Trading intensity, volatility, and arbitrage activity

Nicholas Taylor

Department of Accounting and Finance, Cardiff Business School, Cardiff University, Cardiff CF10 3EU, United Kingdom

Abstract

The objective of this paper is to uncover the determinants of trading intensity in futures markets. In particular, the time between adjacent transactions (referred to as *transaction duration*) on the FTSE 100 index futures market is modeled using various augmentations of the basic autoregressive conditional duration (ACD) model introduced by Engle and Russell [Econometrica 66 (1998) 1127]. The definition of transaction duration used in this paper is an important variable as it represents the inverse of instantaneous conditional return volatility. As such, this paper can also be viewed as an investigation into the determinants of (the inverse of) instantaneous conditional return volatility. The estimated parameters from various ACD models form the basis of the hypothesis tests carried out in the paper. As predicted by various market microstructure theories, we find that bid–ask spread and transaction volume have a significant impact upon subsequent trading intensity. However, the major innovation of this paper is the finding that large (small) differences between the market price and the theoretical price of the futures contract (referred to as *pricing error*) lead to high (low) levels of trading intensity in the subsequent period. Moreover, the functional dependence between pricing error and transaction duration appears to be non-linear in nature. Such dependence is implied by the presence of arbitragers facing non-zero transaction costs. Finally, a comparison of the forecasting ability of the various estimated models shows that a threshold ACD model provides the best out-of-sample performance.
© 2003 Elsevier B.V. All rights reserved.

E-mail address: taylorn@cardiff.ac.uk (N. Taylor).

1. Introduction

Accurate forecasts of the conditional volatility of asset returns is of extreme importance in the areas of risk management and option pricing. In the former, the concept of Value-at-Risk (VaR) has become a widely used method for measuring the market risk of portfolios. It is defined as the tolerable amount of capital that can be lost in the next period for a given predetermined probability. Clearly an accurate forecast of volatility is important here if the solvency of the portfolio owner is to be maintained. Likewise, as the price of an option contract is a function of the volatility of the underlying asset returns, a similar degree of accuracy is required by buyers and sellers of such contracts. Given this importance it is not surprising that a vast literature exists concerning appropriate ways to model conditional volatility.

The conventional approach to modeling conditional volatility is almost always based on the autoregressive conditional heteroscedasticity (ARCH) model introduced by Engle (1982), or one of many generalizations of this model (see Bollerslev (1986) and Nelson (1991), for examples within this vast literature). A common feature of these models is that conditional volatility in the next period is a function of current and previous conditional volatility and/or the square of unexpected returns. In using this form of temporal dependence one necessarily imposes restrictions concerning the sampling frequency of the data. This inevitably leads to ad hoc frequency selection and thus information loss. For example, return volatility measured at hourly intervals could imply zero return volatility even though returns within the interval are highly volatile. At the other extreme, selection of too high a frequency may result in many intervals with no new information and, hence, may induce various forms of heteroscedasticity into the data. To avoid these problems the current paper makes use of various duration models capable of modeling (the inverse of) instantaneous conditional return volatility.

Duration models focus on the times between events and, therefore, do not impose any sampling frequency assumptions. In the current context the event is defined as a *non-zero price impact trade* on the FTSE 100 index futures market with the times between these events being referred to as *transaction duration*.[1] As is shown explicitly in Engle and Russell (1998), there exists an inverse relationship between the conditional expectation of this type of transaction duration and instantaneous conditional return volatility. This follows from basic intuition whereby if transaction duration is expected to be low (high) then it follows that prices are expected to change (in absolute terms) rapidly (slowly). Therefore, by using this particular definition of transaction duration one is able to interpret the models used as models of (the inverse of) instantaneous conditional return volatility.

---

[1] A non-zero price impact trade is defined as a trade which contains a price that is different from the price observed in the previous trade.

The exact form of the model used is based on the autoregressive conditional duration (ACD) model of Engle and Russell (1998). [2] In particular, expected transaction duration is assumed to be a linear function of past transaction duration, past expected transaction duration and a set of explanatory variables thought to have an impact on expected transaction duration as predicted by various market microstructure theories.

Apart from being the first application of ACD models to futures trading, the paper innovates in two other ways: First, the set of explanatory variables is expanded to include past mispricing between spot and futures prices (referred to as *pricing error*) as implied by the cost-of-carry model. Second, the assumption of linear dependence between expected transaction duration and past pricing error is relaxed. In particular, the functional dependence between expected transaction duration and past pricing error appears to be non-linear is nature – a result most likely caused by arbitrage activity. In allowing for non-linear dependence, we introduce a new class of ACD model similar in nature to the threshold ACD (TACD) model of Zhang et al. (2001). However, the new model differs in that the threshold dependence occurs between expected transaction duration and past pricing error and not between expected transaction duration and past transaction duration as in the TACD model. A further related innovation is that we consider a smooth threshold version of the ACD model with the non-linear part of the model applied to past pricing error. This latter assumption is compatible with the observation that arbitragers face heterogeneous transaction costs. Of all the models considered, it is the non-linear threshold models that appear to exhibit the best out-of-sample performance.

The paper is organized as follows: the next section contains a description of the ACD model introduced by Engle and Russell (1998) and motivates and describes the augmented version of the ACD model considered in this paper. Section 3 contains the empirical results of the paper and includes a description of the data used, some summary statistics, results pertaining to various estimated ACD models, and a description of the forecasting ability of the various models. Concluding remarks are provided in the final section. An appendix is also provided containing summary information concerning details of all variables and parameters used, the models considered, the specifications of the transition functions used, and all distributions considered.

2. Modeling duration

Before commencing with a description of the various ACD models considered in this paper, it is worth discussing the choice of whether to model trading intensity, that is, the number of trades per period of time or transaction duration, that is, the time between trades. In this paper we choose to model the latter. There are two reasons for

---

[2] See Bauwens et al. (2000), Drost and Werker (2000), Fernandes and Grammig (2000), Gerhard and Hautsch (2000), Hafner (2000), and Zhang et al. (2001) for various applications and modifications of the ACD model of Engle and Russell (1998).

this choice: First, time-series models of intensity are restrictive in terms of allowable specification. As the observations are integer-valued then one has to ensure that the fitted dependent variables are also integer-valued. In the case of integer-valued ARMA models (see MacDonald and Zucchini (1997) for a review of these models), this is achieved by using the binomial thinning operator of Steutal and van Harn (1979). However, in using this operator the likelihood functions become extremely cumbersome making inclusion of explanatory variables almost impossible. Second, as with the Gaussian ARMA model, there remains the problem of selecting the appropriate sampling frequency. For these reasons we model transaction duration using the ACD model of Engle and Russell (1998). This model is highly flexible in terms of specification and does not rely on any sampling frequency assumptions.

2.1. The basic ACD model

The motivation for the ACD model is the observation that trades appear to be clustered over time. While this clustering can occur even when trades occur randomly over time, the degree of clustering appears to be excessive. For this reason Engle and Russell (1998) introduce a time-series model of transaction duration similar in nature to the ARCH specification used in the context of Gaussian-type data. [3] A major difference between ACD and ARCH-type models is that transaction duration cannot be negatively valued. As such, alternative distributional assumptions are required.

The specification of the ACD model relies on three main assumptions, the first of which states that

$$Y_n/\psi_n = \epsilon_n \sim \text{i.i.d. } \mathscr{D}(\boldsymbol{\kappa}), \tag{1}$$

where Y_n is the transaction duration associated with the nth trade defined as $\tau_n - \tau_{n-1}$ with price impact trades occurring at times $\{\tau_0, \tau_1, \ldots, \tau_n, \ldots\}$, $\psi_n \equiv \mathsf{E}[Y_n|Y_{n-1}, \ldots, Y_1, \psi_{n-1}, \ldots, \psi_1]$ is the conditional expectation of the transaction duration for the nth trade given past realized and expected transaction durations and $\mathscr{D}$ is a general distribution over $(0, \infty)$ with mean equal to one and parameter vector $\boldsymbol{\kappa}$. The second assumption made is that ψ_n is a linear function of past Y_n and past ψ_n,

$$\psi_n = \alpha_0 + \sum_{m=1}^{p} \alpha_m Y_{n-m} + \sum_{m=1}^{q} \beta_m \psi_{n-m}, \tag{2}$$

where up to p and q lags of Y_n and ψ_n are allowed. The final assumption concerns the distribution of ϵ_n. As negative transaction durations cannot occur then distributions with support over $(0, \infty)$ must be used. Of the many distributions available, the exponential, the Weibull, and the generalized gamma densities have be used in this context. [4]

---

[3] At first glance, this particular specification may appear unusual. It may be more obvious to specify an ARMA model of transaction duration. However, just as an ARCH model can be represented as an ARMA model so can the ACD model; see Eq. (17) in Engle and Russell (1998) for the specification of an ARMA model using the coefficients from an ACD model.

[4] Engle and Russell (1998) make use of the exponential and Weibull distribution while Zhang et al. (2001) use a generalized gamma distribution.

One of the advantages of using an ARCH-type specification for transaction duration is that explanatory variables (possibility real-valued in nature) can be introduced into (2) without substantially complicating the estimation process. Therefore, in allowing such variables, (2) becomes

$$\psi_n = \alpha_0 + \sum_{m=1}^{p} \alpha_m Y_{n-m} + \sum_{m=1}^{q} \beta_m \psi_{n-m} + \boldsymbol{\theta}' \mathbf{X}_{n-1}, \tag{3}$$

where $\mathbf{X}_{n-1}$ is a vector of explanatory variables with associated parameter vector $\boldsymbol{\theta}$. [5] Two innovations of this paper are to allow past pricing error into (3) and to allow this variable to be non-linearly related to ψ_n via various parametric transition functions. The economic arguments concerning these innovations are described in the following subsection.

