Modelling bid-ask spreads in competitive dealership markets

Koopman, S.J.M.; Lai, H.N.

Publication date:
1998

Citation for published version (APA):
Koopman, S. J. M., & Lai, H. N. (1998). *Modelling bid-ask spreads in competitive dealership markets*. (CentER Discussion Paper; Vol. 1998-032). Econometrics.

entER
for
nomic Research

Discussion paper

Tilburg University

Center
for
Economic Research

No. 9832

MODELLING BID-ASK SPREADS IN COMPETITIVE DEALERSHIP MARKETS

By Siem Jan Koopman and Hung Neng Lai

March 1998

ISSN 0924-7815

Modelling Bid-Ask Spreads in Competitive Dealership Markets

Siem Jan Koopman
CentER, Tilburg University,
PO Box 90153, 5000 LE Tilburg, The Netherlands
s.j.koopman@kub.nl

Hung Neng Lai
Department of Accounting and Finance,
London School of Economics,
Houghton Street, London WC2A 2AE, UK.

March 1998

Abstract

This paper introduces a new approach to the estimation of the bid-ask spread in competitive dealership markets. In such markets, several transactions may take place simultaneously within the same time interval so that observed prices can not be ordered sequentially. Therefore, the standard approach of estimating spread using differenced prices can not be used without losing a considerable amount of information. We propose an alternative method to model the price behaviour for which data differencing is not needed. This new method is flexible in two ways. Firstly, it can accommodate linear and non-linear cross-sectional properties of the data. Secondly, it can explore the time series properties for data with missing values and for data which may not be ordered sequentially over time. We consider a statistical model for security prices which explicitly takes account of the empirical features of the spread such as trade size and intra-day effects. The model is put into state space form in order to use the Kalman filter smoother to estimate the fundamental price and the spread from the observed prices. The new methodology is illustrated by transaction data of three stocks which are heavily traded on the London Stock Exchange. The model fits the data satisfactory and we conclude that (i) the signed trading volumes strongly affect the fundamental price process, (ii) the size of the trade has a large significant impact on the spread component and (iii) the intra-day effect on the spread appears to be less important than often is suggested in the financial literature.

Key words: microstructure; pricing; spreads; Kalman filter.

JEL classification: C51; G12.

Introduction

The price of securities in financial markets consists of two main components: the fundamental price and the spread. Both components cannot be observed. The size of spread is possibly incurred by order processing costs, inventory holding costs and adverse selection costs of the trade; see the discussion in O'Hara (1995). Also, the spread may vary with factors such as trade size and time of day. The standard approach of estimating the spread is based on the autocovariance structure of the differenced price data or on specific regression techniques. It follows that the prices must be ordered sequentially over time: at each time period only one price is observed.

There are two notions of data sequentiality: one refers to the market and the other refers to the available data set. It can be argued that securities traded at markets with a monopolistic market maker, such as the specialist broker on the New York Stock Exchange, and equities traded at markets in which liquidity is provided by the public limit order book, such as Paris Bourse and Tokyo Stock Exchange, have a price for each time period when a transaction takes place. Data collected from such markets usually maintain the property of data sequentiality.

For competitive dealership markets, such as Chicago Mercantile Exchange and Chicago Board Option Exchange, several dealers negotiate and complete multiple trades simultaneously. Therefore, different prices of the same security float within the market at the same time. The property of price sequentiality may still hold for data sets collected from such markets because of the method of data collection or data manipulation. For example, the average of traded prices within a period can be considered. Data sequentiality remains in this approach but it may lead to a serious loss of information. Different prices are associated with different quantities which have a considerable effect on the spread as we will show. Moreover, such a practical solution is theoretically not satisfactory and we prefer a model-based approach.

In this paper we consider transaction data from the London Stock Exchange (LSE). Most of the securities at the LSE have multiple market makers and several trades of the same security occur simultaneously on a regular basis. It is not straightforward to identify the fundamental price in such markets due to the market itself and due to the lack of sequentiality of the data. Firstly, the working of the London market is not very transparent: security trades are negotiated between several dealers on the phone. Participants in the market do not necessarily observe most of the trading in the market. Also, delays in reporting new transactions take place regularly. Therefore, the fundamental price is affected by so-called "coloured noise". Secondly, data collected from the LSE is intrinsically non-sequential. For example, there were six transactions of the stock of British Telecom taking place on 1 April 1996 at 11:09, while there were ten transactions at 11:10 with five different observed prices. So there is no unique price within the time interval of one minute. The standard approach to estimate the spread using differenced price data can clearly not be used in these cases. This has been our main motivation to develop a new statistical approach to model bid/ask spreads for non-sequential trade markets.

The rest of the paper is organised as follows. Section 1 gives an overview of the contributions in the literature on the empirical features of bid-ask spreads. The statistical model for bid-ask spreads is presented in Section 2 in which we include discussion of the technical features of the model. We also discuss some possible extensions of the model. Section 3 presents an application of the model by using the transaction data of Glaxo Wellcome, British Telecom and Shell Transport which are traded on the London Stock Exchange. It is shown that our model is capable of identifying the fundamental price and the spread in a straightforward manner. Section 4 considers options for further research. The appendices include technical references to regression splines, the Kalman filter smoother and parameter estimation.

