Electronic versus open outcry markets: The case of the bund futures contract

*Francis Breedon**

and

*Allison Holland***

*Bank of England, Threadneedle Street, EC2R 8AH and Lehman Brothers, 1 Broadgate, London, EC2M 7HA.
**Bank of England, Threadneedle Street, EC2R 8AH.

Any views expressed are those of the authors and not necessarily those of the Bank of England or of Lehman Brothers. We would like to thank Jo Corkish, Rachel Charlton, Matt Davies, Michelle Hart, Ravi Pindoria, E Tim Sahlin, and John Tutt for their help. We would like to thank Ralf Herklotz, Jim Moser, and Anne Vila for their comments. Any errors and omissions are of course our own.

Issued by the Bank of England, London, EC2R 8AH to which requests for individual copies should be addressed: envelopes should be marked for the attention of the Publications Group. (Telephone: 0171-601 4030).

ISSN 1368-5562

Contents

Abstract

The Bund (ten-year German government bond) futures contract is the most actively traded bond contract in Europe; it is traded in both London (LIFFE) and Frankfurt (DTB) on open outcry and electronic trading platforms respectively. This paper evaluates the relative liquidity and price discovery roles of these markets and seeks to reconcile the conflicting results of earlier studies. We find these conflicting results are largely a product of the price data used. Using both transactions prices and quotes data, we find that variable transaction costs and the contribution to price formation of each market is similar. The main differences between the two markets are the larger trade size on the open outcry market (compensating perhaps for the higher running cost of an open outcry operation) and a tendency for trading to move toward the open outcry market during volatile periods.

1 Introduction

The German Bund futures contract (based on ten-year federal government and Treuhand bonds) is the largest bond futures contract in Europe. It is also unusual in that there are in fact two separate contracts trading in different markets (one on LIFFE in London and one on the DTB in Frankfurt)[1] which have an almost identical specification. This feature of the contract makes it a useful vehicle for analysing various market microstructure hypotheses and so it is unsurprising that there is a rapidly growing literature in this area. What is surprising is the dramatically contradictory results found in all other studies so far, such that no two papers agree on any aspect of the relative performance of these two markets (see Table A). By using an extensive database for the June 1995 contract, this paper re-visits the issues addressed by these other papers in an attempt to discover the reasons for the contradictions. By analysing both quoted prices and transactions prices we find that many of the contradictions can be explained by the differences in these two data sources. This not only resolves the puzzling contradictions in the study of the Bund contract, it also serves as a cautionary tale for the analysis of other markets.

Table A
Previous results on the Bund contract

	Data and frequency	Spreads: *Which market has wider bid-ask spreads*	Price formation: *Do price movements in one market lead to movements in the other*	Resilience: *Which market operates most successfully in volatile periods*
Shyy and Lee (1995)	Quotes 1 minute	DTB>LIFFE	DTB LIFFE	n/a
Pirrong (1996)	Transactions prices 15 minutes	LIFFE>DTB	n/a	DTB>LIFFE
Franke and Hess (1995)	Transactions prices Daily	n/a	n/a	LIFFE>DTB
Kofman and Moser (1996)	Transactions prices + LIFFE quotes trade-by-trade/ 1 minute	LIFFE=DTB	LIFFE⇔DTB	n/a

Note: n/a= not applicable.

(1) Another example of dual trading of a major financial futures contract is the Nikkei contract that trades in SIMEX and Osaka. See Fremault Vila and Sandmann (1995) for an analysis of those markets.

Overall, we find that many of the differences found in these papers can be explained by differences between transactions and quote data and that neither in isolation is sufficient to give a true picture of the interaction of these two markets.

2 The Bund contract

In both Frankfurt and London the Bund futures contract is specified as delivery of DM250,000 of Bund with a 6% notional coupon within the maturity range 8_ to 10 years on the tenth day (or first business day after) of the delivery month (either March, June, September, or December). In practice this means that there is a basket of Bunds that can be delivered (though the amount delivered is adjusted by a price factor to create equivalence) which are in the right maturity range. Within this basket there is one Bund that is most likely to be delivered - the cheapest to deliver (CTD) - since the short has the option of which Bund to deliver. As noted above, the basket of deliverables is the same in both contracts. Table B outlines the main differences between the two markets.

Table B
A comparison of LIFFE and DTB Bund contracts

	DTB	LIFFE
Trading differences		
Trading mechanism[a]	Electronic	Open outcry
Trading hours[b]	8.00 am - 5.30 pm	7.30 am - 4.15 pm
Turnover[c]	1,410,500 contracts (28%)	3,637,500 contracts (72%)
Contract differences[d]		
Last trading day	4 trading days before delivery	5 trading days before delivery
Price factors	different treatment of weekends	
Penalty for late delivery	Lombard rate + 1% per day	Potentially unlimited

Notes:
(a) LIFFE does, however, offer an electronic trading facility out of normal hours.
(b) London time for LIFFE, Frankfurt time for DTB (Germany is usually one hour ahead but there are a few weeks in spring and autumn when one country moves to or from daylight saving before the other which means that there is no time difference).
(c) January 1996.
(d) All described in more detail in Breedon (1996).

Data

Our dataset consists of a full transactions record (from the Exchanges) and minute-by-minute quotes (collected from a Reuters screen) for both the LIFFE and the DTB over the period 10 April and 2 June 1995 for the June 1995 contract. In our analysis we restrict the dataset to the period when both markets were open (for our purposes, after-hours trading in LIFFE is excluded) and exclude inter-day price changes. The transactions record gives a complete record of prices and times of trade for both markets but the trades are not signed in either market (ie we cannot tell directly whether a trade is a buy or a sell). Also, reliable intra-day data on the size of each transaction are not available for LIFFE and so we do not use this size information in our analysis. However, we do have reliable data on the number and volume of trades at a daily frequency for both markets.

Table C
Summary statistics

	LIFFE	**DTB**
Average number of contracts traded per day	125,806	44,148
Average number of trades per day	2,338	2,203
(per minute)	(4.5)	(4.2)
Average trade size	53 contracts	20 contracts

Trading on both exchanges has almost doubled relative to the period in 1992 examined by Kofman and Moser (1996). Interestingly, Kofman and Moser (1996) found the average trade size to be roughly equal on both exchanges, at about 23 contracts. In the period we examine, while the average size of trade on the DTB has remained approximately the same, the average size of trade on LIFFE has more than doubled. This suggests that LIFFE's capacity to execute large trades has increased and that large trades are more likely to be executed on LIFFE, while there may be more splitting of trades on the DTB.

