j}} AV_{i,t,j}$ is the ask pressure.
- $BPR_{i,k,j} = \frac{1}{bn_{i,k,j}}\sum_{t=1}^{bn_{i,k,j}} BV_{i,t,j}$ is the bid pressure.

6. **Informational Events in the form of time series** ($IETS_{k,j}$).

This is the variable used in the intra day time frames:

- *Backwardation of the Euribor future* ($backwardation_{k,j}$).
The backwardation of the Euribor future revealed in the k interval of day j is

$$backwardation_{k,j} = F_{may2008_{k,j}} - F_{dec2007_{k,j}} \quad (9)$$

where
- $F_{may2008_{k,j}}$ = May 2008 Euribor future close price in the k interval, of day j;
- $F_{dec2007_{k,j}}$ = December 2007 Euribor future close price in the k interval, of day j.

Keeping in mind that the implicit interest rate in the Euribor future is the difference between 100 and the future price F, the subject variable is positive when the interest rates corresponding to the shorter maturity are higher than the interest rates corresponding to the longer maturity. More directly, *backwardation* k, j is positive when market participants perceive a lack of the underlying asset (short term interbank deposits). In fact, backwardation k, j, on a daily basis, is always negative from the starting date of our analysis (the 15^{th} of May 2007) until the 16^{th} of August 2007; after this date it becomes steadily more positive. This variable helps us to capture, on an intra day basis, when it was that the banks became aware of the spread of the sub-prime crisis (proxy of a form of private information, because retail traders are not supposed to be able to obtain these data).

- *TED spread* (ted_j).
It is calculated as

$$ted_j = 3mLIBOR_j - 3mtr_j \quad (10)$$

where
- $3mLIBOR_j$ = 3-months LIBOR, in the day j;
- $3mtr_j$ = 3-months treasury rate, in the day j.
It is the signal of the difference of credit risk premium between a risky asset ($3mLIBOR$) and a risk-free asset ($3mtr$), in the day j.

7. **Informational Events in the form of dummies for news** ($IEN_{k,j}$).

Together with $backwardation_{k,j}$, on a daily basis we also consider the following informations in the form of dummies (relevant news from the italian provider ANSA):

- 23/06/2007 ISE and LSE merger announcement ($d23jun$). On this date, the Board of Borsa Italiana S.p.A. and of the London Stock Exchange Group plc announced its agreement about a recommended offer from the LSE to the shareholders of Borsa Italiana. Starting from that date the two Stock Exchanges gradually moved together toward a merger;
- 16/08/2007 Sub prime earthquake: down also Fiat and Autogrill ($d16ago$);
- 18/09/2007 FED, relevant cut of interest rate to 4,75 percent ($d18sep$);

- 31/10/2007 FED, cut of interest rate to 4,5 percent ($d31oct$).

4.2 The empirical equation

Our general empirical equation can be written as:

$$\begin{aligned} logts_{i,k,j} &= a_i + \sum_{t=1}^{T} \alpha_t logts_{i,k-t,j} + \beta_1 OF_{i,k,j} + \beta_2 rbas_{i,k,j} \\ &+ \beta_3 R_{i,k,j} + \beta_4 logoi_{i,k,j} + \beta_5 IETS_{k,j} + \beta_6 IEN_{k,j} + \varepsilon_{i,k,j} \end{aligned} \quad (11)$$

where the error term, $\varepsilon_{i,k,j}$, is supposed to be both arbitrary heteroskedastic and arbitrary intra-groups correlated (the errors are not independent within individuals, although they are independent between individuals). Note that we use the logarithm transformation for trade size, order flows and order imbalance indicators. This in order to reduce positive skewness in the distributions, and interpreted all the parameter estimates as elasticities. To made results comparable, the riskiness indicator measured by the Garman-Klass Range is multiplied by 100.

Depending on the time frame considered, some differences in the model (11) specification emerge. As far as dynamics is concerned, we use four lags (T=4) in the 30 and 450 minutes frames, and two lags (T=2) in the 150 minutes frame. We believe these lags capture the quite weekly operating period in the daily frame, and the variations over the course of the day (opening, middle of the day and closing) in the 30 and 150 minutes frames. Risk is measured by the Garman-Klass Range in the 30 minutes frame, and by statistical volatility indicator in the 150 and 450 minutes frames. Specific days dummies (included in the variable IEN) are, of course, considered in the daily frame only. In particular, the news considered are: the 23^{rd} of June in Book1, the 16^{th} of August in Book 2 and the 31^{th} of October in Book 3 (we checked also the 18^{th} of September, but the first reduction of interest rate is not significant).

Existing literature offers little consensus in selecting the indicator which better represents the riskiness of a stock. Moreover, the order flows thickness may be expressed in terms of new bid/ask orders, in a time unit, as well as in terms of time distance between one order and the following. We define a benchmark model (called model A) in which order flows are measured by the order intensity indicator. We then carry out a number of robustness checks by varying the measures of OF, R, and IETS. In particular, in the 30 minutes frame, we also try statistical volatility indicator in place of the Garman-Klass range indicator (model A_VOL); vice-versa, in the 150 and 450 minutes frames, we replace vol for gkr (model A_GKR). In all the time frames we alternatively replace the order intensity measure of order flows with either the average or median or minimum or weighted average interquote duration indicators (models A_IQD, A_IQDMED, A_IQDMIN, A_WIQD, respectively). Finally, in the daily frame, we check the effect of replacing backwardation by TED spread (model A_TED).

The estimation method is the fixed effects panel estimator with robust standard errors to both arbitrary heteroskedasticity and arbitrary intra-groups correlation. It is well known that some of the variables affecting the trade size (e.g. relative bid-ask spread, riskiness and order imbalance) are simultaneously affected by the trade size as well. These endogenous variables are instrumented with proper lags (up to four lags in 30 and 450 minutes frames; up to two lags in 150 minutes frame). Dynamics could also create an endogeneity problem: the fixed effects estimator applied to dynamic panels is biased downwards; it is inconsistent as N becomes large; the inconsistency is of order $1/(T-1)$. Since we have, on average in each sub-samples, $T > 470$, we are in the case the fixed effects estimator becomes consistent as T becomes large (Nickell, 1981). We also try IV (2SLS) of Anderson and Hsiao (1982) on the first-differencing transformation. In this case, the lagged dependent variable is instrumented by the level in $t-4$ of trade size. Results are robust, despite, as well known, less precise.

5 Results

Our results provide empirical findings and also form the basis for a discussion of some other related issues. The main empirical results can be summarized as follows (estimates are shown in Appendix 3):

1. main drivers of trade size: we found robust relationships between the trade size and our explanatory variables. In Book1, these relationships are not always significant but, when the PTT is enhanced (in Book2 and especially in Book3), their sign remains resiliently stable, independent of the time frame considered. The trade size is:

   (a) positively linked to its lags (always significant, except for some lags in the daily estimates), as was expected.

   (b) positively linked to the order intensity (always significant). This could mean that when order flows increase, in the time unit, we might expect to observe an increase in the trade size, because: a) some traders may possess private information; in order to exploit it, he/she inserts more orders into the book, that are larger in size, thus concluding with bigger trades; b) the increasing number of orders is 'emotionally' interpreted by traders as a sign that some of them could possess private information; so they assume an imitative behaviour and take part in the trading pressure.

   (c) negatively linked to the bid-ask spread (often significant; always in Book3), as was expected.

   (d) negatively linked to the risk indicator (nearly always significant in the intra day time frames, with the only exception being in Book 1 in the 30 minutes time frame. This result, opposite compared to the existing literature, is strongly confirmed on the intra-day time frame, while on the daily basis the link is rarely significant. We could argue that the persistent (negative) relationship uncovered for the intra day basis indicates that Italian traders are mainly averse to risk and that as risk increases they tend to freeze their trading activity, reducing their trade size.

   (e) positively linked to backwardation (very often significant in the intra day time frames, but less significant on the daily basis).