2.2. The augmented ACD model

The first innovation of this paper involves the introduction of past pricing error into the ACD model given by (3). The motivation for the inclusion of this variable is based on previous studies of the relationship between futures (and spot) returns and past pricing error (see Dwyer et al., 1996; Martens et al., 1998, and Taylor et al., 2000 for examples). In particular, one would expect the trading intensity of arbitragers to be high (low) when past pricing error is large (small) in absolute terms. This is because the larger (smaller) the profit opportunity (as measured by the absolute value of the pricing error) the larger (smaller) the intensity of trades will be such that prices revert back to their equilibrium levels. In terms of transaction duration, one would expect to find a negative relationship between transaction duration and past pricing error.

As with previous studies, the pricing error considered in this paper is defined as the difference between the market price of the futures contract and the theoretical price implied by the cost-of-carry model. This model specifies a contemporaneous relationship between spot and forward prices. [6] In the absence of arbitrage opportunities and transaction costs, we have

$$F_t^* = S_t e^{r(M-t)} - \sum_{h=1}^{H} D_h e^{r(M-\tau_h)}, \tag{4}$$

where $t = \{t \in \mathbb{Z}^+ : 1 \leqslant t \leqslant T\}$, F_t^* is the theoretical (or fair) stock index futures price observed at (non-stochastic) time t for delivery at time M (referred to as a t-type event), S_t is the level of the index, r is the risk-free continuous interest rate applicable

---

[5] Note that we have imposed the restriction that only the first lagged value of the explanatory variable enters (3). This is purely to economize on the degree of notation used and is relaxed in the empirical section of the paper.

[6] This model is capable of describing the relationship between spot and futures prices providing that the term structure of interest rates is flat and constant.

over the contract life, $(M - t)$ is the time to maturity of the futures contract and D_h is the expected cash dividend paid at time τ_h where $t < \tau_h \leqslant M$. [7]

The trading decision rule assumed in this paper is based on the cost-of-carry model given by (4). Arbitragers are assumed to regularly monitor the market (at a frequency sufficient to avoid loss of information) during each of s time periods (referred to as an s-type event), where $s = \{s \in \mathbb{R}^+ : 1 \leqslant s \leqslant T\}$. In particular, arbitragers compare the regularly observed F_t^* revealed via the spot market with the irregularly observed market value of the futures contract revealed by market makers in the futures market. As these market makers will post bid and ask quotes around this price we assume that the market price lies exactly halfway between these quotes. [8] Letting the bid and ask quotes of the futures contract (referred to as i-type and j-type events, respectively), be denoted $F_{\tau_i}^B$ and $F_{\tau_j}^A$, where $\tau_i = \{\tau_i \in \mathbb{R}^+ : 1 \leqslant \tau_i \leqslant T\}$, $\tau_j = \{\tau_j \in \mathbb{R}^+ : 1 \leqslant \tau_j \leqslant T\}$, $i = \{i \in \mathbb{Z}^+ : 1 \leqslant i \leqslant I\}$, and $j = \{j \in \mathbb{Z}^+ : 1 \leqslant j \leqslant J\}$, it follows that

$$\widetilde{F}_{N_s(s)} = \frac{F_{N_i(s)}^A + F_{N_j(s)}^B}{2}, \tag{5}$$

where $N_s(s)$, $N_i(s)$ and $N_j(s)$ are the number of s-type, i-type, and j-type events that have occurred at time s, and $\widetilde{F}_{N_s(s)}$ denotes the market price of the futures contract as revealed by market makers in the futures market. Using this information, arbitragers construct the pricing error,

$$Z_{N_s(s)} = \ln\left(\frac{F_{N_s(s)}^*}{\widetilde{F}_{N_s(s)}}\right), \tag{6}$$

and trade accordingly. In particular, if $Z_{N_s(s)}$, is sufficiently positive (negative) then arbitragers should buy (sell) the futures contract and sell (buy) the index. This will imply a negative relationship between transaction duration and past absolute values of $Z_{N_s(s)}$.

Having defined the trading rule used by arbitragers, the next stage of the analysis involves building a model of arbitrage activity based on this rule. The conventional method of modeling this activity is via the price process observed in the spot and futures market. This ultimately leads to the imposition of a fixed sampling interval. In doing this, important information may be discarded as informational events within the sampling frequency are ignored. Likewise, selection of too high a frequency will induce severe forms of heteroscedasticity into the data. To avoid these problems, this paper models arbitrage behaviour in the context of an ACD model. Hence, ignoring other explanatory variables for the moment, (3) becomes

---

[7] Note that F_t^* is observed at non-stochastic time t because S_t is calculated every minute during the trading day.

[8] Under the new Connect trading system currently operating on LIFFE, a similar assumption can be made where the market price is assumed to lie exactly halfway between the prices specified in the best bid and ask orders currently sitting on the limit-order book. This assumption would be perfectly reasonable given the fact that the majority of orders placed on Connect are placed by dealers trading on their own account, i.e., marker makers.

$$\psi_n = \alpha_0 + \sum_{m=1}^{p} \alpha_m Y_{n-m} + \sum_{m=1}^{q} \beta_m \psi_{n-m} + \omega_1 |Z_{N_s(\tau_{n-1})}|, \tag{7}$$

where $Z_{N_s(\tau_{n-1})}$ represents the last available observation of pricing error at the time of the $(n-1)$th trade. [9]

Thus far we have assumed a linear relationship between expected transaction duration and the absolute value of past pricing error. In making this assumption we are explicitly assuming that transaction costs equal zero. This is clearly an unrealistic assumption. If we relax this assumption then, in the presence of proportional transaction costs, c, and with no short-selling restrictions, arbitrage activity will not take place when the following conditions holds:

$$|Z_{N_s(\tau_{n-1})}| \leqslant c, \tag{8}$$

where c equals the sum of (i) round-trip spot and futures trading costs; (ii) market impact costs from trading in the spot and futures markets and (iii) 'stamp tax' of 0.5% which is charged when investors purchase UK equities. [10] If the no short-selling restriction assumption is relaxed then the no arbitrage activity condition becomes

$$-c_1 \leqslant Z_{N_s(\tau_{n-1})} \leqslant c_2, \tag{9}$$

where c_2 is expected to be greater than c_1 if short-selling restrictions are imposed on trading the underlying asset.

If the condition given by (8) (or (9)) holds then one would not expect past pricing error to have any impact upon transaction duration. However, if it does not hold then one would expect a dramatic decrease in transaction duration when one moves from the condition holding to not holding (referred to as a *regime switch*). Indeed, given that transaction duration cannot be negative, this decrease may be sufficiently large such that changes in pricing error within the 'profit opportunity regime' have no impact upon subsequent transaction duration. These arguments imply that the constant in (7) should be allowed to take different values depending on whether (8) (or (9)) holds or not. To allow for testing of the hypothesis that past pricing error has no impact within the two regimes, the coefficient on past pricing error is also allowed to vary across the regimes. Thus (7) is augmented as follows:

$$\psi_n = (1 - F_i(Z_{N_s(\tau_{n-1})}))(\alpha_0 + \omega_1 |Z_{N_s(\tau_{n-1})}|) + F_i(Z_{N_s(\tau_{n-1})})(\alpha_{0,1} + \omega_{1,1}|Z_{N_s(\tau_{n-1})}|)$$
$$+ \sum_{m=1}^{p} \alpha_m Y_{n-m} + \sum_{m=1}^{q} \beta_m \psi_{n-m}, \tag{10}$$

---

[9] We have assumed that only the first lagged value of pricing error is allowed to enter (7). This is for ease of notation only and is relaxed in the empirical section of the paper.

[10] Arbitragers, however, can, and do, unwind their spot and futures positions before maturity (Sofianos, 1993; Neal, 1996). This is because arbitragers will close out their positions when it is profitable to do so rather than at the maturity of the futures contract – unless of course it is profitable to wait for the contract to mature. Brennan and Schwartz (1988, 1990) model this behaviour as arbitragers having an option and this ultimately leads to a lowering in the transaction cost bound. Indeed, Dwyer et al. (1996) argue that c represents approximately one half the total round-trip transaction costs incurred by arbitragers.

where

$$F_1(Z_{N_s(\tau_{n-1})}) = \begin{cases} 0 & \text{if } |Z_{N_s(\tau_{n-1})}| \leqslant c, \\ 1 & \text{otherwise}, \end{cases} \tag{11}$$

in the case of the no short-selling restriction model, and

$$F_2(Z_{N_s(\tau_{n-1})}) = \begin{cases} 0 & \text{if } -c_1 \leqslant Z_{N_s(\tau_{n-1})} \leqslant c_2, \\ 1 & \text{otherwise}, \end{cases} \tag{12}$$

in the case of the short-selling restriction model. In both cases $F_i(Z_{N_s(\tau_{n-1})})$ is referred to as the *transition function* with c, c_1, and c_2 being parameters to be estimated. To distinguish between the two models, the former will henceforth be referred to as the *symmetric threshold model* and the latter will be referred to as the *asymmetric threshold model*.