3 Transactions prices versus quoted prices

Let us first examine how closely transactions prices are related to quoted prices on the two exchanges. Table D shows some summary statistics for the transactions prices series and the mid-quote series. Statistically the two series are very similar. As we would expect, all series have a unit root present. We examine the data for any lead/lag relationship between the mid-quote series and the transaction price series. Over the whole sample period, there is strong evidence of bi-directional Granger causality between our mid-quote series and our transaction price series, leading us to conclude that both are determined simultaneously.

Table D
Comparison of quotes and transaction prices

	LIFFE		DTB	
	mid-quotes	transaction prices	mid-quotes	transaction prices
Average	93.66	93.66	93.63	93.63
Standard deviation	1.0706	1.0709	1.0651	1.0650
Skewness	0.3458	0.3462	0.3635	0.3638
Kurtosis	1.8984	1.9003	1.9321	1.9326
Correlation between the two series	0.9999	-	0.9999	-

Both series are distributed similarly on the two exchanges and are perfectly correlated with each other. Given that the transaction price series and the mid quote are virtually identical, one might expect any results on liquidity and information transmission to be independent of the data source. Surprisingly, we will show that such results are in fact very sensitive to the type of data used.

4 Measuring bid/ask spreads

Economic theory would predict that if market participants were free to move between markets, then all order flow would eventually move to the ‘cheaper’ market; ie the market with the lowest execution costs. However, it is not clear which of the DTB and LIFFE is the cheapest. The total cost of conducting business on both exchanges is made up of a number of factors. First, the cost of running the business on each exchange is different. A floor operation requires a great deal of staff and infrastructure to maintain whilst an electronic market is comparatively simple, once the initial investment has been undertaken. However, the initial technology costs of the latter may be substantial. Second, there is a greater range of interest rate futures products traded on LIFFE than on the DTB, so there may be economies of scale. The scale of these will depend on the number of products you wish to trade. Third, both exchanges calculate initial and variation margin along similar lines; however, the London Clearing House remunerates any surplus margin in members’ accounts thereby reducing the cost of transferring funds between institutions on a daily basis.(2) The discussion that follows looks at only one aspect of costs, the spread, which is a variable trading cost and gives an indication of relative liquidity in the two markets.

The spread compensates the market-maker for a number of things: (i) the cost of processing the order, (ii) the cost of holding inventory and (iii) the potential loss s/he would make if they dealt with an informed trader (compensation for 'adverse selection'). George, Kaul and Nimalendran (1991) estimate that the order-processing component is the largest part of the spread, accounting for around 90% of the total, whilst the remainder is made up of the adverse selection component. They estimate that inventory holding costs are insignificant. In the case of interest rate futures, it is generally held that the adverse selection component is small as there are unlikely to be any investors with superior or private information. Also, these markets are highly liquid suggesting that inventory costs will be insignificant. One important aspect of the competition between electronic and open outcry markets is the prediction that order-processing costs on electronic markets will be lower. This prediction follows from the observation that an open outcry trading pit is more costly to run relative to an electronic operation. Clearly if this cost differential is significant, this should be passed on to investors in the form of a higher order-processing component of the spread.

Table A gives a somewhat depressing picture of the ability of empirical research to estimate spreads. Of the three studies that have estimated relative spreads at DTB and LIFFE we have all three possible orderings of these two markets. However, closer analysis indicates that the real distinction here may be due to the type of data used. Shyy and Lee (1995) use quotes data only and Pirrong (1996) uses transactions data only. Kofman and Moser (1996) on the other hand use a combination of the two (though they use only LIFFE quotes data). It is also worth noting that Pirrong (1996) finds that he can only reject equality of spreads in the two markets in four out of twelve cases. As a result, it seems that the differences in results may be due to differences in quote and transactions-based estimates of spreads. Using our dataset of both quoted prices and transactions prices, we can assess (a) the extent to which our data replicates the difference between quote and transaction-based spreads found in other papers and (b) which type of data give the best estimates of true variable transactions costs. Though transactions data should in principle give a better measure of variable trading costs, spread estimators based on such data are subject to a number of

(2) Members can hold a surplus balance on their LCH accounts to meet variation margin calls as they arise, thereby requiring less frequent fund transfers between institutions.

potential biases (see, for example, Smith and Whaley (1994)[(3)]). Quote data, on the other hand, gives a simple and accurate measure of the spread as long as transactions are actually undertaken at the quoted prices. Given that quotes on both LIFFE and DTB are firm (ie must be honoured over the period for which they hold[(4)]) one might expect quote data to give a better estimate of variable trading costs in this case.

To assess these questions we estimated a variety of measures of spreads: (i) The simple Roll (1984) measure based on transactions data, which exploits the negative autocovariance induced in transactions prices by the bid-ask 'bounce', (ii) the Thompson-Waller (1988) estimator, also based purely on transactions data, (iii) effective spreads estimated using a combination of transactions data and quote data as proposed by Huang and Stoll (1996) and simple quote-based spreads.[(5)] Ideally we would like to adjust for differences in order size but, given data constraints, we cannot do so. We also estimated explicit measures of the order-processing component of the spread: the method proposed by George, *et al* (1991), henceforth GKN, and the realised spread measure proposed by Huang and Stoll (1996).

The Roll estimator is defined as

$$S_R = 2\sqrt{-(\mathrm{cov}(\Delta P_t, \Delta P_{t-1})} \qquad \textbf{(1)}$$

where ΔP_t = change in log prices (over a minute in our sample).

This measure relies on the covariance of price changes being due purely to order-processing costs (ie pure bid-ask bounce), so inventory risk and adverse selection risk are negligible. However, as GKN demonstrate this measure may be biased if expected returns are time-varying and positively autocorrelated.