2. consequences of the increases in PTT. Moving from B1 to B3, independent of the time frame considered, our models are better instrumented (as far as the specification tests are concerned) and show clearer results (for the mentioned interconnections 1, 2, 3, 4 and 5). For the intra day models we observed significant changes in the absolute values of the following significant parameters:

   - the positive relationship between trade size and order intensity becomes weaker: the absolute value of the parameter decreases; generally with a significant difference. Within the overall analysis, this is the only piece of evidence indicated by the intra day estimates and confirmed on the daily basis. When the market is more transparent and, as in 'Book Profondo', all the traders are allowed to see all orders for each price level, we can reasonably expect that: a) the traders who possess private information are more cautious in quoting and trading, breaking up the size of their orders (e.g. by means of iceberg orders) and repressing the size of their trades; b) all the other traders, thanks to better PTT, are able to understand whether the trading pressure is real or fictitious, avoiding increases in the trading size without sufficient grounds; c) moreover, all the traders, especially retail traders, may identify some signals of risk from the order book disclosed (for example, evidence of iceberg orders), choosing to exit from the market and cancelling their quotes. This should mean that, with greater PTT, a smaller number of quotes are transformed into trades. As a proof of this phenomenon, from the descriptive statistics (see Appendix 2) we observe that the proportion of the quotes that become transformed into trades moves close to 23 per cent within the B1 period and to 20 per cent within the B2 and B3 periods.
   - the negative relationship between the trade size and the bid-ask spread becomes feebler: the absolute value of the parameter decreases and the difference is significant for both of the intra day general models. Even if the inventory and the asymmetrical information paradigms are not appropriate for an order driven market, we can argue that when the market conditions are less transparent, the bid-ask spread widens (Biais, 1993). It appears that PTT works as a mean of this transparency because we can see from the descriptive statistics that the movement from B1 to B3 causes the overall average bid-ask spreads to reduce. In relation to this, our estimates show that the increase in PTT contributes to the bringing to an end of the negative effects that bid-ask spreads have upon trade sizes. Definitely the transparency of the book seems to be compensating for the 'indirect opacity' related to a high bid-ask spread. That is to say that high PTT levels may also allow wide trades for illiquid stocks.
   - the negative relationship between the trade size and the risk indicator (GKR or VOL) reduces its effects: the absolute value of the parameter decreases, even if a significant difference is only detected in the 150 minutes time frame. When a market is not transparent, we can argue that traders, especially the retail ones, may also repress their trading size to compensate uncertainty.

When a market improves its transparency, it is reasonable that traders are able to downsize their risk aversion. A more informative and reliable order book may cause traders to be more confident about their understanding of the trading process. This could lead them to consider trades of higher trade sizes even for risky stocks. Italian traders have mainly been described as being risk averse; as risk increases they tend to freeze their trading sizes; nevertheless, as the market becomes more transparent Italian investors appear to be more willing to assume more 'informed' risky-positions.

3. relevance of informational events. The backwardation is often significant and always attests a positive contribution to the trade size, despite the fact that this indicator inverts its sign in the the first half of August (see Appendix 2 for the descriptive statistics): both the abundance and the reductions of interbank liquidity can, in the short term, provide an informative signal that institutional traders are able to exploit, resulting in increases in trade size. Proof of this informative power of backwardation, in the intra day time frame, comes from observing the daily models where this power is often disturbed by information event dummies (especially 'bad news dummies', like the news of strong downtrends published in the ISE and/or Federal Reserve announcements).

Marginally, our estimates shed some light on the appropriateness of the indicators used in our estimates. Comparing the different models, we have provided evidence on which indicators are interchangeable, and which ones are not. More precisely, for the indicators of order flows, the order intensity and the inter quote duration are completely interchangeable, no matter how they are calculated (mean, median, minimum or weighted), as they capture the relationship between the order vivacity and the trade size with the same effectiveness. We selected order intensity as it is easier to comprehend. Conversely, as far as the risk indicators are concerned, GKR and statistical volatility are not interchangeable; at least on the very short term. In the 30 minutes time frame, the MODEL A_VOL reveals weaker specification tests. Moreover, some important relationships (i.e. between trade size and the bid-ask spread in B2 and B3) loose significance when compared with all the other Models based on the GKR [2]

6 Concluding remarks and future research

Our main empirical findings show that intra day models are more effective than the daily models at describing the behaviour of traders after an enhancement of PTT. The main drivers of the trade size, used as an expression of this behaviour, are generally significant and most of the signs of the relationships are coherent with previous literature (in particular, for order flows and bid-ask spread). On the contrary, we show risk to have

[2] The estimates of the MODEL A_VOL were conducted on a smaller number of observations ($n = 41,072$) compared to the others ($n = 64,094$) due to the specific way in which the statistical volatility was computed; we also ran, therefore, the estimates within the same subset of observations and the results absolutely confirmed the previous findings of MODEL A and of MODEL A_VOL. On the contrary, in the 150 minutes time frame, the benchmark Model exploits risk via statistical volatility while the MODEL A_GKR shows weaker estimates.

a negative link with trade size, and this represents the aversion of the Italian traders to risk. This result is significant on the intra day basis and seems to be in contrast with the findings of a previous study that referred to US stock markets. In the short term, both the abundance and the reductions of the interbank liquidity, measured by backwardation, are proved to act as an informative signal that institutional traders exploit to increase their trade size. Information events, in the form of news, are shown to affect the trade size in an unconventional manner: before the spread of news (e.g. the ISE-LSE merger) or only after an unusual shock (e.g. the second unexpected interest rate cut issued by the FED).

As an overall effect of the enhancement of PTT in the ISE equity markets, we argue that (intra day) traders operating in a more transparent market become more reticent and less risk averse: if they are informed (institutional) traders, they tend to hide their intentions, prudently repressing the size of their trades; if they are non-informed (retail) intra day traders, they are able to better understand if a trading pressure is real or fictitious, in turn controlling emotional imitative behaviours; the latter type of trader also runs the chance of refusing opaque trading and cancelling orders. Moreover, the increased level of disclosure of the order book permits traders to downsize their level of risk aversion, reducing that uncertainty that is able, under the condition of opacity, to break trading activity. A more informative order book induces traders to act confidently on their knowledge of the trading process, dealing with higher trade sizes even for risky or illiquid stocks.

In the near future, we intend to further this research by identifying the specific commonalities (as described by Chordia, Roll and Subrahmanyam, 2000) within the overall list of the shares used; we will look for differences due to market capitalization, business sector and beta. Moreover, we aim to develop further the investigation into the unexpected negative relationship between trade size and risk in order to try to understand whether or not it is due to a time frame peculiarity, or if it is a consequence of a specific national behaviour.

References

Amihud, Y. and H. Mendelson (1980), "Dealership markets: Market making with inventory", *Journal of Financial Economics*, 8, pp. 31–53.

Anderson, T. and C. Hsiao (1982;), "Formulation and Estimation of Dynamic Models using Panel Data", *Journal of Econometrics*, 18, pp. 47–82.

Bandi, F. and J. R. Russell (2006), "Separating microstructure noise from volatility", *Journal of Financial Economics*, 79, pp. 655–692.

Barber, B. and T. Odean (2008), "All that glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors", *Review of Financial Studies*, 21(2), pp. 785–818.

Baum, C., M. Schaffer and S. Stillman (2003), "Instrumental variables and GMM: Estimation and testing", *The Stata Journal*, 3(1), pp. 1–31.

Baum, C., M. Schaffer and S. Stillman (2007), “Enhanced routines for instrumental variables/GMM estimation and testing”, *Boston College Economics Working Paper*, 667.

Biais, B. (1993), “Price formation and equilibrium liquidity in fragmented and centralized markets”, *Journal of Finance*, 48, pp. 157–185.

Bohemer, E., G. Saar and L. Yu (2005), “Lifting the Veil: An Analysis of the Pre-trade Transparency at the NYSE”, *Journal of Finance*, 60(2), pp. 783–815.

Brennan, M. and A. Subrahmanyam (1998), “The determinants of average trade size”, *Journal of Business*, 71(1), pp. 1–25.

Chan, Y. (2000), “The price impact of trading on the stock exchange of Hong Kong”, *Journal of Financial Markets*, 3, pp. 1–16.

Chordia, T., R. Roll and A. Subrahmanyam (2000), “Commonality in liquidity”, *Journal of Financial Economics*, 56, pp. 3–28.

Chordia, T., R. Roll and A. Subrahmanyam (2001), “Market Liquidity and Trading Activity”, *Journal of Finance*, 56(2), pp. 501–530.

Chowdhry, B. and V. Nanda (1991), “Multi-market trading and market liquidity”, *Review of Financial Studies*, 4, pp. 483–511.

Engel, R. and J. Russel (2002), “Autoregressive conditional duration: a new model for irregularly spaced transaction data”, *Econometrica*, 66(5), pp. 1127–1162.

