ELSEVIER

Journal of Financial Markets 7 (2004) 145–185

Journal of FINANCIAL MARKETS

www.elsevier.com/locate/econbase

Expandable limit order markets☆

Leslie Boni[a,*], Chris Leach[b,c]

[a] University of New Mexico, USA
[b] University of Colorado, Boulder, USA
[c] Stockholm Institute of Financial Research, Sweden

Abstract

In expandable limit order markets, quotes represent a firm-price commitment to trade but also serve as an invitation for one-on-one quantity negotiation. Analyzing interdealer Treasury data, we find evidence that expandable limit orders offer benefits by reducing costs associated with information leakage and stale limit orders. They also create queues for other traders who must wait to transact, however. We conclude that viable electronic communication network designs which incorporate expandable limit orders to encourage block-trading must also provide escape valves, such as the periodic enforcement of strict price-time priority, to limit the negative externality imposed on those waiting to trade.

JEL classification: G14; G24

Keywords: Limit orders; Hidden orders; Treasury securities; Interdealer markets; Electronic communication networks

0. Introduction

For more than 20 years, brokers in the interdealer Treasury market allowed dealers to use *expandable* limit orders. Specifically, any trader whose limit order was

☆An earlier version of this paper was titled "Depth Discovery in a Market with Expandable Limit Orders: An Investigation of the U.S. Treasury Market." The authors thank an anonymous referee, Maureen O'Hara, Michael Fleming, Eric Hughson, Michael Goldstein, Janet Rogers, seminar participants at New York University, the Securities and Exchange Commission, the University of Colorado at Boulder, the Stockholm Institute of Financial Research, the 2000 Nasdaq-Notre Dame Microstructure Conference, and the 2001 European Financial Management Association Conference for their comments and suggestions.

*Corresponding author. 450, 5th Street NW, Washington, DC 20549-1105, USA. Tel.: +1-202-942-8038; fax: +1-202-942-9657.

E-mail address: boni@unm.edu (L. Boni).

1386-4181/$ - see front matter
doi:10.1016/j.finmar.2003.10.001

just executed earned the right-of-refusal to trade additional volume, even though limit orders at the same price had already been placed by others. We use a unique dataset for this brokered Treasury market to answer the following questions. First, how do traders trade when limit orders are expandable? Second, what are the implications for depth and price discovery? Third, how can expandable limit orders be incorporated in electronic communication networks (''ECNs'') to compete with upstairs markets for large (''block'') trades? Fourth, what are the implications for researchers who use Treasury market quote and transaction data?[1,2]

When limit orders are expandable, traders can use small limit orders to search for trading counterparties, then negotiate quantity beyond the initial limit order. Traders enter into quantity negotiations only with those who have already expressed their willingness to trade. This may reduce information free-riding by other market participants who could interpret a placement of a large limit order as a signal of valuable private information and seek to free-ride on the signal.[3]

Limit orders can become ''stale'' as traders update their own willingness to trade. Traders incur ''stale limit order costs'' when orders are executed with regret before they can be cancelled.[4] In markets that permit expandable limit orders, the *initial* limit order is exposed to stale order risk, but the trader updates his willingness to trade additional volume in real time. Even minimum-size limit orders earn the right-of-refusal option. Thus, expandable limit orders potentially reduce the costs associated not only with information free-riding but also with stale limit orders.

Expandable limit orders are distinct from ''hidden limit orders'' that have been analyzed for other markets. Biais et al. (1995) and Panchapagesan (2000) describe hidden limit orders as the electronically *staged* limit orders of the Paris Bourse (''CAC'') and Toronto Stock Exchange (''CATS''), respectively. Other types of hidden limit orders have been defined for the NYSE. McInish and Wood (1995) define NYSE hidden limit orders as the limit orders specialists have chosen not to display. Sofianos and Werner (1997) describe how NYSE floor brokers sometimes ''work'' large limit orders: the brokers ask the specialist to display only part of the order. The hidden component of an expandable limit order is different from that of CAC and CATS, where traders pre-commit to the demand that will be revealed in stages, and on the NYSE, where traders fully communicate their willingness to trade to the specialist or floor broker. With expandable limit orders, the only component

[1] Although this right-of-refusal feature has been used in the brokered interdealer Treasury market for more than 20 years (Garbade, 1978), it is rarely mentioned in the literature. Exceptions are Stigum (1990), U.S. General Accounting Office (1990), Umlauf (1991), and Fleming (1997). Boni and Leach (2002) examines expandable limit orders in the context of supply contraction in the Treasury market in the late 1990s and 2000. Because they use a coarser dataset, they must rely on a noisier proxy for expandable order usage and cannot analyze price impacts or the effects of asymmetric quote depth.

[2] Expandable limit orders are also an important feature of eSpeed (formerly the Cantor Financial Futures Exchange) and BrokerTec.

[3] As discussed in detail in Section 1.1, the limit order traders' counterparties are dealers who issue ''responding'' orders. These responding order traders also receive the right-of-refusal to negotiate quantity.

[4] See Stoll (1992) for examples of traders' exposure to stale limit order risk.

of demand that is communicated is the actual willingness (or lack thereof) to trade the next incremental quantity. Because there is no pre-commitment for incremental quantities, the trader's true desire to trade can expand or contract based on negotiations and market conditions.

We document evidence consistent with the hypothesis that traders use expandable limit order strategies to reduce costs associated with stale limit orders and information leakage. Because traders can expand limit orders, many transactions (56.5% of dollar volume) discover greater-than-quoted depth. Traders use expandable order strategies more frequently when continuous monitoring of the market is relatively costly and when information asymmetry is greater, such as during hours outside the New York business day and for trading off-the-run issues. Consistent with what we would expect to find for electronic "pure" limit order books (i.e., those that enforce strict price-time priority), we document that few of the trades of liquidity demanders (0.6% of dollar volume) receive price improvement.

We document that, despite the ability of traders to use expandable limit order strategies to camouflage the extent of their willingness to trade, asymmetric depth (i.e., greater quoted depth on one side of the book than on the other) is correlated with the sign of the next trade. This finding is consistent with what would be expected for transparent limit order books without expandable or hidden limit orders.[5] As a result, we conclude that expandable limit orders do not completely eliminate pre-trade information leakage or the potential for information free-riding.

Consistent with Huang et al. (2002) and Green (2003), we find that large trades in the interdealer broker Treasury market do not have a greater price impact than small trades. Green (2003) shows that limit orders generally receive negative compensation for providing liquidity, however. Using the sequential return approach of Koski and Michaely (2000), we show that liquidity providers experience lower costs if they expand limit orders. An alternative interpretation is that expanded orders *cause* stabilization of ex-post prices. Cohen and Shin (2003) and Green (2003) suggest that traders may watch order flow to help forecast future prices and decide on the direction of their next trades. Market participants may interpret traders' willingness to expand limit orders as an indication of balanced order flow.

We also contribute to the discussion of whether price discovery in the interdealer broker Treasury market is structurally different when trading is extremely rapid and volatile ("stressful"). For example, Furfine and Remolona (2001) find that signed order flow had a stronger effect on ex-post prices during the Long Term Capital Management crisis of autumn 1998. Cohen and Shin (2003) document "positive feedback" trading (i.e., positive returns elicit trading in the same direction) during hectic trading periods in 1999–2000. They note that this is consistent with traders' reluctance to attempt to try to "catch a falling knife". Similarly, we document a stronger impact of trades on returns for October 27 and 28, 1997, when U.S. equity

[5] See Parlour (1998), for example.

markets plummeted and rebounded. With the exception of large purchases, however, the option to expand limit orders does not appear to provide dealers who try to catch the falling knife with much of an advantage on extremely volatile days. And perhaps not surprisingly, on stressful days, the workup process is more likely to invite price competition from aggressive limit and responding orders. Time-overlapping trades of opposite sign, possibly the result of dealers' canceling their limit orders and issuing responding orders, are also more common on stressful days.

While a market design that allows expandable limit orders offers benefits, it also introduces an important friction, first recognized in the literature by Garbade (1978). Expandable limit orders create queues for other traders who must wait to transact. In the brokered interdealer market, we find evidence consistent with traders' willingness to use another broker at worse prices rather than wait in the queue of a broker with quantity negotiations in progress. We conclude that the existence of multiple brokers in this market provides an escape valve to remaining in the queue of any single broker firm. We argue that viable ECN designs which incorporate expandable limit orders to encourage block-trading must also provide escape valves, such as the periodic enforcement of strict price-time priority, to limit the negative externality imposed on those waiting to trade: attempts to use the expandable order protocol as the *only* market-clearing mechanism will invite competition from alternative trading venues.

Finally, the use of expandable limit orders in the brokered interdealer Treasury market has important implications for researchers who use historical quote and transaction data from the broker-dealer joint venture GovPX to study market microstructure impacts on Treasury pricing. As our results document, GovPX quoted depth data can be an incomplete measure of discoverable depth in this market.

The organization of the remainder of the paper is as follows. Section 1 describes the expandable limit order protocol and the dataset. Section 2 analyzes the depth discovery and price improvement characteristics of the market. Section 3 provides empirical tests of why traders use expandable order strategies, pre-trade information leakage, and the impact of trades of ex-post prices. Section 4 investigates differences in price discovery, price competition, and traders' usage of the workup process during stressful and quiet trading periods. Section 5 presents concluding remarks and the implications from our investigation for the use of expandable limit orders in ECN.

1. Description of the expandable limit order protocol and the dataset

1.1. The expandable order protocol

The expandable limit order protocol consists of the following steps:

Step 1: *Limit order* placement.

Step 2: *Responding order* placement.

Step 3: A negotiating period, during which the paired traders are offered right-of-refusal options to transact additional volume at the initial-volume price.

Limit orders are binding indications to buy (or sell) a particular quantity at a particular price. Placing a limit order also establishes the trader's priority for entering into the paired quantity negotiations of Step 3. Limit orders can be cancelled at any time until they have been matched with a responding order. Buy limit orders are called *bids*. Sell limit orders are *offers* or *asks*.

Responding orders are requests to take the opposite side of the transaction at the displayed limit-order price. A responding order is a *hit* if it is an order to sell at the price shown in the limit-order bid. A *take* is a responding order to buy at the limit-order offer price. Thus, responding-order traders "hit the bid" or "take the offer" of limit-order traders.

As an incentive for traders to provide liquidity through limit-order placements, only responding orders pay commissions. As a result, a limit order from Trader A to buy at price Y and a limit order from Trader B to sell at price Y are displayed as a "locked" market, with a bid–ask spread of zero.

The third step in the market-clearing protocol is the quantity negotiation period, or "workup process". During the workup process, the matched pair of limit-order and responding-order traders is given the right-of-refusal option to transact additional quantity at the just-agreed price, conditional on the other party's agreement to trade more. A trader who exercises the option to trade another increment continues to receive a right-of-refusal option to trade again at that price, until the counterparty interest is exhausted. As a limit- or responding-order trader completes his trading, the next limit- or responding-order trader, if one exists at that price, is brought into the workup process. If one counterparty completes his trading, the other counterparty continues to retain a right-of-refusal option as long as he is willing to keep trading at that price.

To inform other traders that volume negotiations are in progress, the trading system displays the initially matched quantity and each subsequent increment in flashing fields on trading screens. Each volume increment (including the initial quantity) is called a *workup*. When all trading interest is exhausted at the initially-agreed price on at least one side of trading, the completed transaction is posted.

Fig. 1 shows an example of the expandable limit order protocol. It illustrates that the posted transaction volume ($4 M face value) is the total accumulation of 3 individual workups at a price of $100\frac{1}{64}\%$ of face value. The sizes of the first, second, and third workups in the transaction are $1 M, $2 M, and $1 M, respectively. The transaction is complete and added to the aggregate trading volume for the day only after exhausting willingness to trade on at least one side of the market at the $100\frac{1}{64}$ price. Trader C, who submits his limit order later than Trader A, starts a queue behind Trader A. Because Trader A does not exhaust Trader D's demand for trade, Trader C is brought into the volume negotiations with Trader D. The transaction, therefore, represents pairwise negotiations of Trader A and Trader D as well as Trader C and Trader D.

		Limit order book displayed on trading screens: Best Bid (Price x Quantity)	Workup field	Best Ask (Price x Quantity)
Time 0:	No limit orders placed.	-		-
Time 1:	Trader A places a BUY limit order for a quantity of \$1 million face value at a price of $100\,{}^{1}/_{64}$% of face.	$100\,{}^{1}/_{64}$ x 1		-
Time 2:	Trader B places SELL limit order for \$1 M face value at $100\,{}^{3}/_{64}$.	$100\,{}^{1}/_{64}$ x 1		$100\,{}^{3}/_{64}$ x 1
Time 3:	Trader C places BUY limit order for \$1 M face value at $100\,{}^{1}/_{64}$, causing aggregate depth at best bid to increase by \$1 M face value.	$100\,{}^{1}/_{64}$ x 2		$100\,{}^{3}/_{64}$ x 1
Time 4:	Trader D places responding order to sell \$5 M face value at $100\,{}^{1}/_{64}$. Trader D is informed he has sold \$1 M face value at $100\,{}^{1}/_{64}$. Trader A is informed he has bought \$1 M face value at $100\,{}^{1}/_{64}$. Trader A is given the right-of-refusal to buy more at this price. *{Workup field flashes price and total transaction volume so far.}*	$100\,{}^{1}/_{64}$ x 2	*$100\,{}^{1}/_{64}$ x 1*	$100\,{}^{3}/_{64}$ x 1
Time 5:	Trader A indicates he wants to buy \$ 2 M more at $100\,{}^{1}/_{64}$. Trader D is informed he has sold another \$ 2 M. *{Workup field flashes price and total transaction volume so far.}*	$100\,{}^{1}/_{64}$ x 2	*$100\,{}^{1}/_{64}$ x 3*	$100\,{}^{3}/_{64}$ x 1
Time 6:	Trader A indicates he does not want to buy more at $100\,{}^{1}/_{64}$. Trader A's initial limit order is cleared from the limit order display. *{Workup field flashes price and total transaction volume so far.}*	$100\,{}^{1}/_{64}$ x 1	*$100\,{}^{1}/_{64}$ x 3*	$100\,{}^{3}/_{64}$ x 1
Time 7:	Trader D is informed he has sold \$1 M face value at $100\,{}^{1}/_{64}$. Trader C is informed he has bought \$1 M face value at $100\,{}^{1}/_{64}$. Trader C is given the right-of-refusal to buy more at this price. *{Workup field flashes price and total transaction volume so far.}*	$100\,{}^{1}/_{64}$ x 1	*$100\,{}^{1}/_{64}$ x 4*	$100\,{}^{3}/_{64}$ x 1
Time 8:	Trader C indicates he does not want to buy more at $100\,{}^{1}/_{64}$. Trader D learns there is no more buying interest at $100\,{}^{1}/_{64}$. The flashing workup field clears from the display. The transaction for \$4 M is added to the total volume on the day.	-		$100\,{}^{3}/_{64}$ x 1

Fig. 1. Example of the expandable limit order protocol for Treasury note trading.