---

(3) Traditional measures assume transaction prices are negatively serially correlated but trade splitting will violate this assumption. Also, they assume that the volatility of the 'true' price is zero and that all observed volatility in prices is due to the bid/ask bounce, thus over-estimating the spread.
(4) In the case of LIFFE, the quote is valid for as long as 'the breath is warm', whilst the DTB quotes are firm for as long as they appear on the trading screen. Following on from that, the time required to revise quotes on LIFFE is likely to be shorter than that on the DTB. For our sample period, the average time between revisions in the quotes recorded was 17 seconds on LIFFE and 22 seconds on the DTB. Note also that quotes on LIFFE are recorded by pit observers and may not cover the entire sample of quotes available, so some revisions to quotes may not actually be recorded, leading to an overstatement of the average time between revisions.
(5) For a fuller discussion of measures of spreads see the Appendix.

The Thompson-Waller estimator is defined as

$$S_{TW} = \frac{2}{T}\sum_{t=1}^{T}\left|\Delta P_t\right| \qquad \textbf{(2)}$$

This measure relies on the assumption that both the expected underlying price change[(6)] and the variance of the underlying price change are zero. If this assumption does not hold, it will tend to be an upwardly biased measure of the true spread.

The effective spread combines quote and transactions data to give a cleaner estimate of the spreads. It is defined as

$$S_E = 2 * \left|P_t - \left(A_t + B_t\right)/2\right| \qquad \textbf{(3)}$$

A_t = Quoted ask price at time t.
B_t = Quoted bid price at time t.

It is similar to the Thompson-Waller measure except that it uses the quote midpoint $((A_t+B_t)/2)$ rather than P_{t-1} as a reference point. As a result it is less subject to bias since quoted prices are likely to incorporate changes in the underlying price that occur over time. However, there may be problems associated with the fact that we cannot exactly match the time at which transactions are undertaken with the time a quote was recorded, given that our quotes are recorded once a minute whilst transactions can be recorded at any time within that minute. In practice, we use the most recently recorded quote before any given transaction (ie the quote data is slightly lagged by an average of about 30 seconds). As this measure incorporates more information than those measures based purely on transactions costs, we believe it to give a better approximation of the ‘true’ spread. This view is supported by our simulation results reported in the appendix.

(6) By underlying price change we mean the true value excluding changes due to bid-ask bounce.

The quoted spread is

$$S_Q = \left(A_t - B_t\right) \qquad \textbf{(4)}$$

In this case bid and ask are recorded at precisely the same time. Table E1 details our findings.

Table E1
Estimated spreads for DTB and LIFFE

	Spread measure (percentage)			
	Roll	Thompson-Waller	Effective	Quoted
LIFFE	0.00636	0.01722	0.01472	0.01044
DTB	0.00478	0.01492	0.01406	0.01322
Percentage difference (LIFFE/DTB)	+33%	+20%	+5%	-21%
T-test for difference	n/a	n/a	5.7*	-53.2*

Notes: A * indicates that the spreads were significantly different at the 5% significance level.
n/a = not applicable.

Comparing the transactions-based estimates with the effective spread in Table E1 shows our predicted ordering. The Roll estimates are the lowest, whilst the Thompson-Waller estimate is the highest with the effective spread lying between the two. These results seem to show the distinction found in earlier studies with the TW measure indicating wider spreads on LIFFE, the effective spread showing almost equal spreads and the quote measure showing narrower spreads on LIFFE. Given that quotes on both the DTB and LIFFE are firm, it is perhaps surprising that the effective spread for LIFFE is significantly wider than the quoted spread at LIFFE; this may however simply reflect the fact that the quotes and transactions prices are non-synchronous.

We now examine explicitly the order-processing component of the spread. The GKN approach is related to the effective spread in that it uses data on quotes as an estimate of price changes due to changes in the true value of the security. They create a returns series (*RD*) that is simply the returns (log ΔP) measured by transactions data minus the returns measured using bid prices subsequent to the trades. Since the bid prices are subsequent to trades, they should reflect any information contained in those trade so that *RD* should contain only order-processing effects. Using GKN's notation,

$$RD = \pi S_q \left(Q_t - Q_{t-1}\right)^{(7)} \quad \textbf{(5)}$$

As a result, even without signed trade data, it is possible to derive a measure of the pure order-processing cost spread using a covariance estimator of the Roll (1984) type:

$$S_{GKN} = 2 * \sqrt{-\operatorname{cov}(RD_t, RD_{t-1})} = \pi S_q \quad \textbf{(6)}$$

Although the GKN measure has the advantage of not requiring signed trades, its use of quote data means that differences in the relationship between quotes and transactions in the two markets may lead to some mis-measurement. As an alternative measure we also calculate the realised spread as proposed by Huang and Stoll (1996).

This is simply a measure of the post-trade revenue of the market-maker and allows for price movements after a trade caused by adverse selection. It is defined simply as the subsequent price change conditional on the last trade being a buy or a sell and attempts to capture the revenue of the market-maker as he reverses his position. It has the same interpretation as the GKN spread; a measure of the pure order-processing cost component of the spread. The surplus of the effective spread over the realised spread will reflect the adverse selection component.

$$S_{R|B} = 2 * \left(P_{t+1} - P_t\right) \big| P_t = B$$

and **(7)**

$$S_{R|A} = 2 * \left(P_{t+1} - P_t\right) \big| P_t = A$$

One complication with this realised spread measure is that trades must be signed in order to calculate the conditional price change. In order to sign trades we use the approach of Lee and Ready (1991) so that a trade is classified as a sell order if the transaction price is closer to the quoted bid than the ask. In the few cases when the transaction price was equidistant between bid and ask then we used the 'tick rule' that an increase in price since the last trade was classified as a buy order.

(7) See the Appendix for full derivation.

Table E2 details our findings on order-processing costs.

Table E2
Order-processing costs components of LIFFE and DTB spreads

	Spread measure (percentage)		
	GKN spread	Realised spread (bid)	Realised spread (ask)
LIFFE	0.01350	0.00806	0.00738
DTB	0.01198	0.00790	0.00702
Percentage difference (LIFFE/DTB)	+13%	+2%	+5%
T-test for difference	n/a	0.3476	0.8445

Notes: The t-test indicates that the difference between the realised spreads on both exchanges is insignificant. n/a = not applicable.

So, on the GKN measure the order-processing component of LIFFE spreads is higher, as predicted. However, order-processing costs based on the realised measures are the same on both exchanges.

If we take it that the effective spread is our best estimate of the true spread,[8] then we can calculate the adverse selection component as the difference between this and our estimates of the order-processing component. Table E3 shows our results.