The transition functions in (11) and (12) are indicator functions taking values of zero and one depending on the value of $Z_{N_s(\tau_{n-1})}$. It could be argued that the transition from one regime to another is actually smooth in nature, thereby allowing an infinite number of regimes. The economic rationale for this smoothness is that arbitragers face heterogeneous transaction costs. For instance, arbitragers facing low transaction costs are likely to trade even for small (absolute) values of the pricing error. By contrast, arbitragers facing large transaction costs will only trade when the pricing error is large in absolute terms. These arguments imply that the transition function is continuous with support on [0,1]. A common assumption made in this context is that the function has the following exponential specification: [11]

$$F_3(Z_{N_s(\tau_{n-1})}) = 1 - \exp\left[-\delta Z_{N_s(\tau_{n-1})}^2\right], \tag{13}$$

where $\delta > 0$ and measures the speed of transition from no impact ($F_3(\cdot) = 0$) to full impact ($F_3(\cdot) = 1$). This model will henceforth be referred to as the *smooth threshold model*.

3. Empirical results

This section contains a description of the data used, various summary statistics associated with the variables considered, and an examination of various estimated ACD models in terms of both in-sample and out-of-sample performance.

3.1. Data

We make use of various pieces of information concerning trades in FTSE 100 futures contracts traded on LIFFE. In addition to the pricing error series, we require the time of the trade, the price of the trade, the bid–ask quote available at the time of the trade, and the number of contracts traded. This information is collected for every trade in the nearest FTSE 100 futures contract carried out between January 5 and

---

[11] See Terasvirta (1994) for more details.

April 24, 1998. These data were obtained from LIFFE. [12] To be able to construct the pricing error we require the (spot) level of the FTSE 100 index over this sample period. These data were obtained from FTSE International. The trading hours of the futures market and the spot market are, 8.35 am to 4.10 pm and 8.00 am to 4.30 pm, respectively. [13] Thus one can obtain overlapping futures and spot data over the period, 8.35 am to 4.10 pm.

The validity of the constructed pricing error series relies heavily on the use of appropriate ex ante dividends and interest rates. To this end we make use of data supplied by Goldman Sachs. These data are used by arbitragers employed by Goldman Sachs when making judgements about the mispricing (or otherwise) of FTSE 100 futures contracts. Goldman Sachs construct ex ante dividends by making individual forecasts for each of the dividends paid by companies in the FTSE 100 index and then weight these by market capitalization. The interest rate applicable over the contract life used by Goldman Sachs is the interpolated LIBOR rate. For instance, if a 25 day interest rate is required then Goldman Sachs interpolate between the two week and the four week rates.

3.2. Summary statistics

The focal point of this paper concerns the determinants of transaction duration. Moreover, we concentrate on trades which have price impact. There are two reasons for this. First, as described in Engle and Russell (1998), there exists an explicit relationship between the inter-arrival times of such trades, i.e., transaction duration, and (the inverse of) instantaneous conditional return volatility. As such, the models considered can be interpreted as models of (the inverse of) instantaneous conditional return volatility. Second, by considering such trades we are able to draw inferences on the speed of price adjustment in the futures market. Given this focus we require a specific definition of a non-zero price impact trade. The problem here is that the choice of what constitutes a trade with 'price impact' is somewhat arbitrary. [14] We make the simple assumption that a non-zero price impact trade is defined as a trade containing a price that is different from the price contained in the previous trade. This amounts to including trades with absolute price changes greater than or equal to 0.25 index points and results in 99,399 observations.

The second issue to be addressed prior to the analysis commencing concerns the treatment of intraday periodicity in the data. Many of the variables used, including

---

[12] The contract is changed when the volume of trading in the next nearest contract is greater than the volume of trading in the nearest contract. The volume cross-over method of changing futures contracts results in one change. The change involves a switch from the March 1998 contract to the June 1998 contract on March 11, 1998. On this day the volume of trading in the March contract was 6312 contracts and the volume of trading in the June contract was 13,355 contracts.

[13] The futures market re-opens at 4.32 pm under the automated pit trading (APT) system. However, this additional period of trading is not considered because of the lack of data between 4.11 pm and 4.31 pm.

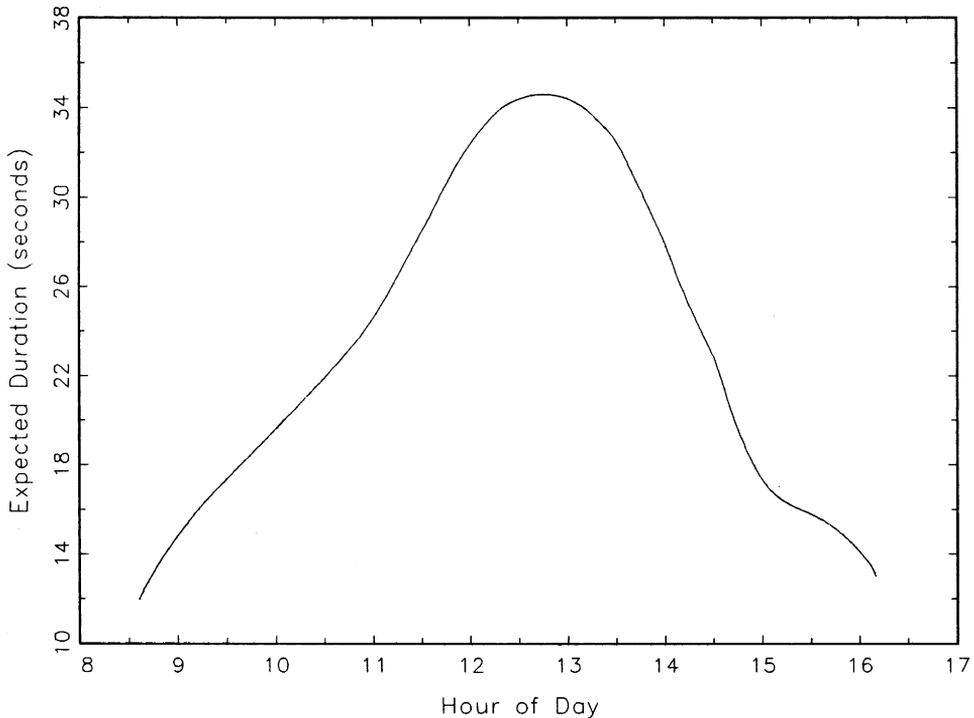[14] This choice is equivalent to selecting a value for the constant, *a*, in Eq. (32) in Engle and Russell (1998).

Fig. 1. A non-parametric estimate of periodicity in transaction duration.

transaction duration, exhibit a strong intraday periodic component. [15] This is exemplified in Fig. 1 which shows a non-parametric estimate of the intraday periodicity of transaction duration. A clear periodicity is apparent with low durations (high intensities) occurring at the beginning and the end of the trading day, and high durations (low intensities) occurring during the middle of the trading day. To correct for this, we make use of the method used by Zhang et al. (2001). In particular, we use the SPLUS function *supsmu* to compute the daily periodic component, denoted $\phi_{\tau_{n-1}}$. The original data is then divided by $\phi_{\tau_{n-1}}$ to obtain a series that does not contain an intraday periodic component – a series that is referred to as a *diurnally adjusted* series. [16]

Summary statistics for unadjusted raw and diurnally adjusted transaction duration are given in Table 1. The mean unadjusted transaction duration is 21.1229 seconds with a standard deviation of 36.8119 seconds. The fact that the standard deviation is greater than the mean suggests that either the durations are time-varying

---

[15] Tse (1999) also documents this periodicity for several variables observed in the FTSE 100 futures market.

[16] Engle and Russell (1998) use spline functions to calculate $\phi_{\tau_{n-1}}$. Moreover, they jointly estimate these functions and the parameters of the ACD model. However, as shown by Engle and Russell (1995), the two-step approach adopted in this paper produces consistent estimates.

Table 1
Summary statistics

	Statistic					
	N	Mean	SD	Min	Max	LB test
Panel A: Raw data						
Y_n	99,399	21.1229	36.8119	0	3742	67511.8010
B_n	99,399	1.6704	1.1306	0.5000	21.0000	92982.4140
V_n	99,399	7.2304	33.8722	1	7444	572.2744
Z_n	99,399	0.0002	0.0016	−0.0154	0.0182	1105148.0000
Panel B: Diurnally adjusted data						
Y_n	99,399	0.9796	1.4679	0.0000	123.6481	49892.5710
B_n	99,399	0.9994	0.6733	0.2825	12.3381	90281.2790
V_n	99,399	1.0017	4.6313	0.1273	1028.0122	574.7592

This table contains summary statistics for transaction duration (Y_n), pricing error (Z_n), bid–ask spread (B_n), and the number of contracts traded per transaction (V_n). The column entitled 'LB test' gives the Ljung–Box test statistic and has an associated 95% critical value of 24.996. Transaction duration is measured in seconds.

and/or the durations cannot be adequately described by an exponential distribution. This is also implied by the unconditional distribution of unadjusted durations given in Fig. 2. This figure shows that the mode of the distribution is greater than zero, a result that is not consistent with an exponential distribution. Moreover, the fact that the mean of the diurnally adjusted durations is greater than their standard deviation suggests that this result is not due to periodic time-variation. Further evidence of the inappropriateness of the exponential distribution is given in Fig. 3. This figure shows the unconditional distribution of diurnally adjusted durations in addition to, an exponential distribution with a mean matched to the sample mean, and the more flexible Weibull distribution with parameters matched to the data. [17] It is clear that the Weibull distribution appears more satisfactory as the mode of such a distribution can be greater than zero. For this reason we use this distribution in the context of the various ACD models estimated in the subsequent analysis.