As shown in Fig. 1, the size of the *first* workup (\$1 M) is not greater than Trader A's initial limit order. Even though Trader D's responding order (\$5 M) is for a larger amount, Trader D waits while Trader A considers exercising the option to trade more. Because the trading screen displays the aggregate depth of Trader A and Trader C's limit orders, the *initial* workup of \$1 M is for less than the displayed depth of \$2 M.

1.2. The dataset

We use GovPX transaction, quote, and workup data for 5-year U.S. Treasury notes for October 1–29, 1997, to obtain the unique dataset for this study. GovPX is a joint venture of primary dealers and interdealer brokers. It was formed in 1990 largely in response to Congressional pressure to increase transparency in the Treasury markets. GovPX captures quote and transaction data electronically from the brokered interdealer market and displays it on a real-time basis to corporations, institutional investors, and other subscribers.

At the time of the sample period, five GovPX-member brokers provided real-time feeds from their proprietary screens to the GovPX electronic system: Garban, Hilliard Farber, Intercapital (formerly RMJ), Liberty, and Tullett & Tokyo. One of the large interdealer brokers, Cantor Fitzgerald, was not involved in the joint venture or reporting. Using data from the Federal Reserve Bank of New York, we estimate GovPX-member firms brokered 71% of the brokered interdealer volume of Treasury coupon securities with maturity of 5 years or less. Cantor Fitzgerald brokered the remainder.[6,7]

At the time of the sample period, all GovPX-reporting brokers used the expandable limit order protocol. These brokers acted as agents for dealers, displaying their limit orders, matching responding orders, and orchestrating the workup process. None of the brokerage firms traded for its own account. Dealers were under no obligation to trade. Strict anonymity was maintained throughout the workup process and even after transactions were completed. Netting, title transfer, and payment were handled through the Government Securities Clearing Corporation.[8]

The salient features of the GovPX raw data for this study are:

- The data are time-stamped records of limit orders aggregated across the brokers and individual workups and transactions, signed *hit* or take consistent with the sign of the responding order in the expandable limit order protocol terminology in Section 1.1.

[6] After GovPX's startup, the price and quantity details of many customer–dealer transactions remained opaque throughout the 1990s. Many customers subscribed to information services that provided GovPX real-time quote and transaction data so they could use this information when they negotiated their own transactions with dealers. Therefore, we believe at the time of this study, much of the price discovery took place in the GovPX-reported market.

[7] In the year 2000, following several years of government surpluses, Treasury trading volume declined heavily. Interdealer brokerage firms consolidated (Garban and Intercapital merged; Tullett & Tokyo and Liberty merged) and several moved to electronic trading platforms to cut staffs. Some brokerage firms that reported voice-brokered activity to GovPX believed they were under no obligation to report activity from their ECNs. [See Fleming (2003) for reduced trading volumes and GovPX reporting percentages.] Fourteen primary dealers formed an electronic trading platform for interdealer trading (“BrokerTec”). Cantor Fitzgerald discontinued voice-brokered operations in the Treasury market as a result of the destruction of its office in the World Trade Center on September 11, 2001, and rebuilt its operations around its electronic trading platform (“eSpeed”).

[8] Fleming and Garbade (2002) provide an in-depth look at settlements in the brokered interdealer market.

- The data do not reveal the identities of the dealers or brokers.
- The data include only the highest-priced bid limit orders and lowest-priced ask limit orders of the 5 brokers. The bid (ask) limit order size is the sum across all broker inputs of bid (ask) limit orders at that price. The data are, thus, the inside spread, as displayed by GovPX. The data are coarser than dealers observed, however, as they could observe best bids and asks from each broker's screen.[9]

The uniqueness of our dataset is that we match workups with transactions.[10] Although GovPX does not indicate to which transaction a workup belongs, the nature of the expandable order protocol and the manner in which GovPX captures the raw data allow us to partition the raw data into clusters, greatly simplifying the job of matching workups with transactions. A sample of the raw data, clusters, and examples of how we match workups with transactions are shown in the appendix. As detailed in the appendix, 82.5% of clusters have only one possible way of matching workups with transactions. For the remainder, we adopted two approaches: one maximizes the number of workups associated with a given transaction and one minimizes the number. These approaches give rise to 2 datasets with almost indistinguishable results. The results presented here are those for the dataset generated from the approach that minimizes the number of workups associated with a given transaction and, therefore, biases against finding evidence of expanded orders.

Table 1 provides summary statistics for transactions, quotes, and workups for the three most recently auctioned 5-year issues as of October 1, 1997. Market participants refer to the most recently auctioned issue as the *on-the-run* issue. The rest of the issues are called *off-the-run* issues. During the sample period, only one 5-year issue, Cusip 9128273J9 ("J9"), was on-the-run. This issue was designated on-the-run from September 25–October 29, 1997. Prior to that, Cusip 9128273G5 ("G5") was the on-the-run issue from August 28–September 24, 1997. Cusip 9128273C4 ("C4") was on-the-run from July 24–August 27, 1997. Table 1 shows that, in comparison with the on-the-run issue, the two off-the-run issues were rarely traded.

The minimum transaction size brokers allow in 5-year notes is \$1 million face value. Table 1 shows that average transaction size is \$ 8.3 million for the on-the-run

[9] For example, we may observe a depth of \$5 million at the best bid of 99.5% of face value in the GovPX data. We cannot observe whether this is \$5 million at a single brokerage firm's screen or whether it is split across several or more brokerage firms. Nor can we observe any brokerage firm's best bid that is inferior to the 99.5% quote(s). In contrast, dealers can observe how the depth is split across brokers and the best bid and depth at each brokerage firm.

[10] Other studies that use GovPX raw data include Fleming (1997), Fleming and Remolona (1997), Elton and Green (1998), Fleming and Remolona (1999), Balduzzi et al. (2001), Green (2003), Fleming (2003, 2002), and Huang et al. (2002). None of these studies attempts to match workups and transactions. Boni and Leach (2002) analyze "clusters" of workups and transactions and, as a result, must rely on a noisier proxy for expandable order usage and cannot analyze pre-trade information leakage or ex-post price impacts.

Table 1
Summary statistics for Treasury note trading in the brokered interdealer market for October 1–29, 1997

This table reports summary statistics for transactions, quotes, and workups for the three most-recently auctioned 5-year issues as of October 1, 1997.

Cusip 9128273	J9	G5	C4
Trading status	On-the-run	Off-the-run	Off-the-run
Issue date	9/30/1997	9/02/1997	7/31/1997
Maturity date	9/30/2002	8/31/2002	7/31/2002
Interest rate (%)	$5\frac{7}{8}$	$6\frac{1}{4}$	6
Par amount issued by Treasury (million \$)	12,807	12,732	12,231
Average daily volume traded (million \$/day)	8,704	70	7
Average daily closing price (% of par)	99.756	101.116	100.079
Transactions per day	1,045	16	1
Transaction size per day (million \$)			
Mean	8.3	4.4	6.1
10%	1	1	1
Median	5	4	4
90%	19	10	20
Ratio takes / hits			
Million \$ transacted	1.07	1.05	2.17
Number of transactions	1.06	0.97	1.40
Quote spread			
% of time both bid and ask are present	89.0%	67.9%	25.9%
Time-weighted spread conditional on both bid and ask present (% of par)	0.01540	0.03097	0.02862
Quote revisions			
Average time between revisions (seconds)	17.9	42.3	156.3
Ratio revisions/number of transaction	4.6	127.3	456.7
Workup statistics			
Average daily number of workups	1,934.0	28.4	1.9
Median workup size	3	1	3
Average time (seconds) from 1st workup to transaction posting for trades of:			
1 workup	13.4	15.1	26.4
2 workups	19.7	19.0	24.5
More than 2 workups	29.2	20.6	29.3

issue and \$ 4.4 million and \$ 6.1 million for the other issues. Ninety percent of transactions are for \$ 20 million or less.

Both bid and ask limit orders are present for the on-the-run issue 89% of the 24-hour trading day. Many of the absences of a two-sided spread occur at the end of the New York day when trader activity shifts geographically from New York to Tokyo. Two-sided spreads are absent more of the time for the off-the-run issues. Conditional on the presence of both bid and ask, the time-weighted average spread for the on-the-run issue is narrower than for either of the off-the-run issues. To calculate the time-weighted average spread conditional on the existence of both bid and ask, we calculate the best ask less best bid in effect for every second both are present, sum these differences, and then divide by the number of seconds when both bid and ask are present, as in Blume and Goldstein (1997).[11]

On average, for the on-the-run issue, every 18 seconds there is a change in quote prices or sizes, causing an updated GovPX record. The off-the-run issues average less frequent quote revision but substantially higher ratio of quote revisions to number of transactions.

Table 1 shows there are an average of 2 workups for every transaction. For the on-the-run issue, the average time between first workup display and transaction volume posting is 13 seconds for single-workup transactions and increases for multiple-workup transactions.

Fig. 2 shows daily prices and trading volumes for the on-the-run issue. Consistent with the GovPX "trading day" designation, a day's closing price is the price of the last transaction recorded before 6:00 p.m. (New York time) that day (and after 6:00 p.m. the previous day). While the on-the-run issue (J9) traded every day, the most-recently-designated off-the-run issue (G5) did not trade on October 27. The second-most-recently designated off-the-run issue (C4) traded only on 10 of the 20 trading days. The largest single-day price changes occurred on:

- October 8, when Federal Reserve Chairman Alan Greenspan's remarks on employment suggested interest rate tightening,
- October 27, when the NYSE first halted trading under Rule 80B following a 550-point drop in the Dow Jones Industrial Average (DJIA), and
- October 28, as the DJIA rebounded 337 points from the previous day.

Clearly, the activity in the U.S. equity market on October 27 and 28, 1997, was extraordinary. Previous authors (e.g., Cohen and Shin, 2003; Furfine and Remolona, 2001) find that the effects of trade flows on price discovery in the U.S. Treasury market depend on amount of trading activity and price volatility. In Section 4, we examine whether dealers use different trading strategies on these extraordinary days.

Interestingly, despite the inclusion of October 27 and 28 in the dataset, almost sixty percent (59.9%) of all on-the-run transactions were at the same price as the previous transaction. Another 25.7% of transactions were posted with a price change that was the minimum tick size that brokers allowed: a positive or negative $\frac{1}{128}$ of one percent of face value.

2. Depth discovery and price improvement

2.1. Price improvement

Liquidity demanders receive "price improvement" when they transact *inside* the quoted spread. Blume and Goldstein (1992), Lee (1993), and Petersen and

[11] Locked market conditions (i.e., best bid equals best ask) are commonplace in this market. For these observations, the bid-ask spread equals zero.

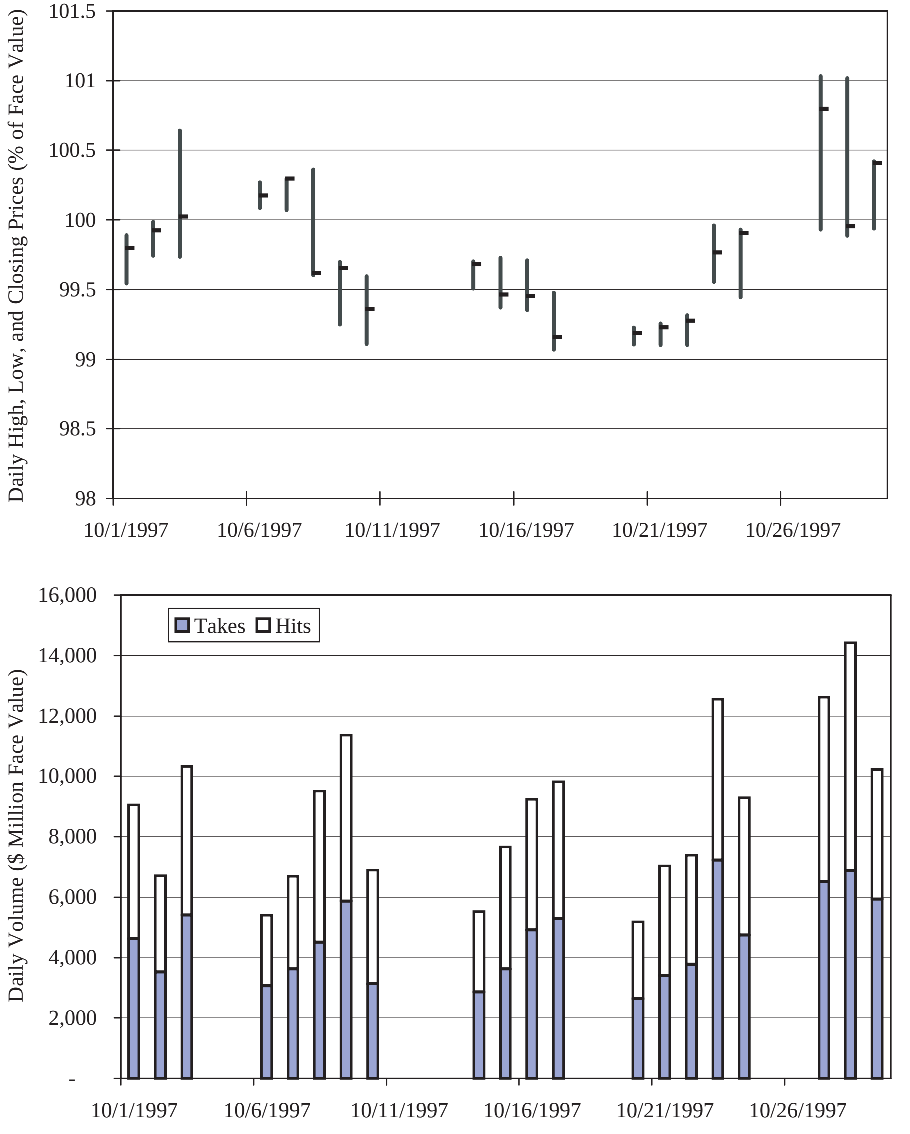

Fig. 2. Daily prices and trading volumes for the on-the-run five-year Treasury note for October 1–29, 1997.