Focault, T., S. Moinas and E. Theissen (2007), “Does Anonymity matter in Electronic Limit Order Markets?”, *Review of Financial Studies*, 20, pp. 1707–1747.

Forster, M. and T. George (1992), “Anonymity in securities markets”, *Journal of Financial Intermediation*, 2, pp. 168–206.

Garman, M. and M. Klass (1980), “On the estimation of security price volatilities from historical data”, *Journal of Business*, 53(1), pp. 67–78.

Hansen, L. (1982), “Large sample properties of generalized method of moments estimators”, *Econometrica*, 50(3), pp. 1029–1054.

Hansen, P. and A. Lunde (2006), “Realized variance and market microstructure noise”, *Journal of Business and Economic Statistics*, 24(2), pp. 127–161.

Harris, J. and P. Schultz (1997), “The importance of firm quotes and rapid executions: evidence from the January 1994 SOES rule change”, *Journal of Financial Economics*, 45, pp. 135–166.

Harris, L. (1996), “Does a large minimum price variation encourage order exposure?”, Unpublished manuscript, University of Southern California.

Hasbrouck, J. (2007), *Empirical Market Microstructure*, Oxford University Press.

Huang, R. and H. Stoll (1996), "Dealer versus auction markets: a paired comparison of execution costs on NASDAQ and the NYSE", *Journal of Financial Economics*, 41, pp. 313–357.

Kleibergen, F. and R. Paap (2006), "Generalized reduced rank tests using the singular value decomposition", *Journal of Econometrics*, 127(1), pp. 97–126.

Kleibergen, F. and M. Schaffer (2007), "RANKTEST: Stata module to testing the rank of a matrix using the Kleibergen-Paap rk statistic", Available at: http://ideas.repec.org/c/boc/bocode/s456865.html.

Lin, J., G. Sanger and G. Booth (1995), "Trade size and components of the bid-ask spread", *Review of Financial Studies*, 8, pp. 1153–1183.

Lo, A. and W. Wang (2001), "Trading volume: definitions, data analysis and implications of portfolio theory", *The Review of Financial Studies*, 13(2), pp. 257–300.

Madhavan, A. (2000), "Market Microstructure: A survey", *Journal of Financial Markets*, 3, pp. 205–258.

Nickell, S. (1981), "Biased in dynamic Models with Fixed Effects", *Econometrica*, 49, pp. 1417–1426.

O'Hara, M. (1995), *Market microstructure theory*, Basil Blackwell, Cambridge.

Simaan, Y., D. Weaver and D. Withcomb (2003), "Market Maker Quotation Behaviour and Pre-trade Transparency", *Journal of Finance*, 50, pp. 1247–1267.

Stoll, H. (1976), "Dealer inventory behavior: An empirical investigation of Nasdaq stocks", *Journal of Financial and Quantitative Analysis*, 11, pp. 359–380.

Stoll, H. (1978), "The supply of dealer services in securities markets", *Journal of Finance*, 33, pp. 1133–1151.

Theissen, E. (2003), "Trader anonymity, Price Formation and Liquidity", *European Finance Review*, 7, pp. 1–26.

Appendix 1: Filtering rules to convert a tick by tick data set into a panel structure

Table 2: Cleaning guidelines: Cases for tt=0 (trades)

Case	Stock	Day_j	Time $_t$	P_{it}	V_{it}	TT_{tt}	
1:	PG.MIL	15/05/2007	9.05.33	0.4615	28103	0	At that time insert:
tt=0							- average of prices;
same price							- sum of volumes.
different volume	PG.MIL	15/05/2007	9.05.33	0.4615	9995	0	
2:	PG.MIL	15/05/2007	9.49.00	0.46	44467	0	At that time insert:
tt=0	PG.MIL	15/05/2007	9.49.00	0.4605	16525	0	- average of prices;
different price	PG.MIL	15/05/2007	9.49.00	0.46	39700	0	- sum of volumes.
different volume	PG.MIL	15/05/2007	9.49.00	0.4605	19762	0	
	PG.MIL	15/05/2007	9.49.00	0.4605	27295	0	
	PG.MIL	15/05/2007	9.49.00	0.46	100	0	
	PG.MIL	15/05/2007	9.49.00	0.46	89000	0	
	PG.MIL	15/05/2007	9.49.00	0.46	10000	0	
	PG.MIL	15/05/2007	9.49.00	0.4605	50000	0	
	PG.MIL	15/05/2007	9.49.00	0.4605	59000	0	
	PG.MIL	15/05/2007	9.49.00	0.46	89171	0	
	PG.MIL	15/05/2007	9.49.00	0.4605	7738	0	
3:							More difficult to find but
tt=0							when it happens it can
same price							be due to iceberg
same volume							orders. For this reason
							filtering rules are identical
							to those previously shown:
							- average of prices;
							- sum of volumes.

Table 3: Cleaning guidelines: Cases for tt=1 (bid) or tt=2 (ask)

Case	Stock	Day_j	Time $_t$	P_{it}	V_{it}	TT_{tt}	
1: tt=1 or 2; REPEATED same price same volume	PG.MIL	15/05/2007	9.05.34	0.461	3467	1	At that time insert: - average of prices; - sum of volumes.
	PG.MIL	15/05/2007	9.05.34	0.461	3467	1	
2: tt=1 or 2; NOT REPEATED same price different volume							At that time insert: - average of prices; - sum of volumes.
	PG.MIL	15/05/2007	9.28.37	0.462	37009	1	
	PG.MIL	15/05/2007	9.28.37	0.462	45749	1	
3: tt=1 or 2; NOT REPEATED: different price different volume or same volume (ex. below)							At that time insert: IF tt=1 - take all the information of the row with the highest price; delete other information. IF tt=2 - take all the information of the row with the lowest price; delete other information.
	PG.MIL	15/05/2007	9.07.08	0.462	100000	1	
	PG.MIL	15/05/2007	9.07.08	0.4625	8993	1	
4:	PG.MIL	15/05/2007	9.07.32	0.463	8993	1	
tt=1 or 2;	PG.MIL	15/05/2007	9.07.32	0.4625	8993	1	
NOT REPEATED:	PG.MIL	15/05/2007	9.07.40	0.463	8993	1	
different price	PG.MIL	15/05/2007	9.08.03	0.463	8993	1	
same volume	PG.MIL	15/05/2007	9.08.09	0.463	8993	1	
	PG.MIL	15/05/2007	9.08.21	0.463	8993	1	

Appendix 2: Descriptive statistics

Table 4: Descriptive statistics - 30 minutes

[ts=trade size; n=total number of trade ticks; TV= trading volume; vol=statistical volatility; gkr=Garman-Klass range; APR=ask pressure; BPR=bid pressure; INT=ordern intensity; bn=number of ticks for bid orders; an=number of ticks for ask orders; wiqd=weighted interquote duration; iqd=interquote duration; iqdmed=median interquote duration; iqdmin=minimum interquote duration; backward=backwardation. In Book 1 the statistics are computed on NT=106,820 (total number of observations), N=267 (total number of stocks), T = 400 (average number of time periods); in the statistical volatility indicator case the total number of observations fall to 76861 (about 287 average number of time periods) because lags are used in the computation.]