If we focus on the GKN measures we see that the order-processing component accounts for 92% of the spread on LIFFE and 81% of the spread on the DTB. This is similar to the findings of GKN. The adverse selection component based on the GKN measure is almost twice as large on the DTB as on LIFFE. This might be expected given the greater anonymity associated with the DTB's electronic order book[9] and the longer length of time needed to revise a quote (or any other aspect of an order) on the DTB. However, the difference between the

(8) Based on the simulation results in the Appendix.
(9) Although your identity is masked, by posting an order you are exposing yourself, revealing information about your position and increasing the possibility of a more informed trader executing against you.

Table E3
Adverse selection components of LIFFE and DTB spreads

	Spread measure (%)		
	GKN Spread	Realised Spread (bid)	Realised Spread (ask)
LIFFE	0.00122	0.00651	0.00766
DTB	0.00208	0.00610	0.00752
Percentage difference (LIFFE/DTB)	-41%	+6.7%	+1.9%
T-test for difference	n/a	0.8049	0.7780

Notes: The t-test iindicates that the difference between the realised spreads on both exchanges is insignificant. n/a = not applicable.

measures based on the realised spreads is insignificant, suggesting that neither the order-processing component nor the adverse selection component is different on the two exchanges.

5 Price discovery

An important aspect of price formation in fragmented markets is the location of price discovery and the speed with which one market reacts to price discovery in the other. As mentioned in Section 1, there has been some analysis of this issue but, as with bid-ask spreads, results have been contradictory with Shyy and Lee (1995) finding one-way causality from DTB to LIFFE and Kofman and Moser (1996) finding bi-directional causality. Once again the use of transactions data versus quote data is the most important issue, with transactions data having the disadvantage that timing of data is not synchronised leaving the higher turnover market, LIFFE, with more up-to-date observations on average than the DTB. This clearly biases most test toward accepting causality from LIFFE to DTB. There are also some differences in methodology between the two studies - Kofman and Moser (1996) allow for time-varying volatility - but in both cases a cointegration approach is used.

The most common method for testing lead/lag relationships is the simple Granger Causality test where non-stationary series are differenced to induce stationarity. For example, a Granger Causality test of whether variable X caused variable Y in the bivariate case with I(1) variables[(10)] would consist of running the following regression.

(10) I(1) variables must be differenced once to induce stationarity.

$$\Delta Y_t = \alpha + \sum_{i=1}^{T} \beta_i \Delta Y_{t-i} + \sum_{j=1}^{J} \gamma_j \Delta X_{t-j} \qquad \textbf{(8)}$$

and testing to see if all γ's are jointly insignificant.

As is well known, this test is problematic if X and Y are cointegrated(11) since it does not exploit this common trend between the variables in testing. Clearly, in the case of DTB and LIFFE Bund contracts, a cointegrating relationship does exist since both set of observed prices are influenced by an underlying common trend.(12)

As a result both Shyy and Lee (1995) and Kofman and Moser (1996) estimate an error correction relationship between the two prices:

$$\Delta Y_t = \alpha + \sum_{i=1}^{T} \beta_i \Delta Y_{t-i} + \sum_{j=1}^{J} \gamma_j \Delta X_{t-j} + \chi (Y_{t-1} - X_{t-1}) \qquad \textbf{(9)}$$

and test if all γ's and λ are jointly insignificant. The error correction term then exploits the levels relationship between the variables.

Hasbrouck (1995) goes one step further and argues that it is precisely this cointegrating relationship that is of most importance in the issue of price discovery. He suggests that the common trend between market prices can be thought of as the underlying efficient price and so the contribution of each market to innovations in this common trend is a measure of their contribution to price discovery. Using this logic, Hasbrouck suggests a measure of price discovery based on the contribution of prices in a given market to the total variance of the common trend.

---

(11) An I(0) variable can be created from a linear combination of these I(1) variables.

(12) Breedon (1996) shows that the LIFFE contract has a tendency to trade slightly more expesive than the DTB contract. However, since the gap between the prices is stable over time this does not influence the cointegrating relationship. We test for the cointegrating relationship between the two markets; as expected we find one cointegrating vector.

Following Hasbrouck we can define the vector moving average representation of prices

$$\Delta P_t = \Psi(L)e_t \qquad \textbf{(10)}$$

where p_t is a vector of prices, $\Psi(L)$ is a polynomial in the lag operator and e_t are serially uncorrelated disturbances with a covariance matrix Ω. Although the prices are non-stationary, the differences between them are, implying that they share a common trend. Stock and Watson (1988) propose a common trends representation of such a system as

$$P_t = P_0 + \psi\left(\sum_{s=1}^{t} e_s\right) + \Psi^*(L)e_t \qquad \textbf{(11)}$$

where p_0 is a constant vector and $\Psi^*(L)$ is a polynomial in the lag operator. The important component here is the common trend (the second term). This is the element that determines permanent changes in p_t. The share of each market's prices in underlying price determination is therefore its contribution to this common trend. Assuming that the variance-covariance matrix Ω is diagonal, this share is simply

$$S_j = \frac{\psi_j^2 \Omega_{jj}}{\psi\Omega\psi'} \qquad \textbf{(12)}$$

which is the contribution of each market's price to the total variance of the underlying trend. In practice this statistic can be derived from the impulse responses of an estimated VECM (vector error correction model).

Three further complication arise in these tests.

(i) Even with the high-frequency data used here, much price formation will appear contemporaneous. This means that the price of both contracts will appear to move simultaneously giving them a strong contemporaneous correlation. In Granger causality tests this means we may fail to find causality in the sense of one price moving before the other even though prices are closely linked. In the case of the Hasbrouck test, this contemporaneous correlation means that we can only give an upper and lower bound for contributions to the common trend but no unique value can be found. These bounds represent the contribution of one market to price formation under the two extreme assumptions; first, all contemporaneous price movements are actually due to that market and secondly, none is (this effectively means changing the ordering of the VECM for the impulse responses). The width of the range reflects the degree of contemporaneous correlation between the two markets; a narrow range means that there is little contemporaneous co-movement between the two markets.