Fialkowski (1994), among others, provide evidence that trades often receive price improvement for NYSE-listed securities. Handa et al. (1999) and Reiss and Werner (1996) find price improvement evidence for AMEX and the London Stock Exchange, respectively. Huang and Stoll (1996) compare price improvement for NYSE and Nasdaq stocks.

The GovPX-reported market is a decentralized market of five limit order books, each of which is operated by one of the GovPX-reporting broker firms. The liquidity demanders in this market are the responding orders. They do not expect price improvement as they expect to execute at the price of the displayed limit orders.

We define *negative* price improvement as execution at prices *outside* the best quotes aggregated across the five brokers.[12] Negative price improvement is possible if dealers place responding orders through brokers who do not have the best bid or offer relative to the other broker firms. There are two possible explanations for this behavior: (1) dealers may expect to receive preferential treatment from these brokers or (2) the expandable order protocol may bottleneck trading at the broker with the best bid or ask, causing impatient dealers to go elsewhere.

Three important elements of the brokered interdealer Treasury market work against the preferential treatment argument for explaining negative price improvement. They are: (1) competition among broker firms, (2) competition among employees at each broker firm, and (3) dealers' desires for anonymity. Dealers have the screens of all broker firms in their trading rooms. In a real sense (through the installation of all the screens), the dealers have created important execution options. As the quotes, workup increments, and completed transactions of each broker are immediately transparent, each dealer quickly knows whether he is getting best execution. Additionally, each broker has multiple employees, each looking out for the interests of his designated dealer. Competition among the broker firm's employees helps to assure that time priority at the broker firm is maintained. Consequently, the opportunity to provide special treatment (in return for order flow) is limited, and it seems unlikely that reputation for special treatment can survive in this market. Finally, one reason that dealers usually use brokers (instead of trading direct) is to preserve anonymity: dealers want to protect private order flow information about their inventory positions. It seems doubtful, therefore, that any dealer would use only a single broker firm as doing so would jeopardize anonymity. Although our data do not permit us to identify individual brokers or dealers, dealers have told us that the ability to use different brokers at different times is an important regularly exercised option.

The more likely explanation for negative price improvement is that the right-of-refusal feature causes impatient dealers who are not first in the queue to cancel orders at the broker firm with workups in progress and to submit orders at a competing broker firm. In addition, a dealer who has not yet placed an order can observe workups in progress and decide to trade. Instead of queuing behind other dealers at the broker with trading in progress, he might try to execute through another broker, even at worse prices. In transparent decentralized limit order markets where multiple orders can be worked at any given time, we would also expect to observe time- overlapping transactions. We hypothesize that incorporation

[12] Clearly, *negative* (*positive*) price improvement from the viewpoint of the responding order trader (i.e., the liquidity demander) is *positive* (*negative*) price improvement from viewpoint the limit-order trader (i.e., the liquidity provider).

of expandable limit orders creates greater opportunities for trade-matching bottlenecks that lead to negative price improvement.

Table 2 shows that we find only a small fraction of trades (0.6% of dollar volume) have positive price improvement. These observations may result if (1) a broker errs when inputting the limit order price and reconciles the error prior to inputting the workup price or (2) a broker simultaneously receives a responding order and a limit order bettering the best price displayed, with the workup concluded at the better price but the best bid or offer not updated. Interestingly, the transactions with positive price improvement are significantly larger. This suggests a third explanation for positive price improvement: brokers may sometimes facilitate transactions inside the quoted spread and these transactions are more likely to be large transactions.

Table 2 shows 36.7% of transactions time-overlap at least one other transaction. We define a transaction as time-overlapping if its initial workup occurs after the initial workup of another transaction but before the posting of that transaction's final volume. For example, suppose an initial workup is posted for Transaction #1 and several seconds later, but prior to the posting of Transaction #1's final volume, Transaction #2's initial workup is posted. Transaction #2 is the overlapping transaction.

Table 2 shows that 6.9% of transactions (6.4% by volume) have negative price improvement. Comparison of overlapping and non-overlapping transactions in Table 2 indicates that a higher fraction of the overlapping transactions execute with negative price improvement. Only 2.9% of non-overlapping transactions have negative price improvement versus 13.7% for overlapping transactions. These results are consistent with the hypothesis that the right-of-refusal option sometimes bottlenecks trading, causing impatient traders to execute through brokers who display worse-than-best quotes. Table 2 also shows negative price improvement is not simply due to responding order dealers who demand larger transactions. Transactions with negative price improvement are not significantly larger.

2.2. *Depth discovery*

In limit order book markets that allow neither expandable nor hidden limit orders, responding (or market) orders do not expect to realize depth in excess of the quotes. Instead, a large market order can quickly run through a series of limit orders at a variety of prices, creating a series of transactions at various prices. In contrast, in an expandable order protocol, "quoted depth" can understate realizable depth if limit-order dealers exercise their right-of-refusal options for additional volume. Depth may be *discovered* sequentially by responding order dealers during pairwise negotiations without its being quoted first to all dealers.

Table 3 shows that 38.8% of transactions (58.9% by volume) have more than one workup. Table 3 also shows that 34.7% of transactions (56.5% by volume) discover depth that exceeds that quoted via limit orders just prior to the trade.

Unfortunately, transaction volume in excess of quoted depth is a noisy proxy for expanded limit orders. Table 3 may underestimate the extent dealers expand orders because we observe only depth aggregated across brokers and aggregated across

Table 2
Price improvement

For the sample period October 1–29, 1997, this table reports price improvement for non-overlapping and overlapping transactions for 5-year Treasury note trading in the GovPX-reported brokered interdealer market. Transactions have *negative* price improvement if transaction price > best ask for take transactions and transaction price < best bid for hit transactions. Transactions have *no* price improvement if transaction price = best ask for take transactions and transaction price = best bid for hit transactions. Transactions have *positive* price improvement if transaction price < best ask for take transactions and transaction price > best bid for hit transactions. A transaction is *overlapping* if its initial workup occurs after the initial workup of another transaction but before the posting of that transaction's final volume. "∗" indicates we reject $H_0 : \mu_i = \mu_j$ at the 5% significance level. "=" indicates we cannot reject.

	Negative price improvement	No price improvement	Positive price improvement	All
Non-overlapping transactions				
Transaction volume, millions (face value)	\$ 2,749	\$ 109,970	\$ 700	\$ 113,419
Number of transactions	386	12,987	61	13,434
Average transaction size, millions (face value)	\$ 7.12	\$ 8.47	\$ 11.48	\$ 8.44
	*	*		
Overlapping transactions	=	*	*	*
Transaction volume, millions (face value)	\$ 8,504	\$ 53,398	\$ 303	\$ 62,205
Number of transactions	1,072	6,711	14	7,797
Average transaction size, millions (face value)	\$ 7.93	\$ 7.96	\$ 21.64	\$ 7.98
	=	*		
Non-overlapping + overlapping transactions				
Transaction volume, millions (face value)	\$ 11,253	\$ 163,368	\$ 1,003	\$ 175,624
Number of transactions	1,458	19,698	75	21,231
Average transaction size, millions (face value)	\$ 7.72	\$ 8.29	\$ 13.37	\$ 8.27
	*	*		

Table 3
Depth discovery

For the sample period October 1–29, 1997, this table reports transactions by whether or not they discover greater depth than that quoted immediately prior to the first workup of the transaction. Transactions are also broken out by whether they consist of more than a single workup.

	Number of transactions	Column percent	Dollar volume of transactions, \$M	Column percent
Single-workup transactions with:				
Size $\leqslant$ quoted depth	11,652	54.9	57,265	32.6
Size > quoted depth	1,328	6.3	14,991	8.5
Multiple-workup transactions with:				
Size $\leqslant$ quoted depth	2,216	10.4	19,076	10.9
Size > quoted depth:				
& initial workup > quoted depth	487	2.3	11,269	6.4
& initial workup $\leqslant$ quoted depth	*5,548*	*26.1*	*73,023*	*41.6*
All transactions	21,231	100.0	175,624	100.0

dealer limit orders. It may overstate expanded orders because *new* limit-order traders may arrive *during* the workup process. It is also worth noting that 8.6% of all trades are single-workup trades with volume exceeding quoted depth or multiple-workup trades with volume of the first workup exceeding quoted depth. Possible explanations are (1) brokers may accumulate negotiated quantity increments prior to posting the first workup, or (2) we may misidentify workups with transactions other than for the 82.5% of data clusters that have unique identification.

3. Why do traders use expandable limit orders?

Previous authors have identified two major risks associated with limit orders. The first is non-execution risk, which can result if price trends away from the limit order before it is filled. The second is the adverse selection risk that the limit order will be executed to the trader's later regret. In this section, we test whether trading activity is consistent with attempts to use expandable limit order strategies to reduce these risks. We then examine whether the option to use these strategies reduces the risk of non-execution by eliminating pre-trade information leakage. Finally, we examine ex-post prices in a direct test of whether the option to expand limit orders reduces non-execution and adverse selection costs.

3.1. Trading activity to reduce non-execution and adverse selection risks

There is always a potential for information free-riding when market participants interpret a large order as a signal of valuable private information and seek to free-ride on the signal.[13] In the Treasury market, large limit orders may signal private information about customer order flow and about future trades.[14] Thus, placing a large limit order may increase the risk of non-execution if it invites competition from other dealers. Dealers in an expandable order market can attempt to reduce information leakage about the extent of their willingness to trade by placing small initial limit orders and then working up volume in paired negotiations. Large transactions will then tend to consist of more than a single workup and discover greater than initially quoted depth. Once paired negotiations are underway, the dealer sequentially divulges the extent of his willingness to trade to dealers who have already expressed a willingness to trade. Therefore, the potential for free-riding is reduced during the workup process and dealers may be more willing to trade larger increments during paired negotiations than in an initial limit order.

In addition to the free-riding problem, in which a dealer must consider whether others will trade ahead of his order, a dealer must also consider the adverse selection or ‘‘staleness’’ problem, in which his limit order is executed to his regret.[15] In the Treasury market, dealers update their private information sets throughout the day as they talk and trade with customers and the Federal Reserve Bank of New York. As a result, limit orders can become stale. By placing a small initial limit order and working volume up in increments, a dealer reduces the value his initial order provides to those who may react more quickly than he can. Upon having his initial limit order executed, the dealer can update his information set and decide whether to trade again. Once paired in the workup process, a dealer commits his willingness to

[13] Burdett and O’Hara (1987) discuss information free-riding as a potential problem for block trading in the upstairs market. Fishman and Longstaff (1992) analyze the free-riding problem that traders face when they use brokers to trade in futures markets. Pagano and Roell (1993) describe the free-riding problem in the context of traders whose orders are front-run by their stockbrokers. Various authors have suggested that free-rider problems for block traders could be mitigated by using upstairs markets or by splitting orders across time, across brokers, or across exchanges.

[14] Madigan and Stehm (1994) report that dealers indicate an important reason for maintaining anonymity in the brokered Treasury market is to protect proprietary speculative positions and inventories. Based on discussions with primary dealers, Meulendyke (1998) reports dealers trade ‘‘to achieve inventories consistent with customer demands and with interest rate expectations’’. Meulendyke indicates dealers also pursue ‘‘arbitrage’’ transactions in the sense of Dow (1998), trying to ‘‘capitalize on the price differences between cash markets and futures and options markets.’’ Speculative positions may also take the form of attempts to corner or ‘‘squeeze’’ a particular Treasury issue (Department of the Treasury et al., 1992; Jordan and Jordan, 1996; Wall Street Journal, 2000). Duffie (1996) analyzes ‘‘repo specialness’’ that can result from dealer-orchestrated or inadvertent squeezes.

[15] Stoll (1992) provides examples of the stale limit order risk in the context of a ‘‘free-option’’ problem, where limit orders are ‘‘picked off’’ across exchanges [by floor traders using the Intermarket Trading System (ITS)] and across dealers [by Small Order Execution System (SOES) traders]. Stoll argues that continuous market monitoring alleviates the free-option problem but that market structures that allow market orders to run quickly through the limit order book, such as automated exchanges, will exacerbate the problem. Stoll further argues that, because a large limit order provides a large free option, traders wanting to transact large quantities (institutional traders) will avoid placing limit orders, possibly choosing to ‘‘hire someone to ‘work’ the order’’. O’Hara (1995) discusses the free-option characterization as applied to markets without designated market makers and reasons that designing electronic limit-order books may be difficult: ‘‘few traders submit large limit orders, preferring instead to work any intended trade via a series of small market orders’’.

trade for a short and finite time period, thus incurring less stale order risk than when he places the initial limit order. So, as is the case when dealers use expandable order strategies to reduce free-riding, dealers, concerned about stale limit order risk, may be more willing to trade larger increments during paired negotiations than in initial limit orders.

Formally, we consider the hypothesis:

Hypothesis. Traders use expandable order strategies to lessen the costs associated with information leakage and stale limit orders.

In summary, if dealers use expandable order strategies to reduce the costs of stale limit orders and information leakage:

- Large transactions will consist of multiple workups.
- Transactions will discover greater depth than that quoted in the initial limit order.
- The size of initial workups will be smaller than the average size of the rest of the workups in transactions.
- Expandable order strategies will be most valuable when costs of market monitoring are highest and information asymmetry is greatest.

There are several alternative hypotheses.