			Book1			
	1st Q	Median	3rd Q	IQR	Mean	SD
ts	482.5	1171.068	3191.187	2708.687	3691.635	10154.67
n	4	11	31	27	28.74771	50.34952
TV	2500	13790	72808.5	70308.5	184875.7	847899.4
vol	0.0006612	0.0016091	0.0032996	0.0026384	0.0026108	0.0033366
gkr	0	3.64E-06	0.0000221	0.0000221	0.0000718	0.0007316
APR	417.6026	1045.479	3002.716	2585.113	4647.283	21345.28
BPR	405.3917	1029.599	2994.167	2588.775	4450.505	16086.12
INT	24	66	159	135	122.9534	155.3525
bn	24	66	159	135	122.9644	155.3533
an	24	66	158	134	122.9424	155.3675
wiqd	10.70605	25.76458	71.17729	60.47124	77.51958	195.7756
iqd	11.2625	26.86765	73	61.7375	76.58201	188.8773
iqdmed	4	10	28	24	44.16685	172.4986
iqdmin	1	1	1	0	12.02583	137.7318
backward	-0.1000061	-0.0800018	-0.0400009	0.0600052	-0.0753538	0.0345086
			Book2			
	1st Q	Median	3rd Q	IQR	Mean	SD
ts	433.9474	1096	2940.444	2506.497	3352.912	9105.54
n	3	10	30	27	26.29057	45.192
TV	2096	11581	61004	58908	147854.4	667266.8
vol	0.0007534	0.0018728	0.0038905	0.0031371	0.0030788	0.0039713
gkr	0	3.68E-06	0.0000281	0.0000281	0.0000879	0.0007382
APR	385.419	946.26	2612.214	2226.795	3830.747	15275.03
BPR	381.3044	940.6102	2600.735	2219.43	3792.491	14125.75
INT	23	66	166	143	126.1742	159.056
bn	23	66	166	143	126.187	159.046
an	23	66	166	143	126.1614	159.0848
wiqd	10.25704	25.85086	73.89018	63.63315	83.61446	223.5503
iqd	10.76024	26.85075	75.6087	64.84846	82.31102	216.1881
iqdmed	4	10	29	25	48.24927	199.6314
iqdmin	1	1	1	0	14.67284	166.3191
backward	-0.1100006	-0.0599976	0.0800018	0.1900024	-0.0318872	0.0895535
			Book3			
	1st Q	Median	3rd Q	IQR	Mean	SD
ts	466.6667	1097.833	2813.345	2346.678	3233.554	9909.999
n	3	10	29	26	25.64137	44.59837
TV	2100	11378.5	56665	54565	144724.9	759018.4
vol	0.0007698	0.0018893	0.0038912	0.0031213	0.0030605	0.0039362
gkr	0	4.01E-06	0.0000291	0.0000291	0.0000949	0.0008102
APR	422.2113	995.9361	2640.783	2218.571	4555.043	26175.03
BPR	399.4898	961.875	2562.55	2163.06	4289.12	18328.26
INT	21	63	163	142	121.6524	154.2853
bn	21	63	164	143	121.6666	154.2811
an	21	63	164	143	121.6383	154.3078
wiqd	10.38315	26.99399	80.47823	70.09508	92.82041	259.3969
iqd	10.88724	28.01563	82.25	71.36276	91.45066	252.7157
iqdmed	4	10	30	26	54.32902	233.5966
iqdmin	1	1	1	0	18.33301	197.9085
backward	0.1600037	0.1999969	0.2300034	0.0699997	0.1911549	0.0595456

Table 5: Descriptive statistics - 150 minutes

[ts=trade size; n=total number of trade ticks; TV= trading volume; vol=statistical volatility; gkr=Garman-Klass range; APR=ask pressure; BPR=bid pressure; INT=ordern intensity; bn=number of ticks for bid orders; an=number of ticks for ask orders; wiqd=weighted interquote duration; iqd=interquote duration; iqdmed=median interquote duration; iqdmin=minimum interquote duration; backward=backwardation. In Book 1 the statistics are computed on NT=106,820 (total number of observations), N=267 (total number of stocks), T = 400 (average number of time periods); in the statistical volatility indicator case the total number of observations fall to 76861 (about 287 average number of time periods) because lags are used in the computation.]

			Book1			
	1st Q	Median	3rd Q	IQR	Mean	SD
ts	585.6617	1275.583	3260.35	2674.688	3518.609	7274.547
n	21	58	161	140	143.8251	233.1228
TV	16322	79549	397656	381334	926663.6	3861745
vol	0.0015584	0.0024244	0.0038831	0.0023248	0.0031862	0.0087289
gkr	3.07E-06	0.0000141	0.00004	0.000037	0.0000683	0.0004338
APR	500.3556	1129.221	3153.039	2652.684	4667.841	16391.23
BPR	490.3733	1135.523	3191.605	2701.232	4505.454	13966.09
INT	134	353	793	659	610.0701	721.7915
bn	134	353	793	659	610.0575	721.7534
an	134.5	353	793	658.5	610.0827	721.842
wiqd	10.65024	23.63223	61.51342	50.86318	59.9017	110.7211
iqd	11.13768	24.62158	63.46023	52.32255	60.36752	106.5999
iqdmed	4	8	20	16	23.77514	61.1656
iqdmin	1	1	1	0	1.398687	12.93779
backward	-0.0999985	-0.0800018	-0.0499954	0.0500031	-0.0767093	0.0332695
			Book2			
	1st Q	Median	3rd Q	IQR	Mean	SD
ts	545.3256	1219.146	3065.895	2520.569	3312.222	6979.244
n	19	55	157	138	131.9297	201.8179
TV	14125	68603	338912	324787	750686.6	2985258
vol	0.0017944	0.0028458	0.0046359	0.0028416	0.0036766	0.0030291
gkr	2.75E-06	0.0000178	0.0000544	0.0000517	0.0000869	0.0003663
APR	465.4796	1046.091	2772.343	2306.863	3907.44	12160.38
BPR	456.8173	1030.408	2724.44	2267.623	3860.286	11931.65
INT	132	361	854	722	631.2035	727.2923
bn	132	361	854	722	631.1893	727.2572
an	132	361	854	722	631.2177	727.3407
wiqd	9.825535	23.15388	63.78135	53.95581	65.72079	141.8265
iqd	10.31501	24.16936	65.06522	54.75021	65.70221	136.4272
iqdmed	4	8.5	21	17	26.78179	92.83074
iqdmin	1	1	1	0	1.797501	24.60997
backward	-0.1100006	-0.0599976	0.0899963	0.1999969	-0.0314927	0.0899512
			Book3			
	1st Q	Median	3rd Q	IQR	Mean	SD
ts	577.6316	1216.118	2942.643	2365.011	3166.304	7134.682
n	17	53	155	138	128.6791	198.4082
TV	14191	66789	313115	298924	734536.2	3284482
vol	0.0018276	0.0028577	0.0045771	0.0027495	0.0036856	0.0032091
gkr	3.12E-06	0.0000186	0.0000557	0.0000525	0.0001033	0.0005964
APR	503.9208	1086.522	2782.509	2278.588	4663.406	24175.77
BPR	478.5	1050.448	2709.689	2231.189	4397.265	16496.15
INT	120	347	843	723	607.0928	701.6279
bn	120	347	843	723	607.0897	701.6177
an	120	347	843	723	607.0958	701.6453
wiqd	9.895412	23.94417	69.52753	59.63212	72.49693	156.1541
iqd	10.35321	24.97778	71.50085	61.14764	72.69948	150.7166
iqdmed	4	8.75	22	18	29.72715	100.2532
iqdmin	1	1	1	0	2.4651	39.87992
backward	0.1600037	0.1999969	0.2399979	0.0799942	0.1913558	0.0597493

Table 6: Descriptive statistics - daily

[ts=trade size; n=total number of trade ticks; TV= trading volume; vol=statistical volatility; gkr=Garman-Klass range; APR=ask pressure; BPR=bid pressure; INT=ordern intensity; bn=number of ticks for bid orders; an=number of ticks for ask orders; wiqd=weighted interquote duration; iqd=interquote duration; iqdmed=median interquote duration; iqdmin=minimum interquote duration; backward=backwardation. In Book 1 the statistics are computed on NT=106,820 (total number of observations), N=267 (total number of stocks), T= 400 (average number of time periods); in the statistical volatility indicator case the total number of observations fall to 76861 (about 287 average number of time periods) because lags are used in the computation.]