3.3. Model estimation

Six different ACD models are estimated in this paper each of which is estimated by quasi-maximum likelihood estimation. More specifically, the likelihood function associated with each of the models is maximized using the BHHH algorithm. [18] For each model, the dependent variable is diurnally adjusted transaction duration and the basic assumption given by (1) is maintained with the Weibull distribution imposed.

---

[17] Table 4 in Appendix A gives the probability density function (pdf) of the exponential and Weibull distributions. It is clear from these pdf's that the exponential distribution represents a restricted version of the Weibull distribution. In particular, the former distribution is equivalent to the latter distribution when the restriction $\gamma = 1$ is imposed.

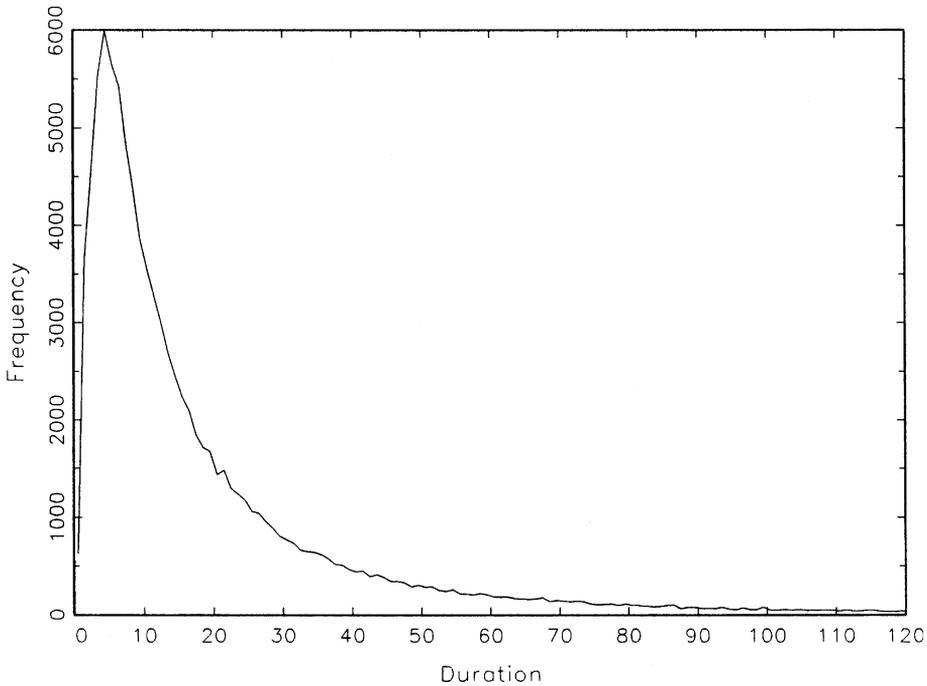[18] See Berndt et al. (1974) for more details.

Fig. 2. The distribution of transaction duration.

The first model considered is based on the basic specification of ψ_n given by (2). This model (denoted M1) selects the lag structure by grid search over p and q. For the allowable parameter space $p = \{1, 2, 3\}$ and $q = \{1, 2, 3\}$, we find that $p = 1$ and $q = 2$ delivers the minimum value of the Akaike information criterion, thus M1 is given by

$$\psi_n = \alpha_0 + \alpha_1 Y_{n-1} + \beta_1 \psi_{n-1} + \beta_2 \psi_{n-2}. \tag{14}$$

The estimated parameters together with their associated heteroscedastic-consistent standard errors and the results of various diagnostic tests are given in Table 2. The results indicate that transaction duration is stationary but exhibits a significant amount of time dependency as indicated by the significant coefficients on past expected transaction duration and past transaction duration. [19]

To assess the quality of this model a series of diagnostic tests are conducted using the transformed residuals from the models. Given the relationship between the exponential and Weibull distributions (see Appendix A), the residuals

$$\hat{\epsilon}_n^\gamma = \left(\frac{Y_n}{\psi_n} \right)^\gamma \tag{15}$$

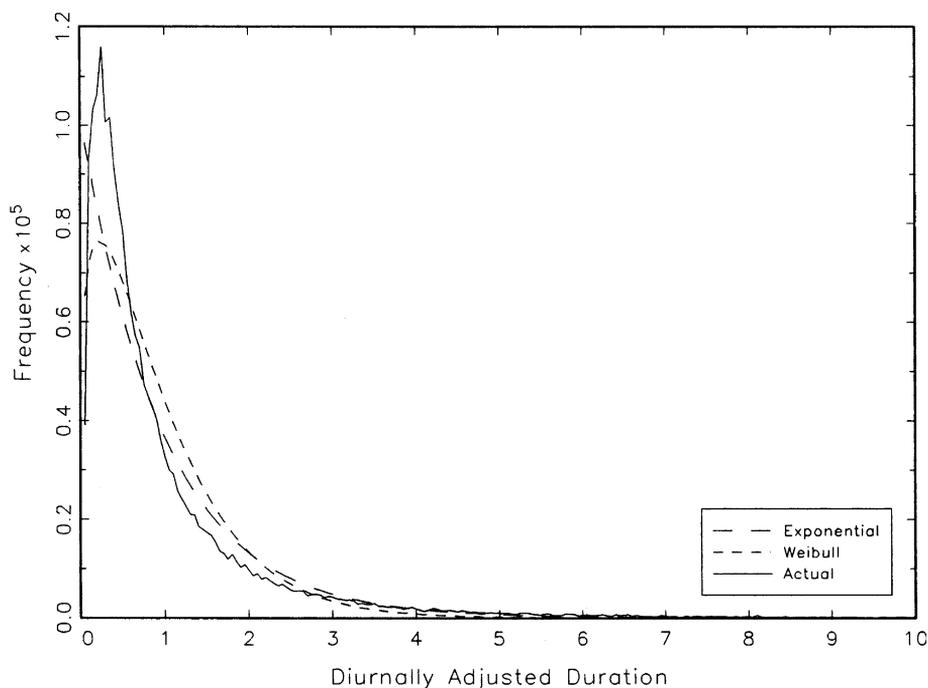[19] The series is stationary as the coefficients sum to less than unity.

Fig. 3. A comparison of transaction duration distributions.

should be i.i.d. and distributed as a unit exponential under the null hypothesis of model adequacy. [20] The diagnostic tests applied in this paper are designed to test for these particular characteristics. First, we test the hypothesis that the standard deviation of the transformed residuals is equal to unity – as implied by a unit exponential distribution. Using the test described by Engle and Russell (1998), the results of this test for excess dispersion are reported in Table 2. The results indicate that M1 fails in this respect. To test the i.i.d. characteristic we perform a linear and a non-linear test for serial correlation. The (linear) Ljung–Box test is applied using up to 15 lags of the dependent variable. The results given in Table 2 indicate that the transformed residuals exhibit a degree of serial dependence that is inconsistent with model adequacy. However, comparing these results with the extremely large test statistics obtained when using the actual transaction durations (as given in Table 1) then we can claim some success. Finally, a non-linear test for serial dependence is conducted as described in Engle and Russell (1998). Again the results indicate an excessive level

---

[20] The *Weibull parameter*, γ, measures the degree of excess dispersion in the model residuals. Under the restriction that $\gamma = 1$, the Weibull distribution is equivalent to the exponential distribution. As such, under this restriction, the mean and standard deviation of the errors are forced to equate. If this restriction is not imposed, i.e., the Weibull distribution is used, then the commonly observed feature of excess dispersion is allowed. Indeed, as shown in Table 2, the estimated value of γ is greater than unity implying excess dispersion in the residuals.

Table 2
Parameter estimates of various ACD models

	Model					
	M1	M2	M3	M4	M5	M6
$\hat{\alpha}_0$	0.0267	0.0302	0.0316	0.0314	0.0316	0.0318
	(0.0016)	(0.0015)	(0.0007)	(0.0014)	(0.0012)	(0.0006)
$\hat{\alpha}_{0,1}$				0.0265	0.0264	0.0267
				(0.0018)	(0.0014)	(0.0003)
$\hat{\alpha}_1$	0.1159	0.1159	0.1160	0.1166	0.1169	0.1167
	(0.0043)	(0.0027)	(0.0035)	(0.0029)	(0.0024)	(0.0030)
$\hat{\beta}_1$	0.6810	0.6813	0.6805	0.6658	0.6626	0.6638
	(0.0402)	(0.0233)	(0.0490)	(0.0293)	(0.0163)	(0.0242)
$\hat{\beta}_2$	0.1771	0.1762	0.1766	0.1909	0.1938	0.1928
	(0.0356)	(0.0216)	(0.0437)	(0.0275)	(0.0153)	(0.0215)
$\hat{\theta}_1$		−0.0025	−0.0023	−0.0023	−0.0023	−0.0023
		(0.0003)	(0.0001)	(0.0003)	(0.0002)	(0.0002)
$\hat{\theta}_2$		−0.0005	−0.0005	−0.0005	−0.0005	−0.0005
		(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
$\hat{\omega}_1$			−1.0338	−0.7947	−0.9660	−0.9583
			(0.3526)	(0.2069)	(0.1344)	(0.0951)
$\hat{\omega}_{1,1}$				0.0099	0.0010	0.0039
				(0.0010)	(0.0020)	(0.0016)
γ	1.0288	1.0291	1.0292	1.0294	1.0294	1.0293
	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)
Log likelihood	−0.891328	−0.891103	−0.891050	−0.890979	−0.890975	−0.890988
Residual mean	1.0120	1.0121	1.0122	1.0122	1.0122	1.0122
Residual SD	1.1710	1.1710	1.1706	1.1704	1.1704	1.1705
Dispersion test	41.3806	41.3779	41.2724	41.2357	41.2340	41.2496
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Ljung–Box test	113.5424	110.3866	109.9462	112.2315	112.3545	112.4438
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Non-linear test	2.7944	2.7798	2.7850	2.7675	2.7619	2.7651
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)

