Price Discovery in Spot and Futures Markets: A Reconsideration*

Erik Theissen**

October 2005

Abstract: We reconsider the issue of price discovery in spot and futures markets. We use a threshold error correction model to allow for arbitrage operations to have an impact on the return dynamics. We estimate the model using quote midpoints, and we modify the model to account for time-varying transaction costs. We find that the futures market leads in the process of price discovery. The lead of the futures market is more pronounced in the presence of arbitrage signals. Thus, when the deviation between the spot and the futures market is large, the spot market tends to adjust to the futures market.

JEL classification: G13, G14

Keywords: Futures Markets, Threshold error correction, Information shares, Common factor weights

* I thank Bloomberg, L.P. for providing the data. I thank Alexander Kempf, Olaf Korn and seminar participants at the university of Cologne for valuable comments.

** Erik Theissen, University of Bonn, BWL I, Adenauerallee 24-42, 53113 Bonn, Germany, Email: theissen@uni-bonn.de.

1 Introduction

Which market impounds new information faster into prices, the index futures market or the spot market for the constituent stocks of the index? Transaction costs are likely to be lower in the futures market. Given that the magnitude of the transaction costs determines whether a trader can profitably trade on a given piece of information, the adjustment of prices to market-wide information (e.g. announcements of macroeconomic variables) should be faster in the futures market. On the other hand, traders possessing information about the value of individual stocks will most likely trade that stock rather than the whole index. Consequently, stock-specific information should be reflected in the spot market first.

The issue of the relative contributions of spot and futures markets to the process of price discovery is of obvious importance, and consequently has received considerable attention in the literature. The by now common methodology is to estimate an error correction model. There are, however, several problems which make straight estimation of the model troublesome.

First, the constituent stocks of the index trade infrequently. Consequently, index values are partially based on stale prices. The infrequent trading effect together with bid-ask bounce introduces distinct serial correlation patterns into the time series of index returns which may induce a spurious lead of the futures market. Although Stoll and Whaley (1990) have proposed a method to purge the return data of the infrequent trading effects, it is much less clear how the index level data needed in the estimation of the ECM can be purged of those effects. Second, the cointegrating relation between index levels and index futures prices implied by the cost-of-carry model is not constant over time but rather changes daily. Third, the standard error correction model implies that the speed of adjustment of prices to deviations from the long-run equilibrium relation is independent of the size of the deviation. This is not necessarily the case, however, because arbitrageurs will start trading when the deviation is larger than the expected roundtrip transaction cost. Their trading activity is likely to speed the adjustment.

One potential solution to the infrequent trading (and bid-ask bounce) problem, first proposed by Shyy, Vijayraghavan and Scott-Quinn (1996), is to use quote midpoints rather than prices. The time-variability of the cointegrating relationship can be accounted for by either demeaning the log price series as proposed by Dwyer, Locke and Yu (1996) or by using discounted futures prices as is done by Kempf and Korn (1996) and Martens, Kofman and Vorst (1998). Finally, a threshold error correction model allows the adjustment coefficients to depend on the magnitude of the deviation from the long-run equilibrium relation and is thus able to account for the presence of arbitrageurs (Dwyer, Locke and Yu 1996).

The present paper contributes to this line of research. We use data from the German blue chip index DAX and the DAX futures contract traded on the EUREX to assess both markets' contributions to price discovery. As suggested above, we use quote midpoint data, we use demeaned log price series, and we use a threshold error correction model. The contribution of our paper is threefold. First, we modify the threshold error correction model to allow for time-varying transaction costs. Previous papers (Dwyer, Locke and Yu 1996, Martens, Kofman and Vorst 1998) have estimated the threshold transaction costs (i.e., the size of the deviation of prices from their long-run equilibrium that allows arbitrageurs to break even) and implicitly assumed the costs to be constant. It is, however, well established that bid-ask spreads follow a distinct intradaily pattern. We allow for this time-variation by making the threshold dependent on the bid-ask spreads in the two markets. Second, this is the first paper to estimate a threshold error correction model using midquote data. This is potentially important because arbitrage signals should be based on tradable prices (i.e., bid and ask quotes) rather than on past transaction prices - even more so as the index values are affected by the infrequent trading problem alluded to above. Finally, we use data at a higher frequency than previous papers (15 seconds as compared to 1 or 5 minutes). This allows a more precise estimation of the contribution of the cash and the futures markets to the process of price discovery.[1] The increased number of observations further allows us to estimate separate models for each trading day. Another distinctive feature of our paper is that both markets under scrutiny are electronic limit order markets. Consequently, the results are unlikely to be caused by differences in market microstructure.

Our results can be summarized as follows. The futures market clearly dominates the price discovery process. Returns in the cash market depend much more heavily on lagged returns in the futures market than vice versa. When measuring the contributions to price discovery using the information shares or the common factor weights we also find that the futures market leads. We further find that the dynamics of the adjustment process is different when arbitrage opportunities exist. In these cases, the leading role of the futures market is even more pronounced.

The paper is structured as follows. Section 2 provides a brief survey of the literature. Section 3 describes the data set and presents some descriptive statistics. Methodology and results of our empirical analysis are presented in section 4. Section 5 concludes.

2 A Brief Review of the Literature

Empirical analysis of the relation between stock index values and index futures prices is complicated by methodological problems. Stocks in the spot market are not traded simultaneously. Consequently, the index is partially calculated from stale prices. This introduces positive serial correlation in the index returns which, in turn, may introduce a spurious lead-lag

1 Note that when estimating the Hasbrouck (1995) information shares, the contemporaneous correlation between the return innovations is arbitrarily assigned to one market. By reversing the order of markets, upper and lower bounds for the information shares can be obtained. The higher the frequency of observations the

relation. Further, bid-ask bounce may induce negative serial correlation in the return series. Stoll and Whaley (1990) propose to estimate an ARMA model for the index returns and to use the innovations from the model rather than the index returns to analyze the lead-lag relation between the spot and the futures market. Using a VAR model they find that the futures market leads the stock market by about 5 minutes. The general result that the futures market leads the spot market has, despite all methodological differences, almost universally been confirmed in subsequent research.[2]

The VAR approach does not take into account that index values and futures prices are cointegrated. What is required instead is an error correction model (ECM). Different approaches at estimating an ECM have been proposed. Some authors have estimated the cointegrating relationship (e.g. Shyy, Vijayraghavan and Scott-Quinn 1996) but the more common approach is to use a pre-specified cointegrating vector based on the theoretical cost-of-carry relation (e.g. Fleming, Ostdiek and Whaley 1996, Dwyer, Locke and Yu 1996, Kempf and Korn 1996, Martens, Kofman and Vorst 1998, Booth, So and Tse 1999).