Alternative 1. Traders express the full extent of their willingness to trade in their initial limit orders. They do not expand limit orders. Multiple-workup transactions result from large responding orders requiring execution against the limit orders of multiple traders. Responding orders discover greater-than-quoted depth when new traders submit limit orders during the workup process.

Alternative 2. Traders express the full extent of their willingness to trade in their initial limit orders. Expanded limit orders result when execution of the initial limit order randomly coincides with an event that increases the limit order trader's demand for trade at that price.

The key distinction between our hypothesis and each of the alternatives is that in the alternatives, traders express in their initial limit order the full extent of their willingness to trade.

To distinguish among the hypotheses, we estimate two separate probit models. In the first specification, we model the probability that transactions discover greater than quoted depth by setting a dependent dummy variable to one if they do and zero otherwise. In the second specification, we model the probability that multiple-workup transactions have an initial workup that is smaller than the average size of the rest of the workups in the transaction. We set a dependent dummy variable to one if the transaction has a smaller first workup than the average of the rest of the workups in that transaction. The dummy is set to zero otherwise. The second model is estimated using only multiple-workup transactions.

The explanatory variables in each regression are:

SIZE	≡ Transaction size (in million dollars face value).
SPREAD	≡ Ask price less bid price just prior to the transaction (as % of face value).
DNIGHT	≡ Dummy set to 1 if trading is during New York business hours; 0 otherwise.
DOFF	≡ Dummy set to 1 if transaction is for an off-the-run security; 0 otherwise.
DMACRO	≡ Dummy set to 1 if transaction occurs following a scheduled macroeconomic announcement; 0 otherwise.
IND	≡ Dummy set to 1 if transaction is an overlapping trade; 0 otherwise.

For both probit specifications, a positive (negative) coefficient for an explanatory variable increases (decreases) the probability of observing a value of one for the dependent variable. Each model is estimated separately for hits and takes to allow for the possibility that traders adopt different strategies depending on whether they are buying or selling.

The probability that transactions discover greater than quoted depth should increase with transaction size regardless which of the hypotheses hold. We can, however, distinguish among hypotheses with the second model. If traders use expandable order strategies to reduce the costs of stale limit orders and information leakage, then the probability that the initial workup in multiple-workup transactions is smaller than the other workups should increase with transaction size. If traders express the full extent of their willingness to trade in their initial limit orders, however, there is no reason for this to be so.

Narrower spreads suggest less information asymmetry between traders. Limit order traders then face reduced adverse selection costs and should be more likely to negotiate additional volume beyond the initial order. If the hypothesis holds, the coefficient estimates for *SPREAD* should be negative for both models.

The variables *DNIGHT*, *DOFF*, and *DMACRO* are included as proxies for the three possible regimes of greater information asymmetry that have been identified by previous researchers for the brokered interdealer Treasury market. Madigan and Stehm (1994) and Fleming (1997) indicate that trading during the New York business day is structurally different from trading conducted outside the New York hours. There are fewer analysts, traders, and brokers working during non-New York hours. These structural differences present opportunities for some dealers to obtain and react more quickly to valuable information than others can. For example, Madigan and Stehm report that a higher fraction of trades is done directly between dealers and therefore not necessarily disseminated through the broker network.

Elton and Green (1998) show off-the-run issues are subject to pricing errors relative to on-the-run issues as the result of liquidity and tax-timing effects. Off-the-run issues trade infrequently in the brokered market, and customer-dealer trading is less transparent than in the brokered market. As a result, dealers who trade off-the-run issues more frequently may be able to update or react to new information obtained from the customer-dealer market more quickly than other dealers can.

Balduzzi et al. (2001) and Green (2003) conclude that there is information asymmetry up to fifteen minutes following macroeconomic announcements. They conclude some dealers are better informed by virtue of being better able to interpret the just-released information.

The variables *DNIGHT* and *DOFF* are also of interest as they also represent regimes when market monitoring is relatively more costly. Outside New York business hours, fewer traders cover securities. It is also more costly to monitor the market for off-the-run securities as each issue trades infrequently. As with greater information asymmetry, more costly market monitoring for overnight and off-the-run trading would drive the coefficient estimates for *DNIGHT* and *DOFF* to be positive if the hypothesis holds.

We add the dummy variable *IND* to control for the possibility that overlapping trades, which are the result of traders' placement of responding orders with other brokers while a trade is being worked up, may be driven by different trading strategies than are non-overlapping trades. For example, because negotiated workups are reported quickly, an overlapping trade can free-ride on the price and depth discovery of the trade it overlaps. Or because not all brokers necessarily have limit orders at the inside quote, an overlapping trade can result from an aggressive responding order at a worse price than the trade it overlaps.

Maximum likelihood estimates of coefficients, *p*-values, and slopes for the probit regressions are shown in Table 4. Slopes in Panel B are first derivatives of the fitted probability function, or marginal effects, calculated at the explanatory variable means (Greene, 1997, pp. 877–879).

Estimates for SIZE are positive as expected for each regression. The marginal effect of transaction size is an increase in probability of 2.3% per million-dollar face value for the first regression. The probability of discovering greater-than-quoted depth is 57% for a \$19 million hit trade (i.e., a 90th-percentile size trade in the on-the-run issue) versus 20% for a \$5 million hit trade (i.e., a median size on-the-run trade).[16] The marginal effect of transaction size is smaller in the second regression: less than 1% per million-dollar face value.

Estimates for *DNIGHT* and *DOFF* are also positive for each regression, consistent with traders' greater usage of expandable limit orders when information asymmetry is greater and market monitoring is more costly. The marginal effects of these variables are substantial. Trading outside New York business hours increases the

[16] Probabilities reported in this section and in the following section are calculated using the relevant fitted probit model and mean values for the other explanatory variables. For example, for the model in the first column of Table 4, panels A and B, the mean values for *SPREAD*, *DNIGHT*, *DOFF*, *DMACRO*, and *IND* are 0.00771, 0.0855, 0.0166, 0.0330, and 0.378, respectively. The intercept is -0.9946. Using notation consistent with Greene (1997), pp. 877–9, for $SIZE = 19$ million, $\beta'\mathbf{x} = 0.1689$. Thus, the probability of discovering greater-than-quoted depth [i.e., $\Phi(\beta'\mathbf{x})$] is 57%. For $SIZE = \$5$ million, $\beta'\mathbf{x} = -0.8447$ and the probability is 20%. Thus, the difference in probabilities is 37% (i.e., 57%–20%). Alternatively, we can use the slope from Panel B of 0.023 to calculate more simply the linear approximation of this difference: (0.023)(\$19million − \$5 million) = 32%.

Table 4
Determinants of expandable order usage

This table reports maximum likelihood regression coefficients, p-values, and slopes for two probit models. In the first model, the choice variable is a dummy variable that has the value of one if a transaction discovers greater than quoted depth and zero otherwise. In the second, the choice variable is a dummy variable that has the value of one if a multiple-workup transaction has an initial workup that is smaller than the average size of the rest of the workups in the transaction and zero otherwise. The second model is estimated using only multiple-workup transactions. In each regression, a positive regression coefficient indicates a positive relationship between the explanatory variable and the probability that the choice variable takes on the value of one. The explanatory variables in each regression are: *SIZE*, which is transaction size in million dollars face value; *SPREAD*, which is the ask price less the bid price just prior to the transaction as percent of face value; *DNIGHT*, which is a dummy variable which takes on the value of one if trading is during New York business hours and zero otherwise; *DOFF*, which is a dummy variable which takes on the value of one if the transaction is for an off-the-run security and zero otherwise; *DMACRO*, which is a dummy variable which takes on the value of one if the transaction occurs in the 15 min following a scheduled macroeconomic announcement and zero otherwise; and *IND*, which is a dummy variable which takes on the value of one if the transaction is an overlapping trade and zero otherwise. New York business hours are 7:30 a.m. to 6:00 p.m. New York time. The scheduled macroeconomic announcements are National Association of Purchasing Managers' Survey, Construction Spending, Leading Indicators, Factory Orders, Existing Home Sales, Consumer Confidence, Weekly Initial Jobless Claims, Consumer Price Index, Durable Goods Orders, Housing Starts, Producer Price Index, Retail Sales, Employment Report for September, and Trade Balance. Each model is estimated separately for hits and takes to allow for the possibility that traders adopt different strategies depending on whether they are buying or selling. Numbers in parentheses are p-values. The symbols (***), (**), and (*) indicate statistical significance at the 1, 5, and 10 percent level, respectively. χ^2 values are Wald statistics. Slopes (Panel B) are first derivatives of the fitted probability function evaluated at the explanatory variable means (Greene, 1997, pp. 877–879). Panel C reports coefficient estimates when the models are estimated using only on-the-run observations and *VOLATILITY* is included as an explanatory variable.

Choice variable:	Transaction discovers greater than quoted depth		Transaction has initial workup smaller than the average of the rest of the workups in the transaction	
	Hits	Takes	Hits	Takes
Panel A. Coefficient estimates				
Explanatory variable:				
SIZE	0.0724	0.0726	0.0110	0.0132
	(*0.0000*)***	(*0.0000*)***	(*0.0000*)***	(*0.0000*)***
SPREAD	−19.9492	−20.0406	−10.2007	−8.4505
	(*0.0000*)***	(*0.0000*)***	(*0.0005*)***	(*0.0029*)***
DNIGHT	0.7699	0.6958	0.7205	0.7144
	(*0.0000*)***	(*0.0000*)***	(*0.0000*)***	(*0.0000*)***
DOFF	0.5671	0.5087	0.4593	0.2744
	(*0.0000*)***	(*0.0000*)***	(*0.0039*)***	(*0.0863*)*
DMACRO	0.0793	0.1116	0.1098	0.0736
	(*0.4240*)	(*0.2257*)	(*0.3466*)	(*0.5614*)
IND	−0.3602	−0.3855	0.0538	0.0434
	(*0.0000*)***	(*0.0000*)***	(*0.2312*)	(*0.3327*)
χ^2 test $H_0 : \beta = 0$	1,156.5***	1,592.7***	141.6***	168.9***
Observations	9,873	10,519	3,765	4,079

Table 4 (*continued*)

Choice variable:	Transaction discovers greater than quoted depth		Transaction has initial workup smaller than the average of the rest of the workups in the transaction	
	Hits	Takes	Hits	Takes
Panel B. Slopes				
Explanatory variable:				
SIZE	0.023	0.023	0.004	0.005
SPREAD	−6.455	−6.378	−4.062	−3.370
DNIGHT	0.249	0.222	0.287	0.285
DOFF	0.184	0.161	0.183	0.109
DMACRO	0.026	0.036	0.044	0.0294
IND	−0.117	−0.123	0.021	0.0173
Panel C. Coefficient estimates when VOLATILITY is included, on-the-run only				
Explanatory variable:				
SIZE	0.0718	0.0725	0.0107	0.0130
	(*0.0000*)***	(*0.0000*)***	(*0.0000*)***	(*0.0000*)***
SPREAD	−20.1032	−20.4231	−11.0432	−10.4391
	(*0.0000*)***	(*0.0000*)***	(*0.0002*)***	(*0.0004*)***
DNIGHT	0.7655	0.7113	0.7152	0.7391
	(*0.0000*)***	(*0.0000*)***	(*0.0000*)***	(*0.0000*)***
DOFF	*N/A*	*N/A*	*N/A*	*N/A*
DMACRO	0.0722	0.0428	0.0873	−0.0365
	(*0.4699*)	(*0.6538*)	(*0.4626*)	(*0.7812*)
IND	−0.3593	−0.3900	0.0517	0.0384
	(*0.0000*)***	(*0.0000*)***	(*0.2510*)	(*0.3826*)
VOLATILITY	0.0009	0.0071	0.0030	0.0103
	(*0.7198*)	(*0.0008*)***	(*0.3094*)	(*0.0004*)***

probability of discovering greater than quoted depth by 25% for take transactions (specifically, the probability is 23% for New York hours versus 48% overnight) and by 28% for hit transactions (24% for New York hours versus 52% overnight). It increases the probability of the first workup's being smaller than the average of the rest of the workups in multiple-workup transactions by 27%. Trading off-the-run issues increases the probability of discovering greater than quoted depth by 18% and by 21% for take and hit transactions respectively. It increases the probability of the first workup's being smaller than the average of the rest of the workups in multiple-workup transactions by 11% and by 18% for take and hit transactions respectively. Estimates for *DMACRO* are not significant at conventional levels possibly because we have not attempted to distinguish between major and non-major announcements or to incorporate a measure of the "surprise" contained in the announcement relative to the pre-announcement consensus.

The coefficient estimates for the bid-ask spread are negative for all regressions. The marginal effect of a change in spread is small relative to the marginal effects for *DNIGHT* or *DOFF*, however. The minimum price increment (i.e., a "tick") in the 5-year notes is $\frac{1}{128}$th of one percent of face value. As shown in Table 1, the time-weighted average bid-ask spread for the on-the-run issue is about two ticks, or

0.0154% of face value. A one-tick narrowing of the bid-ask spread increases the probability of discovering greater than quoted depth about 5%. It increases the probability of the first workup's being smaller than the average of the rest of the workups in multiple-workup transactions about 3%.

The coefficient estimates for the overlapping transaction indicator are significant for the first model, indicating that overlapping transactions are less likely to discover greater than quoted depth. The coefficient estimates for the second model are not significant at conventional levels, however. We examine traders' decisions to use overlapping transactions in Section 4.

If the risks of non-execution and adverse selection increase with price volatility, the value of the option to expand limit orders should increase during volatile trading periods. Fig. 3 (top panel) shows on-the-run intraday price volatility, as characterized by the mean absolute value of the difference between the last bid-ask midpoint in the half hour of trade and the last bid-ask midpoint in the previous half hour, consistent with Fleming (1997). Consistent with Fleming as well, volatility increases and average bid-ask spreads decrease (middle panel) during New York trading hours. Fig. 3 (lower panel) also shows three proxies for the usage of expandable limit order strategies: (1) trades that realize greater depth than initially quoted, (2) trades that have an initial workup of minimum size, and (3) trades that consist of multiple workups. Casual observation suggests usage of expandable order strategies is more strongly driven by whether trading occurs in the overnight market or by spreads than by volatility.