This table contains the parameter estimates of the ACD models described and denoted in the text (and in Table 4 in Appendix A) as M1–M6. The numbers in the parentheses are the heteroscedastic-consistent standard errors associated with the parameter estimates or the *p*-values associated with diagnostic tests applied to the transformed residuals from the models.

of serial correlation in the transformed residuals. Having applied all three tests it is somewhat disappointing, though not unusual, to find that all three tests indicate a rejection of model adequacy. [21]

The second model (M2) considered in this paper augments M1 by allowing explanatory variables in (2) as given by (3). The explanatory variables considered in this paper are motivated by various market microstructure theories. A key implication of market microstructure theory (see O'Hara, 1995, for a review) is that condi-

---

[21] Engle and Russell (1998), and Zhang et al. (2001) also find, in their studies of the US stock market, that these tests of model adequacy are rejected.

tional return volatility is positively related to the extent of information asymmetry. [22] Given the inverse relationship between transaction duration and instantaneous conditional return volatility, this implication is equivalent to stating that transaction duration and information asymmetry are negatively related. It follows that a measure of information asymmetry would seem to be a useful explanatory variable. The problem here is that information asymmetry cannot be readily observed in financial markets. The approach taken in this paper is to include variables that are theoretically related to the extent of information asymmetry. One such variable is the bid–ask spread. Glosten and Milgrom (1985) and Easley and O'Hara (1992b) argue that market makers will widen spreads when the extent of information asymmetry increases. Such action is taken by market makers to lessen the probability of trading with an informed agent. The second variable included is trading volume. Inclusion of this variable is motivated by the results of Easley and O'Hara (1987, 1992a) and Lee et al. (1993). They argue that, for a given price, informed traders have an incentive to trade a larger quantity of shares than non-informed traders. [23] By including these variables, M2 has the following specification:

$$\psi_n = \alpha_0 + \alpha_1 Y_{n-1} + \beta_1 \psi_{n-1} + \beta_2 \psi_{n-2} + \theta_1 B_{N_{s(\tau_{n-1})}} + \theta_2 V_{n-1}, \tag{16}$$

where $B_{N_{s(\tau_{n-1})}}$ is the bid–ask spread associated with the latest quote available at the time of the previous trade (hence the time subscript) and V_{n-1} is the number of contracts traded in the previous trade. [24]

The results given in Table 2 are in accordance with the above market microstructure theories. In particular, transaction duration is negatively (and significantly) related to both the bid–ask spread and trading volume. Hence, given the above arguments, there would appear to be a significant relationship between return volatility and information asymmetry – a result revealed by transaction duration and proxies for information asymmetry. As with M1, the diagnostic tests applied to the transformed residuals from M2 indicate model inadequacy.

---

[22] Information asymmetry measures the amount of informed trading versus non-informed trading in a market at any point in time.

[23] The relationship between trading volume, the bid–ask spread, and information asymmetry remains a contentious issue in the market microstructure literature; see Karpoff (1987), Stickel and Verrecchia (1993), and Jones and Kaul (1994) for surveys of the various arguments posited in this area. As such, the results of the tests performed in this paper should be treated with some caution.

[24] There remains four features of M2 to be discussed. First, it is necessary to diurnally adjust both bid–ask spreads and volume using the same methodology used to adjust transaction duration. Summary statistics relating to these market microstructure data are given in Table 1. Second, the selection of p and q is again achieved by way of a grid search. As in M1 (and all subsequent models) we find that using $p = 1$ and $q = 2$ gives the best optimization hence the specification given by (16). Third, we conduct a similar grid search over the number of lagged values of the market microstructure variables. In doing this for this model (and all subsequent models), we find that one lagged value of both market microstructure variables provides the best fit as implied by the value of the Akaike information criterion. Fourth, we find that the results are robust to changes in the definition of trading volume. For instance, when the dollar value of trading volume is used instead of the number of contracts traded, the results are virtually unchanged. These results are available upon request.

The final four models estimated in this section all include the above explanatory variables and past pricing error in their specification. The third model (M3) assumes that the absolute value of past pricing error is linearly related to expected transaction duration, that is,

$$\psi_n = \alpha_0 + \alpha_1 Y_{n-1} + \beta_1 \psi_{n-1} + \beta_2 \psi_{n-2} + \theta_1 B_{N_{s(\tau_{n-1})}} + \theta_2 V_{n-1} + \omega_1 |Z_{N_{s(\tau_{n-1})}}|, \qquad (17)$$

where $Z_{N_{s(\tau_{n-1})}}$ denotes the latest pricing error observed at the time of the last trade. [25]

Given the arguments described in the previous section we expect ω_1 to be negative. The results in Table 2 confirm this prediction. The coefficient is significantly less than zero and takes a value of approximately (negative) unity. The fit of this model (as given by the value of the log likelihood function) appears to be superior to the previous models though the diagnostic tests still imply model inadequacy.

In the presence of non-zero trading costs, the assumed linearity between transaction duration and past pricing error becomes questionable and necessitates use of a non-linear model. The first two non-linear models considered (M4 and M5) are based on the threshold models given by (10) and (11), (10) and (12), respectively, but with explanatory variables included,

$$\psi_n = (1 - F_i(Z_{N_{s(\tau_{n-1})}}))(\alpha_0 + \omega_1 |Z_{N_{s(\tau_{n-1})}}|) + F_i(Z_{N_{s(\tau_{n-1})}})(\alpha_{0,1} + \omega_{1,1} |Z_{N_{s(\tau_{n-1})}}|)$$

$$+ \alpha_1 Y_{n-1} + \beta_1 \psi_{n-1} + \beta_2 \psi_{n-2} + \theta_1 B_{N_{s(\tau_{n-1})}} + \theta_2 V_{n-1}, \qquad (18)$$

where $F_1(Z_{N_{s(\tau_{n-1})}})$ is given by (11) and $F_2(Z_{N_{s(\tau_{n-1})}})$ is given by (12). The parameters c, c_1, and c_2 are determined by grid search using values of c, c_1, and c_2 from 0.001 to 0.003 with increments of 0.0001.

An important feature of such threshold models concerns the estimated value of the transaction cost parameters c, c_1, and c_2. In the case of the symmetric threshold model M4, using the grid search method we find that $\hat{c}$ equals 0.0021, equivalently 21 basis points. This is reassuring close to estimated transaction costs obtained using threshold models of the relationship between futures returns and past pricing error. [26] For instance, using UK data Garrett and Taylor (2001) find that their estimate of transaction costs is 23 basis points. In the case of the asymmetric threshold model M5, we find that c_1 equals -0.0020 and c_2 equals 0.0022. This result lends support to the argument that short-selling the underlying asset is more costly than taking a long position in the underlying asset.

The second important feature of M4 and M5 concerns the hypothesis associated with the values of α_0, α_1, ω_1, and $\omega_{1,1}$. In particular, given the arguments described in the previous section, we expect α_0 to be greater than $\alpha_{0,1}$ and ω_1 and $\omega_{1,1}$ to be equal to zero. Recall that we would expect no arbitrage activity when the pricing error is less than the cost of trading, hence $\omega_1 = 0$. When profitable opportunities do occur ($|Z_{N_{s(\tau_{n-1})}}| > c$) we expect an increase in arbitrage activity across regimes, hence

---

[25] The lag structure associated with pricing error in this model (and all subsequent models) is determined by grid search.

[26] Such models are often referred to as threshold error correction models, see Dwyer et al. (1996) for an example.
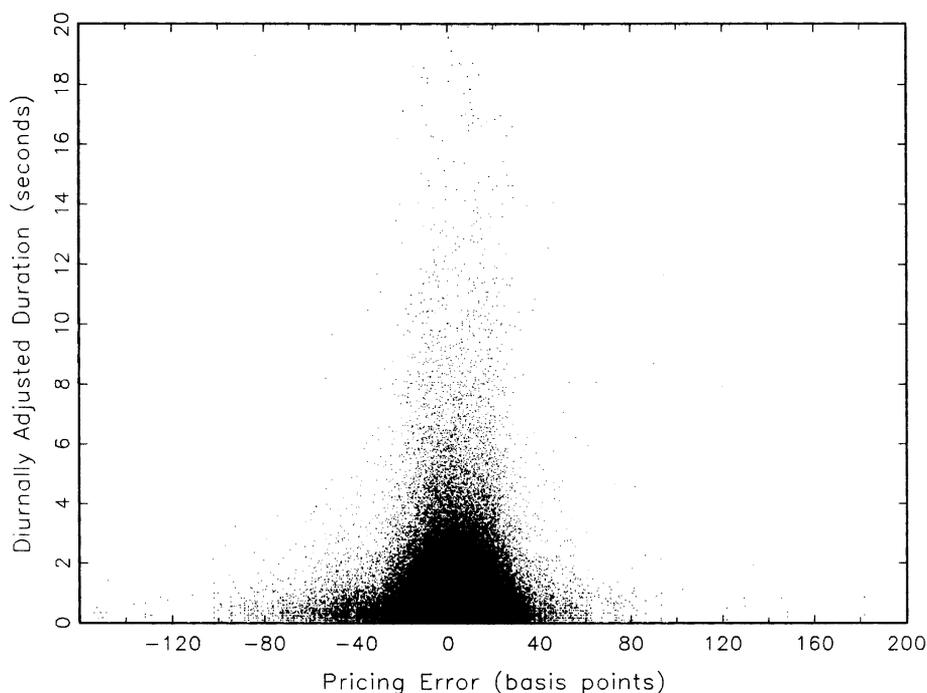
Fig. 4. The relationship between transaction duration and pricing error.