(ii) As Kofman and Moser (1996) point out, there is clear evidence of heteroskedasiticty in these data and so assuming constant variance, as Shyy and Lee (1995) do, may reduce the efficiency of the price discovery tests. To look at this issue we present results for both constant variance and GARCH(1,1) representations of all the regressions. For example, when we estimated (8) we also estimated

$$\Delta Y_t = \alpha + \sum_{i=1}^{T} \beta_i \Delta Y_{t-i} + \sum_{j=1}^{J} \gamma_j \Delta X_{t-j} + e_t \qquad (8a)$$

$$\sigma_t^2 = \varpi + \lambda \sigma_{t-1}^2 + \rho e_{t-1}^2$$

where σt2 is the variance of the dependent variable ΔYt up to time t.

(iii) As was noted above, undertaking causality tests with transactions data is problematic, given that the observations are not synchronised. In order to undertake testing we first follow the approach of Kofman and Moser who create minute-by-minute data for each market by taking the last observed transations price in each minute. Whilst this approach creates a conventional time series, it does have the implication that the data for LIFFE will, on average, lead the data from DTB. This occurs simply because there are 4,469 more trades on LIFFE in our sample and so the last transaction price in the minute on LIFFE will tend to be more up-to-date than the equivalent DTB price. This effect may bias us to accept causality from LIFFE to DTB. As a crude method of removing this bias, we also present causality results for a reduced LIFFE series(13). Here we simply randomly remove observations from the LIFFE series until there are as many transactions prices for LIFFE as for DTB. We then construct the minute-by-minute series using the matched samples. Clearly this is a crude technique, but it may give some insight into the extent of the bias induced by more frequent sampling in the LIFFE series.(14)

For all representations, the lag length was determined by the Schwarz information criterion.

---

(13) As a further robustness check, we also test for causality using data extracted with a five-minute grid; the results were qualitatively similar.

(14) We also ran the tests using the mid quote; again there was strong bi-directional causality between the two exchanges. Under the assumption of constant variance, LIFFE's share in price discovery ranged from 16%-72% while the DTB's share ranged from 21%-77%, indicating a slightly stronger role for LIFFE relative to full quotes.

Table F
Tests of price discovery

		Simple Granger causality test		ECM causality test		Hasbrouck Share Statistic	
		H_0: LIFFE does not cause DTB	*H_0: DTB does not cause LIFFE*	*H_0: LIFFE does not cause DTB*	*H_0: DTB does not cause LIFFE*	*LIFFE contribution to price formation*	*DTB contribution to price formation*
Quote data	Constant variance[(a)]	66.5**	87.3**	61.6**	82.0**	15%-61%	39%-85%
	GARCH(1,1)[(b)]	1214.8**	1505.0**	1224.6**	1533.3**	21%-66%	34%-79%
Transaction data	Constant variance[(a)]	125.9**	62.7**	118.3**	58.5**	36%-78%	22%-64%
	GARCH(1,1)[(b)]	2570.7**	1521.4**	2660.7**	1417.4**	32%-73%	27%-68%
Matched transaction data	Constant variance[(a)]	114.9**	69.2**	108.1**	64.6**	35%-77%	23%-65%
	GARCH(1,1)[(b)]	2366.7**	2022.8**	2453.4**	1879.5**	32%-73%	27%-68%

Notes: The first two columns show F-statistics for exclusion of one market's price from the equation of the other. ** indicates rejection of the null hypothesis at the 1% significance level. Final column shows Hasbrouck's information share variable ie the contribution to the common trend. The range allows for contemporaneous correlation and indicates, for example, that if we assume all contemporaneous price movement are due to the DTB, the DTB's contribution to price formation using quote data is 85%.
(a) The test statistic is F-distributed.
(b) The test statistic is χ^2-distributed.

All the tests shown in Table F indicate a strong two-way causation between LIFFE and DTB prices.[(15)] Lags of the other market’s prices are highly significant in both the simple Granger causality and the VECM specifications for quote and transactions data and the share statistic is well above zero for both markets, even at the lower range. It is also clear that the bias induced by more frequent trading on LIFFE is minimal for our sample (ie the results for the full transactions sample and the matched transactions sample are almost identical). The share statistics show some differences between quote and transactions data based measures in that, for quote data, DTB prices make a stronger contribution to price formation (in line with the results of Shyy and Lee (1995) for quote data) whilst for transactions data the contributions of the two markets are roughly equivalent (in line with the results of Kofman and Moser (1996) for transactions data). This seems to

(15) This is consistent with the findings of Fremault Vila and Sandmann (1995).

indicate that the DTB has greater pre-trade transparency since its quotes are more informative to prospective traders (in line with the results on effective and quoted spreads above), though once again the non-synchronous nature of the quotes and transactions data and the lack of data on contract size traded intra-day on LIFFE makes it difficult to draw firm conclusions.

6 Market resilience

An important aspect of the debate concerning open outcry versus electronic markets is their relative performance in periods of stress. Here, once again there appears to be some disagreement in the literature though in this case the hypotheses tested are not exactly comparable. Pirrong (1996) tests the hypothesis that predictable changes in volume influence prices. He argues that if volume - particularly very high volume - does influence prices, that indicates that the market has limited depth since market-making capacity cannot enter the market and so price volatility increases (presumably because of wider spreads - though this is not made clear). He then tests this hypothesis in a two-stage model where he models expected volume as a function of past volume and volatility and then looks at the relationship between expected volume and current price variation (absolute change in prices). He finds that expected increases in volume have a smaller effect on price variability on the DTB than on LIFFE, suggesting that the electronic system is deeper.

Franke and Hess (1995) test a slightly different hypothesis. They look at how the share of turnover in each market changes with market volatility and find that in volatile periods LIFFE's share increases, indicating that the open outcry market is more attractive in volatile periods. Although this is a somewhat different test to Pirrong's, its results *do not* seem consistent with the idea that market-making capacity is limited on LIFFE (volatile periods are usually ones of high turnover as Pirrong's own results show). As an additional test in this area we looked at the relationship between spreads and volatility and turnover.(16) If Pirrong's hypothesis is correct we should find that LIFFE spreads widen more during periods of market stress (though this would then leave the puzzle of why traders move to LIFFE during these periods).

Our data partly confirm Franke and Hess' results. On a daily basis, rank correlation tests show that there is a significant positive relationship between market volatility (as measured by the weighted average of the daily standard deviation of transactions prices in the two markets) and the total number of contracts traded, confirming that periods of market stress are associated with high turnover. However, although LIFFE's share of this turnover is 40%

(16) In this analysis we use the number of trades as a proxy for turnover.

correlated with this measure of volatility, the rank correlation test shows that the relationship is not statistically significant.