To clarify this, we estimate the same probit model as above except that we add explanatory variable *VOLATILITY*. Because two-sided spreads are often lacking in the off-the-run issues, the model is estimated using the on-the-run observations only. Coefficient estimates for *VOLATILITY* (Table 4, Panel C) are positive (i.e., the probability of observing expanded limit orders increases with volatility) for purchases but not significant for sales. We examine the relationship between expandable limit orders and volatility further in Section 4.

In summary, the results of the probit analyses are consistent with dealers' using expandable limit order strategies to attempt to lessen the risks of non-execution and adverse selection. We next examine whether expandable order strategies are successful in eliminating pre-trade information leakage.

3.2. Pre-trade information leakage

In markets that allow expandable limit orders, traders may expect limit orders to be such a poor proxy for true depth that the potential for free-riding on the information contained in quoted depth may be eliminated. We define a limit order book as having asymmetric depth if it has greater quoted depth on one side of the book than on the other. If information leakage of the full extent of willingness to trade is eliminated pre-trade, asymmetric depth in the limit order book and the direction of the next trade (i.e., "trade sign") will be uncorrelated.

Testing whether trade direction and asymmetric depth are uncorrelated is more than just a test of the potential for information free-riding. It is also a test of whether

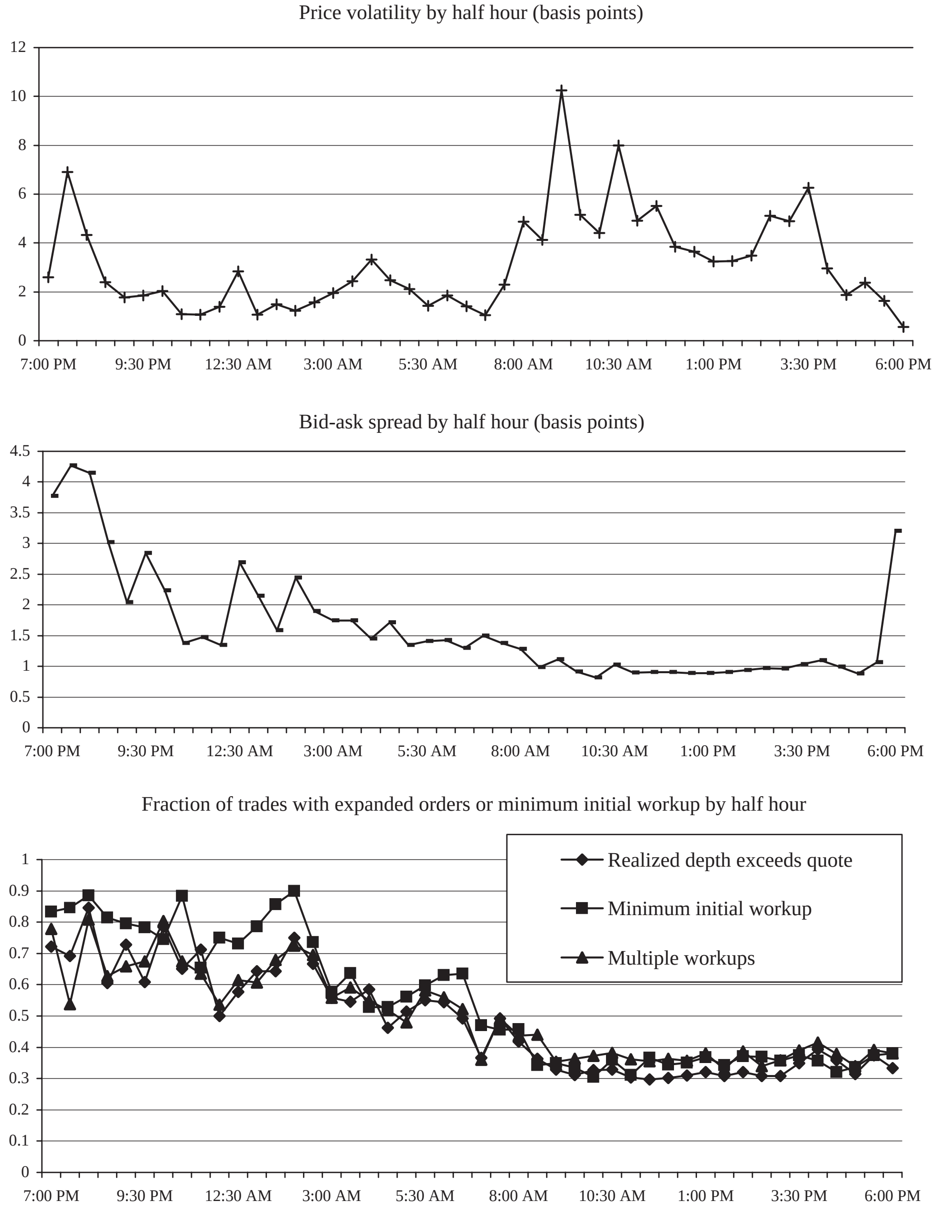

Fig. 3. Intraday trading statistics for the on-the-run five-year Treasury note for October 1–29, 1997.

thickness on one side of the limit order book "crowds out" the placement of additional limit orders on that side of the book, as suggested by Parlour (1998). Parlour models a transparent limit order book with small orders that are not expandable and with time priority strictly enforced as a rationing rule. Parlour shows that, even in markets without asymmetric information, systematic patterns in trade direction emerge. Traders know that new limit orders placed on the thicker side of the limit order book have a lower probability of execution. As a result, all else equal, traders are more likely to place market sell (buy) orders than limit sell (buy) orders when the book is thicker at the ask (bid). Therefore, in a transparent limit order book market without hidden or expandable limit orders, hit (take) transactions will be more likely than take (hit) transactions when the book is thicker at the ask (bid). In comparison, in an expandable limit order market, because true depth is unknown by any participant, asymmetric depth could be uncorrelated with future trade direction.

We test for a relationship between asymmetric depth and trade sign using a probit model. The dependent variable has a value of one if the transaction (trade_k) is a "take" transaction (i.e., results from a responding order that takes a limit order ask) and zero if the transaction is a "hit" (i.e., the responding order hits a limit order bid). The explanatory variables in the probit specification are:

RELDEPTH $\equiv$ bid size less ask size just prior to trade_k (in million dollars face value).
IND0SIGN $\equiv (1 - IND_k)(IBS_{k-1})$.
IND0DELP $\equiv (1 - IND_k)(Z_{k-1})$.
IND1SIGN $\equiv (IND_k)(IBS_{k-1})$.
IND1DELP $\equiv (IND_k)(Z_{k-1})$,
where
IND_k $\equiv$ dummy variable set to 1 if trade_k is an overlapping trade; 0 otherwise,
IBS_{k-1} $\equiv +1$ if trade_{k-1} is a *take* and -1 if trade_{k-1} is a *hit*, and
Z_{k-1} $\equiv$ the price of trade_{k-1} less the price of trade_{k-2} (as a percent of face value).

A positive (negative) coefficient for an explanatory variable increases (decreases) the probability of observing a take transaction.

If the potential for information free-riding exists or if asymmetric book thickness crowds out limit order placement on that side of the book, the probability of a take transaction will increase with *RELDEPTH*. We include IBS_{k-1} to control for the possibility that trade signs are serially correlated, Z_{k-1} to control for the possibility that prior trade price change influences trade sign, and IND_k to control for the possibility that overlapping and non-overlapping trades are influenced differently by the sign or price change of the previous trade. As discussed in the previous section, dealers appear to be more likely to use expandable order strategies outside New York business hours and for off-the-run issues. Therefore, we fit the model separately for these different trading regimes.

Table 5 provides the maximum likelihood estimates of coefficients, *p*-values, and slopes for the probit regressions. As in the previous section, slopes are first derivatives, or marginal effects, calculated at the explanatory variable means.

The coefficient estimate for the asymmetric depth variable (*RELDEPTH*) is positive and significant for each trading regime. The probability of a take transaction increases by 0.3% for each million-dollar face value that quoted bid depth exceeds quoted ask depth for on-the-run trading during the New York business day. This

Table 5
Determinants of trade sign

This table reports maximum likelihood regression coefficients, p-values, and slopes for a probit regression that has as the choice variable a dummy variable that has the value of one if the transaction (trade_k) results from a responding order that takes a limit order ask and zero if the responding order hits a limit order bid. A positive regression coefficient indicates a positive relationship between the explanatory variable and the probability that the choice variable takes on the value of one. The explanatory variables are: $RELDEPTH \equiv$ bid size less ask size just prior to trade_k (in million dollars face value), $IND0SIGN \equiv (1 - IND_k)(IBS_{k-1})$, $IND0DELP \equiv (1 - IND_k)(Z_{k-1})$, $IND1SIGN \equiv (IND_k)(IBS_{k-1})$, and $IND1DELP \equiv (IND_k)(Z_{k-1})$, where: $IND_k \equiv$ dummy variable set to 1 if trade_k is an overlapping trade; 0 otherwise, $IBS_{k-1} \equiv +1$ if trade_{k-1} is a *take* and -1 if trade_{k-1} is a *hit*, and $Z_{k-1} \equiv$ the price of trade_{k-1} less the price of trade_{k-2} (as a percent of face value). The model is fit separately for on-the-run transactions during and outside New York business hours. New York business hours are defined as 7:30 a.m. to 6:00 p.m. New York time. The model is also fit for off-the-run trades. Numbers in parentheses are p-values. The symbols (***), (**), and (*) indicate statistical significance at the 1, 5, and 10 percent level, respectively. χ^2 values are Wald statistics. Slopes are first derivatives of the fitted probability function evaluated at the explanatory variable means (Greene, 1997, pp. 877–879).

Trading regime	On-the-run New York hours	On-the-run Non-New York hours	Off-the-run trading
Panel A. Coefficient estimates			
Explanatory variable:			
RELDEPTH	0.0083	0.0192	0.0282
	(*0.0000*)***	(*0.0000*)***	(*0.0852*)*
IND0SIGN	0.3746	0.4312	0.2379
	(*0.0000*)***	(*0.0000*)***	(*0.0009*)***
IND1SIGN	0.9375	1.3220	
	(*0.0000*)***	(*0.0000*)***	
IND0DELP	8.5922	3.7950	3.5653
	(*0.0000*)***	(*0.2119*)	(*0.0082*)***
IND1DELP	27.4222	68.7084	
	(*0.0000*)***	(*0.0000*)***	
χ^2 test $H_0: \beta = 0.$	4,023.4***	308.0***	23.4***
Observations	18,174	1,891	323
Panel B. Slopes			
Explanatory variable:			
RELDEPTH	0.0033	0.0070	0.0112
IND0SIGN	0.1494	0.1564	0.0949
IND1SIGN	0.3738	0.4796	
IND0DELP	3.4258	1.3766	1.4223
IND1DELP	10.9336	24.9235	
Panel C. Bid size less ask size just prior to transactions (in million dollars face value)			
Mean	0.24	−0.50	−0.21
10%	−14	−9	−5
Median	0	0	0
90%	15	6	5

marginal effect is higher for on-the-run trading outside New York business hours and highest for off-the-run trading. Panel C of Table 5 shows a simple size distribution for the *RELDEPTH* observations. For *RELDEPTH* equal to the 90th-percentile point ($15 million), the probability of a take transaction increases 5% for on-the-run trading during New York business hours relative to a symmetric book. For 90th-percentile point *RELDEPTH* comparisons with the symmetric books for overnight ($6 million) and off-the-run ($5 million) transactions, the probability of take transactions increase 4% and 6% respectively.

Marginal effects are greater for previous trade sign. Coefficient estimates are positive and significant for each regime. Take transactions tend to follow take transactions, particularly when the trades are time-overlapping. As Biais et al. (1995) note in their study of the centralized limit order book of the Paris Bourse, this finding is consistent with order splitting, informed traders' repeated trading, uninformed traders' imitating the trading of informed traders, and similar reactions to new information. The exception is for off-the-run transactions, for which there are no overlapping trades.

For time-overlapping transactions, the probability of a take transaction increases when the previous trade had a price increase relative to its prior trade. For a price change of one tick ($\frac{1}{128}$th of one percent of face value), the probability of a take transaction increases 8% for on-the-run New York hours trading and 19% overnight. When trades are not overlapping, the marginal effect of price change is small, however (and not significant for on-the-run trading outside New York business hours).

The results of the probit analysis suggest that, despite traders' use of expandable limit order strategies to camouflage the *extent* of their willingness to trade, the potential for information free-riding and the "crowding out" of limit orders exists. The value of using asymmetric depth alone as a predictor is small, however. A better predictor for trade sign is the sign of the previous trade, particularly if that trade is still being worked up.

3.3. The value of the option to expand limit orders: an examination of ex-post prices

So far, we have shown that trading activity is consistent with the usage of expandable order strategies to reduce non-execution and adverse selection risks. The option to expand limit orders does not eliminate pre-trade information leakage, however. We now examine ex-post prices as a direct test of whether expandable order strategies are successful in reducing costs.

Consistent with Green (2003), we examine ex-post prices using the sequential return approach of Koski and Michaely (2000). As documented by Koski and Michaely for equity markets, using transaction prices to measure ex-post returns leads to biased inferences due to the bid-ask spread. Therefore, to examine the price impact of each purchase (sale), we calculate returns using the prevailing ask (bid) quotes for the sequence of trades prior to and following the trade of interest. For example, the transaction-to-transaction return for the purchase (sale) of interest, defined as the return for event 0, is the return from a purchase (sale) at the transaction price and an offsetting sale (purchase) at the next transaction's prevailing ask (bid) quote. As our data include some very volatile trading days, we adjust returns using a benchmark of the average of transaction-to-transaction returns for the 11 to 20 trades prior to the trade of interest. Data are for on-the-run trading during New York trading hours.

Like Green (2003), we sort trades by 33rd and 67th percentile size groups. Fig. 4 shows the cumulative returns, excess to the benchmark, for five trades prior to five trades after the trade of interest. Because we are interested in whether the option to expand limit orders reduces costs, we segregate single and multiple workup trades. Clearly, multiple workup trades are a noisy proxy for dealers' expanded orders because multiple dealers may have been brought sequentially into the workup process. It is, however, the only proxy observable by the other dealers as anonymity is preserved by brokers.