Two issues deserve attention. First, the cost-of-carry relation $F_t = S_t e^{r(T-t)}$ implies that the cointegrating relation is not constant over time but rather changes daily.[3] Many previous papers do not take that into account. There are, however, some notable exceptions. Dwyer, Locke and Yu (1996) subtract the daily mean from the time series of log prices before esti-

---

lower the contemporaneous correlation. Consequently, higher frequency of observations allows for more accurate estimation of the information shares.

2 A notable exception is Shyy, Vijayraghavan and Scott-Quinn (1996). They confirm the result of a lead of the futures markets when basing their estimates on price data. Estimation based on quote midpoints, on the other hand, leads to the conclusion that the cash market leads. Frino, Walter and West (2000) who also use quote midpoints do not confirm this result.

3 If, as is usual, the model is estimated using logs, the relation becomes $\ln(F_t) = \ln(S_t) + r(T-t)$. This implies that, in a regression of $\ln(F_t)$ on $\ln(S_t)$, the slope is constant and equal to one, whereas the intercept changes daily. Note that we do not include the expected dividend yield in the cost-of-carry relation. The reason is that the DAX is a performance index, i.e., calculation of the index is based on the presumption that dividends are reinvested.

mating the ECM. Kempf and Korn (1996) and Martens, Kofman and Vorst (1998) use a pre-specified cointegrating vector that takes the cost-of-carry relation explicitly into account.[4]

The second issue is related to the infrequent trading problem. The ECM is usually estimated using simple log returns. These returns do, however, suffer from the infrequent trading problem. Some authors (e.g. Fleming, Ostdiek and Whaley 1996, Kempf and Korn 1996, Pizzi, Economopoulos and O'Neill 1998) have used ARMA residuals rather than log returns when estimating the ECM. The problem with this approach is that it combines an error correction term directly derived from the index and futures price levels with the ARMA residuals in one model, thereby introducing a sort of inconsistency into the model.

Two possible solutions have been proposed. Jokivuolle (1995) develops a procedure, based on the Beveridge-Nelson decomposition, that allows estimation of the true index level. Using these estimates rather than the observed index levels allows to formulate an ECM in which both the error correction term and the lagged returns are purged of infrequent trading effects. To the best of our knowledge this procedure has not yet been applied to test the lead-lag relation between spot and futures markets. Alternatively, the estimation can be based on quote midpoints rather than on prices (see Shyy, Vijayraghavan and Scott-Quinn 1996). Midpoints are based on firm quotes and thus should not suffer from an infrequent trading problem. Further, there is no bid-ask bounce in quote data.

The general ECM specification implies that, whenever prices deviate from the long-run equilibrium relation (which, in turn, is given by the cost-of-carry relation), there is a tendency for prices to adjust. The speed of adjustment is independent of the magnitude of the deviation. Several authors have argued that this is likely to be an incomplete description of the adjustment process. When deviations from the long-run equilibrium are larger than the round-trip

---

4 Specifically, their error correction term at time t is $z_t = \ln(F_t) - \ln(S_t) - (r_{t,T} - q_{t,T})(T - t)$ where r is the

transaction costs, arbitrageurs step in, thereby speeding the adjustment process. The resulting dynamics can be captured by a threshold error correction model (TECM). This approach was pioneered by Yadav, Pope and Paudyal (1994) and subsequently adopted by Dwyer, Locke and Yu (1996), Kempf and Korn (1996) and Martens, Kofman and Vorst (1998).

In these papers the TECM is estimated using transaction price data. Thus, it is assumed that a sufficiently large deviation between lagged futures prices and lagged cash index values triggers an arbitrage signal. However, arbitrageurs can not trade at these prices. This is particularly true for the cash index because the calculation of the index value is partially based on stale prices. It would be preferable to construct the arbitrage signal from quotation data because trades can actually be executed at these prices. Data on bid and ask quotes is, however, not usually available from open outcry futures markets.

A second implicit assumption made in previous papers is that the transaction cost and, consequently, the price difference triggering an arbitrage signal, is constant. This is not necessarily the case, however. The most important determinant of the transaction cost is the bid-ask spread. The spread, however, is time-varying. Some of the variation is caused by distinct intradaily patterns. Consequently, a model that assumes constant roundtrip transaction costs may fail to fully capture the dynamics of the adjustment process. The methodology used in the present paper takes the time-varying nature of transaction costs explicitly into account.

3 Data

We use data for the German blue chip index DAX. The DAX is a value-weighted index calculated from the prices of the 30 most liquid German stocks. The prices are taken from Xetra,

---

risk-free interest rate, q is the expected dividend yield and T is the maturity date of the futures contract.

the most liquid market for German stocks.[5] Index values are published in intervals of 15 seconds. The DAX is a performance index, i.e., the calculation of the index is based on the presumption that dividends are reinvested. Consequently, the expected dividend yield does not enter the cost of carry relation.

Besides an index calculated from the most recent transaction prices, Deutsche Börse AG also calculates an index from the current best ask prices (ADAX) and an index from the current best bid prices (BDAX). These indices are value-weighted averages of the inside quotes, and the difference between them is equivalent to a value-weighted average bid-ask spread.

Futures contracts on the DAX are traded on the EUREX. The contracts are cash-settled and mature on the third Friday of the months March, June, September and December. The DAX futures contract is a highly liquid instrument. In the first quarter of 1999 (our sample period), more than 1,150,000 transactions were recorded. The open interest at the end of the quarter was more than 290,000 contracts.[6]

Both Xetra and EUREX are electronic open limit order books. Therefore, the results of our empirical analysis are unlikely to be affected by differences in the microstructure of the markets.[7] The trading hours in the two markets differ. Trading in Xetra starts with a call auction held between 8.25 am and 8:30 am. After the opening auction, continuous trading starts and extends until 5 pm, interrupted by an intraday auction which takes place between 1:00 pm and 1:02 pm. Trading of the DAX futures contract starts at 9 am and extends until 5 pm.

---

5 During our sample period, the first quarter of 1999, Xetra accounted for 79.9% of the total order book turnover in the constituent stocks of the DAX on all German exchanges. See the fact book 1999 of Deutsche Börse AG, p. 33. Note that, during our sample period, Deutsche Börse AG also calculated DAX values based on the prices of the Frankfurt Stock Exchange.

6 See the fact book 1999 of Deutsche Börse AG, p. 88.

7 Some previous papers, most notably Grünbichler, Longstaff and Schwartz (1994), Kempf and Korn (1998) and Frino and McKenzie (2002), analyze spot and futures markets with different trading protocols. The focus of these papers is to assess the implications of the trading protocol for price discovery.

We obtained all data from Bloomberg. Our sample period is the first quarter of 1999 and covers 61 trading days. For this period we obtained the values of the DAX index and the two quote-based indices ADAX and BDAX at a frequency of 15 seconds. From the quote-based indices we calculated a midquote index

$$MQDAX_t = \frac{ADAX_t + BDAX_t}{2}$$

and a time series of percentage bid-ask spreads

$$S_t = 100 \frac{ADAX_t - BDAX_t}{MQDAX_t}$$

We further obtained a time series of all bid and ask quotes and all transaction prices of the nearby DAX futures contract.