$\alpha_0 > \alpha_{0,1}$. However, given the bounded nature of transaction duration, we expect the activity to reach some maximum level within the profit opportunity regime. As such, increases (decreases) in the pricing error within this regime should have no impact upon arbitrage activity, hence $\omega_{1,1} = 0$. If this particular prediction is found to exist then it implies that when profitable opportunities exist, no relationship will exist between pricing error and transaction duration. This is equivalent to stating that arbitrage activity is at its greatest and cannot be increased even when the size of the profit opportunity increases.

The results given in Table 2 lend some support to the above arguments. For both models there appears to be an increase in activity when one moves from the no profit opportunity regime to the profit opportunity regime. In particular, $\hat{\alpha}_0 > \hat{\alpha}_{0,1}$, a result that supports intuition and theory. [27] However, contrary to the above predictions, there does appear to be some variation in activity within the regimes, in particular, $\hat{\omega}_1 < 0$. A graphical description of this variation is provided in Fig. 4. This figure gives a scatter plot of transaction duration against the first lagged value of pricing error. It is clear from this diagram that there is variation in activity within regimes and across regimes. The 'within regime' variation is particularly apparent in the no profit opportunity regime. There appears to be a wide variety of transaction

---

[27] The standard errors associated with these coefficients suggest that the difference is significant.
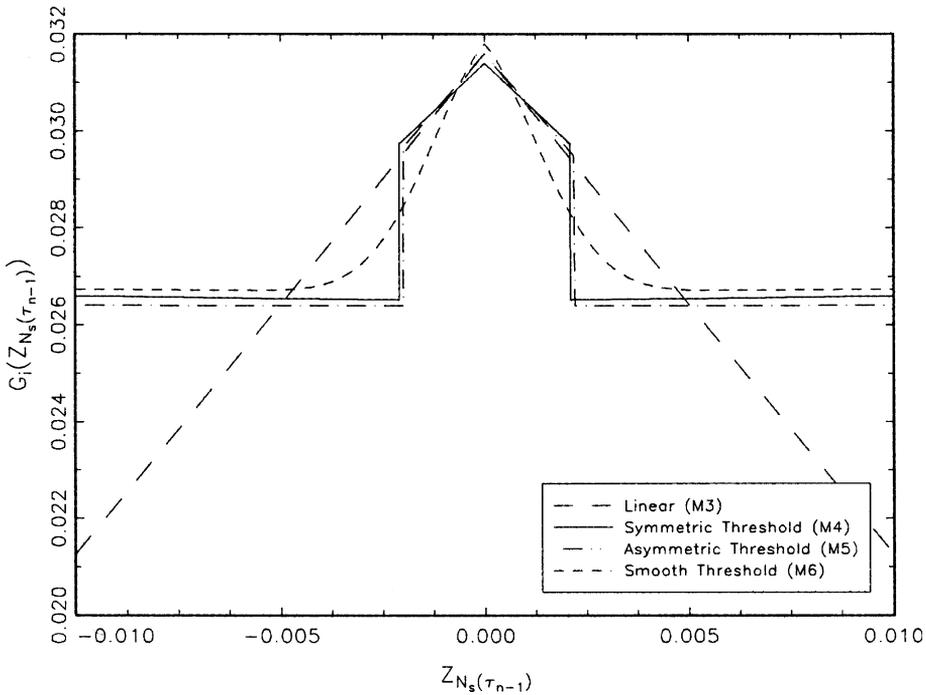
Fig. 5. The impact of pricing error on expected transaction duration.

durations and also a negative relationship between mean transaction duration and absolute pricing error within this regime. [28] To give some idea as to how the estimated parameters capture the latter type of variation, Fig. 5 shows a plot of the following function against past pricing error:

$$
\begin{aligned}
G_i(Z_{N_{s(\tau_{n-1})}}) = {} & (1 - F_i(Z_{N_{s(\tau_{n-1})}}))(\hat{\alpha}_0 + \hat{\omega}_1 |Z_{N_{s(\tau_{n-1})}}|) \\
& + F_i(Z_{N_{s(\tau_{n-1})}})(\hat{\alpha}_{0,1} + \hat{\omega}_{1,1} |Z_{N_{s(\tau_{n-1})}}|),
\end{aligned}
\tag{19}
$$

where $G_i(\cdot)$ measures the adjustment in expected duration due to previous pricing error, that is, adjustment due to arbitrage activity. The function is calculated using the estimated parameters from M3, M4 and M5. In the case of the former (linear) model, $F_i(\cdot)$ is set equal to zero for all pricing error values. One can see from this figure that, as $\hat{\alpha}_0 > \hat{\alpha}_{0,1}$, there is an abrupt decrease in expected transaction duration when one switches to the profit opportunity regime – a feature that cannot be picked up by the linear model. One can also see that, as $\hat{\omega}_{1,1} \approx 0$, there is very little variation in expected transaction duration in the profit opportunity regime. By contrast, as $\hat{\omega}_1 < 0$, there appears to be a large degree of variation in expected transaction duration in the no profit opportunity regime. This may be because arbitragers face

---

[28] The former type of variation is unsurprising given that there could be substantial clustering of trading activity due to the clustered nature of information arrival when no arbitrage profits are available.

heterogeneous trading costs. To investigate this possibility we estimate an alternative type of threshold model.

The final model (M6) assumes the same specification for ψ_n as given by M5 but differs by allowing $F_i(Z_{N_{s(\tau_{n-1})}})$ to be given by (13), with δ determined by grid search over an appropriate parameter space. The resulting coefficient estimates are given in Table 2. These estimates are very similar to the M4 and M5 estimates. However, the models differ in terms of the mode of transition across states. This can be readily observed in Fig. 5 which plots the smooth threshold adjustment function $G_i(\cdot)$ against past pricing error. The shape of the function suggests that transaction cost heterogeneity is most apparent for arbitragers facing small transaction costs. When the profit opportunity becomes large the heterogeneity becomes negligible with activity reaching a maximum asymptotic level. One can also see from this figure that the M4 and M5 coefficients also pick up this effect but in a linear piecewise fashion as opposed to the continuous non-linear nature of M6. This latter model also has appeal when one compares the shape of the M6 activity function given in Fig. 5 with the scatter plot given in Fig. 4. There is clearly a correspondence between the functional dependence between expected transaction duration and past pricing error as implied by M6 and the actual dependence given by the data.

In terms of the fit of M4, M5 and M6, the results suggest that they offer a superior representation of the data over previous models. However, diagnostic tests of model adequacy fail to support these models. A more detailed analysis of the universal failure of the models in this respect is given in the following subsection.

Before proceeding to the next subsection it is worth considering the implications of the estimated models in terms of the pricing efficiency of the FTSE 100 futures market. The evidence presented in the paper shows that there is rapid adjustment of prices to information arrival. In this paper, 'adjustment' is measured via the trading process with 'information' defined as pricing error. Figs. 4 and 5 perhaps best illustrate this compatibility with market efficiency. Both figures clearly show that in the profit opportunity regime there is a very short period of time between price impact trades. In other words, prices are changing rapidly within this regime. By contrast, in the no profit opportunity regime prices are less rapidly adjusting. This is to be expected given the difference in information flow over the two regimes.

3.4. Model adequacy

The most disappointing feature of the models considered in this paper is their failure to pass various diagnostic tests of model adequacy. Such inadequacy is not unique to the dataset used in this paper. Both Engle and Russell (1998) and Zhang et al. (2001) find similar results when using data covering an individual stock listed on the New York Stock Exchange. Efforts were made by these authors to correct for this failure. For instance, Zhang et al. (2001) introduce an alternative error distribution and perform a subperiod analysis in an attempt to provide adequate models of transaction duration. While such modifications produce some success, the underlying result from such papers is that passing tests of model adequacy remains an issue in the context of ACD models. An alternative approach to model adequacy is taken in this

paper. Rather than modify the models and apply the usual diagnostic tests, we consider an alternative approach to model performance.

Given the importance of accurate forecasts of the volatility of asset returns in areas such as risk management and option pricing, it seems reasonable to assess model performance via a comparison of forecasting ability. In conducting such a comparison we are explicitly conceding the fact that the 'best' model may still be 'inadequate' in terms of model misspecification. However, as a universal set of currently available ACD models is used, the approach will give some indication as to which model most accurately represents the data and, hence, is of most importance to users of financial markets.

The six models considered in the previous subsection are re-estimated using the first half of the dataset. [29] The estimated coefficients and the second half of the dataset are then used to generate time-consistent 1-step ahead forecasts of transaction duration. These forecasts are then compared to realized transaction duration observed in the second half of the sample. The mean squared forecast errors associated with these forecasts are given in Table 3. [30] In addition to the six models considered in the previous section, we also consider a naïve model where forecasts are set equal to the mean transaction duration observed in the first half of the dataset – this model is henceforth denoted M0. The results indicate that M0 is not surprisingly the poorest of the models considered. The best model appears to be M1 followed by M2 and M6.