Consistent with Green (2003), as measured by ex-post returns, limit orders generally receive *negative* compensation for providing liquidity in this market. In other words, prices generally continue to trend against the limit order after the trade of interest. Green notes that possible explanations are: (1) heterogeneous valuations by dealers, (2) offsetting (presumably compensating) trades by dealers in the customer-dealer market, and (3) the fact that only responding orders pay interdealer broker fees. Furthermore, to put these excess returns in context, it is worth emphasizing that the minimum one-tick price change allowed by brokers in this market is one 128th of one percent of face value. Thus, the cumulative excess return across the entire trading sequence in each panel falls within just one to two ticks.

Consistent with Huang et al. (2002) and Green (2003), large trades do not have a greater price impact than small trades; medium trades show a slightly greater impact. Fig. 4 (middle and lower panels) shows that, conditional on trade size, however, dealers appear to reduce costs when they negotiate size through multiple workups. An alternative interpretation is that expanded orders *cause* stabilization of ex-post prices. Previous authors (e.g., Cohen and Shin, 2003; Green, 2003) suggest that traders may watch order flow to help forecast future prices and decide on the direction of their next trades. Market participants may interpret willingness of limit order traders to expand orders as an indication of a more balanced order flow. It is worth noting, however, that prices *before* multiple-workup medium and large trades are more stable than prior to single workup trades and that this accounts for at least some of the apparent ex-post stabilization.

The ex-post returns for single-workup small size trades (top panels of Fig. 4) are presumably indicative of the returns that dealers experience when they place minimum size initial limit orders and then decide not to expand their orders. They still experience negative compensation for liquidity, but because they decided not to expand their limit orders, the cost is on a smaller volume.

What remains unclear is why dealers *ever* use anything other than expandable order strategies. In other words, why do we ever observe initial workups greater than minimum size? We address this question in the next section.

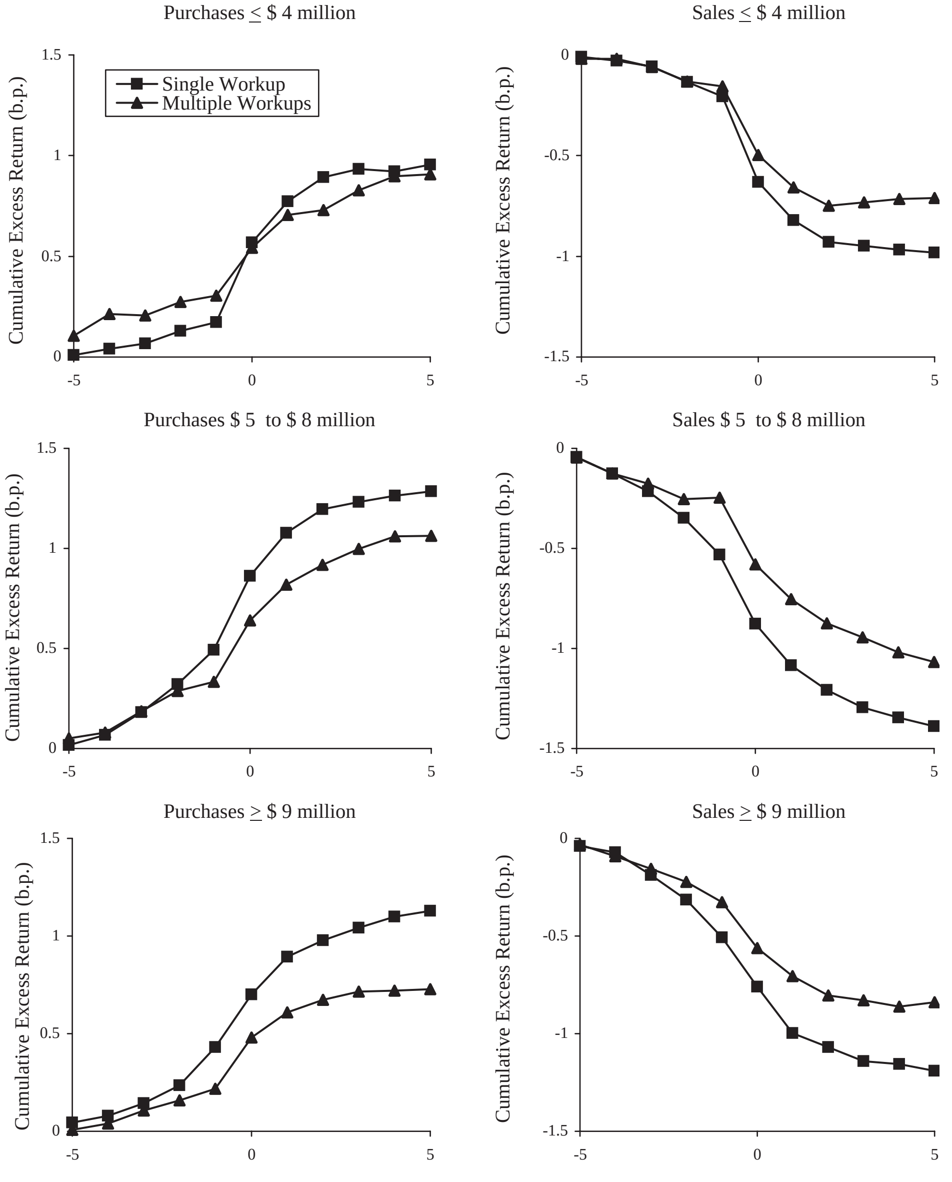

Fig. 4. The impacts of trades on ex-post prices.

4. Expandable limit order strategies in volatile markets

Several authors have suggested that price discovery in the interdealer broker Treasury market is structurally different when trading is extremely rapid and volatile. Furfine and Remolona (2001) find that during the Long Term Capital

Management (LTCM) crisis of autumn 1998, signed order flow had a stronger effect on ex-post prices and trades against shallower depth had a stronger impact on prices. Cohen and Shin (2003) find that the impact of signed order flow on prices is more strongly positive and the impact of lagged returns on signed order flow is also positive during hectic trading periods in 1999–2000. They note that this ‘‘positive feedback trading is consistent with the market adage that one should not try to ‘catch a falling knife’—that is, one should not trade against a strong trend in price.’’

4.1. Price discovery and the workup process on quiet and stressful days

Table 6 shows summary statistics for the 10% of trading days for which price range and volatility, bid-ask spreads, and volume were highest (‘‘stressful days’’) and lowest (‘‘quiet days’’) along with the full sample for the on-the-run issue during New York trading hours. The stressful days are October 27 and 28, 1997, when U.S. equity markets plummeted and then rebounded. Interestingly, although trading volume on these stressful days is more than twice that of quiet days, the first workup of trades is slightly more likely to be of minimum size and trades are slightly less likely to consist of multiple workups.

The impacts of trades on ex-post prices for the quiet- and stressful-day subsamples are plotted in Fig. 5 using the same methodology as in the previous section. Consistent with Furfine and Remolona (2001) and Cohen and Shin (2003), signed order flow generally has a stronger effect on ex-post prices during stressful days. In the previous section for the full sample, the use of expanded limit orders, conditional on trade size, lessened the impact of signed order flow on ex-post prices. Fig. 5 shows that this difference in impact between single and multiple workup trades disappears on quiet days, however. Additionally, with the exception of large purchases, the option to expand limit orders does not provide dealers with much of an advantage if they attempt to ‘‘catch the falling knife’’ during stressful days.

The results in Fig. 5 suggest several explanations for why we ever observe initial limit orders of greater than minimum size. Remembering that a one-tick price movement in this market is 0.78 basis points (b.p.), it becomes clear that on quiet days, regardless of size, trades have little impact on prices. This suggests that market participants are less likely to interpret greater quoted depth as a signal of informed trading and attempt to front-run the signal on quiet days. During these times, limit orders that display a greater willingness to trade (i.e., greater than minimum size initial orders) are less likely to invite free-riding price competition and therefore introduce less risk of non-execution. Similarly, consistent with observed narrower spreads and lower volatility, limit order traders likely face lower adverse selection risk on quiet days.

Fig. 5 also suggests an explanation for why we observe greater than minimum-size initial workups on volatile days. On stressful days, medium and large purchases (sales) follow price run-ups (declines). (The exception is for multiple-workup large purchases, and these trades have little impact on ex-post prices.) Presumably limit order traders have inferred that a greater degree of information asymmetry exists but

Table 6
Summary statistics for on-the-run five-year note Treasury notes and trading day subsamples

This table presents summary statistics for the on-the-run five-year Treasury note traded October 1–29, 1997, 7:00 a.m. to 5:00 p.m. The "quiet day" subsample and "stressful day" subsample are the 10% of trading days with the least and greatest price volatility, respectively.

	Full sample	Quiet day subsample	Stressful day subsample
Transactions per day	947	556	1,338
% buys	51.1	53.1	47.7
Time between trades (minutes)	0.62	1.07	0.45
Volume per trade (million $)			
Mean	8.34	8.58	8.47
Standard deviation	9.88	9.29	11.35
Transaction returns (basis points)			
Mean	−6.36E − 06	1.55E − 04	−5.04E − 05
Mean absolute value	4.69E − 03	2.52E − 03	7.12E − 03
Standard deviation	−9.73E − 03	5.00E − 03	1.22E − 02
Average daily price range (% of par)	0.450	0.158	1.115
Price volatility (basis points)			
Mean	4.54	1.64	10.68
Standard deviation	6.29	1.47	10.59
Bid-ask spread (basis points)	0.96	0.71	1.29
Fraction of trades that realize greater than quoted depth	0.33	0.33	0.33
Fraction of trades with initial workup of minimum size	0.35	0.31	0.38
Fraction of trades that have more than one workup	0.37	0.41	0.36
Correlation between trade_j and trade_{j-1}:			
Sign	0.45	0.50	0.43
Multiple-workup trades of same sign:			
Buys	0.11	0.10	0.11
Sells	0.10	0.13	0.09

have decided to buck the short-term trend. They have accepted the adverse selection risk and have little worry about non-execution.

4.2. Price competition and the workup process on quiet and stressful days

We further compare trading activity on quiet and stressful days to understand price competition during the workup process and traders' decisions to use time-overlapping trades. It is worth emphasizing that at the time of our sample, each

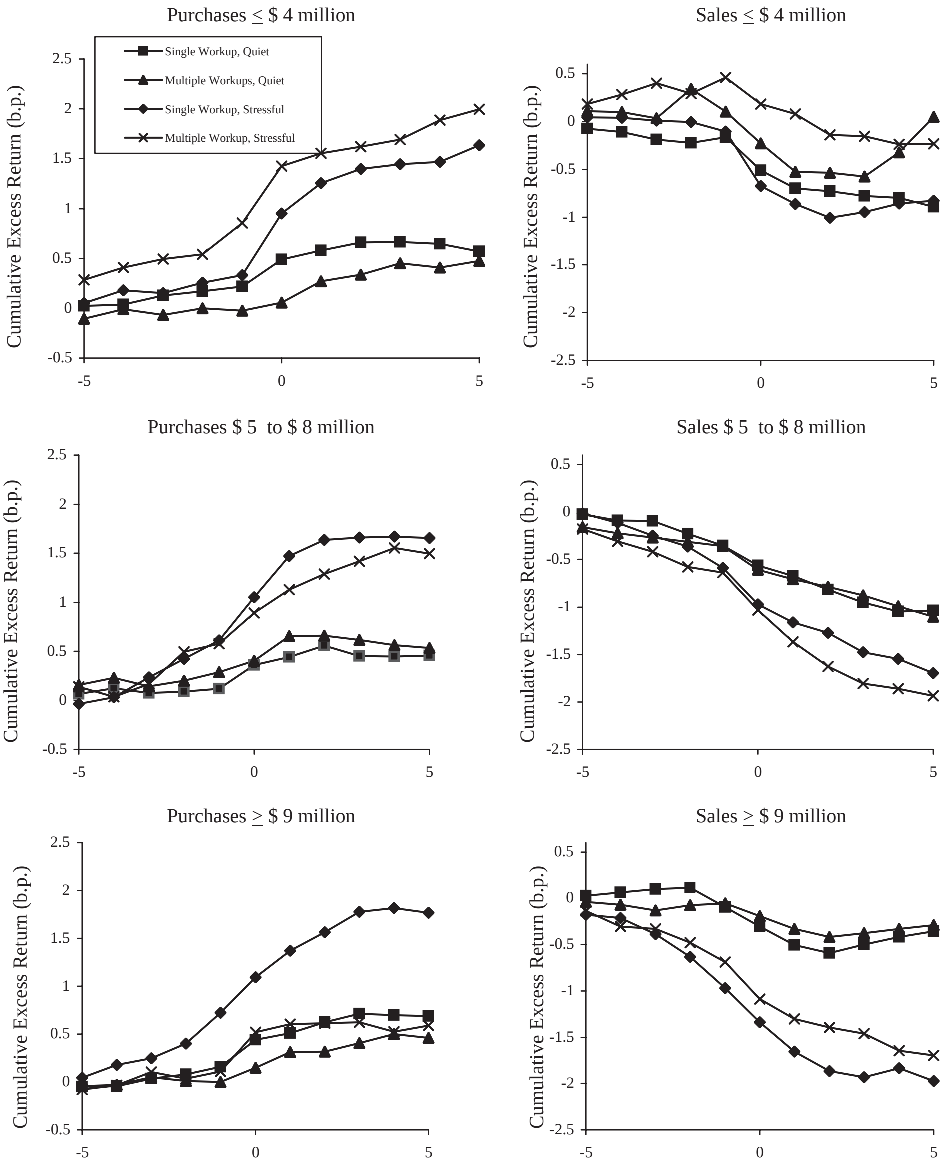

Fig. 5. The impacts of trades on ex-post prices for quite and stressful day subsamples.

brokerage firm orchestrated the workup process for only one pair of traders at any point in time. Therefore, a time-overlapping trade must be facilitated by a different brokerage firm than the trade it overlaps.