We only use data for the period of simultaneous operation of both markets. We further discard all observations before 9 am and from 4:55 pm onwards. We also discard all observations within 5 minutes from the time of the intraday call auction (held between 1:00 pm and 1:02 pm). When estimating the ECM we assure that all lagged returns are from the same trading day.

In order to synchronize the data from the cash and the futures market we proceeded as follows. For each index level observation we identify the most recent transaction price and the most recent quote midpoint from the DAX futures data. Thus, in each pair of observations the observation from the futures market is older (though by some seconds only) than the matched observation from the cash market. This procedure clearly works to the disadvantage of the futures market.

The cost-of-carry relation implies that the cash index and the futures contract are cointegrated. In order to eliminate the time-variation of the cointegrating relation we follow the procedure introduced by Dwyer, Locke and Yu (1996). We calculate the mean of the log price series for each trading day and subtract the mean from the original series. This procedure leaves the intraday returns unaffected but eliminates the average daily level difference between the futures prices and the cash index level.[8] All error correction models are estimated using these de-meaned series.

One distinguishing feature of our dataset is its high frequency. However, increasing the frequency of observations will only increase the precision of the estimates when the frequency of events (transactions or quote changes) in the market is sufficiently high. A simple way to assess the frequency of events is to consider the fraction of zero returns. Table 1 shows these frequencies for the four return series under scrutiny. Zero returns for the DAX are observed in 5% of the return intervals. For the midquote returns this frequency is substantially lower, amounting to only 0.53%. These low values are not too surprising because a transaction or a quote change, respectively, will be observed whenever there is a transaction or a quote change in at least one of the 30 constituent stocks. Things look a bit differently for the futures market. Here, we observe zero returns in 21.1% of the case when we consider returns calculated from prices and in 16.7% of the cases when considering midquote returns. These figures, also being considerably higher than those for the DAX, are still low enough to suggest that the higher frequency of observations is warranted.

Insert Table 1 about here

Besides the frequency of zero returns Table 1 provides a variety of further descriptive statistics. The return standard deviation is higher in the futures market, and in both markets it is higher for the price returns than for the midquote returns. This is not surprising because price returns are affected by bid-ask bounce whereas midquote returns are not. All four series ex-

---

8 As noted previously, an alternative procedure would be to use discounted futures prices (as in Martens, Kofman and Vorst 1998). However, if futures prices deviate systematically from the values implied by the cost of carry relation (as is suggested by several empirical papers, including Bühler and Kempf (1995) for the

hibit negative skewness and excess kurtosis. Both characteristics are more pronounced in the cash market.

The DAX returns exhibit positive serial correlation ($\rho = 0.12$). This comes as no surprise given that the constituent stocks of the index trade infrequently and non-synchronously. What is a surprise, however, is the observation that the first order serial correlation of the midquote returns is even higher, amounting to 12.9%.[9] The pattern for the futures market is more in line with what one would expect. The returns calculated from prices are negatively correlated, most likely because of bid-ask bounce. The midquote returns are weakly positively correlated ($\rho = 0.04$).

The last line of Table 1 shows the average bid-ask spreads. These amount to 0.28% for the DAX but to only 0.03% for the DAX futures contract. These figures are consistent with results for the UK reported in Berkman, Brailsford and Frino (2005) and substantiate our earlier claim that transaction costs are lower in the DAX futures market.

Arbitrage requires to either sell in the cash market and buy in the futures market or to do the reverse. In both cases the transaction cost is the sum of the half-spread in the spot market and the half spread in the futures market. In passing, we note that this measure may overstate the true transaction costs for two reasons. First, arbitrageurs do not necessarily have to trade all 30 DAX stocks. They can instead trade a tracking portfolio consisting of fewer stocks (thereby, of course, introducing tracking error). As this portfolio is likely to be tilted towards liquid stocks, the average spread will be lower than the average spread of all DAX stocks.

---

the German market), this procedure will produce biased arbitrage signals. De-meaning, on the other hand, removes any systematic deviation of futures prices from the cost of carry relation.

9 This contrasts with the negative serial correlation at the individual stock level documented by Hasbrouck (1991) and others. A possible explanation for the positive serial correlation is that a quote change in one stock may trigger a quote change in other stocks. This would induce positive serial correlation in the returns of the midquote index. This correlation, then, would be a characteristic feature of the modus operandi of the spot market. We therefore did not attempt to remove the serial correlation by applying an ARMA filter to the data.

Second, there is a positive probability that the arbitrageur will be able to unwind his position early at a profit. The value of the early unwinding option (Brennan and Schwartz 1988, 1990) reduces the price differential necessary to make arbitrage profitable. Dwyer, Locke and Yu (1996, p. 312) suggest "that the trigger for index arbitrage is about one-half of the round-trip transaction costs". We will return to this issue in section 4.

Figure 1 shows boxplots of the transaction costs. We sample the transaction costs at hourly intervals, starting at 9.30 am and ending at 4.30 pm, resulting in 61 observations for each point in time. The differences between the boxplots are representative of the intraday pattern of our transaction cost measure. Apparently, transaction costs follow a J-shaped pattern. The individual boxplots provide evidence that there is also considerable variation in the transaction costs across trading days.

Insert Figure 1 about here

As a prerequisite for our empirical analysis we have to establish that the time series are I(1) and are cointegrated. Table 2 presents the results of augmented Dickey-Fuller tests and Phillips-Perron tests applied to the log of the levels and their first differences. Four time series are considered, the DAX index itself, the DAX midquote index and the prices and the quote midpoints of the DAX futures. The results of the stationarity tests clearly suggest that all series are I(1).

Results of Johansen tests (not shown) applied to pairs of log time series (DAX level and DAX futures prices, DAX midquote index and DAX futures midquotes) provide clear evidence that the time series are cointegrated.

Insert Table 2 about here

4 Methodology and Results

Having established that the time series are I(1) and cointegrated we can proceed by estimating the error correction model

$$
\begin{aligned}
r_t^X &= \alpha^X + \sum_{\tau=1}^{k} \beta_\tau^X r_{t-\tau}^X + \sum_{\tau=1}^{k} \gamma_\tau^X r_{t-\tau}^F + \delta^X \left(p_{t-1}^X - p_{t-1}^F \right) + \varepsilon_t^X \\
r_t^F &= \alpha^F + \sum_{\tau=1}^{k} \beta_\tau^F r_{t-\tau}^F + \sum_{\tau=1}^{k} \gamma_\tau^F r_{t-\tau}^X + \delta^F \left(p_{t-1}^X - p_{t-1}^F \right) + \varepsilon_t^F
\end{aligned} \tag{1}
$$

where *p* denotes a de-meaned log price series and *r* denotes a log return. The indices X and F identify observations and coefficients relating to the cash market (X, Xetra) and the futures market (F). We follow the literature (e.g., Dwyer, Locke and Yu 1996) by using a pre-specified cointegrating vector.