1 Bid-ask spreads

1.1 Estimation using differenced prices

The fundamental price and the spread of a security are unobserved components of the trade price. By making appropriate assumptions about the dynamic behaviour of the fundamental price and the spread, they can be estimated from the observed price. Inference about the spread can be based on different approaches, but most of them require some sequential ordering of the data. Roll (1984) initiated the standard approach of estimating the spread using the autocovariance structure of the price differences. Improvements on the original model of Roll are presented by, among others, Stoll (1989) and George, Kaul and Nimalendran (1991). The approach taken by Glosten and Harris (1988) is based on regression techniques using trade indicators. Hasbrouck (1988, 1991a, 1991b) uses vector autoregressive (VAR) models to measure the inventory and information components of trade and quote prices. The statistical analysis of Madhavan and Smidt (1991) is taken from a Bayesian perspective and they have used data sets which are sequentialised using record numbers of transactions. Huang and Stoll (1997) give a very detailed overview of these methods and they unify them to a general approach. These approaches require the use of differenced data, including prices, quotes and trade-direction indicators, as dependent or independent variables. Most data in these studies are ordered sequentially: daily data is ordered by day while intra-daily data is usually ordered by unique time stamps. Examples of studies based on daily data are the original study of Roll (1984), Stoll (1989) who used the last three prices of the day and the work of Affleck-Graves et al. (1994). Studies of foreign exchange markets can adopt the standard approach because the quotes are sequentially ordered; see, for example, Bollerslev and Domowitz (1993), Bollerslev and Melvin (1994) and Bollerslev, Domowitz and Wang (1997). Note that trade records are rarely available for these markets. Studies of futures or options markets analyse trades which are recorded sequentially by the clerks in the exchange; see Chung (1991) and Locke and Venkatesh (1997). It should be stressed here that the observed data sequence may not necessarily be identical to the sequence of trading. For example, members of the New York Stock Exchange are responsible to enter at least 90% of the trades to the Consolidated Tape System withing 90 seconds of execution, which may distort the reporting sequence such that it is not consistent with the trade sequence; see Hasbrouck et al. (1993). However, the data sequence enables the use of differenced prices.

1.2 Non-sequential data

In a competitive dealership market, one may expect dealers to negotiate the trades simultaneously. Multiple deals of a liquid security may be executed nearly at the same time as a result. The price series is not sequential and estimation based on differenced data is not possible. One notable and important example is the trade records retrieved from the Stock Exchange Automated Quotation (SEAQ) system of the London Stock Exchange (LSE). Investigations of the LSE trades are carried out by, among others, Hansch and Neuberger (1993) who study block trading and by de Jong, Nijman and Roëll (1995) who study the spreads of French stocks traded at SEAQ International. The time series properties in these studies have not been considered in full: the mid-quote price is treated as an approximation of the fundamental price and the spread is defined as twice the difference between the transaction price and the mid-quote price. Thus the spread is not measured from changes in prices. It may be argued that the mid-quote is not persistently above or below the fundamental price; see Hansch et al. (1996). On the other hand, Reiss and Werner (1996) argue that the mid-quote does not represent the average quotes of market makers and it may be very different from the fundamental price. Moreover, mid-quote prices may not be available outside the mandatory quote period which is one of the main reasons why these outside trades are excluded in all the studies of the LSE market.

This paper presents a model which preserves the time series properties while data sequentiality is not required. The time series properties are modelled explicitly in order to obtain estimates of the underlying fundamental prices and spreads. Also, the model takes explicitly account of a number of effects which causes the variation of spread such as size of trade and time of trade. The structure of our model is general and it is easy to include more effects into the model. The remainder of this section discusses explanatory factors and other empirical features affecting the spread which have been emphasized by other contributions in the literature.

1.3 Trade size effect

The size of the trade is closely related to its order processing, inventory and information costs. Easley and O'Hara (1987) have argued that the spread increases with the size of the trade because of the adverse selection effect. The effect exists when customers have more information than market makers which implies that large orders indicate good or bad news known by the customer. The spread needs to compensate the loss of the market maker. On the other hand, the market maker deviates from the optimal portfolio by taking the order from the customer, so the spread is increasing in the size of the trade to compensate the loss; see Amihud and Mendelson (1980) and Ho and Stoll (1981, 1983). Furthermore, other authors have emphasized that spread size decreases with trade size because fixed costs are associated with each trade. Striking a deal takes about the same time for the market maker regardless of the order size; see Stoll (1978b), Board and Sutcliffe (1995), de Jong, Nijman and Roëll (1995) and Reiss and Werner (1996). Therefore, it is argued that the size effect is expected to have a so-called U-shape: the spread is big for small and big trades, and it is small for medium trades.

1.4 Intra-day effect

The spread may vary over time in a general fashion but the dominant feature in many data sets is the intra-day variation. Several contributions in the literature have studied the importance of intra-day variation on the New York Stock Exchange (NYSE); see Wood, McInish and Ord (1985), Wood and McInish (1992) and Brock and Kleidon (1992). Most contributions agree that traders have strong demands to re-establish their optimal portfolios at the start of the trading period. Also, market makers may fear that investors have private information at the start of the trading period so they buy or sell shares at the expense of other market makers. At the closing of the market, traders prepare for the non-trading period by adjusting their portfolios in an appropriate way. Brock and Kleidon (1992) summarise other reasons for a higher spread at the end of the trading period: (i) brokers are instructed to execute orders at discretion over the day and, as time passes by, the desire to fill the remaining orders will increase; (ii) the performance of the fund manager is often evaluated on the value of the portfolio at closing time which gives an incentive to execute trades at closing time; (iii) the sales of shares are generally based on closing prices such that mutual funds managers prefer to trade close to closing time.

Outside the NYSE, empirical evidence of the intraday pattern of spreads is mixed. Wang et al. (1994) use data obtained from the Chicago Mercantile Exchange and they find that the spreads of the S&P 500 index futures have an intra-day U-shape. Lee et al. (1993) find that effective spreads of stocks from NYSE and American Stock Exchange (AMEX) exhibit U-shape patterns. On the other hand, Werner and Kleidon (1996) study cross-listed stocks in the United States (US) and the United Kingdom (UK) and they conclude that the spreads in the UK decline during the day. Chan, Chung and Johnson (1995) observe that the spreads of the Chicago Board Option Exchange are smaller at the end of the trading period. Furthermore, Chan, Christie and Schultz (1995) find the inside spreads of NASDAQ stocks remain relatively constant in the morning and decline in the afternoon, and they suggest that differences of intra-day patterns are due to institutional differences. Finally, van Ravenswaaij (1997) finds that only the first hour of trading on Paris Bourse has a considerable impact on the size of the spread.

1.5 Other effects

The level of competition of market makers is negatively related to the spreads; see Demsetz (1968). Competition level may be measured by the number of market makers of a security. The spread is positively related to the risk of returns to compensate market makers when market participants are risk-averse; see Stoll (1978a) for theoretical arguments and Bollerslev and Melvin (1994) for empirical evidence. Furthermore, Stoll (1978a) argues that factors such as the wealth of market makers and holding periods of securities also affect the spread. Easley and O'Hara (1992) argue that high volume of trades signals an information event so that market makers have to increase the spread to protect themselves from adverse selection. On the other hand, Dutta and Madhavan (1997) argue that the positive relationship between volume and spread may result from the monopolistic power of market makers.