As in Section 2.1, the trade of interest (trade$_j$) is a time-overlapping trade only if its first workup is posted before the prior trade (trade$_{j-1}$) is completed. We further partition these overlapping trades into 6 categories by trade sign and price relative to the prior trade.[17] Table 7 (panel A) reports the breakdown of on-the-run New York hours trades by category for quiet and stressful days as well as for the full sample. As could be expected, the frequency of overlapping trades increases with the daily level of trading activity.

On quiet days, traders are more likely to experience locked spreads (Panel B) when they trade. And more of the overlapping trades that do occur take place at the same price and sign as the trade they overlap (Column 2).

In contrast, on stressful days, the workup process is more likely to invite price competition from aggressive responding orders (Column 3) or from aggressive limit orders (Column 6). Overlapping trades of opposite sign, possibly the result of dealers' canceling their limit orders and issuing responding orders, are also more common on stressful days. Perhaps not surprisingly, overlapping trades resulting from limit orders on the opposite side at a more aggressive price than the trade they overlap (Column 7) are extremely rare, regardless of the pace of trading activity.

As shown in Section 3, a market design that allows expandable limit orders offers benefits for traders. But as the price competition analysis suggests, it also introduces an important friction, first recognized in the literature by Garbade (1978). Expandable limit orders create queues for other traders who must wait to transact. As discussed in Section 2.1, time-overlapping transactions are expected in decentralized markets where multiple orders can be worked at any given time. Therefore, we cannot conclude that all time-overlapping transactions (almost 40% of trades) represent queue-jumping. Quotes, workups, and transactions *across all brokers* are extremely transparent in this market, however. As a result, at a minimum, we interpret time-overlapping transactions with negative price improvement (13.7% of overlapping trades) as some evidence of dealers' decisions to jump the queue. Additionally, the analysis of ex-post prices (Figs. 3 and 5) suggests that the costly option to leave the queue and trade through a competing broker via aggressive responding orders is sometimes in the money, validating the need for competing brokers (and the important execution options they present) in a market with queuing.

5. Concluding remarks and implications for ECN design

Despite the proliferation of ECNs in recent years, they largely coexist with, but do not replace, upstairs markets for the same securities. Barclay et al. (2003) document that large trades of Nasdaq stocks receive better execution costs on ECNs than via market makers. Large trades on ECNs are rare, however. They find that ECNs capture 30% of dollar volume in their sample overall but only 2% of large trades ($>10,000$ shares). They conclude that ECNs seldom offer sufficient depth to handle large trades. Weston (2001) finds that competition from ECNs on the Nasdaq has been beneficial for small trades but that "demanders of liquidity still rely heavily on

[17] Because we cannot observe quote revisions outside the inside quote as aggregated across brokers by GovPX, we focus on orders that result in trades, not all quote revisions, in this analysis.

Table 7

Price competition during the workup process

This table shows price competition during the workup process for 7:00 a.m. through 5:00 p.m. New York time for the on-the-run 5-year Treasury note for October 1–29, 1997. Each trade (trade_j) is partitioned into one of 7 trade categories (Columns 1–7). If trade_j begins before the workup process of the previous trade (trade_{j-1}) is completed, it is categorized as an overlapping trade. Non-overlapping trades are shown in column 1. Overlapping trades (columns 2–7) are further categorized by whether sign (buy or sell) of trade_j equals the sign of trade_{j-1} and its relative price. Panel A reports observations (as %) by trade categories and type of day, and each row percentage sums to 100%. Panel B reports observations (as %) by bid-ask spread category when trade_j begins: positive (ask price > bid price), locked (ask = bid), or negative (bid > ask). Each column sums to 100% for each type of day.

	Trade_j does not overlap trade_{j-1}	Trade_j overlaps trade_{j-1}					
		$\text{Sign}_j = \text{Sign}_{j-1}$ $P_j = P_{j-1}$	$\text{Sign}_j = \text{Sign}_{j-1}$ If J = buy, $P_j > P_{j-1}$. If J = sell, $P_j < P_{j-1}$.	Sign_j n.e. Sign_{j-1} $P_j = P_{j-1}$	Sign_j n.e. Sign_{j-1} If J = buy, $P_j > P_{j-1}$. If J = sell, $P_j < P_{j-1}$.	$\text{Sign}_j = \text{Sign}_{j-1}$ If J = buy, $P_j < P_{j-1}$. If J = sell, $P_j > P_{j-1}$.	Sign_j n.e. Sign_{j-1} If J = buy, $P_j < P_{j-1}$. If J = sell, $P_j > P_{j-1}$.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Observations by trade category and type of day (%)							
Quiet days	64.2	29.2	3.1	2.2	0.4	0.9	0.0
Full sample	60.4	26.6	5.1	3.8	1.7	1.8	0.6
Stressful days	58.7	23.6	7.1	3.5	3.3	3.1	0.8
Panel B. Observations by bid-ask spread category (%)							
Quiet days							
Positive	61.2	72.6	32.4	4.0	0.0	30.0	0.0
Locked	38.0	27.1	55.9	96.0	100.0	70.0	0.0
Negative	0.8	0.3	11.8	0.0	0.0	0.0	0.0
Full sample							
Positive	69.1	74.7	53.5	7.6	80.5	48.4	5.5
Locked	30.0	24.5	41.1	90.3	16.8	43.3	9.9
Negative	0.9	0.8	5.4	2.1	2.7	8.4	84.6
Stressful days							
Positive	79.0	83.0	60.3	6.4	83.0	56.1	4.8
Locked	19.4	16.1	32.8	89.4	15.9	36.6	9.5
Negative	1.6	1.0	6.9	4.3	1.1	7.3	85.7

traditional dealers for execution of large block trades.'' Conrad et al. (2003) examine institutional trading in U.S. equity markets and find that the average order size is small for ECNs relative to traditional broker executions.

Using data from the Paris Bourse, Bessembinder and Venkataraman (2003) find that while upstairs brokers can provide lower execution costs for large trades by distinguishing among informed and uninformed traders, upstairs markets also provide opportunities for large traders to ''tap into pools of 'hidden' or 'unexpressed' liquidity.'' They conclude that electronic exchanges may be better able to compete with upstairs markets in the future by providing ''a greater degree of strategic interaction among orders.'' We conclude that the incorporation of expandable limit orders is one way to provide this very flexibility.

The ECNs BrokerTec (recently acquired by ICAP) and eSpeed (formerly the Cantor Financial Futures Exchange) have incorporated an expandable order feature in their trading rules. The Commodity Futures Trading Commission (''CFTC'') reports in their 2001 ''Rule Enforcement Review of the Cantor Financial Futures Exchange'' that Cantor estimated 20% of trades were executed during a right-of-refusal period.

In the Treasury market, interdealer voice-brokers administered the right-of-refusal. The time to expiration for this option could vary depending on market conditions and broker-dealer relationship. Clearly, when used in ECNs, the right-of-refusal options have to have predetermined durations, but these could vary by security, by trading hours, and even by trader classification. The CFTC notes that the right-of-refusal was 6 seconds on the Cantor Financial Futures Exchange but that it was changed from time to time. BrokerTec limits the right-of-refusal to 5 seconds.

Given the inherent tendency of the right-of-refusal protocol to create queues for traders, ECNs that allow expandable limit orders have to provide ''relief valves'' to queuing. In the brokered interdealer Treasury market, the existence of competing broker firms provided such relief valves. One possible solution to the queuing problem in ECN design is suggested by BrokerTec and the Cantor Financial Futures Exchange: the right-of-refusal is extended only to the traders who place the *first* limit order and *first* responding order in a queue; after that, traders' orders received strict price-time priority.[18]

An alternative solution to the queuing problem could be to structure the ECN as a set of parallel limit order books in an analogy to the multiple-broker structure of the brokered interdealer Treasury market. In the primary limit order book, orders would be expandable. Traders could have the option of routing their orders to the parallel (secondary) limit order book, in which orders would *not* be expandable. In other words, the secondary limit order book would enforce price-time priority. Traders could have the option of routing orders to the secondary limit order book only when a paired negotiation in the primary limit order book was in progress.

[18] The CFTC report (2001) notes that the right-of-refusal was given back and forth repeatedly, however, to the traders who placed the initial limit and responding orders as long as both traders exercised it.

Appendix. Matching workups with transaction postings in GovPX data

This appendix describes how we match GovPX data for workups with transactions. Appendix Table A.1 shows a sample of the GovPX raw data for the on-the-run five-year Treasury note on October 14, 1997, from 7:20:46 a.m. to 7:33:41 a.m. Each broker firm provides its customers (dealers) with real-time screens of the quotes, workups, and transactions they are working. When a broker inputs data to this system, it is fed to the GovPX system, causing a new observation (i.e., a new line) to be generated in the GovPX dataset. The automated GovPX system updates each of the following variables, as appropriate, to reflect this new information.

- *RTC*: the number that GovPX assigns to the updated line of information. RTC monotonically increases with time.
- *TIMES:* the time the update occurs to the nearest second.
- *BIDPRICE:* the highest price at which any broker firm has a dealer bidding the note (% of face value).
- *BIDSIZE:* the depth aggregated across all broker firms who have bids at that price ($ million face value).
- *ASKPRICE*: the lowest price at which any broker firm has a dealer offering the note (% of face value).
- *ASKSIZE:* the depth aggregated across all broker firms who have offers at that price ($ million face value).
- *LTHORT:* "H" if the transaction most recently posted is a "hit" and "T" if the most recently posted transaction is a "take".
- *LTP:* the price of the most recently posted transaction (% of face value).
- *LTS:* the size of the most recently posted transaction ($ million face value).
- *CURHWP:* the price of the most recently posted "hit" workup (% of face value).
- *CURHWS:* the size of the most recently posted "hit" workup ($ million face value).
- *CURTWP:* the price of the most recently posted "take" workup (% of face value).
- *CURTWS:* the price of the most recently posted "take" workup
- ($ million face value).
- *AGVOL:* the running daily aggregate transaction volume ($ million face value).

We assign the variables *TRANSNO* (to identify individual transactions) and *HITNO* and *TAKNO* (indicating the match for each workup with a transaction sale or purchase) using the logic described below.

Whenever a line's *AGVOL* value exceeds the prior line's *AGVOL* value, the line indicates a transaction observation. Specifically, a broker has finished with workup negotiations and is posting the sum of the workups in the form of a completed

Table A.1
Matching workup observations with transaction observations

RTC	TIMES	BIDPRICE	BIDSIZE	ASKPRICE	ASKSIZE	LTHORT	LTP	LTS	TRANSNO	CURHWP	CURHWS	HITNO	CURTWP	CURTWS	TAKNO	AGVOL	
495	7:20:46	99.5313	1	99.5391	6	T	99.5313	15	.	0.0000	0	.	0.0000	0	.	443	
496	7:21:04	99.5313	1	99.5391	5	T	99.5313	15	.	0.0000	0	.	0.0000	0	.	443	
497	7:21:31	99.5313	1	99.5391	5	T	99.5313	15	.	0.0000	0	.	99.5391	5	36	443	
499	7:21:45	99.5313	1	99.5391	5	T	99.5313	15	.	0.0000	0	.	99.5391	2	36	443	*Cluster A*
500	7:21:47	99.5313	1	99.5391	5	T	99.5313	15	.	0.0000	0	.	99.5391	7	36	443	
501	7:22:06	99.5313	1	99.5469	1	T	99.5391	14	36	0.0000	0	.	0.0000	0	.	457	
502	7:23:10	99.5313	1	99.5469	26	T	99.5391	14	.	0.0000	0	.	0.0000	0	.	457	
503	7:23:57	99.5313	1	99.5391	1	T	99.5391	14	.	0.0000	0	.	0.0000	0	.	457	
504	7:27:06	99.5313	1	99.5391	11	T	99.5391	14	.	0.0000	0	.	0.0000	0	.	457	
505	7:27:07	99.5313	1	99.5391	35	T	99.5391	14	.	0.0000	0	.	0.0000	0	.	457	
506	7:27:36	99.5313	1	99.5391	25	T	99.5391	14	.	0.0000	0	.	99.5391	6	37	457	
507	7:27:40	99.5313	1	99.5391	10	T	99.5391	14	.	0.0000	0	.	0.0000	0	.	457	*Cluster B*
508	7:27:41	99.5313	1	99.5391	29	T	99.5391	14	.	0.0000	0	.	0.0000	0	.	457	
509	7:27:41	99.5313	1	99.5391	29	T	99.5391	6	37	0.0000	0	.	0.0000	0	.	463	
510	7:30:29	99.5313	1	99.5391	10	T	99.5391	6	.	0.0000	0	.	0.0000	0	.	463	
511	7:30:31	99.5313	1	99.5313	6	T	99.5391	6	.	0.0000	0	.	0.0000	0	.	463	
512	7:30:47	99.5313	1	99.5313	16	T	99.5391	6	.	0.0000	0	.	0.0000	0	.	463	
513	7:31:17	99.5313	1	99.5313	16	T	99.5391	6	.	99.5313	1	21	0.0000	0	.	463	
514	7:31:19	99.5313	1	99.5313	6	T	99.5391	6	.	99.5313	1	.	0.0000	0	.	463	
515	7:31:23	99.5313	1	99.5313	6	T	99.5391	6	.	99.5313	9	21	0.0000	0	.	463	*Cluster C*
516	7:31:39	99.5313	1	99.5313	6	T	99.5391	6	.	0.0000	0	.	0.0000	0	.	463	
517	7:31:40	99.5313	1	99.5313	6	H	99.5313	10	21	0.0000	0	.	0.0000	0	.	473	
518	7:31:47	99.5313	1	99.5313	6	H	99.5313	10	.	0.0000	0	.	99.5313	6	38	473	
519	7:31:55	99.5313	2	99.5313	6	H	99.5313	10	.	0.0000	0	.	99.5313	6	.	473	*Cluster D*
520	7:31:55	99.5313	2	99.5547	1	T	99.5313	6	38	0.0000	0	.	0.0000	0	.	479	
521	7·32·09	99.5469	1	99.5547	1	T	99.5313	6	.	0.0000	0	.	0.0000	0 ..	.	479	
522	7:32:15	99.5469	1	99.5547	1	T	99.5313	6	.	99.5469	1	22	0.0000	0	.	479	*Cluster E*
523	7:32:30	99.5313	1	99.5547	1	H	99.5469	1	22	0.0000	0	.	0.0000	0	.	480	
524	7:33:41	99.5313	1	99.5625	2	H	99.5469	1	.	0.0000	0	.	0.0000	0	.	480	

transaction. Therefore, the transaction information represents that of a single broker. GovPX updates *LTHORT* to indicate the transaction's sign (i.e., whether the completed transaction is a hit or take). GovPX updates *LTP* and *LTS* to indicate the completed transaction's price and size. Five of the lines shown in Appendix Table A.1 are transaction observations: *RTC* #501, *RTC* #509, *RTC* #517, *RTC* #520, and *RTC* #523.