We estimate model (1) using OLS, for both prices and quote midpoints.[10]

The Schwarz information criterion suggests to include 16 lags in the price model and 12 lags in the quote midpoint model. We decided to include 20 lags in both models. This corresponds to 5 minutes.

Two approaches have been proposed to assess the contributions to price discovery.[11] Hasbrouck (1995) introduced the information share (IS). The information share relates the contribution of an individual market's innovation to the total innovation of the common efficient price by decomposing the variance of the error term. The information shares are not unique whenever the error terms in the two equations are correlated. A Cholesky factorization is used which arbitrarily attributes the covariance contribution to the market which is defined to be the first market in the system. This procedure thus maximizes the information share of the

---

10 Using both prices and midpoints allows us to check whether we can replicate the result obtained by Shyy et al. (1996), i.e., to check whether prices and quote midpoints yield different conclusions as to which market leads in the process of price discovery.

11 For a discussion of the relative merits of these two methods see Baillie et al. (2002), de Jong (2002), deB Harris et al. (2002), Hasbrouck (2002) and Lehman (2002).

first market and, consequently, minimizes the share of the second market. By permuting the order of the markets, upper and lower bounds for each market's information share are obtained.

The second measure of the contribution to price discovery is the common factor weight (CFW). It has first been proposed by Schwarz and Szacmary (1994) on intuitive grounds. A formal justification, based on the work of Gonzalo and Granger (1995), has been provided by Booth et al. (2002), deB Harris, McInish and Wood (2002) and Theissen (2002). The common factor weights are easily obtained from the coefficients on the error correction terms in (1):

$$CFW^X = \frac{\delta^F}{\delta^F - \delta^X}, CFW^F = \frac{-\delta^X}{\delta^F - \delta^X} \tag{2}$$

The results are presented in Table 3. To conserve space we only report coefficients for the first four lags. Considering the model estimated from transaction price data first, we note that the independent variables have considerable explanatory power for the cash market returns, as is evidenced by an adjusted R^2 of 0.18. They have much less explanatory power for the returns in the futures markets. The adjusted R^2 for the futures market equation is a mere 0.01. Returns in both markets depend negatively on their own lagged values. This may be due to bid-ask bounce. We further find that returns in both markets depend positively on lagged returns in the other market. The F statistic indicates bi-directional causality. A look at the values of the F statistics and at the coefficient values and their t statistics reveals, however, that the impact of lagged futures returns on the cash market is far stronger than the impact of cash market returns on the futures market.

In both equations the coefficient on the error correction term has the expected sign and is significant. Thus, both markets contribute to price discovery. Apparently, however, the futures market dominates the process of price discovery. The information share for the futures market is in the range from 85.12 % to 93.95% (lower and upper bound, respectively) as compared to a range from 6.05% to 14.88% for the cash market. The common factor weight is somewhat more favorable for the cash market, assigning it a 28.39% contribution, but the qualitative implication is the same. The futures market is the clear leader in the process of price discovery.

Insert Table 3 about here

The results obtained when estimating (1) with quote midpoint data are comparable. The R^2 for the cash market equation is higher at 0.23 whereas the R^2 for the futures market equation drops to 0.008. Midquote returns in the cash market depend negatively on their own lagged values. We do not observe a similar pattern for the futures market. Returns in both markets depend positively on lagged returns in the other market. Although the F statistic again indicates bi-directional causality it is obvious from the estimation results that the futures market dominates.

When proceeding to the measures of the contribution to price discovery, we note that both measures assign the cash market a slightly higher contribution than in the transaction price model. Still, both measures confirm that the futures market leads in the process of price discovery. This contrasts with the results of Shyy et al. (1996) who find that the cash market leads in the process of price discovery when the estimation is based on quote midpoints. When interpreting our results it should be kept in mind that the construction of our dataset puts the futures market at a disadvantage. Thus our results are likely to even understate the role of the futures market in the process of price discovery.

To check the robustness of our results we estimate model (1) for each day separately. A summary of the results is presented in Table 4.[12] They are very similar to those obtained for the

---

[12] In some cases the estimate of the coefficient on the error correction term in the futures market equation was negative. This implies that returns in the futures market to not adjust to deviations of price levels from their long-run equilibrium. In these cases the common factor weight as defined in equation (2) would assign a

pooled data set and clearly confirm the finding that the futures market leads in the process of price discovery.

Insert Table 4 about here

We use the daily estimates to test whether the contributions of the spot and the futures market to price discovery are different on days with positive versus negative index returns, and on days with high versus low volatility. We do not find any significant differences.

As noted previously, model (1) assumes that the speed of adjustment to deviations of the price levels from their long-run equilibrium relation is independent of the size of these deviations. This is unlikely to be the case, however, as arbitrageurs stand ready to take opportunity of any profits available. Thus, when the deviations are large enough to make arbitrage profitable (i.e., when they are larger than the transaction costs) we should expect faster adjustment.

In order to pursue this issue further we first have to define an arbitrage signal. Previous papers assumed that arbitrage will set in when the price deviation exceeds a constant threshold level. However, it is well known (and was documented in Figure 1) that transaction costs are time varying. In order to take advantage of profit opportunities, arbitrageurs have to trade fast. They are thus likely to use market orders and consequently have to pay the spread. An arbitrage trade consists of either selling shares at the bid in the cash market and buying the futures at the ask, or of selling futures at the bid and buying shares at the ask. In both cases, the total transaction cost is the half spread in the cash market plus the half spread in the futures market.

We assume that arbitrage is profitable when the price deviation exceeds this threshold. We thereby assume that there are no other relevant transaction costs besides the spread, and we assume that the position is either held until maturity or can be unwound at zero cost. This

negative weight to the cash market and a weight larger than 1 to the futures market. When calculating the average common factor weight we replaced these values with 0 and 1, respectively.

corresponds to the conjecture by Dwyer, Locke and Yu (1996, p. 312) that "the trigger for index arbitrage is about one-half of the round-trip transaction costs".

As both markets under scrutiny are fully automated, arbitrage trades may be executed as program trades. We therefore do not consider the possibility of delays between the occurrence of price deviations and the onset of arbitrage.[13] We thereby implicitly assume that the reaction time is no more than our data frequency, i.e., 15 seconds.

Table 5 takes a closer look at the arbitrage opportunities. Overall, the deviation between the (de-meaned) cash and futures market quote midpoints exceeds the transaction costs in about 5.46% of the cases. In 2.42% of the observations, the cash index is larger than the futures price whereas in 3.03% the reverse is true.[14] In most cases, the price deviation exceeds the transaction cost only by a small amount. The average value is 1.83 index points. Larger deviations do occur, however, as is evidenced by a maximum value of almost 19 points.