Table G1
Rank correlation coefficients

Hypothesis		
High market volatility and high turnover	Correlation coefficient Rank correlation coefficient T-stat	0.69 0.60 4.29
High market volatility and high market share	Correlation coefficient Rank correlation coefficient T-stat	0.40 0.09 0.51

Note: Figures in bold indicate that the relationship was statistically significant at the 5% significance level.

At the high frequency level we find that increased market volatility (measured both using the conditional volatility from our GARCH estimates above and average absolute changes in DTB bid prices over the previous 15 minutes) results in a significant increase in the share of the total number of trades undertaken on LIFFE.(17)

However, as Table G2 below indicates the movement of trades onto LIFFE does not appear to be because LIFFE spreads are less influenced by volatility. We find that although LIFFE quoted spreads do not increase as much as those on the DTB during volatile periods, the effective spreads on both markets increase similarly during volatile times. Once again, however, lack of information on LIFFE trade sizes means that we cannot draw strong inferences from these results, though Pirrong's conjecture that LIFFE has less depth seems inconsistent with our (and that of Franke and Hess) finding on market share. It may be that in periods of market stress, orders flow to the larger market, in this case LIFFE, regardless of the trading platform. This would be consistent with the findings of Fremault Vila and Sandmann (1995) who find that in volatile periods, the market share of the larger market (in this case the electronic market) increases.

(17) A 10% increase in the average absolute price change and conditional volatility result in 0.6% and 0.2% decline in DTB's market share respectively. These results are derived from impulse responses of a bi-variate VAR of market share and the volatility/trade intensity measure; ie the impulse responses from a VAR in the variables $\Delta share$ and $\Delta volatility$.

Table G2
Impact of a 10% increase in volatility/volume measure on the spread

	Measures of spreads			
	Quoted spread (percentage change)		Effective spread (percentage change)	
Measures of market stress	**LIFFE**	**DTB**	**LIFFE**	**DTB**
Absolute price change	-0.000001	0.000005	0.000013	0.000013
Conditional volatility	0.000001	0.000003	0.000010	0.000010
Volume traded	-0.000001	-0.000001	0.000006	0.000006

Notes: Absolute price change is the average absolute change in DTB bid prices over the previous 15 minutes. Conditional volatility is the estimate of h from the GARCH equation for DTB quoted prices shown in Table F. Volume traded is the total number of trades in the previous 15 minutes. In all cases, the results are derived from impulse responses of a bi-variate VAR of spreads and the volatility/trade intensity measure; ie the impulse responses from a VAR in the variables Δspread and Δvolatility.

7 Conclusion

Dually traded futures have become increasingly popular test beds for theories of market microstructure. They are highly liquid and informationally efficient and are often, as is the case with the Bund contract, traded on very different trading systems. Overall, we have found that the two markets are highly integrated and that despite being the smaller market, the DTB has as much of a role in underlying price discovery as LIFFE. In fact, DTB quotes seem to be more informative than LIFFE ones both in the sense of contributing to price discovery and in that the quoted spread is more closely aligned to the effective spread. Generally, variable trading costs on the two markets are similar with the larger average trade size on LIFFE making firmer comparison difficult. There is also some evidence to support the proposition that the open outcry market (LIFFE) performs better in periods of volatility than the electronic market (DTB) since greater volatility seems to be associated with an increase in the market share for LIFFE.

While transactions prices and quotes seem to be very similar on average, we find that the conflicting results of past studies can be explained by the choice of data set. It is apparent that results are conditional on the data used and may not hold over all specifications of the data. Our preferred measure of spread is the effective spread of Huang and Stoll (1996), which uses both quotes and transactions prices; ie a larger information set than more traditional measures such as the Roll (1984) and the Thompson-Waller

(1988) estimates, which provide a lower and upper bound for the effective spread.

Ideally, we would also like to condition all our results on trade size. This would make our comparisons of spreads more meaningful and would allow for a fuller analysis of market resilience. However, until this data becomes available, that avenue of research remains unexplored.

Appendix
Measures of the bid-ask spread

Theoretical models view the bid-ask spread as compensation to market-makers for three costs: (i) order-processing costs; (ii) inventory risk and (iii) adverse selection risk. As reliable quote data are often not available on an *ex post* basis, financial economists have developed a number of methods to estimate the bid-ask spread from the available times series of transactions prices. These measures have been widely applied and investigated in the context of equity markets, but they also hold in futures markets. Below we present an overview of a variety of methods used to measure the bid-ask spread. The objective is to measure the actual cost incurred and not necessarily the quoted spread. We then attempt to rank each method by using a Monte Carlo approach, applying each measure to a generated time series of transactions prices with a known spread.

Measures of the spread

Let us assume an efficient market with an equal number of buyers and sellers and a homogenous pool of market-makers. Consider the representative market-maker. Let us assume that orders arrive randomly and the market-maker is indifferent to the size and sign of the order, therefore his quoted price will by symmetric about the 'true' price of the asset, p_t*, prior to receipt of an order. Explicitly we are assuming that inventory costs for the market-maker are zero.

Let the spread, s, be the difference between the bid, p_b and the ask, p_a. So, at time t, a transaction will either be done at p_b or p_a with equal probability. Call this transaction price, p_{kt}. If we define Q as an indicator of sign of trade, with Q = 1 if its a customer buy order and Q = -1 if its a customer sell order, then $k = a$ if Q = 1 and $k = b$ if Q = -1. So,

$$\left.\begin{aligned} p_{at} &= p_t^* + \frac{s}{2} \\ p_{bt} &= p_t^* - \frac{s}{2} \end{aligned}\right\}.$$

Now we can decompose the spread into two components; the order-processing cost and compensation for adverse selection, in proportion π, and

(1-π) respectively. Let

$$p_t = p_t^* + (1-\pi)\frac{s}{2}Q_t, \qquad \textbf{(A1)}$$

ie the market-maker updates his valuation of the asset on receipt of the order. This takes into account the adverse selection component and reflects all available information up to that point.