For a line in the dataset to represent a hit (take) "workup observation", it must have nonzero values for *CURHWP* (*CURTWP*) and *CURHWS* (*CURTWS*). As with transaction observations, workup observations represent new information from a single broker. Workup observations replace, not increment, information from prior data lines.

Not all data lines with nonzero *CURHWP*, *CURHWS*, *CURTWP*, and *CURTWS* values are workup observations. *CURHWP*, *CURHWS*, *CURTWP*, and *CURTWS* fields are not zeroed out by GovPX until the transaction which the workup is part of is posted. As a result, if *CURHWP* and *CURHWS* are nonzero and identical to the prior line's values, it is possible that the line is not an indication that a broker is posting another *hit* workup, identical to the last. Instead, the line may indicate that a broker is posting a transaction, updating bid depth or price or ask depth or price, or posting a *take* workup. If *CURHWP* and *CURHWS* are nonzero and identical to the prior line's values but any other variable's value is different from that of the prior line, then we do *not* treat the line as a hit workup observation. For example, in Appendix Table A.1, RTC #513 is a hit workup observation: *CURHWP* and *CURHWS* are nonzero and the prior line is not a hit workup observation. For the next line, *RTC* #514, *CURHWP* and *CURHWS* are nonzero but identical to *RTC* #513's values. *ASKSIZE* is different: $6 million face value for *RTC* #514 versus $16 million face value for *RTC* #513. Therefore, *RTC* #514 is an observation of a broker posting an updated ask depth. It is not a hit workup observation.

Eight of the lines in Appendix Table A.1 are workup observations. Three of these are hit workup observations: *RTC* #513, *RTC* #515, and *RTC* #522. The take workup observations are *RTC* #497, *RTC* #499, *RTC* #500, *RTC* #506, and *RTC* #518.

After identifying transaction and workup observations, we parse the data into *clusters*. The first line in a cluster is the first workup observation following a transaction observation that zeroes out *CURHWP* and *CURHWS* (or *CURTWP* and *CURTWS*) values. The last observation in the cluster is the transaction observation that zeroes out *CURHWP* and *CURHWS* (or *CURTWP* and *CURTWS*) values. The data shown in Appendix Table A.1 have five clusters, labeled "A" through "E".

The following are necessary but not sufficient conditions for workup observations to be matched with a transaction:

- The workup observation(s) must be in the same cluster as the transaction observation.
- The sign ("hit" or "take") of the workup observation(s) must be the same as the sign of the transaction observation.

- The price of the workup observation(s) must be the same as the price of the transaction observation.
- If the workup observation is the only workup matched to the transaction observation, then its size must equal the size of the transaction. If multiple workup observations are matched with a transaction, then the sum of these workup observations' size values must equal the size of the transaction.
- The workup observations must occur prior to the transaction observation: *RTC* and *TIMES* of the workup observation(s) must be less than *RTC* and *TIMES* for the transaction observation.
- A workup observation can be matched with only one transaction.

Given these conditions, each of the workups in Appendix Table A.1 is uniquely matched with a transaction. Matching can become problematic, however, when clusters are overlapping. If overlapping clusters contain transactions at different prices or signs, workups are uniquely matched with transactions. But as shown in Appendix Table A.2, the raw data contain a single cluster with two transaction observations (flagged with *TRANSNO* set to 1) and four workup observations (flagged with *TAKNO* set to 1). All have the same price and sign. Thus, for this cluster, there are 4 possible ways of matching workups with transactions.

For the on-the-run five-year note for October 1–29, 1997, we find that for 12,714 clusters (82.5% of the total) workups and transactions are uniquely matched. Seventeen percent of all clusters (or 2,611 clusters) present more than one way to match workups with transactions.[19] To select a single match for these clusters, we apply the following logic.

Step 1: From the set of all possible matches for the cluster, create the subset of matches that match the *minimum number* of workups in the cluster with transactions in the cluster. If this set contains only one match, use this match for the dataset. Otherwise, go on to Step 2.

Step 2: From the Step 1 subset of matches, create the subset of the matches that have the earliest-posted workups. If this set contains only one match, use this match for the dataset. Otherwise, go on to Step 3.

Step 3: From the Step 2 subset of matches, create the subset of the matches that have the earliest-posted transactions. If this set contains only one match, use this match for the dataset. Otherwise, go on to Step 4.

Step 4: From the Step 3 subset of matches, select the match that matches the earliest- posted workups with the earliest-posted transactions.

The logic for steps 1 through 4 biases the resulting dataset *against* finding evidence of multiple-workup transactions. It is also important to note that we assume brokers post *each* negotiated volume increment as an individual workup and that the volume of the first workup posted is the first quantity agreed between dealers. Any tendency of a broker to accumulate negotiated volume increments and

[19] For 78 clusters (0.5% of the total), the cluster workups do not meet the necessary conditions to be matched with transactions and these observations are omitted from the final dataset.

Table A.2
Example of overlapping trades with same price and sign

RTC	TIMES	BIDPRICE	BIDSIZE	ASKPRICE	ASKSIZE	LTHORT	LTP	LTS	TRANSNO	CURHWP	CURHWS	HITNO	CURTWP	CURTWS	TAKNO	AGVOL
3871	14:23:53	99.6016	1	99.6094	1	T	99.6016	3	.	0	0	.	99.6094	1	1	4520
3872	14:23:55	99.5938	4	99.6094	1	T	99.6016	3	.	0	0	.	99.6094	1	.	4520
3873	14:24:05	99.5938	4	99.6094	1	T	99.6016	3	.	0	0	.	99.6094	1	1	4520
3874	14:24:08	99.5938	4	99.6094	1	T	99.6016	3	.	0	0	.	99.6094	1	1	4520
3875	14:24:09	99.5938	4	99.6094	1	T	99.6016	3	.	0	0	.	99.6094	1	1	4520
3876	14:24:10	99.6016	5	99.6094	1	T	99.6016	3	.	0	0	.	99.6094	1	.	4520
3877	14:24:16	99.6016	6	99.6094	1	T	99.6016	3	.	0	0	.	99.6094	1	.	4520
3878	14:24:19	99.6016	6	99.6094	1	T	99.6094	3	1	0	0	.	99.6094	1	.	4523
3879	14:24:19	99.6016	6	99.6250	15	T	99.6094	1	1	0	0	.	0.0000	0	.	4524

Note. Data are for the on-the-run five-year note for October 14, 1997, 2:23:53 p.m. to 2:24:19 p.m.

post the accumulation as a single workup biases against finding evidence of expanded orders.

To check the sensitivity of our findings to our matching algorithm, we repeat the analysis for the dataset that results from *maximizing* the number of workups matched with transactions in Step 1. This approach biases *in favor of* finding evidence of expanded orders. Comparing the results from both datasets, we found them almost indistinguishable. In fact, choice of dataset does not affect our conclusions in any way.

References

Balduzzi, P., Elton, E.J., Green, T.C., 2001. Economic news and bond prices: evidence from the U.S. Treasury market. Journal of Financial and Quantitative Analysis 36, 523–544.

Barclay, M.J., Hendershott, T., McCormick, D.T., 2003. Competition among trading venues: information and trading on electronic communication networks. Journal of Finance, forthcoming.

Bessembinder, H., Venkataraman, K., 2003. Does an electronic stock exchange need an upstairs market? Journal of Financial Economics, forthcoming.

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

Blume, M.E., Goldstein, M.A., 1992. Displayed and effective spreads by market. Unpublished paper, University of Pennsylvania.

Blume, M.E., Goldstein, M.A., 1997. Quotes, order flow, and price discovery. Journal of Finance 52, 221–244.

Boni, L., Leach, J.C., 2002. Supply contraction and trading protocol: an examination of recent changes in the U.S. Treasury market. Journal of Money, Credit, and Banking 34, 740–762.

Burdett, K., O'Hara, M., 1987. Building blocks: an introduction to block trading. Journal of Banking and Finance 11, 193–212.

Cohen, B.H., Shin, H.S., 2003. Positive feedback trading under stress: evidence from the U.S. Treasury securities market. BIS Working Papers No. 122.

Commodity Futures Trading Commission, 2001. Rule enforcement review of the Cantor Financial Futures Exchange. www.cftc.gov/files/tm/tmcxrer031901.pdf, March 28.

Conrad, J., Johnson, K.M., Wahal, S., 2003. Institutional trading and alternative trading systems. Journal of Financial Economics 70, 99–134.

Department of the Treasury, U.S. Securities and Exchange Commission, and Board of Governors of the Federal Reserve, 1992. Joint report on the government securities market.

Dow, J., 1998. Arbitrage, hedging, and financial innovation. Review of Financial Studies 11, 739–755.

Duffie, D., 1996. Special repo rates. Journal of Finance 51, 493–526.

Elton, E.J., Green, T.C., 1998. Tax and liquidity effects in pricing government bonds. Journal of Finance 53, 1533–1562.

Fishman, M.J., Longstaff, F.A., 1992. Dual trading in futures markets. Journal of Finance 47, 643–671.

Fleming, M.J., 1997. The round-the-clock market for U.S. Treasury securities. Federal Reserve Bank of New York Economic Policy Review 3, 9–32.

Fleming, M.J., 2002. Are larger Treasury issues more liquid? Evidence from bill reopenings. Journal of Money, Credit, and Banking 34, 707–735.

Fleming, M.J., 2003. Measuring Treasury market liquidity. Federal Reserve Bank of New York Economic Policy Review 9, 83–108.

Fleming, M.J., Garbade, K.D., 2002. When the back office moved to the front burner: settlement fails in the Treasury market after 9/11. Federal Reserve Bank of New York Economic Policy Review 8, 35–57.

Fleming, M.J., Remolona, E.M., 1997. What moves the bond market? Federal Reserve Bank of New York Economic Policy Review 3, 31–50.

Fleming, M.J., Remolona, E.M., 1999. Price formation and liquidity in the U.S. Treasury market: the response to public information. Journal of Finance 54, 1901–1915.
Furfine, C.H., Remolona, E.M., 2001. Price discovery in a market under stress: the U.S. Treasury market in fall 1998. Unpublished paper, Bank for International Settlements.
Garbade, K.D., 1978. The effect of interdealer brokerage on the transactional characteristics of dealer markets. Journal of Business 51, 477–498.
Green, T.C., 2003. Economic news and the impact of trading on bond prices. Journal of Finance, forthcoming.
Greene, W.H., 1997. Econometric Analysis. Prentice-Hall, Upper Saddle River, New Jersey.
Handa, P., Schwartz, R.A., Tiwari, A., 1999. Price improvement and price discovery on a primary market. Journal of Portfolio Management 25, 55–64.
Huang, R.D., Stoll, H.R., 1996. Dealer versus auction markets: a paired comparison of execution costs on NASDAQ and the NYSE. Journal of Financial Economics 41, 313–357.
Huang, R.D., Cai, J., Wang, X., 2002. Information-based trading in the Treasury note interdealer broker market. Journal of Financial Intermediation 11, 269–296.
Jordan, B.D., Jordan, S.D., 1996. Salomon Brothers and the May 1991 Treasury auction: analysis of a market corner. Journal of Banking and Finance 20, 25–40.
Koski, J.L., Michaely, R., 2000. Prices, liquidity, and the information content of trades. Review of Financial Studies 13, 659–696.
Lee, C.M.C., 1993. Market integration and price execution for NYSE-listed securities. Journal of Finance 48, 1009–1038.
Madigan, B., Stehm, J., 1994. An overview of the secondary market for U.S. Treasury securities in London and Tokyo. Federal Reserve Board Finance and Economics Discussion Paper, 94-17.
McInish, T.H., Wood, R.A., 1995. Hidden limit orders on the NYSE. Journal of Portfolio Management 21, 19–26.
Meulendyke, A., 1998. U.S. monetary policy and financial markets. Federal Reserve Bank of New York, NY.
O'Hara, M., 1995. Market microstructure theory. Blackwell Publishers, Cambridge, MA.
Pagano, M., Roell, A., 1993. Front running and stock market liquidity. In: Conti, V., Hamaui, R. (Eds.), Financial Markets' Liberalisation and the Role of Banks. Cambridge University Press, Cambridge.
Panchapagesan, V., 2000. What if time priority is not enforced among traders? Unpublished paper, Washington University.
Parlour, C.A., 1998. Price dynamics in limit order markets. Review of Financial Studies 11, 789–816.
Petersen, M.A., Fialkowski, D., 1994. Posted versus effective spreads. Journal of Financial Economics 35, 269–292.
Reiss, P.C., Werner, I.M., 1996. Transaction costs in dealer markets. In: Lo, A. (Ed.), The industrial organization and regulation of the securities industry. University of Chicago Press, Chicago.
Sofianos, G., Werner, I.M., 1997. The trades of NYSE floor brokers. New York Stock Exchange Working Paper # 97-04.
Stigum, M., 1990. The Money Market. Irwin, New York, NY.
Stoll, H.R., 1992. Principles of trading market structure. Journal of Financial Services Research 6, 75–107.
Umlauf, S.R., 1991. Information asymmetries and security market design: an empirical study of the secondary market for U.S. government securities. Journal of Finance 46, 929–953.
U.S. General Accounting Office, 1990. How broker trading systems operate. In: U.S. government securities: more transaction information and investor protection are needed, GAO/GGD-90-114, Appendix III, 97–100.
Weston, J.P., 2001. Competition on the Nasdaq and the growth of electronic communication networks. Unpublished paper, Rice University.
Wall Street Journal, 2000. SEC is probing if two firms tried 'squeeze' in Treasury bills. February 25, 2000.