Generally, the structure of an individual market may determine the spreads. Competitive dealership markets are not transparent and, therefore, market makers increase the spread to protect themselves from information advantage; see Pagano and Roëll (1992). In contrast, Naik et al. (1994) suggest that market makers may narrow the spread to solicit informed trades.

1.6 Volatility

The variances for the disturbances may not be time invariant. It has been emphasized in the econometric and financial literature that for many financial time series, the disturbances are heteroscedastic and that it can be modelled by autoregressive conditional heteroscedastic (ARCH) structures or by stochastic volatility (SV) specifications. However, in this paper we concentrate on the intra-day volatility. Therefore, we do not treat volatility as a stochastic process but as a deterministic effect on variances of different time periods within the day.

Many contributions in the literature have found a U-shape for the intra-day volatility of security returns; see Wood, McInish and Ord (1985), Park (1993), Chan, Chung and Johnson (1995), Chan, Christie and Schultz (1995) and Werner and Kleidon (1996). The work of Admati and Pfliderer (1988) and Foster and Viswanathan (1990) indicates that volatility and volume are correlated. Furthermore, French and Roll (1986) suggest that trading itself creates volatility and Slezak (1994) argues that the information asymmertry during market closure contributes the uncertainty to the opening. Empirical evidence of the positive relationship between volume and volatility is provided by Jain and Joh (1988). All these indications hint that the heavy trading at the beginning and at the closing of the trading period may be the source of the U-shape volatility. Finally, Jones, Kaul and Lipson (1994) point out that the volatility-volume relationship is essentially a volatility-transaction relationship: the number of transactions is correlated more closely to the volatility than the volume is.

1.7 Discontinuity of trading hours

Trades take place during a limited part of the day and they only take place during weekdays. To reflect the uncertainties caused by the discontinuity of the trading period, the fundamental equity price at the beginning of each trading day is treated as a parameter which must be estimated. The procedure for estimating these initial prices can be implemented within our framework without any additional computational cost.

2 Statistical model for bid-ask spreads

The statistical model for prices of competitive dealership markets with multiple market makers presented below can be used generally for observations which can not necessarily be sequentially ordered over time for all periods.

2.1 The main structure of the model

Suppose that N_t transactions have occured at time t. The trade prices of a security at time t are stacked into the vector y_t and the associated trade quantities are stacked into the vector x_t in the same order as y_t. The elements of vectors y_t and x_t are denoted by $y_{t,i}$ and $x_{t,i}$, respectively, for $i = 1, \ldots, N_t$. When observations are not available for some time point $t = \tau$, the observation τ is treated as missing. The estimation methods which we employ later handle missing observations in a straightforward manner.

The underlying fundamental price of an equity is denoted by the scalar μ_t and it applies to all trade prices within time period t. The model is given by

$$\begin{aligned} y_{t,i} &= \theta_t + s_{t,i} + \varepsilon_{t,i}, & \varepsilon_{t,i} &\sim \mathrm{N}\left(0, \sigma_\varepsilon^2\right), & i &= 1, \ldots, N_t \\ \mu_t &= \mu_{t-1} + q_t + \eta_t, & \eta_t &\sim \mathrm{N}\left(0, \sigma_\eta^2\right), & t &= 1, \ldots, n, \end{aligned} \tag{1}$$

with signal $\theta_t = \mu_t$. The specification for the bid-ask spread $s_{t,i}$ and the adverse selection effect q_t are discussed below. The normal distributed disturbances $\varepsilon_{t,i}$ are mutually independent and uncorrelated with the normal distributed disturbances η_t. The signal θ_t is set equal to the fundamental price μ_t but this can be generalised to more general specifications such as $\theta_t = \mu_t + \rho_t$ with the stationary autoregressive effect ρ_t given by

$$\rho_t = \phi\rho_{t-1} + \xi_t, \qquad \xi_t \sim \mathrm{N}\left(0, \sigma_\xi^2\right), \qquad |\phi| < 1, \tag{2}$$

where the normal distributed disturbances ξ_t are mutually independent and uncorrelated with $\varepsilon_{t,i}$ and η_t. The structure of the model is similar to the ones used by Glosten and Harris (1988) and Huang and Stoll (1997) but model (1) allows the number of trades to vary with time t. Also, the particular specifications for the spread and the adverse selection effect are different; see below.

The assumption of normality for $\varepsilon_{t,i}$ may not be very realistic due to the nature of the observations which have been subjected to rounding functions. Problems related to rounded data have been given some attention in the finance literature; see Harris (1991, 1994), Hausman et.al. (1992) and Chordia and Subrahmanyam (1995). Furthermore, statistical techniques have been developed to deal with the rounding problem and they can be applied to our model. Hasbrouck (1996) considers a discrete bid-ask price model and estimate the model using a nonlinear filtering technique. Manrique and Shephard (1997) extend this analysis by presenting a Bayesian treatment using Markov chain Monte Carlo methods. We believe that in view of our aim and the large sample size, the rounding problem is not the crucial issue in our analysis.

When s_t and q_t are assumed to be zero for all time periods, the model reduces to a simple multivariate random walk plus noise model. Moreover, when $N_t = 1$, for $t = 1, \ldots, n$, the model is the univariate random walk plus noise. The statistical characteristics of this nonstationary time series model are discussed in detail by Harvey (1989). The model is a classical one: the forecast function takes the form of an exponentially weighted moving average (EWMA) and it can be written as the autoregressive integrated moving average model of order (0,1,1), that is the ARIMA(0,1,1) model. Finally, the random walk plus noise model is the most basic structural time series model of Harvey (1989). This allows the model to be extended, for example, by including seasonals and cycles. For example, in our model (1), the random walk component can be extended with the stationary AR(1) component (2) to allow for the short term dynamics in the series.