To formally test the comparative accuracy of the model-based forecasts we make use of the asymptotic test introduced by Diebold and Mariano (1995). This test allows use of an arbitrary loss function instead of the usual squared forecast error loss, and is robust to non-zero mean forecast errors, non-normally distributed forecast errors, and serially correlated forecast errors. In the current application it is the robustness of the Diebold and Mariano statistic to the non-normality assumption that is most attractive. Indeed, when the forecast errors are tested for normality the null is rejected at the 1% significance level on every occasion. [31]

Diebold and Mariano show that the following test statistic is (asymptotically) normally distributed with zero mean and unit variance:

$$S_1 = \frac{\bar{d}}{\sqrt{\frac{2\pi\hat{f}_d(0)}{T}}}, \tag{20}$$

where

$$\bar{d} = \frac{1}{T} \sum_{t=1}^{T} \left[g(e_t) - g(e'_t) \right] \tag{21}$$

---

[29] Details of the estimated models are available upon request. Other subsets of the dataset are considered. For instance, we use the first quarter and the first three quarters of the dataset in the estimation section of the exercise. The results obtained using such sample periods produce similar results to those presented in this paper. These results are also available upon request.

[30] Similar results are obtained when mean absolute forecast errors are considered. These results are available upon request.

[31] These results are available upon request.

Table 3
Testing the statistical significance of duration forecasts

		Model						
		M0	M1	M2	M3	M4	M5	M6
Panel A: symmetric quadratic loss								
MSFE		2.6510	2.1535	2.1553	2.1608	2.1599	2.1582	2.1571
D–M test	M0		7.4481	7.5243	7.5042	7.5072	7.5786	7.5375
	M1	−7.4481		−1.8208	−4.4492	−4.0385	−2.2484	−2.4988
	M2	−7.5243	1.8208		−7.8862	−6.4761	−2.2642	−3.1174
	M3	−7.5042	4.4492	7.8862		2.3732	2.6162	6.9671
	M4	−7.5072	4.0385	6.4761	−2.3732		1.6485	7.5835
	M5	−7.5786	2.2484	2.2642	−2.6162	−1.6485		1.1927
	M6	−7.5375	2.4988	3.1174	−6.9671	−7.5835	−1.1927	
D–M test score		0	5	5	1	2	2	3
Panel B: asymmetric linex loss								
D–M test score	$\rho = 0.02$	0	3	3	1	2	3	3
	$\rho = 0.04$	0	3	3	1	2	4	3
	$\rho = 0.06$	0	2	4	1	2	6	3
	$\rho = 0.08$	0	1	4	1	2	6	4
	$\rho = 0.10$	0	1	4	1	2	6	4

Panel A of this table gives the mean squared forecasts errors (MSFE), the test statistics associated with the Diebold–Mariano test of comparable forecast performance and the associated number of successful forecast comparisons achieved at the 5% significance level. Panel B of this table gives the number of successful forecast performance comparisons under the (asymmetric) Linex loss function at the 5% significance level. For descriptions of models, M0–M6, see Table 4 in Appendix A.

is the sample loss differential and $\hat{f}_d(0)$ is a consistent estimate of the spectral density of the loss differential at frequency zero,

$$f_d(0) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \gamma_d(\tau), \tag{22}$$

where e_t is the h-step ahead forecast error associated with one particular model, e_t' is the h-step ahead forecast error associated with a competing model, $g(\cdot)$ is the loss function, $\gamma_d(\tau) = E[(d_t - \mu)(d_{t-\tau} - \mu)]$ is the autocovariance of the loss differential at displacement τ, and μ, is the population mean loss differential. Following Diebold and Mariano, we use a uniform lag window of size $h - 1$ to estimate $f_d(0)$. The loss functions used are the squared function (MSFE) and the absolute function (MAFE) though only results pertaining to the former function are presented in the paper. [32]

A summary of the results obtained when the Diebold and Mariano test is performed on all available forecasts is given in Table 3. We report the test statistics associated with the Diebold–Mariano test where each model-based set of forecasts are

---

[32] Results associated with the latter function are available upon request.

compared with each other. In addition, we report the number of times for which one particular model-based set of forecasts is significantly better (at the 5% level) than the other model-based forecasts. Thus for each model there will be six such tests performed. The results indicate that M1 and M2 each successfully beat the other models on five occasions. The non-linear models, M4–M6, are less successful. Of these models the most successful is M6 with three successes.

In using a symmetric loss function we are explicitly assuming that positive and negative forecast errors (of the same absolute magnitude) receive equal weight in the loss function. In the current context this may not be the case. Both Brailsford and Faff (1996) and Bystrom (2000) argue that underprediction of return volatility should be more heavily penalized than overprediction. The argument is demonstrated by consideration of a seller of a call option. If such a trader underpredicts the underlying asset return volatility then there will be a downward bias in the estimate of the call price. As such, the trader will be prepared to accept less for their option than it is actually worth. The need to protect against underprediction can also be demonstrated in the context of the VaR methodology. It is obvious in this context that underprediction of return volatility is to be avoided if the solvency of the portfolio owner is to be maintained. Therefore, as the above ACD models can be interpreted as models of the inverse of instantaneous conditional return volatility then some form of asymmetric loss function is required.

A commonly used asymmetric loss function is the Linex loss function introduced by Varian (1974) and used by Zellner (1986). It is given by the following expression:

$$g(e_t) = \exp[\rho e_t] - \rho e_t - 1. \tag{23}$$

When $\rho > 0$ ($\rho < 0$), positive (negative) forecast errors are weighted more heavily than negative (positive) forecast errors. Another important feature of this loss function is that the function becomes asymptotically equivalent to the symmetric quadratic loss function when $\rho \to 0$. Given the above arguments we allow $\rho = \{0.02, \ 0.04, 0.06, 0.08, 0.10\}$. Note that we are penalizing overprediction more heavily than underprediction. This may at first appear to contradict the above arguments. However, recall that the forecasts of transaction duration generated in this paper are essentially forecasts of the inverse of instantaneous conditional return volatility. As such, the asymmetry of the loss function is reversed accordingly.

The Diebold–Mariano test is applied using the Linex loss function for all combinations of possible comparisons. The results are given in Table 3. Not surprisingly, when ρ is close to zero, the number of successes closely resembles the results obtained using the symmetric quadratic loss function. However, as ρ becomes large and, hence, the degree of asymmetry increases, the number of successes for each model changes. Most notably, the asymmetric threshold model M5 becomes increasingly successful. Indeed, for $\rho \geqslant 0.06$ this model achieves the maximum of six successes and can clearly be referred to as the dominant model under these conditions.

4. Concluding remarks

This paper has shown that trading activity on the FTSE 100 futures market is determined by several factors. First, there exists a strong time dependent component in intensity. Thus, the observed clustering in trading activity is greater than that predicted by an independent process. Second, three explanatory variables play a key role in explaining levels of trading activity. The first two of these effects are market microstructure in nature. In particular, we find that bid–ask spread and trading volume have a negative effect on transaction duration. The transition mechanism used to explain these results is based on the argument that there is a positive relationship between return volatility and these proxies for the degree of information asymmetry in the market. Given the negative relationship between return volatility and the definition of transaction duration used in this paper, the results are compatible with various market microstructure theories. While these market microstructure effects have been investigated in previous studies of trading activity, the third of these effects has not. In this paper we have argued that arbitrager behaviour will have a significant impact upon trading activity. In particular, we argue that arbitrager trading activity will be a function of the magnitude of the mispricing between spot and futures prices. These arguments are born out by the evidence presented in this paper which shows that the greater the profit opportunity, as implied by non-zero pricing error, the greater the intensity of trading activity. The nature of this dependency is also examined and found to be non-linear in nature. In particular, we find that trading activity dramatically increases when the size of the profit opportunity available to arbitragers exceeds the costs of trading the futures contract and the associated underlying asset. In terms of the trading costs faced by arbitragers, we find evidence to suggest that these costs are different depending on the nature of the trade required and may differ across arbitragers.

An important aspect of this paper is that only transactions that have a non-zero price impact are considered in the analysis. As such, the models considered can be interpreted as models of (the inverse of) return volatility. Given the importance of accurate forecasts of return volatility in areas of risk management and option pricing, it is of some importance that the forecasting performance of the models was assessed. The results indicate that, under certain conditions, the non-linear ACD models developed in this paper provide more accurate forecasts of transaction duration than the conventional linear ACD models. It is the success of the former models that is to be exploited in future research. In particular, it would be a useful exercise to compare the performance of non-linear ACD models with conventional models of return volatility such as ARCH models. Such research would throw light on what factors determine return volatility over high frequencies.

Appendix A. Notation and definitions

Table 4 contains definitions of all variables and parameters (Panel A), presents all models considered (Panel B), gives the specifications of the transition functions used (Panel C), and lists all distributions considered (Panel D).