Insert Table 5 about here

We define a dummy variable D_t taking on the value 1 if there is an arbitrage opportunity as defined above and zero otherwise. We then augment model (1) to obtain

$$\begin{aligned} r_t^X &= \alpha^X + \sum_{\tau=1}^{k} \beta_\tau^X r_{t-\tau}^X + \sum_{\tau=1}^{k} \gamma_\tau^X r_{t-\tau}^F + \delta_1^X \left(p_{t-1}^X - p_{t-1}^F \right) + \delta_2^X D_{t-1} \left(p_{t-1}^X - p_{t-1}^F \right) + \varepsilon_t^X \\ r_t^F &= \alpha^F + \sum_{\tau=1}^{k} \beta_\tau^F r_{t-\tau}^F + \sum_{\tau=1}^{k} \gamma_\tau^F r_{t-\tau}^X + \delta_1^F \left(p_{t-1}^X - p_{t-1}^F \right) + \delta_2^F D_{t-1} \left(p_{t-1}^X - p_{t-1}^F \right) + \varepsilon_t^F \end{aligned} \tag{3}$$

---

13 In contrast, Dwyer, Locke and Yu (1996) use data from open outcry markets. In such an environment delays are likely. Dwyer, Locke and Yu (1996) address the issue empirically and estimate delays ranging from 1 minute to 5 minutes.

14 These figures are clearly lower than the corresponding values in Dwyer, Locke and Yu (1996, p. 324). They report that slightly less than 9% of their observations are in each of the two tail regimes that are associated with arbitrage opportunities.

The coefficients δ_2^X and δ_2^F measure whether the adjustment to price deviations is different in the presence of arbitrage opportunities. We expect these coefficients to have the same sign as δ_1^X and δ_1^F .

As already noted, arbitrage requires to either sell in the cash market and buy in the futures markets or to do the reverse. The price dynamics in the two cases may be different because selling in the cash market may require short sales. We therefore estimate an additional model in which we allow the coefficient on the error correction term to be different in the two cases alluded to above. The model is

$$\begin{aligned} r_t^X &= \alpha^X + \sum_{\tau=1}^{k} \beta_\tau^X r_{t-\tau}^X + \sum_{\tau=1}^{k} \gamma_\tau^X r_{t-\tau}^F \\ &\quad + \delta_1^X \left(p_{t-1}^X - p_{t-1}^F \right) + \delta_2^X D_{t-1}^1 \left(p_{t-1}^X - p_{t-1}^F \right) + \delta_3^X D_{t-1}^2 \left(p_{t-1}^X - p_{t-1}^F \right) + \varepsilon_t^X \\ r_t^F &= \alpha^F + \sum_{\tau=1}^{k} \beta_\tau^F r_{t-\tau}^F + \sum_{\tau=1}^{k} \gamma_\tau^F r_{t-\tau}^X \\ &\quad + \delta_1^F \left(p_{t-1}^X - p_{t-1}^F \right) + \delta_2^F D_{t-1}^1 \left(p_{t-1}^X - p_{t-1}^F \right) + \delta_3^F D_{t-1}^2 \left(p_{t-1}^X - p_{t-1}^F \right) + \varepsilon_t^F \end{aligned} \tag{4}$$

where D_t^1 and D_t^2 are dummy variables identifying those arbitrage opportunities that require selling in the cash market (D_t^1) and selling in the futures market (D_t^2).

The information shares are not properly defined for the augmented models. We can, however, construct suitable extensions of the common factor weights as follows:

$$CFW_2^X = \frac{\left(\delta_1^F + \delta_2^F\right)}{\left(\delta_1^F + \delta_2^F\right) - \left(\delta_1^X + \delta_2^X\right)}, CFW_2^F = \frac{-\left(\delta_1^X + \delta_2^X\right)}{\left(\delta_1^F + \delta_2^F\right) - \left(\delta_1^X + \delta_2^X\right)} \tag{5}$$

CFW_2^X and CFW_2^F measure the contribution to price discovery in the presence of arbitrage opportunities. Analogous to (5) we can also define CFW measures for the two "arbitrage regimes" in model (4).

We have argued earlier that the identification of arbitrage opportunities should be based on quote data rather than on transaction price data. Consequently, we estimate models (3) and (4) using quote midpoint data. To enhance comparability with our previous results we include 20 lages in both models although the Schwarz information criterion suggests to use less (14 for model (3) and 12 for model (4)).

The results are presented in Table 6. They are comparable to those shown in Table 3. The cash market returns depend negatively on their own lagged values and depend strongly and positively on lagged futures returns. Futures returns, on the other hand, depend positively on lagged cash market returns but depend on their own lagged values significantly only at lag 1. As before we find bi-directional causality, and as before we can conclude from the magnitude of the coefficient estimates and the test statistics that the dependence of the cash market on the futures market is much stronger than the reverse dependence. These results hold for model (3) as well as for model (4).

The estimates of the coefficient on the error correction term in the "no-arbitrage regime" have the same sign but are smaller in magnitude than those presented before. This is a plausible result as it suggests that prices adjust slower in the absence of arbitrage. Based on these estimates, the CFW measure attributes both markets almost equal contributions to price discovery (48.7% for the cash market and 51.3% for the futures market). It should be kept in mind, though, that we are likely to understate the contribution of the futures market. The coefficients CFW_2^X and CFW_2^F have the expected sign and are significant. When measuring the contributions to price discovery in the arbitrage regime using (5) we find that the share of the cash market drops to 36.4% whereas the share of the futures market rises to 63.6%. The results thus suggest that the leading role of the futures market in the price discovery process is particularly pronounced when price deviations are large (i.e., when arbitrage opportunities exist).

The estimates of the parameters $\delta_2^X, \delta_3^X, \delta_2^F$ and δ_3^F in model (4) have the expected sign and are significant. The result that the contribution of the futures market to the price discovery process is higher when price deviations are large is confirmed. Additionally, we observe that the share of the cash market is lowest when there are arbitrage opportunities and the cash market index is larger than the futures price. This is the case where arbitrage requires selling in the cash market.

Insert Table 6 about here

We check the robustness of the results by estimating model (3) for individual days. We can not do the same for model (4) because the number of observations in the two arbitrage regimes is very low on some days (see the figures shown in the last line of Table 5). The results, shown in Table 7, are fully consistent with our previous results.

Insert Table 7 about here

To summarize our results, we find that the futures market clearly dominates the price discovery process. Even so we constructed our sample such that the futures market is at a disadvantage, we find that returns in the cash market depend much more heavily on lagged returns in the futures market than vice versa. The measures of the contributions to price discovery also indicate that the futures market leads. We further find that the dynamics of the adjustment process is different when arbitrage opportunities exist. In these cases, the leading role of the futures market is even more pronounced.