2.2 The spread

The specification of the spread effect depend on a number of unknown but fixed (non-stochastic) parameters. The spread at time t of the i-th trade is denoted by $s_{t,i}$ and its specification is given by

$$s_{t,i} = d_{t,i}\left(z_t'\gamma + w_x'\delta\right), \qquad x = x_{t,i}, \tag{3}$$

where $d_{t,i}$ is set equal to unity when the i-th trade at time t is a buyer-initiated trade and it is set equal to minus unity when it is a seller-initiated trade:

$$d_{t,i} = \begin{cases} 1, & \text{trade } t,i \text{ is buyer-initiated} \\ -1, & \text{trade } t,i \text{ is seller-initiated} \end{cases}, \tag{4}$$

When the data set contains information about the identities of the traders, as it is the case in our sample, the direction of the trade can be easily identified. When such information is not available, the mid-point of the quote may be used or a tick rule may be applied; see Lee and Ready (1991). Using μ_{t-1} of our model (1), the direction of the trade can also be identified by,

$$d_{t,i} = \begin{cases} 1, & \text{if } y_{t,i} > \mu_{t-1} \\ -1, & \text{if } y_{t,i} < \mu_{t-1} \end{cases}$$

The size of the i-th trade at time t is denoted by $x_{t,i}$.

The part of (3) in brackets represent a linear regression equation with the parameter vectors γ and δ. The explanatory variables z_t and w_x are constructed vectors which are based on the time-of-day and the size $x_{t,i}$, respectively. This representation allows the introduction of piece-wise regression effects which have different parameters for different intervals within the range of, for example, time indices or trade sizes; see Johnston (1984, Chapter 10, Section 2). This specification can be generalised to regression cubic spline functions which join the discrete jumps of the parameter coefficients to a twice differentiable smooth function; see Poirier (1973, 1976). Figure 1 graphically displays a regression line, a piece-wise regression line and a cubic spline function for an artificial set of random points. The cubic spline regression gives the best fit compared to the piece-wise regression with the same number of parameters. Both regression techniques require a set of benchmarks (or knots) to be choosen by the user a priori. Some technical details of regression cubic splines are discussed in Appendix A. Model (1) can be regarded as a multivariate version of the additive nonparametric regression model with autocorrelated components; see Smith, Wong and Kohn (1998).

The Introduction gives an overview of the theory and previous empirical work about the spread. Our model specification for the spread captures the following two features:

1. The intraday effect $z_t'\gamma$ is modelled by a regression cubic spline, in which the x-scale is time. A limited number of knots are equally distributed between 0:00 and 23:59 hours. The spline is restricted to sum to zero to avoid confounding with the total effect.

2. The size effect $w_x'\delta$ is also modelled by a regression spline but here the x-scale is size. The knots are placed between the minimum and maximum number of possible trades. When no institutional boundaries for trade sizes exist, the minimum and maximum can be determined from the data. The starting point of the size spline may be restricted to zero but in this case a constant term is required in the specification for the spread.

An alternative approach for modelling non-linear effects is to transform the underlying variables into a vector of dummy variables and to estimate the corresponding coefficients by generalised least squares techniques (see Lehman and Modest, 1994 and Werner and Kleidon, 1996) or by general method of moments (see Sheikh and Ronn, 1994 and Chan, Chung and Johnson, 1995). This approach is unsatisfactory because of the discontinuity of parameters for different intervals and it is also cumbersome when a lot of dummy variables are required as is usually the case.

2.3 Adverse selection effect

After trades are executed, market makers revise their beliefs about the fundamental price of the security. When more buys than sells take place, the security is probably undervalued and the fundamental price must be adjusted upwards. Similarly, if there are more sells than buys, the price needs to be adjusted downwards. Since the trade information is not transparent in a competitive dealership market, the adjustments may be slow. The adverse selection is modelled by

$$q_t = r_t'\beta = \sum_{j=1}^{S} \beta_j r_{j,t} \qquad \text{where} \qquad r_{j,t} = \sum_{i=1}^{N_{t-j}} d_{t-j,i}\, x_{t-j,i}, \qquad j = 1, \ldots, S, \tag{5}$$

with $\beta = (\beta_1, \ldots, \beta_S)'$ as a fixed unknown vector of coefficients and S as the maximal time when all the details of trades are published. The vector $r_t = (r_{1,t}, \ldots, r_{S,t})'$ contains the sum of the trade volumes multiplied by trade dummies $d_{t,i}$'s and it can be interpreted as the change of inventory of all market makers. If market makers possess all trade information, the adjustment of their beliefs may be very fast and hence $S = 1$. In this case, equation (5) is similar to the adverse-selection term in equation (1) of Huang and Stoll (1997). The differences between their equation and our equation (5) are: (i) we use the observed trade size instead of trade indicators to measure the adverse selection effect; (ii) equation (5) is set up for a competitive dealership market instead for a specialist market.

2.4 Disturbances

The disturbances in our model (1) are normally distributed. The variance of η_t can be expressed as a ratio of the variance of ε_t which we refer to as the signal-to-noise ratio, that is

$$\omega = \sigma_\eta^2 / \sigma_\varepsilon^2.$$

This unknown parameter is fixed for all time points except for certain time periods. Three different periods are distinguished. Different values for σ_ε^2 and σ_η^2 apply to these different time periods. This approach increases the number of parameters to be estimated. Alternatively, the variances can be multiplied by certain constants for these special periods. This is our preferred method of dealing with disturbances associated with the following three time periods.

1. *First and last half hour of trade period.* In Section 1 it is pointed out that the volatility may be relatively large at the beginning and end of the trading period, so the variance of the disturbance ε_t need to be increased.

2. *Outside official trade period.* Most exchange markets impose official trading sessions but sometimes it is allowed to trade outside these periods. For example, the mandatory quote period for the London market is between 8:30 am and 4:30 pm. Special arrangements exist for clients who operate on other markets abroad and who wish to trade outside the official period. Transactions affected by these special trading facilities can be given a less pronounced role in our model by increasing the variance σ_ε^2.

3. *Period of fundamental price correction.* Price adjustements and large fluctuations in prices are expected when new information becomes available to traders. This requires relatively larger values for η_t and ε_t and therefore the corresponding variances should be increased in these periods. The identification of such periods is done by a prior analysis based on model (1) and on checking the absolute values of the estimated η_t's.

The correction factors for these three periods can be treated as fixed and they can be choosen by the user. The estimation results as presented in Section 3 are not very sensitive to different choices of correction factors.