Now p_t evolves as follows:

$$p_t = p_{t-1} + E_t + (1-\pi)\frac{s}{2}Q_t + u_t \qquad \textbf{(A2)}$$

where E_t is the expected return at time t-1 for the period t-1 to t and u_t is the unpredictable innovation in 'true' prices, with $u_t \sim N(0,\sigma)$.

Then, at time t, transactions will take place at p_{kt} where

$$p_{kt} = p_t + \pi\frac{s}{2}Q_t; \qquad \textbf{(A3)}$$

$\pi\frac{s}{2}Q_t$ reflects the cost of processing the order.

So, the returns at time t based on transactions prices can be written as:

$$R_t = p_t - p_{t-1} + \pi\frac{s}{2}\left[Q_t - Q_{t-1}\right]. \qquad \textbf{(A4)}$$

Substituting in for p_t - p_{t-1}, we get

$$R_t = E_t + \pi\frac{s}{2}\left[Q_t - Q_{t-1}\right] + (1-\pi)\frac{s}{2}Q_t + u_t\,. \qquad \textbf{(A5)}$$

Now, $\mathrm{cov}(R_t, R_{t-1}) = \mathrm{cov}(E_t, E_{t-1}) - \pi^2\frac{s^2}{4}$, so we can then calculate the order-processing component of the spread from:

$$\pi s = 2\sqrt{-\left[\mathrm{cov}(R_t, R_{t-1}) - \mathrm{cov}(E_t, E_{t-1})\right]} \qquad \textbf{(A6)}$$

- Consider first the case with π=1, ie there is no adverse selection risk. Let us assume also that expected returns are constant, so E_t=E for all t. Then equation (A6) simplifies to the well-known Roll (1984) estimate,

  $$s = 2\sqrt{-\mathrm{cov}(R_t, R_{t-1})}\,. \qquad \textbf{(A7)}$$

It is clear that this measure will be biased if $\pi<1$ and/or if expected returns are time-varying, ie $\mathrm{cov}(E_t,E_{t-1}) \neq 0$.(18)

- More generally, if there is serial correlation present in transactions with the probability of a bid (ask) following a bid (ask) equal to ρ, but no time variation in returns, then the estimator becomes

$$s = \left\{\sqrt{-Cov(R_t, R_{t-1})}\right\} / \left(1-\rho\right). \qquad \textbf{(A8)}$$

- If we now allow expected returns to be time-varying *and* relax the constraint on π, ie π is not necessarily equal to 1, then both these estimates will be biased. Expected returns are unobservable, so we need some way to measure them. If we have data on quotes, we can use these to obtain an alternative measure for the difference between realised and expected returns. If bp_t is the bid price following receipt of an order at time t, then

  $bp_t = p_t - \pi\dfrac{s}{2}$ and returns based on bid prices will be

$$RB_t = E_t + (1-\pi)\frac{s}{2}Q_t + u_t\,. \qquad \textbf{(A9)}$$

  So define

$$RD_t = R_t - RB_t = \pi\frac{s}{2}\Delta Q_t\,. \qquad \textbf{(A10)}$$

  Then $\mathrm{cov}(RD_t, RD_{t-1}) = -\pi^2\dfrac{s^2}{4}$. **(A11)**

  So another estimate of the order-processing component of the spread is given by

$$s = 2\sqrt{-\mathrm{cov}(RD_t, RD_{t-1})}\,. \qquad \textbf{(A12)}$$

  This measure is due to George, Kaul and Nimalendran (1991), the GKN spread.

- Next, if we have quotes data and we assume market-makers are indifferent to their inventory position, the mid-quote should proxy the market-makers true valuation of the asset, p^*_t, prior to receipt of an order.

---

(18) Note that this assumes that the conditional probability of a bid at time t, given a bid at time t-1, is equal to the conditional probability of an ask. It is sometimes the case that, with time and sales data, split transactions (ie sequential transactions executed at the same price) are not recorded. In these cases, the probability of a bid following a bid is zero and the Roll estimate should be scaled down by a factor of 2 (see Followill and Helms (1990)).

So, the difference between the transaction price and the mid-quote should, on average, be one half of the spread and will cover all costs. This allows for trades to take place within the publicly quoted spreads (often these are just the starting point for negotiation) at the market-makers' implicit quote. So

$$s = 2 * \frac{1}{T} \sum_{t=1}^{T} \left| p_{kt} - mq_t \right|. \qquad \textbf{(A13)}$$

This is the effective spread due to Huang and Stoll (1996). This should, in theory, be better than the quoted spread as it is based on the prices at which trades occur.

- Similarly, Huang and Stoll (1996) try to capture the revenue of the market-maker from reversing a trade, taking into account that prices will move against him if he deals with an informed trader; ie it is net of the adverse selection cost. The realised return will only reflect order-processing costs.

  So, $$s = \begin{cases} 2 * \left(p_{k,t+1} - p_{k,t}\right) \mid p_{k,t} = p_b \\ 2 * \left(p_{k,t} - p_{k,t+1}\right) \mid p_{k,t} = p_a \end{cases}. \qquad \textbf{(A14)}$$

  It allows for the fact that, on average, the dealer will not be able to realise the effective spread when he trades with an informed trader.

- Finally, if we assume that if a bid (ask) follows a bid (ask), the return is zero (no news arrival so no revision to the true valuation of the asset), otherwise the absolute value of the return is the distance between the bid and the ask price, ie the spread. Then $\left| R_t \right| = s$ with probability _.

  So, $$s = 2\left(\frac{1}{T} \sum_{t=1}^{T} \left| R_t \right|\right). \qquad \textbf{(A15)}$$

  This is the Thompson-Waller (1988) estimate of the spread.

Simulation results

In order to rank the various measures of the spread presented above, we conduct the following experiment. We generate a series of prices (p_t) such that the underlying returns (Δp_t) are normally distributed with mean zero and standard deviation of 0.025, and p_0=100. We then generate a series of transactions prices

(p_{kt}) where :

$$p_{kt} = p_t + s/2 \text{ with probability \_ and } k=a$$
$$p_{kt} = p_t - s/2 \text{ with probability \_ and } k=b$$

where s is the spread. We also generate a series of quotes where the bid (ask) quote is equal to the true price less (plus) half the spread. This assumes market-makers are indifferent to the order flow. We use 10,000 observations and repeat the draw 1,000 times. For each draw we calculate the Roll, Thompson-Waller, GKN, effective and realised spreads. Table 1 shows the results. This is our base case. We calculate the mean error and the mean squared error in order to rank the measures. Under these assumptions the effective measure performs the best and the Thompson-Waller the worst. Table 2 and 3 show the results of a similar experiment where we allow serial correlation in expected returns;
ie $\Delta p_t = \mu + \rho\Delta p_{t-1}+v_t$, with v_t~N(0,0.00525) (implicitly we are setting ρ=0 in the first experiment). Again, the same ranking holds if we consider the absolute mean error, although the mean squared error indicates that the realised spreads perform best.