	Book1					
	1st Q	Median	3rd Q	IQR	Mean	SD
ts	609.2031	1306.758	3367.125	2757.922	3624.847	7319.812
n	44	131	397	353	358.9128	581.5413
TV	36166	188835	974753	938587	2314259	9301036
vol	0.0019659	0.0030109	0.0046434	0.0026775	0.0038384	0.0099262
gkr	7.96E-06	0.0000238	0.0000662	0.0000583	0.0001176	0.0005531
APR	524.65	1148.131	3238.299	2713.649	4366.205	13074.88
BPR	511.5686	1122.988	3185.421	2673.853	4257.039	12224.21
INT	316.5	849	1982	1665.5	1550.955	1866.338
bn	318	846	1980	1662	1551.161	1866.044
an	317	850	1984	1667	1550.75	1866.678
wiqd	12.49434	28.35387	77.72148	65.22714	73.85363	137.4825
iqd	13.00144	29.47513	77.98251	64.98106	74.51243	138.701
iqdmed	4	8	20	16	22.45576	54.38333
iqdmin	1	1	1	0	1.031068	0.4285452
backward	-0.1000061	-0.0899963	-0.0500031	0.0500031	-0.0774627	0.0318625
	Book2					
	1st Q	Median	3rd Q	IQR	Mean	SD
ts	583.5606	1286.733	3348.64	2765.079	3554.39	7552.663
n	38	116.5	370	332	319.1074	469.1148
TV	32306	159375.5	826573	794267	1875229	7117618
vol	0.0020791	0.0031968	0.0050814	0.0030022	0.0039956	0.0028867
gkr	7.30E-06	0.0000257	0.0000766	0.0000693	0.0001108	0.0003499
APR	492.1584	1058.741	2922.657	2430.499	3993.222	12409.65
BPR	470.6563	1053.101	2876.246	2405.589	3994.739	11564.74
INT	268	790.5	2073	1805	1534.219	1789.378
bn	270	792	2073	1803	1534.504	1788.756
an	267	791	2073	1806	1533.933	1790.078
wiqd	12.11436	31.03867	91.6099	79.49555	82.98744	146.0183
iqd	12.57283	31.63609	92.10583	79.533	84.81924	155.1365
iqdmed	4	9	24	20	26.17399	68.38353
iqdmin	1	1	1	0	1.228042	13.28814
backward	-0.1299973	-0.1200027	-0.0800018	0.0499954	-0.1019669	0.0286851
	Book3					
	1st Q	Median	3rd Q	IQR	Mean	SD
ts	585.8146	1227.601	2970.475	2384.661	3207.283	7146.016
n	30	101	351	321	302.658	457.7739
TV	24624.5	138369	708838.5	684214	1710966	7220700
vol	0.002332	0.0035963	0.0056955	0.0033634	0.0045473	0.0036892
gkr	7.97E-06	0.0000304	0.0000923	0.0000843	0.0001738	0.0008069
APR	499.9324	1073.744	2790.612	2290.68	4300.782	19656.62
BPR	474.1987	1038.112	2692.889	2218.691	4100.706	14277.23
INT	228	708.75	2003.75	1775.75	1453.693	1729.193
bn	228	710	2003.5	1775.5	1453.952	1728.883
an	227	708	2003.5	1776.5	1453.434	1729.545
wiqd	12.21616	34.12286	107.3433	95.12711	100.768	190.4392
iqd	12.93282	35.38755	108.837	95.90422	103.7459	202.8949
iqdmed	4	10	26	22	30.73038	85.32622
iqdmin	1	1	1	0	1.260147	8.459883
backward	0.1600037	0.1900024	0.2300034	0.0699997	0.1919607	0.0609217

Appendix 3: Estimates

Table 7: Models in Book1 - 30 minutes

[Estimates are obtained by using the routine *ivreg2*, see Baum *et al.* (2003) and Baum *et al.* (2007). Inside each cell: estimated parameter (first number), robust to both arbitrary heteroskedasticy and arbitrary intra-groups autocorrelation standard error (second number). N is the number of observations; N_clust is the number of stocks. Labels j, jp and jdf indicate the J test statistic of Hansen (1982), its $P-value$, and its degrees of freedom. The Hansen test is a test of overidentifying restrictions; the joint null hypothesis is that the instruments are valid instruments, i.e. uncorrelated with the error term. Labels $idstat$, idp and $iddf$ indicate the LM version of the Kleibergen and Paap (2006) rk statistic, its $P-value$, and its degrees of freedom. The rk statistic is a underidentification test, testing that the instruments are "relevant" (correlated with the endogenous regressors). The Kleibergen-Paap test is a test of the rank of the matrix of reduced form coefficients on the instruments; under the null hypothesis the matrix is not full column rank and the equation is underidentified. See also Kleibergen and Schaffer (2007) for the Stata implementation, ranktest.]

variable	*A*	*A_VOL*	*A_IQD*	*A_IQDMED*	*A_IQDMIN*	*A_WIQD*
gkr	-2287,078		-1959,916	-2249,889	-2425,528	-1912,592
	1778,382		1623,752	1899,420	1751,597	1607,719
rbas	-0,423	0,085	-0,441	-0,622	-0,173	-0,455
	0,670	0,409	0,675	0,741	0,807	0,668
logoi	0,006	-0,012	0,006	0,012	0,017	0,008
	0,015	0,011	0,014	0,016	0,018	0,014
l1logtv	0,132	0,205	0,130	0,152	0,187	0,133
	0,009 ***	0,011 ***	0,009 ***	0,010 ***	0,019 ***	0,009 ***
l2logtv	0,083	0,094	0,085	0,097	0,112	0,086
	0,010 ***	0,008 ***	0,010 ***	0,008 ***	0,008 ***	0,010 ***
l3logtv	0,075	0,089	0,075	0,085	0,095	0,077
	0,009 ***	0,008 ***	0,009 ***	0,011 ***	0,013 ***	0,009 ***
l4logtv	0,059	0,068	0,058	0,067	0,071	0,059
	0,008 ***	0,008 ***	0,008 ***	0,009 ***	0,010 ***	0,008 ***
logint	0,414	0,406				
	0,135 ***	0,036 ***				
backward	0,595	0,119	0,614	0,624	0,265	0,591
	0,163 ***	0,128	0,151 ***	0,174 ***	0,189	0,150 ***
logiqd			-0,388			
			0,125 ***			
logiqdmed				-0,295		
				0,103 ***		
logiqdmin					-0,201	
					0,032 ***	
logwiqd						-0,365
						0,121 ***
vol		-158,266				
		27,234 ***				
N	64094	41072	64094	64094	64094	64094
N_clust	275	256	275	275	275	275
rmse	0,9593	0,5796	0,8906	0,9562	1,0222	0,8818
j	12,095	16,8494	12,8823	10,3016	8,5123	11,5596
jp	0,208	0,0511	0,168	0,3266	0,4835	0,2393
jdf	9	9	9	9	9	9
idstat	9,3116	66,9863	10,0961	9,7693	7,5724	8,7343
idp	0,5028	0	0,4321	0,461	0,6705	0,5575
iddf	10	10	10	10	10	10

Table 8: Models in Book2 - 30 minutes

[Estimates are obtained by using the routine *ivreg2*, see Baum *et al.* (2003) and Baum *et al.* (2007). Inside each cell: estimated parameter (first number), robust to both arbitrary heteroskedasticy and arbitrary intra-groups autocorrelation standard error (second number). N is the number of observations; N_clust is the number of stocks. Labels j, jp and jdf indicate the J test statistic of Hansen (1982), its $P - value$, and its degrees of freedom. The Hansen test is a test of overidentifying restrictions; the joint null hypothesis is that the instruments are valid instruments, i.e. uncorrelated with the error term. Labels $idstat$, idp and $iddf$ indicate the LM version of the Kleibergen and Paap (2006) rk statistic, its $P - value$, and its degrees of freedom. The rk statistic is a underidentification test, testing that the instruments are "relevant" (correlated with the endogenous regressors). The Kleibergen-Paap test is a test of the rank of the matrix of reduced form coefficients on the instruments; under the null hypothesis the matrix is not full column rank and the equation is underidentified. See also Kleibergen and Schaffer (2007) for the Stata implementation, ranktest.]

variable	A	A_VOL	A_IQD	A_IQDMED	A_IQDMIN	A_WIQD
gkr	-303,517		-299,553	-137,489	172,355	-280,321
	78,602 ***		78,717 ***	59,384 **	54,637 ***	76,512 ***
rbas	-60,403	1,531	-60,640	-63,307	-52,076	-59,689
	4,767 ***	12,952	4,735 ***	4,474 ***	4,187 ***	4,676 ***
logoi	-0,015	-0,021	-0,015	-0,009	-0,008	-0,014
	0,008 *	0,010 **	0,008 *	0,008	0,008	0,008 *
l1logtv	0,172	0,253	0,172	0,180	0,196	0,175
	0,008 ***	0,008 ***	0,008 ***	0,008 ***	0,008 ***	0,008 ***
l2logtv	0,113	0,139	0,113	0,119	0,128	0,115
	0,007 ***	0,008 ***	0,007 ***	0,006 ***	0,006 ***	0,007 ***
l3logtv	0,096	0,106	0,096	0,098	0,103	0,097
	0,007 ***	0,007 ***	0,007 ***	0,007 ***	0,007 ***	0,007 ***
l4logtv	0,080	0,108	0,080	0,082	0,085	0,081
	0,006 ***	0,007 ***	0,006 ***	0,006 ***	0,006 ***	0,006 ***
logint	0,282	0,370				
	0,011 ***	0,027 ***				
backward	0,053	0,155	0,050	0,029	-0,051	0,041
	0,052	0,056 ***	0,052	0,050	0,045	0,051
logiqd			-0,278			
			0,011 ***			
logiqdmed				-0,205		
				0,007 ***		
logiqdmin					-0,155	
					0,009 ***	
logwiqd						-0,261
						0,011 ***
vol		-107,107				
		18,104 ***				
N	78024	49456	78024	78024	78024	78024
N_clust	269	245	269	269	269	269
rmse	0,692	0,5513	0,6929	0,6863	0,6865	0,6923
j	11,2473	18,8713	11,8152	11,5353	9,1981	10,7028
jp	0,2591	0,0263	0,2239	0,2408	0,4192	0,2966
jdf	9	9	9	9	9	9
idstat	25,9509	37,2525	25,7875	26,2993	27,9936	25,6301
idp	0,0038	0,0001	0,004	0,0034	0,0018	0,0043
iddf	10	10	10	10	10	10