2.5 State space representation

The model can be represented into state space form; see, among others, Harvey (1993) for a general discussion. The state space form encaptures a wide range of linear time series including (dynamic) regression models, the well-known ARIMA models of Box and Jenkins (1976) and the structural time series models of Harvey (1989). The state space form consists of a transition and a measurement equation; they are respectively given by

$$\begin{aligned} \alpha_t &= T_t \alpha_{t-1} + W_t \lambda_w + R_t \eta_t, \quad \eta_t \sim \mathrm{N}(0, Q_t), \qquad t = 1, \ldots, n, & (6) \\ y_{t,i} &= Z_{t,i} \alpha_t + X_{t,i} \lambda_x + \varepsilon_{t,i}, \quad \varepsilon_{t,i} \sim \mathrm{N}\left(0, \sigma_{t,i}^2\right), \qquad i = 1, \ldots, N_t, & (7) \end{aligned}$$

where α_t is the $m \times 1$ state vector. The observation $y_{t,i}$ for time t and subject i is modelled as a linear function of the state vector α_t, the explanatory variable vector $X_{t,i}$ and the disturbance $\varepsilon_{t,i}$. The state vector follows a vector autoregressive process with transition matrix T_t, explanatory matrix W_t and selection matrix R_t for the disturbance vector η_t. The disturbances are mutually independent and uncorrelated with each other. The parameter vectors λ_x and λ_w associated with explanatory variables $X_{t,i}$ and W_t, respectively, allow the inclusion of fixed effects in the model. The explanatory variables can also be used to include dummy effects, interventions and other fixed effects into the model. The initial specification of the initial state vector is given by

$$\alpha_1 \sim \mathrm{N}\left(a, \sigma^2 P\right), \tag{8}$$

where vector a and matrix P are fixed and known. Nonstationary components within the state vector require a diffuse prior; see, among others, Koopman (1997). The matrices T_t, R_t and Q_t and the vectors $Z_{t,i}$ and $X_{t,i}$ are referred to as system matrices and vectors which are assumed deterministic and known. However, a small number of elements within the system matrices and vectors may be unknown. Let us denote the vector of these elements by ψ. The parameter vector ψ can be estimated by maximum likelihood methods. De Jong (1991) and Koopman and Durbin (1998) are references for a more detailed discussion of the state space form.

The standard Kalman filter recursions evaluate the mean of the state vector α_{t+1} conditional on the vectors obervations $y_1, \ldots, y_t$, that is $a_{t+1} = \mathrm{E}(\alpha_{t+1} | y_1, \ldots, y_t)$, where

$$y_t = \begin{pmatrix} y_{t,1} \\ \vdots \\ y_{t,N_t} \end{pmatrix},$$

together with the variance matrix $P_{t+1} = \mathrm{var}(\alpha_{t+1} | y_1, \ldots, y_t)$; see Anderson and Moore (1979). Koopman and Durbin (1998) argue that considerable and, for cases such as the one considered in this paper, dramatic computing savings can be achieved by treating the vector series $y_1, \ldots, y_n$ as the univariate series $y_{1,1}, \ldots, y_{1,N_1}, y_{2,1}, \ldots, y_{n,N_n}$ and applying the Kalman filter to the univariate series. The exact treatment of diffuse priors within the initial state vector variance matrix for the Kalman filter is also simplified considerably using the univariate approach.

The Kalman filter evaluates one-step and multi-step predictions of the state vector and it evaluates one-step ahead prediction errors including their variances. These predictions are interpreted as minimum mean squares linear estimators. The smoothed estimator of the state vector, that is $\hat{\alpha}_t = \mathrm{E}(\alpha_t | y_1, \ldots, y_n)$, and its variance matrix can be computed using a smoothing algorithm which is associated with the Kalman filter. The likelihood function can be constructed using the prediction errors via the prediction error decomposition; see Harvey (1993). In order to get the maximum likelihood estimate for the parameter vector ψ, numerical optimisation routines are used to maximise the likelihood function with respect to the parameter vector ψ. The parameter vectors λ_x and λ_w can be estimated by generalised least squares methods; see de Jong (1991). A straightforward method is to place the parameter vectors λ_x and λ_w in the state vector α_t and to apply the Kalman filter to the augmented state space form; see Harvey (1993) for more details. The Kalman filter and the associated smoothing algorithm are discussed further in Appendix B.

Model (1) is easily put into state space form. The state vector is the scalar μ_t so that $T_t = 1$, $R_t = 1$ and $Q_t = \sigma_\eta^2$ in (6). The adverse selection effect q_t of (5) is modelled via the regression effect $W_t\lambda_w$ so that $W_t = r_t'$ and $\lambda_w = \beta$. The initial state requires a diffuse prior condition, that is

$$\mu_1 \sim \mathrm{N}\{0, \kappa I\},$$

where the diffuse prior κ represents a large scalar value, for example 10^5. The large variance is required because the fundamental price μ_t is modelled as a nonstationary time series process. The observation equation has $Z_{t,i}$ equals unity and the regression effect $X_{t,i}\lambda_x$ is used to model the spread (3) so that $X_{t,i} = (z_t', w_x')$ and $\lambda_x = (\gamma', \delta')'$ with the transaction size $x = X_{t,i}$. The variance $\sigma_{t,i}^2$ is the constant σ_ε^2.

3 Empirical results for the London Stock Exchange

3.1 The LSE data set

The data set is obtained from the London Stock Exchange (LSE) directly via the CD-ROM *Transaction Data Services*, which contains all of the settlement records between January and June 1996. This period contains 125 trading days. Market makers are obliged to report bid and ask prices in SEAQ between 8:30 am and 4:30 pm every weekday, during which most of the trades occur. However, 3.08% of all trades take place outside this time. Market makers at the LSE trade domestic securities in two different trading systems: the Inter-Dealer Broker (IDB) market and the Stock Exchange Automated Quotation (SEAQ) system. The data used in this paper do not include any inter-dealer trading; every observation in our sample comes from a trade executed in SEAQ by a market maker and another market participant. Most of the price data are rounded: 91.22% of the trade prices are measured in pence so the vast majority of observations is discrete. In our sample period, one penny was approximately equal to 0.0154 US dollars.