5 Summary and Conclusion

In this paper we reconsider the issue of price discovery in spot and futures markets. Its contribution is threefold. First, we modify the threshold error correction model to allow for time-varying transaction costs. Second, we estimate a threshold error correction model using midquote data whereas previous papers used price data. Midquote data is conceptually superior because arbitrage signals should be based on tradable prices (i.e., bid and ask quotes) rather than on past transaction prices. Finally, we use data at a very high frequency (15 seconds as compared to 1 or 5 minutes in previous papers). This allows a more precise estimation of the contribution of the cash and the futures markets to the process of price discovery.

Our basic finding that the futures market leads in the process of price discovery is consistent with most previous results. We do not confirm the finding of Shyy et al. (1996) that the spot market leads when the estimation is based on quote midpoints rather than on transaction prices.

The lead of the futures market is more pronounced in the presence of arbitrage signals. Thus, when the price (or, more precisely, quote midpoint) deviation between the spot and the futures market is large, the spot market tends to adjust to the futures market.

Our results imply that the futures market generally impounds new information faster than the spot market. They also imply that market-wide information (which is likely to be reflected in the futures market first) is more important for returns at the index level than stock-specific information (which is likely to be reflected in the spot market first). As a consequence, researchers investigating into the market response to macroeconomic news, or into informational linkages between markets in different countries, should consider using futures market data rather than spot market data.

References

Baillie, R, G. Booth, Y. Tse and T. Zabotina (2002): Price Discovery and Common Factor Models. Journal of Financial Markets 5, 309-321.

Berkmann, H., T. Brailsford and A. Frino (2005): A Note on Execution Costs for Stock Index Futures: Information versus Liquidity Effects. Journal of Banking and Finance 29, 565-577.

Booth, G., J.-Ch. Lin, T. Martikainen and Y. Tse (2002): Trading and Pricing in Upstairs and Downstairs Stock Markets. Review of Financial Studies 15, 1111-1135.

Booth, G., R. So and Y. Tse (1999): Price Discovery in the German Equity Index Derivatives Markets. Journal of Futures Markets 19, 619-643.

Brennan, M. and E. Schwartz (1988): Optimal Arbitrage Strategies Under Basis Variability. Studies in Banking and Finance 5, 167–180.

Brennan, M. and E. Schwartz (1990): Arbitrage in Stock Index Futures. Journal of Business 63, S7–S31.

Bühler, W. and A. Kempf (1995): DAX Index Futures: Mispricing and Arbitrage in German Markets. Journal of Futures Markets 15, 833-859.

deB Harris, F., Th. McInish and R. Wood (2002): Security Price Adjustment Across Exchanges: An Investigation of Common Factor Components for Dow stocks. Journal of Financial Markets 5, 277-308.

deB Harris, F., Th. McInish and R. Wood (2002): Common Factor Components versus Information Shares: A Reply. Journal of Financial Markets 5, 341-348.

de Jong, F. (2002): Measures of Contributions to Price Discovery: A Comparison. Journal of Financial Markets 5, 323-327.

Deutsche Börse AG: Fact Book 1999.

Dwyer Jr., G., P. Locke and W. Yu (1996): Index Arbitrage and Nonlinear Dynamics Between the S&P 500 Futures and Cash. Review of Financial Studies 9, 301-332.

Fleming, J., B. Ostdiek and R. Whaley (1996): Trading Costs and the Relative Rates of Price Discovery in Stock, Futures and Option Markets. Journal of Futures Markets 16, 353-387.

Freihube, Th. and E. Theissen (2001): An Index Is an Index Is an Index? Schmalenbach Business Review 53, 295-320.

Frino, A. and M. McKenzie (2002): The Impact of Screen Trading on the Link Between Stock Index and Stock Index Futures Prices: Evidence from UK Markets. Working Paper.

Frino, A., T. Walter and A. West (2000): The Lead-Lag Relationship Between Equities and Stock Index Futures Markets Around Information Releases. Journal of Futures Markets 20, 467-487.

Gonzalo, J. and C. Granger (1995): Estimation of Common Long-Memory Components in Cointegrated Systems. Journal of Business & Economic Statistics 13, 27-35.

Grünbichler, A., F. Longstaff and E. Schwartz (1994): Electronic Screen Trading and the Transmission of Information: An Empirical Examination. Journal of Financial Intermediation 3, 166-187.

Hasbrouck, J. (1991): Measuring the Information Content of Stock Prices. Journal of Finance 46, 179-207.

Hasbrouck, J. (1995): One Security, Many Markets: Determining the Contributions to Price Discovery. Journal of Finance 50, 1175-1199.

Hasbrouck, J. (2002): Stalking the Efficient Price. Journal of Financial Markets 5, 329-339.

Jokivuolle, E. (1995): Measuring True Stock Index Value in the Presence of Infrequent Trading. Journal of Financial and Quantitative Analysis 30, 455-464.

Kempf, A. and O. Korn (1996): Preisführerschaft und imperfekte Arbitrage. Zeitschrift für Betriebswirtschaft 66, 837-859.

Kempf, A. and O. Korn (1998): Trading System and Market Integration. Journal of Financial Intermediation 7, 220-239.

Lehman, B. (2002): Some Desiderata for the Measurement of Price Discovery Across Markets. Journal of Financial Markets 5, 259-276.

Martens, M., P. Kofman and T. Vorst (1998): A Threshold Error Correction Model for Intraday Futures and Index Returns. Journal of Applied Econometrics 13, 245-263.

Pizzi, M., Economopoulos, A. and H. O'Neill (1998): An Estimation of the Relationship Between Stock Index Cash and Futures Markets: A Cointegration Approach. Journal of Futures Markets 18, 297-305.

Schwarz, Th. and A. Szakmary (1994): Price Discovery in Petroleum Markets: Arbitrage, Cointegration, and the Time Interval of Analysis. Journal of Futures Markets 14, 147-167.

Shyy, G., V. Vijayraghavan and B. Scott-Quinn (1996): A Further Investigation of the Lead-Lag Relationship Between the Cash Market and Stock Index Futures Market With the Use of Bid/Ask Quotes: The Case of France. Journal of Futures Markets 16, 405-420.

Stoll, H. and R. Whaley (1990): The Dynamics of Stock Index and Stock Index Futures Returns. Journal of Financial and Quantitative Analysis 25, 441-468.

Theissen, E. (2002): Price Discovery in Floor and Screen Trading Systems. Journal of Empirical Finance 9, 455-474.

Table 1: Descriptive Statistics

The table presents descriptive statistics for four return series: DAX returns, DAX midquote returns, DAX futures returns and DAX futures midquote returns. The returns are calculated over intervals of 15 seconds. The last line shows the average quoted bid-ask spread. For the cash market this is the value-weighted average of the spreads of the constituent stocks.