Table 9: Models in Book3 - 30 minutes

[Estimates are obtained by using the routine *ivreg2*, see Baum *et al.* (2003) and Baum *et al.* (2007). Inside each cell: estimated parameter (first number), robust to both arbitrary heteroskedasticy and arbitrary intra-groups autocorrelation standard error (second number). N is the number of observations; N_clust is the number of stocks. Labels j, jp and jdf indicate the J test statistic of Hansen (1982), its $P-value$, and its degrees of freedom. The Hansen test is a test of overidentifying restrictions; the joint null hypothesis is that the instruments are valid instruments, i.e. uncorrelated with the error term. Labels $idstat$, idp and $iddf$ indicate the LM version of the Kleibergen and Paap (2006) rk statistic, its $P-value$, and its degrees of freedom. The rk statistic is a underidentification test, testing that the instruments are "relevant" (correlated with the endogenous regressors). The Kleibergen-Paap test is a test of the rank of the matrix of reduced form coefficients on the instruments; under the null hypothesis the matrix is not full column rank and the equation is underidentified. See also Kleibergen and Schaffer (2007) for the Stata implementation, ranktest.]

variable	A	A_VOL	A_IQD	A_IQDMED	A_IQDMIN	A_WIQD
gkr	-316,414		-320,841	-231,558	0,266	-307,866
	89,962 ***		93,413 ***	68,988 ***	41,354	89,802 ***
rbas	-48,785	43,885	-49,037	-53,979	-45,681	-49,155
	4,152 ***	28,273	4,200 ***	4,267 ***	3,658 ***	4,166 ***
logoi	-0,004	-0,003	-0,004	-0,001	0,004	-0,002
	0,009	0,011	0,009	0,009	0,008	0,009
l1logtv	0,148	0,234	0,147	0,158	0,173	0,150
	0,010 ***	0,011 ***	0,010 ***	0,009 ***	0,009 ***	0,010 ***
l2logtv	0,106	0,129	0,105	0,111	0,119	0,106
	0,008 ***	0,007 ***	0,008 ***	0,007 ***	0,007 ***	0,008 ***
l3logtv	0,077	0,083	0,078	0,080	0,086	0,077
	0,007 ***	0,008 ***	0,007 ***	0,007 ***	0,006 ***	0,007 ***
l4logtv	0,074	0,084	0,075	0,076	0,077	0,075
	0,006 ***	0,009 ***	0,006 ***	0,006 ***	0,006 ***	0,006 ***
logint	0,273	0,439				
	0,014 ***	0,050 ***				
backward	0,320	0,390	0,319	0,320	0,277	0,315
	0,053 ***	0,060 ***	0,053 ***	0,051 ***	0,048 ***	0,052 ***
logiqd			-0,271			
			0,014 ***			
logiqdmed				-0,203		
				0,008 ***		
logiqdmin					-0,150	
					0,008 ***	
logwiqd						-0,256
						0,013 ***
vol		-159,270				
		35,228 ***				
N	83972	52863	83972	83972	83972	83972
N_clust	265	241	265	265	265	265
rmse	0,6888	0,6226	0,6905	0,6858	0,6796	0,6899
j	7,5771	18,2539	7,332	7,4425	7,8942	7,1017
jp	0,5773	0,0323	0,6026	0,5912	0,5448	0,6265
jdf	9	9	9	9	9	9
idstat	18,1102	18,6365	18,0542	19,6531	23,2373	18,1316
idp	0,0531	0,0451	0,0541	0,0327	0,0099	0,0528
iddf	10	10	10	10	10	10

Table 10: Models in Book1 - 150 minutes

[Estimates are obtained by using the routine *ivreg*2, see Baum *et al.* (2003) and Baum *et al.* (2007). Inside each cell: estimated parameter (first number), robust to both arbitrary heteroskedasticy and arbitrary intra-groups autocorrelation standard error (second number). N is the number of observations; N_clust is the number of stocks. Labels j, jp and jdf indicate the J test statistic of Hansen (1982), its $P-value$, and its degrees of freedom. The Hansen test is a test of overidentifying restrictions; the joint null hypothesis is that the instruments are valid instruments, i.e. uncorrelated with the error term. Labels $idstat$, idp and $iddf$ indicate the LM version of the Kleibergen and Paap (2006) rk statistic, its $P-value$, and its degrees of freedom. The rk statistic is a underidentification test, testing that the instruments are "relevant" (correlated with the endogenous regressors). The Kleibergen-Paap test is a test of the rank of the matrix of reduced form coefficients on the instruments; under the null hypothesis the matrix is not full column rank and the equation is underidentified. See also Kleibergen and Schaffer (2007) for the Stata implementation, ranktest.]

variable	A	A_GKR	A_IQD	A_IQDMED	A_IQDMIN	A_WIQD
vol	-99,714		-111,779	-47,600	11,946	-95,839
	30,447 ***		32,824 ***	21,396 **	16,638	31,379 ***
rbas	-63,554	-2,041	-58,194	-103,363	-120,614	-68,194
	32,253 **	2,618	32,686 *	31,539 ***	33,978 ***	33,988 **
logoi	-0,006	0,010	-0,009	0,000	-0,003	-0,009
	0,030	0,029	0,030	0,031	0,031	0,031
l1logtv	0,269	0,201	0,265	0,285	0,290	0,273
	0,021 ***	0,020 ***	0,021 ***	0,021 ***	0,022 ***	0,021 ***
l2logtv	0,120	0,088	0,116	0,130	0,139	0,120
	0,023 ***	0,018 ***	0,023 ***	0,025 ***	0,026 ***	0,023 ***
logint	0,351	0,230				
	0,048 ***	0,033 ***				
backward	-0,078	0,758	-0,035	0,170	0,238	0,005
	0,230	0,201 ***	0,229	0,215	0,213	0,226
gkr		-407,887				
		188,771 **				
logiqd			-0,376			
			0,053 ***			
logiqdmed				-0,212		
				0,029 ***		
logiqdmin					-0,285	
					0,079 ***	
logwiqd						-0,329
						0,048 ***
N	6975	8506	6975	6975	6975	6975
N_clust	268	277	268	268	268	268
rmse	0,4447	0,5191	0,4486	0,459	0,4684	0,4484
j	1,302	6,3933	1,3578	1,0469	0,5473	1,1802
jp	0,7287	0,094	0,7155	0,7899	0,9084	0,7577
jdf	3	3	3	3	3	3
idstat	36,6705	13,9871	34,4787	55,5882	56,624	35,5544
idp	0	0,0073	0	0	0	0
iddf	4	4	4	4	4	4

Table 11: Models in Book2 - 150 minutes

[Estimates are obtained by using the routine *ivreg2*, see Baum *et al.* (2003) and Baum *et al.* (2007). Inside each cell: estimated parameter (first number), robust to both arbitrary heteroskedasticy and arbitrary intra-groups autocorrelation standard error (second number). N is the number of observations; N_clust is the number of stocks. Labels j, jp and jdf indicate the J test statistic of Hansen (1982), its $P-value$, and its degrees of freedom. The Hansen test is a test of overidentifying restrictions; the joint null hypothesis is that the instruments are valid instruments, i.e. uncorrelated with the error term. Labels $idstat$, idp and $iddf$ indicate the LM version of the Kleibergen and Paap (2006) rk statistic, its $P-value$, and its degrees of freedom. The rk statistic is a underidentification test, testing that the instruments are "relevant" (correlated with the endogenous regressors). The Kleibergen-Paap test is a test of the rank of the matrix of reduced form coefficients on the instruments; under the null hypothesis the matrix is not full column rank and the equation is underidentified. See also Kleibergen and Schaffer (2007) for the Stata implementation, ranktest.]