In this paper we consider three securities: Glaxo Wellcome, British Telecom and Shell Transport. They are among the ten most heavily traded stocks on the London market in 1996. The standard procedure of data editing is applied, see Hansch and Neuberger (1993), Board and Sutcliffe (1995), Reiss and Werner (1996, 1997) and Neuberger and Hansch (1996) for details. Some descriptive statistics are given in Table 1. The settlement records are time-stamped in minutes. There are occasions at which many trades are executed, for example, 89 transactions of British Telecom are reported to be executed on 17 May 1996 at 8:46. Including all of the data in the sample will increase the computational burden substantially, so we only allow ten randomly selected trades in our final sample if there are more than ten trades within a minute. As a result the data set is reduced by 2.42%.

3.2 Details of the model

Figure 2 illustrates the data format of the *Transaction Data Service*. Each trade record contains time of the trade t, the security price $y_{t,i}$ and the trade size $x_{t,i}$ (the quantity). As a result, the variables of model (1) N_t and q_t are known. The unit of $y_{t,i}$ is in pennies, the unit of $x_{t,i}$ is in 100 shares and the time t is in minutes. The trade records also contain buy and sell flags of the trading parties (the so-called buy/sell cap) and market makers are marked as "M". The direction of the trade is the action of the counter party of the market maker. This information determines the trade dummy variable $d_{t,i}$ of (4).

We consider model (1) with $\theta_t = \mu_t$. The spread is modelled as two regression spline functions: one for the time-of-day effect and one for the size effect. The spline functions require a set of knots. We have experimented with different numbers of knots and with different knot positions for the two splines. Although a larger number of knots gives a better fit, the optimal number is based on the right balance between fit and parsimony (that is, the desire to reduce the number of parameters in a model). The Akaike information criterion (AIC) is given by

$$\text{AIC} = 2\,(k - \log L)\,,$$

where L is the value of the likelihood function of the estimated model and k is the number of parameters which is defined as the sum of the sizes of the vectors α_t, λ_x, λ_w and ψ. The AIC is often used for selecting the optimal number of parameters in a model. The model with the lowest AIC value is selected as the appropriate model. The initial positions of the knots are equally spread along possible values of time or size unless prior information is available which insists on placing knots in a specific area.

The time spline is restricted to sum to zero and it is based on four joint intervals (knots). The size spline takes account of the level of the spread and it is based on three knots. The knot positions are

$$\begin{array}{cccc} & \text{Glaxo} & \text{BT} & \text{Shell} \\ \text{time spline} & \begin{bmatrix} 0 \\ 360 \\ 720 \\ 1080 \end{bmatrix} & \begin{bmatrix} 0 \\ 360 \\ 720 \\ 1080 \end{bmatrix} & \begin{bmatrix} 0 \\ 360 \\ 720 \\ 1080 \end{bmatrix} \\ \text{size spline} & \begin{bmatrix} 0 \\ 16000 \\ 32000 \end{bmatrix} & \begin{bmatrix} 0 \\ 22500 \\ 45000 \end{bmatrix} & \begin{bmatrix} 0 \\ 9000 \\ 18000 \end{bmatrix} \end{array}$$

The maximum time lag S considered for the adverse selection effect is set to $S = 2$. The period length for reporting transactions is one hour under the publication rule of the London Stock Exchange which implies that the maximum for S is 60. The inclusion of more lags than 2 increases the computational burden and it does not increase the fit of the model. In the same way as for the number of knots for splines, the maximum time lag S can be determined by the AIC decision rule.

Given the details of the model, the signal-to-noise ratio ω is estimated via numerical optimization of the likelihood function. The likelihood function is evaluated using the Kalman filter; see Appendices B and C. After estimating ω, the Kalman filter is used to compute one-step ahead prediction residuals together with their variances. Figure 3 presents the average of absolute standardised errors for intervals of 10 minutes over the day using all data in our sample. It shows that the errors are larger at the opening and at the closure of the trading period as it is suggested in Section 1.4. Therefore we adjust the variances of the model to take account of these empirical phenomenons. Define the variance multipliers a_t and b_t so that the variances of the disturbances in model (1) become time varying and they are given by

$$\sigma^2_{\varepsilon,t,i} = a_t\sigma^2_\varepsilon, \qquad \sigma^2_{\eta,t} = \omega b_t \sigma^2_\varepsilon, \qquad t = 1, \ldots, n, \qquad i = 1, \ldots, N_t. \tag{9}$$

The variance multipliers for our model are as follows:

1. The variance σ^2_ε is doubled ($a_t = 2$) at the opening of the trading period between 8:31am and 8:50 am and at the closure of the trading period between 4:00 pm and 4:28 pm.

2. The variance σ_ε^2 is multiplied by eight ($a_t = 8$) when trades take place before the offical opening of the trade period and it is multiplied by six ($a_t = 6$) when trades take place after the closure of the trade period.

3. Some standardised one-step ahead residuals are relatively large. When the absolute values of more than two consecutive residuals are greater than some benchmark, a_t and b_t are given values larger than one; see Table 2. Although the adjustments are somewhat arbitrary, most of these adjustments coincide with news reports on, for example, earnings or mergers published by the *Financial Times.* This implies that the fundamental price must be given the flexibility to adjust for changes during these limited periods. Hence the variance adjustments are well justified.

These adjustments can be altered by the user but after some experimentation we obtained satisfactory results with these settings. It should be noted that final results were not very sensitive to a different set of adjustments. By taking into account the variance adjustments, model (1) is re-estimated.

3.3 Model specification: time, size and quantity effects

In this section we investigate whether the size and intra-day effects explain the spread and whether the volume effect explain the fundamental price. We test these hypotheses by using the likelihood ratio test and the Akaike information criterion (AIC). The likelihood ratio (LR) test statistic is given by

$$\mathrm{LR} = -2 * \log(L_0/L) = 2 * (\log L - \log L_0)$$

where L is the likelihood of the estimated model (1) and L_0 is the likelihood of the estimated restricted model. Three restrictions are considered: (i) no existence of size effect on the spread, that is $w_x'\delta = 0$, (ii) no existence of intra-day effect on the spread, that is $z_t'\gamma = 0$, and (iii) no existence of volume effect on the fundamental price, that is $q_t = 0$. Under the restricted hypothesis and assuming normality for the disturbances, the likelihood ratio test statistic is chi-square distributed with degrees of freedom equals the extra number of variables for the unrestricted model. It is already mentioned that the assumption of normality is weak for our data set and therefore we regard the LR statistic as an indicative diagnostic rather than as a formal test.