	DAX	MQDAX	FDAX	MQFDAX
Percentage of zero returns	5.00%	0.53%	21.05%	16.7%
Return standard deviation	0.000298	0.000223	0.000404	0.000340
Skewness	-0.0938	-0.9588	-0.1074	-0.1655
Kurtosis	25.62	27.07	6.32	7.65
First order serial correlation	0.120	0.129	-0.079	0.040
Average bid-ask spread	0.2846%		0.0292%	

Table 2: Stationarity Tests

The table presents the p-values from augmented Dickey Fuller tests and Phillips-Perron tests applied to both the levels and to the first differences of the time series.

	level		first difference	
	Augmented DF	Philipps / Perron	Augmented DF	Philipps / Perron
log(xdax)	0.349	0.412	0.000	0.000
log(mqdax)	0.401	0.519	0.000	0.000
log(fdax)	0.439	0.399	0.000	0.000
log(mqfdax)	0.370	0.396	0.000	0.000

Table 3: Error Correction Models - Pooled Data

The table presents the results of the error correction model

$$r_t^X = \alpha^X + \sum_{\tau=1}^{k} \beta_\tau^X r_{t-\tau}^X + \sum_{\tau=1}^{k} \gamma_\tau^X r_{t-\tau}^F + \delta^X \left(p_{t-1}^X - p_{t-1}^F \right) + \varepsilon_t^X$$

$$r_t^F = \alpha^F + \sum_{\tau=1}^{k} \beta_\tau^F r_{t-\tau}^F + \sum_{\tau=1}^{k} \gamma_\tau^F r_{t-\tau}^X + \delta^F \left(p_{t-1}^X - p_{t-1}^F \right) + \varepsilon_t^F$$

where p denotes a de-meaned log price series and r denotes a log return. The indices X and F identify observations and coefficients relating to the cash market (X, Xetra) and the futures market (F). We use a pre-specified cointegrating vector. The model is estimated by OLS with 20 lags, but only the coefficients for lags 1-4 are shown. We report the F-statistic for a test of the null hypothesis that the coefficients for the lagged returns of the other market (i.e., the cash market in the futures equation and vice versa) are jointly zero. The last lines report the measures of the contributions to price discovery. We report the common factor weights and lower and upper bounds for the information shares. The model is estimated based on prices (columns 1 and 2) and quote midpoints (columns 3 and 4).

	Transaction Prices		Quote Midpoints	
	XDAX	FDAX	XDAX	FDAX
Constant	-4.18 E-6 (-4.83)	-1.12 E-6 (-0.87)	-2.60 E-6 (-4.13)	-9.80 E-7 (-0.90)
EC	-0.0540 (-36.63)	0.0214 (9.77)	-0.0278 (-28.46)	0.0190 (11.25)
XDAX(-1)	-0.0104 (-3.11)	0.0651 (13.09)	-0.0744 (-22.93)	0.0443 (7.88)
XDAX(-2)	-0.0374 (-11.19)	0.0500 (10.06)	-0.0649 (-19.94)	0.0441 (7.84)
XDAX(-3)	-0.0362 (-10.85)	0.0446 (8.97)	-0.0515 (-15.80)	0.0486 (8.61)
XDAX(-4)	-0.0412 (-12.36)	0.0270 (5.45)	-0.0413 (-12.65)	0.0407 (7.22)
FDAX(-1)	0.1532 (60.22)	-0.0732 (-19.34)	0.1935 (94.65)	0.0487 (13.76)
FDAX(-2)	0.1264 (48.79)	-0.0311 (-8.08)	0.1413 (66.36)	-0.0046 (-1.25)
FDAX(-3)	0.1113 (42.55)	-0.0195 (-5.02)	0.1066 (49.07)	-0.0062 (-1.64)
FDAX(-4)	0.0884 (33.60)	-0.0078 (-1.99)	0.0849 (38.67)	0.0026 (0.68)
R^2	0.1807	0.0143	0.2281	0.0076
F statistic	244.50	16.72	604.28	10.66
Lags included	20		20	
IS - lower bound	0.0605	0.8512	0.1200	0.7671
IS - upper bound	0.1488	0.9395	0.2329	0.8800
CFW	0.2839	0.7161	0.4060	0.5939

Table 4: Error Correction Models - Daily Estimates

The table presents summary results of error correction models estimated for each day of the sample period separately. We report the mean of the coefficient estimates, the mean R^2 and the mean values of the common factor weights and the lower and upper bounds of the information share. Only the coefficients for lags 1 to 4 are reported. The model is estimated based on prices (columns 1 and 2) and quote midpoints (columns 3 and 4).

	Transaction Prices		Quote Midpoints	
	XDAX	FDAX	XDAX	FDAX
Constant	-5.18 E-6	-4.36 E-7	-3.12 E-6	5.74 E-9
EC	-0.0764	0.0247	-0.0384	0.0239
XDAX(-1)	0.0037	0.0645	-0.0700	0.0441
XDAX(-2)	-0.0261	0.0512	-0.0590	0.0420
XDAX(-3)	-0.0222	0.0436	-0.0480	0.0423
XDAX(-4)	-0.0271	0.0247	-0.0391	0.0398
FDAX(-1)	0.1291	-0.0775	0.1756	0.0470
FDAX(-2)	01084	-0.0323	0.1301	-0.0009
FDAX(-3)	0.0946	-0.0166	0.0977	-0.0013
FDAX(-4)	0.0721	-0.0117	0.0755	0.0056
R^2	0.1848	0.0207	0.2304	0.0153
Lags included	20		20	
IS - lower bound	0.0696	0.8564	0.1390	0.7657
IS - upper bound	0.1436	0.9304	0.2343	0.8610
CFW	0.2376	0.7624	0.3633	0.6367

Table 5: Arbitrage Opportunities

An arbitrage signal, in our definition, occurs when the absolute difference between the de-meaned cash and futures prices is larger than the transaction cost (the sum of the half-spread in the cash market and the half-spread in the futures market). The table shows the number of arbitrage opportunities, the mean and median arbitrage profit and the maximum profit. Profits are measured in index points. The last line shows the lowest number of arbitrage opportunities observed on any individual day of the sample period. Columns 1 and 2 show separate figures for arbitrage opportunities where the cash index value is larger [smaller] than the futures price.