We then change the volatility of our generated 'true' returns. Table 4 assumes that returns are normally distributed with mean zero and standard deviation 0.05, while Table 5 assumes a standard deviation of 0.1. In both these experiments ρ=0. Note how large the errors become in the Thompson-Waller estimate as returns become increasingly more volatile.

Overall, the realised spread measures perform best. However, these measures are designed to only measure order-processing costs. In this experiment these are the only component of the spread; however, if there is inventory risk and a risk of adverse selection, then the effective spread measure gives the best estimate of the total spread. Note that the traditional Roll measure performs quite well.

Table 1
ρ=0, standard deviation of underlying returns 0.025

	Roll	Thompson-Waller	Effective	GKN	Realised spread (bid)	Realised spread (ask)
Average	0.100020	0.119949	0.099994	0.102011	0.099612	0.100044
Absolute mean error (percentage)	0.020266	19.948884	0.006078	2.010807	0.038768	0.043668
Mean squared error (percentage)	0.003659	0.398774	0.000036	0.007600	0.002562	0.002520
True spread	0.100000					

Note: The errors are expressed as a percentage of the true spread.

Table 2
ρ=0.2, standard deviation of underlying returns 0.025

	Roll	Thompson-Waller	Effective	GKN	Realised spread (bid)	Realised spread (ask)
Average	0.099343	0.119913	0.099999	0.101977	0.099927	0.100050
Absolute mean error (percentage)	0.656700	19.912900	0.000226	1.976560	0.000735	0.000499
Mean squared error (percentage)	0.004014	0.397300	0.000039	0.007628	0.000028	0.000024
True spread	0.100000					

Table 3
ρ=-0.2, standard deviation of underlying returns 0.025

	Roll	Thompson-Waller	Effective	GKN	Realised spread (bid)	Realised spread (ask)
Average	0.100579	0.119966	0.099989	0.101995	0.100096	0.100031
Absolute mean error (percentage)	0.579240	19.966090	0.011140	1.995230	0.000956	0.000309
Mean squared error (percentage)	0.003765	0.399410	0.000043	0.007263	0.000027	0.000023
True spread	0.100000					

Table 4
ρ=0.0, standard deviation of underlying returns 0.05

	Roll	Thompson-Waller	Effective	GKN	Realised spread (bid)	Realised spread (ask)
Average	0.100124	0.140780	0.099981	0.102071	0.100034	0.100015
Absolute mean error (percentage)	0.123826	40.779756	0.018878	2.070577	0.000034	0.000015
Mean squared error (percentage)	0.011186	1.664033	0.000039	0.014461	0.000004	0.000004
True spread	0.100000					

Table 5
ρ=0.0, standard deviation of underlying returns 0.1

	Roll	Thompson-Waller	Effective	GKN	Realised spread (bid)	Realised spread (ask)
Average	0.100654	0.196496	0.099992	0.102526	0.100212	0.100136
Absolute mean error (percentage)	0.653710	96.495891	0.008009	2.525730	0.010212	0.000136
Mean squared error (percentage)	0.046319	9.313582	0.000043	0.045349	0.000009	0.000011
True spread	0.100000					

References

Breedon F (1996), 'Why do LIFFE and DTB Bund Contracts trade at different prices?', *Bank of England* Working Paper, No 57.

Followill, R and Helms, B (1990), 'Put-Call-Futures Parity and Arbitrage Opportunities in the Market for Options on Gold Futures Contracts', *The Journal of Futures Markets*, 10, pages 339-52.

Franke, G and Hess, D (1995), 'Anonymous electronic trading versus floor trading', *University of Konstanz,* mimeo.

Fremault Vila, A, and Sandmann, G (1995), 'Floor Trading versus Electronic Screen Trading: An Empirical Analysis of Market Liquidity and Information Transmission in the Nikkei Stock Index Futures Market', *LSE Financial Markets Group,* Discussion Paper No 218.

George T, Kaul, G and Nimalendran, M (1991), 'Estimation of the Bid-Ask Spread and its Components: A New Approach', *Review of Financial Studies*, 4, pages 623-56.

Hasbrouck, J (1995), 'One Security, Many Markets: Determining the Contributions to Price Discovery', *Journal of Finance*, 50, pages 1175-99.

Huang, R and Stoll, H (1996), 'Dealer versus auctions markets: A paired comparison of execution costs on NASDAQ and the NYSE', *Journal of Financial Economics*, 41, pages 313-57.

Kofman, P, and Moser, J (1996), 'Spreads, Information Flows and Transparency across Trading Systems', Monash University, *mimeo*.

Lee, C and Ready, M (1991), 'Inferring Trade Direction from Intraday Data' *Journal of Finance,* 46, pages 733-46.

Pirrong, C (1996), 'Market Liquidity and Depth on Computerised and Open Outcry Trading Systems: A Comparison of DTB and LIFFE Bund contracts', *Journal of Futures Markets*, 16, pages 519-43.

Roll, R (1984), 'A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market' *Journal of Finance*, 39, pages 1127-39.

Shyy, G and Lee, J (1995), 'Price Transmission and Information Asymmetry in Bund Futures Markets: LIFFE vs. DTB', *Journal of Futures Markets*, 15, pages 87-99.

Smith, T and Whaley, R (1994), 'Estimating the Effective Bid/Ask Spread from Time and Sales Data', *Journal of Futures Markets,* 14, pages 437-455.

Stock, J and Watson, A (1988), 'Testing for Common Trends', *Journal of the American Statistical Association,* 83, pages 1097-107.

Thompson, S and Waller, M (1988), 'Determinants of Liquidity Costs in Commodity Futures Markets', *Review of Futures Markets,* 7, pages 110-26.