The distribution of the initial state is diffuse implying that the initial state variance is arbitrarily large. The Kalman filter may behave unstable for the first set of observations. The likelihood function is evaluated via the Kalman filter and it is decided to exclude the first percent of observations from the summation operators of the logged likelihood function; see Appendix C. The LR statistics for the three restrictions are reported in Table 3. The restrictions of no size effect and no volume effect are clearly rejected. The restriction of no time-of-day effect for British Telecom (BT) securities can be accepted and there is also a case to take no account of the time-of-day effect for Glaxo Wellcome (Glaxo) and Shell Transport (Shell).

The AIC statistic is given in the previous section. Table 4 reports the AIC value for the estimated model (1) and the three AIC values for the estimated models with the three subsequent restrictions of no size, no time and no volume effects. The theory of AIC's suggests to select the model with the lowest AIC value and that is the estimated model (1) for Glaxo and Shell but it is the estimated model without the time-of-day effect for BT. However, the AIC values for the estimated model without any restriction and the estimated model without the time-of-day effect are not very different from each other in the case of Glaxo and Shell. The exclusion of other effects causes larger AIC values.

A graphical display of the time-of-day spline $z_j'\gamma$ for $j = 1, \ldots, 1440$ (that is 24 hours of 60 minutes) is presented in Figure 4. In the first column of graphs, the splines are displayed for $j = 1, \ldots, 1440$ and, in the second column, the splines are plotted against $j = 510, \ldots, 990$ which represents the trading period. From the left hand panel of plots we may conclude that the Glaxo time spline has the familiar U-shape, the BT spline is large during the day and the Shell spline is small around noon. The shape of time splines during the trading hours are also different for the three securities as can be observed from the right hand panel of pictures in Figure 4. The spread of Glaxo is large around noon and small at three o'clock, the spread of BT is almost flat everywhere but with a sharp decline in the late afternoon and the time spline of Shell has an inverse U-shape fir the trading period. The variation of the spline is very small in all cases. The difference between the maximal and minimal value of the splines is at most 1.6 pence and only 0.1 pence during trading hours. The reported evidence does not prove the existence of a significant time-of-day effect or an U-shaped time effect within the day.

The size effect for the spread is strong. Figure 5 plots the estimated size spline functions $w_x'\hat{\delta}$ for possible values of the quantity x. The size splines for the three securities exhibit the same pattern: initially, the spread is decreasing with the size increasing and, after some benchmark, the spread is increasing. The phenomenon of a considerable increase of the spread when the size gets large is consistent with the inventory and information theory in the literature. The initial fall of the spread may be attributed to the fixed component of the transaction costs.

3.4 Model misspecification: estimated disturbances

The average of the standardised one-step ahead predictions residuals within each minute are presented in Figure 6 and some summary statistics are shown in Table 6. It can be concluded that the residuals are not normally distributed and exhibit weak autocorrelations. It is mentioned in Section 2 that the assumption of normality is not very realistic and therefore it is not surprising that normality test statistics point to a departure from normality. Figure 7 presents the correlograms of the residuals for the three securities. The first coefficient of the correlogram is modest in all three cases and the coefficients are negligible after ten lags. The existence of serial correlation can partly be explained because the residuals under consideration are an average of a set of residuals within one minute. This type of pooling may introduce some serial correlation in large data sets.

Volatility does not appear to be a problem in our analysis. The squared estimated disturbances η_t possess no serial correlation. Moreover, we have also regressed these squared residuals against the number of transactions N_t and the fit of this regression (measured by the so-called coefficient of determination R^2) was very weak to negligible in all cases. Therefore we can conclude that the imposed multiplication factors for the variances σ_ε^2 and σ_η^2 have successfully dealt with the possible volatility for our three securities.

3.5 Estimated parameter coefficients

Table 5 presents the estimated parameters of model (1) with $\theta_t = \mu_t$ but without the time spline. Firstly, the signal-to-noise ratio ω is reported which is estimated by maximum likelihood. This requires numerical optimisation of the likelihood function which is computed by the Kalman filter; see Appendix C. The estimated ratio ω is roughly the same for Glaxo and Shell and it takes the value of around 0.1. This value is found regularly in a variety of applications. The ratio for BT is much smaller which indicates that the fundamental price evolves more smoothly than the price of the other equities. It may be concluded that new information has less influence on the fundamental price for BT than for Glaxo and Shell. On the other hand, it may also indicate some lack of relevant new information for BT during the period of our sample. The Figures 8, 9 and 10 present the fundamental price μ_t for the three equities and Figure 11 shows an example of the evolution of the fundamental price on a specific day for Shell. This graphical representation of the fundamental price can be very informative for the traders on the market.

Secondly, the estimated coefficients for the size effects are reported together with the corresponding t-statistics. The estimate for parameter vector δ is required to generate a graphical display of the size spline. Again, the size effects appear to be statistically significant since the t-statistics associated with the knot coefficients have values larger than 3.5.

Finally, the estimated coefficients of the adverse selection effect β_1 and β_2 are reported. They are highly significant and they all take positive values as expected. The estimates for β_1 and β_2 can be interpreted as an indication of the depth of the market; see Kyle (1985). For example, the sum of $\hat{\beta}_1$ and $\hat{\beta}_2$ of Glaxo Wellcome is about 0.000075, which means, roughly speaking, that the fundamental price of the security will increase one penny if the volume of buys in the market exceeds that of sells by $1/((\hat{\beta}_1 + \hat{\beta}_2)/100) = 1,333,333$ shares.

4 Discussions and conclusions

The available level of detail in databases of intra-daily transactions data for competitive dealership markets brings mixed blessings to the study of market microstructure. On the one hand, transaction prices and volumes become available for small time intervals, which provide the opportunity for financial analysts to have a better understanding about the market behaviour. On the other hand, the data set is not necessarily ordered sequentially and therefore standard techniques for estimation of the spread may not be applicable.