	MQDAX>MQFDAX	MQFDAX>MQDAX	Both
number of cases	2,658 2.42%	3,331 3.03%	5,989 5.46%
mean arbitrage profit	1.4788	2.1086	1.8291
median arbitrage profit	1.0751	1.2503	1.1559
maximum arbitrage profit	16.9659	18.9944	18.9944
lowest daily number of observations	1	1	9

Table 6: TECM - Pooled Data

The table presents the results of the error correction models

$$r_t^X = \alpha^X + \sum_{\tau=1}^{k} \beta_\tau^X r_{t-\tau}^X + \sum_{\tau=1}^{k} \gamma_\tau^X r_{t-\tau}^F + \delta_1^X \left(p_{t-1}^X - p_{t-1}^F\right) + \delta_2^X D_{t-1} \left(p_{t-1}^X - p_{t-1}^F\right) + \varepsilon_t^X$$

$$r_t^F = \alpha^F + \sum_{\tau=1}^{k} \beta_\tau^F r_{t-\tau}^F + \sum_{\tau=1}^{k} \gamma_\tau^F r_{t-\tau}^X + \delta_1^F \left(p_{t-1}^X - p_{t-1}^F\right) + \delta_2^F D_{t-1} \left(p_{t-1}^X - p_{t-1}^F\right) + \varepsilon_t^F$$

(columns 1 and 2) and

$$r_t^X = \alpha^X + \sum_{\tau=1}^{k} \beta_\tau^X r_{t-\tau}^X + \sum_{\tau=1}^{k} \gamma_\tau^X r_{t-\tau}^F + \delta_1^X \left(p_{t-1}^X - p_{t-1}^F\right) + \delta_2^X D_{t-1}^1 \left(p_{t-1}^X - p_{t-1}^F\right) + \delta_3^X D_{t-1}^2 \left(p_{t-1}^X - p_{t-1}^F\right) + \varepsilon_t^X$$

$$r_t^F = \alpha^F + \sum_{\tau=1}^{k} \beta_\tau^F r_{t-\tau}^F + \sum_{\tau=1}^{k} \gamma_\tau^F r_{t-\tau}^X + \delta_1^F \left(p_{t-1}^X - p_{t-1}^F\right) + \delta_2^F D_{t-1}^1 \left(p_{t-1}^X - p_{t-1}^F\right) + \delta_3^F D_{t-1}^2 \left(p_{t-1}^X - p_{t-1}^F\right) + \varepsilon_t^F$$

(columns 3 and 4). *p* denotes a de-meaned log price series and *r* denotes a log return. The indices X and F identify observations and coefficients relating to the cash market (X) and the futures market (F). We use a pre-specified cointegrating vector. The dummy variable D_t identifies all arbitrage signals. The dummy variables D_t^1 [D_t^2] identify those arbitrage signals where the cash market midquote index is larger [smaller] than the midquote in the futures market. The models are estimated by OLS with 20 lags, but only the coefficients for lags 1-4 are shown. We report the F-statistic for a test of the null hypothesis that the coefficients for the lagged returns of the other market (i.e., the cash market in the futures equation and vice versa) are jointly zero. The last line reports the common factor weights

	Arbitrage signals pooled		Separate arbitrage signals	
	XDAX	FDAX	XDAX	FDAX
Constant	-2.77 E-6 (-4.43)	-8.96 E-7 (-0.82)	5.15 E-7 (0.81)	-1.71 E-6 (1.54)
EC / no arbitarge	-0.0119 (-10.91)	0.0113 (5.96)	-0.0131 (-12.05)	0.0116 (6.11)
EC / arbitrage	-0.0511 (-32.55)	0.0248 (9.09)		
EC / arb. X-F			-0.0923 (-39.60)	0.0350 (8.62)
EC / arb. F-X			-0.0265 (-14.17)	0.0187 (5.73)
XDAX(-1)	-0.0764 (-23.66)	0.0452 (8.06)	-0.0748 (-23.23)	0.0448 (7.99)
XDAX(-2)	-0.0661 (-20.43)	0.0447 (7.94)	-0.0648 (-20.09)	0.0444 (7.89)
XDAX(-3)	-0.0527 (-16.27)	0.0492 (8.73)	-0.0514 (-15.89)	0.0488 (8.66)
XDAX(-4)	-0.0423 (-13.03)	0.0412 (7.31)	-0.0408 (-12.61)	0.0409 (7.25)
FDAX(-1)	0.1894 (92.96)	0.0507 (14.30)	0.1850 (90.68)	0.0518 (14.55)
FDAX(-2)	0.1413 (66.69)	-0.0046 (-1.24)	0.1387 (65.56)	-0.0039 (-1.06)
FDAX(-3)	0.1079 (49.89)	-0.0068 (-1.80)	0.1057 (48.99)	-0.0062 (-1.66)
FDAX(-4)	0.0867 (39.70)	0.0017 (0.45)	0.0850 (38.96)	0.0022 (0.58)
R^2	0.2362	0.0084	0.2405	0.0085
F statistic	594.72	10.97	567.29	10.79
Lags included	20		20	
CFW / no arbitrage	0.4871	0.5129	0.4693	0.5307
CFW / arbitrage	0.3643	0.6357		
CFW / arb. X-F			0.3065	0.6934
CFW / arb. F-X			0.4332	0.5668

Table 7: TECM - Daily Estimates

The table presents summary results of error correction model

$$r_t^X = \alpha^X + \sum_{\tau=1}^{k} \beta_\tau^X r_{t-\tau}^X + \sum_{\tau=1}^{k} \gamma_\tau^X r_{t-\tau}^F + \delta_1^X \left(p_{t-1}^X - p_{t-1}^F \right) + \delta_2^X D_{t-1} \left(p_{t-1}^X - p_{t-1}^F \right) + \varepsilon_t^X$$
$$r_t^F = \alpha^F + \sum_{\tau=1}^{k} \beta_\tau^F r_{t-\tau}^F + \sum_{\tau=1}^{k} \gamma_\tau^F r_{t-\tau}^X + \delta_1^F \left(p_{t-1}^X - p_{t-1}^F \right) + \delta_2^F D_{t-1} \left(p_{t-1}^X - p_{t-1}^F \right) + \varepsilon_t^F$$

estimated for each day of the sample period separately. We report the mean of the coefficient estimates, the mean R^2 and the mean values of the common factor weights. Only the coefficients for lags 1 to 4 are reported. The model is estimated based on quote midpoints. The last line presents the t-statistic for a test of the null hypothesis that the common factor weights in the arbitrage regime and the no arbitrage regimes are equal. Only one t-statistic is given because the common factor weights for the two markets sum to one and are thus linearly dependent.

	XDAX	FDAX
Constant	-3.17 E-6	-6.65 E-7
EC / no arbitrage	-0.0243	0.0168
EC / arbitrage	-0.0825	0.0361
XDAX(-1)	-0.0686	0.0434
XDAX(-2)	-0.0573	0.0411
XDAX(-3)	-0.0461	0.0415
XDAX(-4)	-0.0376	0.0390
FDAX(-1)	0.1637	0.0520
FDAX(-2)	0.1250	0.0013
FDAX(-3)	0.0951	-0.0002
FDAX(-4)	0.0739	0.0059
R^2	0.2475	0.0169
Lags included	20	
CFW / no arbitrage	0.4696	0.5304
CFW / arbitrage	0.3409	0.6591
t-statistic	3.05	

Figure 1: Distribution of transaction costs at hourly intervals