Table 4
Summary information

Notation	Definition
Panel A: Variable definitions	
Y_n	The time between 'price impact' trades (transaction duration).
$B_{N_{s(\tau_n)}}$	The difference between the bid and ask quote (bid–ask spread).
V_n	The number of contracts traded per transaction (trading volume).
$Z_{N_{s(\tau_n)}}$	The difference between the theoretical and market price of the futures contract (pricing error).
ψ_n	The conditional expectation of transaction duration.
$N_{s(\tau_n)}$	The number of s-type events that have occurred by time τ_n.
ϕ_{τ_n}	The daily periodic component.
Panel B: Model specifications	
M0	$\psi_n = \alpha_0.$
M1	$\psi_n = \alpha_0 + \alpha_1 Y_{n-1} + \beta_1 \psi_{n-1} + \beta_2 \psi_{n-2}.$
M2	$\psi_n = \alpha_0 + \alpha_1 Y_{n-1} + \beta_1 \psi_{n-1} + \beta_2 \psi_{n-2} + \theta_1 B_{N_{s(\tau_{n-1})}} + \theta_2 V_{n-1}.$
M3	$\psi_n = \alpha_0 + \alpha_1 Y_{n-1} + \beta_1 \psi_{n-1} + \beta_2 \psi_{n-2} + \theta_1 B_{N_{s(\tau_{n-1})}} + \theta_2 V_{n-1} + \omega_1 \lvert Z_{N_{s(\tau_{n-1})}} \rvert.$
M4	$\psi_n = (1 - F_1(Z_{N_{s(\tau_{n-1})}}))(\alpha_0 + \omega_1 \lvert Z_{N_{s(\tau_{n-1})}} \rvert) + F_1(Z_{N_{s(\tau_{n-1})}})(\alpha_{0,1} + \omega_{1,1} \lvert Z_{N_{s(\tau_{n-1})}} \rvert).$ $+ \alpha_1 Y_{n-1} + \beta_1 \psi_{n-1} + \beta_2 \psi_{n-2} + \theta_1 B_{N_{s(\tau_{n-1})}} + \theta_2 V_{n-1}.$
M5	$\psi_n = (1 - F_2(Z_{N_{s(\tau_{n-1})}}))(\alpha_0 + \omega_1 \lvert Z_{N_{s(\tau_{n-1})}} \rvert) + F_2(Z_{N_{s(\tau_{n-1})}})(\alpha_{0,1} + \omega_{1,1} \lvert Z_{N_{s(\tau_{n-1})}} \rvert).$ $+ \alpha_1 Y_{n-1} + \beta_1 \psi_{n-1} + \beta_2 \psi_{n-2} + \theta_1 B_{N_{s(\tau_{n-1})}} + \theta_2 V_{n-1}.$
M6	$\psi_n = (1 - F_3(Z_{N_{s(\tau_{n-1})}}))(\alpha_0 + \omega_1 \lvert Z_{N_{s(\tau_{n-1})}} \rvert) + F_3(Z_{N_{s(\tau_{n-1})}})(\alpha_{0,1} + \omega_{1,1} \lvert Z_{N_{s(\tau_{n-1})}} \rvert)$ $+ \alpha_1 Y_{n-1} + \beta_1 \psi_{n-1} + \beta_2 \psi_{n-2} + \theta_1 B_{N_{s(\tau_{n-1})}} + \theta_2 V_{n-1}.$
Panel C: Transition function specifications	
$F_1(Z_{N_s(\tau_{n-1})})$	$F_1(Z_{N_s(\tau_{n-1})}) = \begin{cases} 0 & \text{if } \lvert Z_{N_s(\tau_{n-1})} \rvert \leqslant c, \\ 1 & \text{otherwise,} \end{cases}$ where c is the *symmetric threshold* transaction cost bound parameter.
$F_2(Z_{N_s(\tau_{n-1})})$	$F_2(Z_{N_s(\tau_{n-1})}) = \begin{cases} 0 & \text{if } -c_1 \leqslant \lvert Z_{N_s(\tau_{n-1})} \rvert \leqslant c_2, \\ 1 & \text{otherwise,} \end{cases}$ where c_1 and c_2 are the *asymmetric threshold* transaction cost bound parameters.
$F_3(Z_{N_s(\tau_{n-1})})$	$F_3(Z_{N_s(\tau_{n-1})}) = 1 - \exp\left[-\delta Z_{N_s(\tau_{n-1})}^2\right],$ where δ is the *smooth threshold* adjustment parameter.
Panel D: Distribution specifications	
$Y_n/\psi_n = \epsilon_n \sim \text{i.i.d.} \mathscr{D}(\boldsymbol{\kappa})$	Exponential $\mathscr{D}(\mu) : f(\epsilon_n) = \left(\frac{1}{\mu}\right) \exp[-\frac{\epsilon_n}{\mu}],\ \epsilon_n \geqslant 0,\ \mu > 0.$
	Weibull $\mathscr{D}(\mu, \gamma) : f(\epsilon_n) = \left(\frac{\gamma}{\mu}\right)\left(\frac{\epsilon_n}{\mu}\right)^{\gamma-1} \exp[-\left(\frac{\epsilon_n}{\mu}\right)^\gamma],\ \epsilon_n \geqslant 0,\ \mu > 0,\ \gamma > 0.$

References

Bauwens, L., Giot, P., Grammig, J., Veredas, D., 2000. A comparison of financial duration models via density forecasts, Université Catholiqué de Louvain Working Paper.

Berndt, E., Hall, B., Hall, R., Hausman, J., 1974. Estimation and inference in non-linear structural models. Annals of Economic and Social Management 3, 653–665.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroscedasticity. Journal of Econometrics 31, 307–327.

Brailsford, J., Faff, R., 1996. An evaluation of volatility forecasting techniques. Journal of Banking and Finance 20, 419–438.

Brennan, M., Schwartz, E., 1988. Optimal arbitrage strategies under basis variability. Studies in Banking and Finance 5, 167–180.

Brennan, M., Schwartz, E., 1990. Arbitrage in stock index futures. Journal of Business 63, S7–S31.

Bystrom, H., 2000. Orthogonal GARCH and covariance matrix forecasting in a stress scenario: The Nordic stock markets during the Asian financial crisis 1997–1998. Lund University Working Paper.

Diebold, F., Mariano, R., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13, 253–265.

Drost, F., Werker, B., 2000. Efficient estimation in semiparametric time series: The ACD model, Tilburg University Working Paper.

Dwyer, G., Locke, P., Yu, W., 1996. Index arbitrage and non-linear dynamics between the S&P500 futures and cash. Review of Financial Studies 9, 301–332.

Easley, D., O'Hara, M., 1987. Price, trade size, and information in securities markets. Journal of Financial Economics 19, 69–90.

Easley, D., O'Hara, M., 1992a. Adverse selection and large trade volume. Journal of Financial and Quantitative Analysis 27, 185–208.

Easley, D., O'Hara, M., 1992b. Time and the process of security price adjustment. Journal of Finance 47, 577–606.

Engle, R., 1982. Autoregressive conditional heteroscedasticity with estimates of the United Kingdom inflation. Econometrica 50, 987–1007.

Engle, R., Russell, J., 1995. Autoregressive conditional duration: A new model for irregularly spaced transaction data, UCSD Working Paper.

Engle, R., Russell, J., 1998. Autoregressive conditional duration: A new model for irregularly spaced transaction data. Econometrica 66, 1127–1162.

Fernandes, M., Grammig, J., 2000. Non-parametric specification tests for conditional duration models, European University Institute Working Paper.

Garrett, I., Taylor, N., 2001. Intraday and interday basis dynamics: Evidence from the FTSE 100 index futures market. Studies in Nonlinear Dynamics and Econometrics 5, 133–152.

Gerhard, F., Hautsch, N., 2000. Determinants of inter-trade durations using proportional hazard ARM A models, University of Kanstanz Working Paper.

Glosten, L., Milgrom, P., 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. Journal of Financial Economics 14, 71–100.

Hafner, C., 2000. Durations, volume and the prediction of financial returns in transaction time, Electrabel Working Paper.

Jones, C., Kaul, G., 1994. Transactions volume and volatility. Review of Financial Studies 7, 631–651.

Karpoff, J., 1987. The relation between price changes and trading volume: A survey. Journal of Financial and Quantitative Analysis 22, 109–126.

Lee, C., Mucklow, B., Ready, M., 1993. Spreads, depths and the impact of earnings information: An intraday analysis. Review of Financial Studies 6, 345–374.

MacDonald, I., Zucchini, W., 1997. Hidden Markov Models and Other Models for Discrete-valued Time Series. Chapman and Hall, New York.

Martens, M., Kofman, P., Vorst, T., 1998. A threshold error-correction model for intraday futures and index returns. Journal of Applied Econometrics 13, 245–263.

Neal, R., 1996. Direct tests of index arbitrage models. Journal of Financial and Quantitative Analysis 31, 541–562.

Nelson, D., 1991. Conditional heteroscedasticity in asset returns: A new approach. Econometrica 59, 347–370.

O'Hara, M., 1995. Market Microstructure Theory. Basil Blackwell, Oxford.

Sofianos, N., 1993. Index arbitrage profitability. Journal of Derivatives 1, 6–20.

Steutal, F., van Harn, K., 1979. Central limit theorems for finite Walsh–Fourier transforms of weakly stationary time series. Journal of Time Series Analysis 6, 261–267.

Stickel, S., Verrecchia, R., 1993. Evidence that volume sustains price changes, Wharton School Working Paper.

Taylor, N., van Dijk, D., Franses, P., Lucas, A., 2000. SETS, arbitrage activity, and stock price dynamics. Journal of Banking and Finance, 1289–1306.

Terasvirta, T., 1994. Specification, estimation, and evaluation of smooth transition autoregressive models. Journal of the American Statistical Association 89, 208–218.

Tse, Y., 1999. Market microstructure of the FTSE 100 index futures: An intraday empirical analysis. Journal of Futures Markets 19, 31–58.

Varian, H., 1974. A Bayesian approach to real estate assessment. In: Feinberg, S., Zellner, A. (Eds.), Studies in Bayesian Econometrics and Statistics in Honor of L.J. Savage. North-Holland, Amsterdam.

Zellner, A., 1986. Bayesian estimation and prediction using asymmetric loss functions. Journal of the American Statistical Association 81, 446–451.

Zhang, M., Russell, J., Tsay, R., 2001. A non-linear autoregressive conditional duration model with applications to financial transaction data. Journal of Econometrics 104, 179–207.