variable	A	A_GKR	A_IQD	A_IQDMED	A_IQDMIN	A_WIQD
vol	-68,473		-76,228	-42,952	5,225	-64,296
	20,012 ***		22,724 ***	14,788 ***	10,596	19,647 ***
rbas	-18,550	-53,203	-15,834	-33,863	-44,429	-21,342
	12,393	12,117 ***	13,308	10,628 ***	10,351 ***	12,211 **
logoi	-0,035	-0,015	-0,034	-0,030	-0,028	-0,033
	0,023	0,025	0,023	0,022	0,021	0,023
l1logtv	0,338	0,251	0,336	0,345	0,350	0,342
	0,019 ***	0,019 ***	0,019 ***	0,017 ***	0,017 ***	0,019 ***
l2logtv	0,220	0,157	0,218	0,221	0,232	0,220
	0,015 ***	0,023 ***	0,015 ***	0,015 ***	0,014 ***	0,015 ***
logint	0,287	0,303				
	0,044 ***	0,049 ***				
backward	0,160	0,099	0,175	0,095	-0,054	0,146
	0,083 *	0,101	0,088 **	0,075	0,065	0,082 **
gkr		-588,609				
		285,542 **				
logiqd			-0,302			
			0,051 ***			
logiqdmed				-0,185		
				0,025 ***		
logiqdmin					-0,037	
					0,042	
logwiqd						-0,262
						0,042 ***
N	8274	10664	8274	8274	8274	8274
N_clust	261	271	261	261	261	261
rmse	0,4227	0,5986	0,4297	0,4096	0,396	0,4214
j	0,9548	0,6267	1,0564	0,4764	0,3702	1,0576
jp	0,8122	0,8903	0,7876	0,924	0,9463	0,7873
jdf	3	3	3	3	3	3
idstat	13,4582	9,8221	12,3304	19,4823	20,1336	13,8452
idp	0,0092	0,0435	0,0151	0,0006	0,0005	0,0078
iddf	4	4	4	4	4	4

Table 12: Models in Book3 - 150 minutes

[Estimates are obtained by using the routine *ivreg*2, see Baum *et al.* (2003) and Baum *et al.* (2007). Inside each cell: estimated parameter (first number), robust to both arbitrary heteroskedasticy and arbitrary intra-groups autocorrelation standard error (second number). N is the number of observations; N_clust is the number of stocks. Labels j, jp and jdf indicate the J test statistic of Hansen (1982), its $P-value$, and its degrees of freedom. The Hansen test is a test of overidentifying restrictions; the joint null hypothesis is that the instruments are valid instruments, i.e. uncorrelated with the error term. Labels $idstat$, idp and $iddf$ indicate the LM version of the Kleibergen and Paap (2006) rk statistic, its $P-value$, and its degrees of freedom. The rk statistic is a underidentification test, testing that the instruments are "relevant" (correlated with the endogenous regressors). The Kleibergen-Paap test is a test of the rank of the matrix of reduced form coefficients on the instruments; under the null hypothesis the matrix is not full column rank and the equation is underidentified. See also Kleibergen and Schaffer (2007) for the Stata implementation, ranktest.]

variable	A	A_GKR	A_IQD	A_IQDMED	A_IQDMIN	A_WIQD
vol	-46,417		-49,991	-35,258	-1,857	-46,864
	13,852 ***		14,690 ***	12,614 ***	9,587	13,793 ***
rbas	-9,321	-29,263	-8,567	-16,337	-21,391	-9,563
	5,450 *	10,164 ***	5,110 *	8,997 *	12,600 *	5,332 *
logoi	0,025	-0,015	0,025	0,027	0,021	0,027
	0,025	0,030	0,025	0,024	0,023	0,025
l1logtv	0,314	0,212	0,313	0,317	0,323	0,313
	0,017 ***	0,018 ***	0,017 ***	0,017 ***	0,016 ***	0,017 ***
l2logtv	0,166	0,104	0,165	0,169	0,174	0,167
	0,022 ***	0,020 ***	0,022 ***	0,020 ***	0,021 ***	0,022 ***
logint	0,213	0,187				
	0,030 ***	0,050 ***				
backward	0,224	0,284	0,224	0,190	0,154	0,210
	0,076 ***	0,122 **	0,077 ***	0,077 **	0,078 **	0,076 ***
gkr		39,177				
		192,987				
logiqd			-0,221			
			0,032 ***			
logiqdmed				-0,153		
				0,021 ***		
logiqdmin					-0,135	
					0,043 ***	
logwiqd						-0,206
						0,029 ***
N	8898	11881	8898	8898	8898	8898
N_clust	257	268	257	257	257	257
rmse	0,3895	0,554	0,3917	0,3868	0,3815	0,3901
j	7,6863	7,1847	8,0323	5,3018	3,0899	7,8623
jp	0,053	0,0662	0,0453	0,151	0,378	0,0489
jdf	3	3	3	3	3	3
idstat	37,4182	10,4358	34,6751	47,1929	53,7992	36,4034
idp	0	0,0337	0	0	0	0
iddf	4	4	4	4	4	4

Table 13: Models in Book1 - daily

[Estimates are obtained by using the routine *ivreg2*, see Baum *et al.* (2003) and Baum *et al.* (2007). Inside each cell: estimated parameter (first number), robust to both arbitrary heteroskedasticy and arbitrary intra-groups autocorrelation standard error (second number). N is the number of observations; N_clust is the number of stocks. Labels j, jp and jdf indicate the J test statistic of Hansen (1982), its $P-value$, and its degrees of freedom. The Hansen test is a test of overidentifying restrictions; the joint null hypothesis is that the instruments are valid instruments, i.e. uncorrelated with the error term. Labels $idstat$, idp and $iddf$ indicate the LM version of the Kleibergen and Paap (2006) rk statistic, its $P-value$, and its degrees of freedom. The rk statistic is a underidentification test, testing that the instruments are "relevant" (correlated with the endogenous regressors). The Kleibergen-Paap test is a test of the rank of the matrix of reduced form coefficients on the instruments; under the null hypothesis the matrix is not full column rank and the equation is underidentified. See also Kleibergen and Schaffer (2007) for the Stata implementation, ranktest.]

variable	*A*	*A_GKR*	*A_IQD*	*A_IQDMED*	*A_IQDMIN*	*A_WIQD*	*A_TED*
vol	-21.067		-29.093	19.583	30.757	-22.461	-16.617
	35.634		38.665	27.373	20.914	37.439	33.265
rbas	-13.487	-91.191	-13.560	-5.498	-12.501	-19.620	-13.673
	29.336	65.699	28.795	29.385	31.171	29.392	27.897
logoi	0.164	0.067	0.152	0.181	0.146	0.145	0.154
	0.159	0.204	0.159	0.163	0.159	0.158	0.160
l1logtv	0.222	0.176	0.219	0.234	0.221	0.213	0.220
	0.052 ***	0.094 *	0.052 ***	0.054 ***	0.053 ***	0.053 ***	0.053 ***
l2logtv	-0.024	0.000	-0.024	-0.024	-0.029	-0.029	-0.030
	0.044	0.064	0.044	0.046	0.044	0.044	0.044
l3logtv	0.004	-0.038	0.002	0.029	0.026	0.008	0.009
	0.056	0.091	0.057	0.054	0.053	0.057	0.056
l4logtv	0.080	0.061	0.084	0.078	0.084	0.092	0.074
	0.039 **	0.081	0.040 **	0.038 **	0.039 **	0.041 **	0.038 **
logint	0.236	0.328					0.223
	0.098 **	0.111 ***					0.090 **
backward	-0.844	-1.619	-0.836	-0.567	-0.711	-1.022	
	0.602	0.935 *	0.586	0.570	0.624	0.624	
d23jun	-0.090	-0.066	-0.087	-0.095	-0.100	-0.084	-0.096
	0.039 **	0.054	0.039 **	0.039 **	0.043 **	0.039 **	0.039 **
gkr		-334.505					
		305.895					
logiqd			-0.267				
			0.113 **				
logiqdmed				-0.075			
				0.062			
logiqdmin					-0.539		
					0.281 *		
logwiqd						-0.226	
						0.102 **	
ted							0.262
							0.122 **
N	1104	1213	1104	1104	1104	1104	1104
N_clust	264	275	264	264	264	264	264
rmse	0.3141	0.491	0.3124	0.3154	0.3142	0.316	0.3119
j	5.6207	3.9069	5.9415	6.0612	4.3145	5.9704	5.3397
jp	0.7772	0.9174	0.7458	0.7338	0.8895	0.7429	0.8037
jdf	9	9	9	9	9	9	9
idstat	11.8007	13.8309	11.5157	12.0215	11.4934	11.3852	11.8112
idp	0.2986	0.1808	0.3188	0.2836	0.3204	0.3283	0.2979
iddf	10	10	10	10	10	10	10

Table 14: Models in Book2 - daily

[Estimates are obtained by using the routine *ivreg*2, see Baum *et al.* (2003) and Baum *et al.* (2007). Inside each cell: estimated parameter (first number), robust to both arbitrary heteroskedasticy and arbitrary intra-groups autocorrelation standard error (second number). N is the number of observations; N_clust is the number of stocks. Labels j, jp and jdf indicate the J test statistic of Hansen (1982), its $P-value$, and its degrees of freedom. The Hansen test is a test of overidentifying restrictions; the joint null hypothesis is that the instruments are valid instruments, i.e. uncorrelated with the error term. Labels $idstat$, idp and $iddf$ indicate the LM version of the Kleibergen and Paap (2006) rk statistic, its $P-value$, and its degrees of freedom. The rk statistic is a underidentification test, testing that the instruments are "relevant" (correlated with the endogenous regressors). The Kleibergen-Paap test is a test of the rank of the matrix of reduced form coefficients on the instruments; under the null hypothesis the matrix is not full column rank and the equation is underidentified. See also Kleibergen and Schaffer (2007) for the Stata implementation, ranktest.]

variable	*A*	*A_GKR*	*A_IQD*	*A_IQDMED*	*A_IQDMIN*	*A_WIQD*	*A_TED*	*A_16AUG*
vol	1.683		-2.442	13.747	35.115	0.405	-4.938	3.997
	24.660		25.837	18.976	15.977 **	24.908	26.498	23.838
rbas	-56.170	-61.542	-55.054	-62.461	-66.245	-57.047	-56.836	-55.308
	28.433 **	24.788 **	28.618 *	26.781 **	27.466 **	28.268 **	29.141 *	27.757 **
logoi	0.028	0.026	0.028	0.031	0.044	0.043	0.027	0.020
	0.081	0.104	0.081	0.081	0.081	0.081	0.082	0.081
l1logtv	0.138	0.163	0.136	0.141	0.155	0.136	0.150	0.148
	0.035 ***	0.040 ***	0.035 ***	0.035 ***	0.037 ***	0.036 ***	0.036 ***	0.035 ***
l2logtv	0.069	0.095	0.065	0.071	0.095	0.066	0.058	0.068
	0.041 *	0.033 ***	0.041	0.041 *	0.039 **	0.042	0.042	0.040 *
l3logtv	0.047	0.003	0.047	0.053	0.046	0.046	0.051	0.049
	0.040	0.042	0.040	0.040	0.040	0.040	0.040	0.040
l4logtv	0.070	0.026	0.069	0.071	0.061	0.068	0.069	0.069
	0.038 *	0.041	0.038 *	0.038 *	0.039	0.038 *	0.038 *	0.038 *
logint	0.192	0.220					0.205	0.190
	0.067 ***	0.074 ***					0.071 ***	0.063 ***
backward	0.294	0.316	0.314	0.251	0.141	0.277		
	0.132 **	0.210	0.135 **	0.132 *	0.125	0.135 **		
d16aug								0.068
								0.023 ***
gkr		-20.045						
		224.279						
logiqd			-0.206					
			0.071 ***					
logiqdmed				-0.129				
				0.041 ***				
logiqdmin					0.159			
					0.033 ***			
logwiqd						-0.189		
						0.064 ***		
ted							0.087	
							0.030 ***	
N	1464	1686	1464	1464	1464	1464	1464	1464
N_clust	257	272	257	257	257	257	257	257
rmse	0.3474	0.44	0.3463	0.3527	0.3617	0.349	0.3489	0.3454
j	12.4715	10.3283	12.2575	12.0314	13.2649	12.3719	11.4896	12.3683
jp	0.188	0.3246	0.1992	0.2115	0.151	0.1931	0.2436	0.1933
jdf	9	9	9	9	9	9	9	9
idstat	15.9357	9.1703	15.4696	19.867	20.1251	16.8687	15.7164	15.4478
idp	0.1015	0.516	0.1159	0.0305	0.0281	0.0773	0.108	0.1166
iddf	10	10	10	10	10	10	10	10

Table 15: Models in Book3 - daily

[Estimates are obtained by using the routine *ivreg2*, see Baum *et al.* (2003) and Baum *et al.* (2007). Inside each cell: estimated parameter (first number), robust to both arbitrary heteroskedasticy and arbitrary intra-groups autocorrelation standard error (second number). N is the number of observations; N_clust is the number of stocks. Labels j, jp and jdf indicate the J test statistic of Hansen (1982), its $P-value$, and its degrees of freedom. The Hansen test is a test of overidentifying restrictions; the joint null hypothesis is that the instruments are valid instruments, i.e. uncorrelated with the error term. Labels $idstat$, idp and $iddf$ indicate the LM version of the Kleibergen and Paap (2006) rk statistic, its $P-value$, and its degrees of freedom. The rk statistic is a underidentification test, testing that the instruments are "relevant" (correlated with the endogenous regressors). The Kleibergen-Paap test is a test of the rank of the matrix of reduced form coefficients on the instruments; under the null hypothesis the matrix is not full column rank and the equation is underidentified. See also Kleibergen and Schaffer (2007) for the Stata implementation, ranktest.]

variable	*A*	*A_GKR*	*A_IQD*	*A_IQDMED*	*A_IQDMIN*	*A_WIQD*	*A_TED*
vol	-13.134		-16.561	-9.872	-3.519	-16.564	-12.556
	12.6397		13.2869	12.4328	11.6518	13.2778	13.0572
rbas	-27.694	-41.222	-28.342	-24.682	-22.009	-29.231	-26.587
	15.7842 *	29.22	15.951 *	15.3608	15.1234	16.0781 *	16.1479
logoi	-0.157	-0.066	-0.161	-0.161	-0.154	-0.165	-0.167
	0.0996	0.1653	0.1001	0.0949 *	0.0978	0.1008	0.1023
l1logtv	0.144	0.17	0.14	0.148	0.161	0.144	0.141
	0.0424 ***	0.0485 ***	0.043 ***	0.04 ***	0.0394 ***	0.0425 ***	0.043 ***
l2logtv	0.141	0.116	0.14	0.138	0.137	0.142	0.145
	0.0373 ***	0.061 *	0.0377 ***	0.0361 ***	0.0364 ***	0.0385 ***	0.0363 ***
l3logtv	0.068	0.131	0.067	0.071	0.072	0.063	0.068
	0.0307 **	0.1654	0.0307 **	0.0295 **	0.0307 **	0.031 **	0.0299 **
l4logtv	0.052	-0.027	0.053	0.048	0.056	0.054	0.052
	0.031 *	0.0646	0.0309 *	0.0298	0.0316 *	0.0307 *	0.0311 *
logint	0.135	0.174					0.137
	0.0353 ***	0.0476 ***					0.0348 ***
backward	0.17	0.718	0.169	0.098	0.111	0.156	
	0.374	0.7357	0.3741	0.3575	0.3689	0.3699	
d31oct	0.114	0.133	0.104	0.101	0.13	0.111	0.118
	0.0558 **	0.0844	0.0543 *	0.0509 **	0.0557 **		
gkr		-126.257					
		136.0636					
logiqd			-0.177				
			0.0394 ***				
logiqdmed				-0.114			
				0.0288 ***			
logiqdmin					-0.187		
					0.1108 *		
logwiqd						-0.172	
						0.0363 ***	
ted							-0.041
							0.0707
N	1584	1851	1584	1584	1584	1584	1584
N_clust	252	269	252	252	252	252	252
rmse	0.3239	0.4793	0.3227	0.3198	0.322	0.323	0.3239
j	4.6321	10.1297	4.7488	6.102	5.6648	4.6456	4.2747
jp	0.8651	0.3401	0.8556	0.7297	0.7729	0.864	0.8924
jdf	9	9	9	9	9	9	9
idstat	17.2925	8.9217	16.4876	18.4449	20.5848	16.515	14.3
idp	0.0681	0.5396	0.0865	0.0479	0.0242	0.0858	0.1597
iddf	10	10	10	10	10	10	10