This paper has presented a simple model to analyse the data of three heavily traded stocks on the London Stock Exchange. The model includes components which allow for the time series properties of the data and the existence of nonlinear effects. The problem of non-sequentiality is solved by putting the model into state space form and to estimate the model by the Kalman filter. The updating recursions of the Kalman filter do not require the dimension of the observational vector to be constant. Moreover, the Kalman filter and associated algorithms can deal with missing observations in a straightforward manner. The underlying fundamental price of the securities is extracted from the data and at the same time the effect of volumes on the price and the explanatory factors of the spread are estimated. Strong evidence is found to support the fact that spread is a nonlinear function of the trade size. The evidence of intra-daily effects is less strong. Apart from the weak significance of test statistics, the estimated time splines for the securities have three different shapes and the variation of the three splines are small. This has convinced us that the time-of-day effect is not a determinant factor of the spread on the London market. In conclusion, our analyses of the three securities have been successful. Model (1) can generate a wide range of Tables and Figures, such as the ones reported in this paper, which provide detailed information about the available transaction data. Empirical studies based on our model and analysis may contribute to a better understanding of the fundamentals of the bid/ask spread.

The basic model (1) can accommodate almost all features of the fundamental price and the spread. Variations of the model may be used to address other questions of the market microstructure. However, it may be interesting to modify the model specification and to improve the estimation techniques for further research. The strategy of imposing different disturbance variances for specific time periods is not satisfactory and it can be altered. The introduction of an intra-daily variance function can be considered and its specification may rely on dummy variables or smooth functions such as the cubic spline. Also, it is argued that the normality assumption is not realistic because the data is subjected to rounding functions. Furthermore, the erratic behaviour of financial time series may require error distributions with heavier tails. Thus a more in-depth analysis of the data requires more advanced estimation techniques. However, such improvements come with a price. The model specification in this paper consists of one parameter which needs to be estimated by numerical optimisation. The introduction of intra-daily variances will lead to an increase of computational costs.

Appendices

A Regression spline functions

The regression spline function is defined as a smooth function through the data points y_t which are a response to the scalar series x_t, for which $x_t < x_{t+1}$ and $t = 1, \dots, n$. The spline model is

$$y_t = \theta(x_t) + \varepsilon_t, \qquad \mathrm{E}(\varepsilon_t) = 0, \qquad \mathrm{var}(\varepsilon_t) = \sigma^2,$$

where $\theta(\cdot)$ is a smooth function which is based on $k+1$ knot points $(x_0^\dagger, y_0^\dagger), \dots, (x_k^\dagger, y_k^\dagger)$. The smoothness of $\theta(\cdot)$ is created by setting its second derivative with respect to x as a linear function of $k+1$ coefficients, that is

$$\theta_i''(x) = [(x_i^\dagger - x)/d_i]a_{i-1} + [(x - x_{i-1}^\dagger)/d_i]a_i$$

with $d_i = x_i^\dagger - x_{i-1}^\dagger$ and $\theta_i(x) = \theta(x)$ for $x_{i-1}^\dagger < x < x_i^\dagger$ and $i = 1, \dots, k$. The $k+1$ coefficients a_i are assumed fixed and they can be identified by solving a linear set of equations. These regression spline equations are obtained as follows: (i) by considering $\theta_i''(x)$ and using standard integration rules, we get expressions for $\theta_i(x)$; (ii) we enforce the spline function $\theta_i(x)$ at $x = x_i^\dagger$ to be equal to the known value of $y_i^\dagger$; (iii) we restrict the first derivative to be continuous by enforcing $\theta_i'(x_i^\dagger) = \theta_{i+1}'(x_i^\dagger)$ for $i = 1, \dots, k-1$. Step (ii) leads to a linear expression for $\theta_i(x)$ in terms of $y_i^\dagger$ and a_i, for $i = 0, \dots, k$. Step (iii) leads to $k-1$ linear equations for the $k+1$ coefficients $a_0, \dots, a_k$ in terms of $y_0^\dagger, \dots, y_k^\dagger$. The 'natural' restrictions $a_0 = a_k = 0$ allow solving this linear system with respect to the remaining coefficients a_i for $i = 1, \dots, k-1$. The spline function can now be fully expressed in terms of $y_0^\dagger, \dots, y_k^\dagger$ by

$$\theta(x_t) = \theta_i(x_t) = b_{0,t}y_0^\dagger + \dots + b_{k,t}y_k^\dagger, \qquad x_{i-1}^\dagger < x_t < x_i^\dagger, \qquad , t = 1, \dots, n,$$

where the weights $b_{0,t}, \dots, b_{k,t}$ depend on the knot positions $x_0^\dagger, \dots, x_k^\dagger$ and the value for (or the position of) x_t. For a given set of values $y_0^\dagger, \dots, y_k^\dagger$, the spline function can be computed for any $x_0^\dagger < x < x_k^\dagger$. The regression spline can be expressed as

$$\theta(x_t) = b_t' y^\dagger,$$

where $b_t = (b_{0,t}, \dots, b_{k,t})'$ and $y^\dagger = (y_0^\dagger, \dots, y_k^\dagger)'$. Consequently, the spline model can be expressed as the standard regression model

$$y_t = b_t' y^\dagger + \varepsilon_t,$$

where parameter vector $y^\dagger$ can be estimated by least squares techniques. In the case of our model (1), the parameter vectors for the two different splines are estimated by generalised least squares. More details are given by Poirier (1973, 1976). The generalisation of time-varying regression splines within the state space framework are developed by Harvey and Koopman (1993).

B Kalman filter smoother

Consider the state space model (6) and (7). The Kalman filter evaluates the minimum mean squared linear estimator of the state vector, conditional on 'past' observations, together with its variance matrix. We follow the treatment of Koopman and Durbin (1998) and we exclude the regression vectors λ_x and λ_w from the state space model. Define $a_{t,1} = \mathrm{E}(\alpha_t | Y_{t-1})$ and

$a_{t,i} = \mathrm{E}\left(\alpha_t | Y_{t-1}, y_{t,1}, \ldots, y_{t,i-1}\right)$ with $P_{t,1} = \mathrm{var}\left(\alpha_t | Y_{t-1}\right)$ and $P_{t,i} = \mathrm{var}\left(\alpha_t | Y_{t-1}, y_{t,1}, \ldots, y_{t,i-1}\right)$, for $i = 2, \ldots, N_t$, where $Y_t = \{y_{1,1}, \ldots, y_{1,N_1}, y_{2,1}, \ldots, y_{t,N_t}\}$. The filtering equations are given by

$$a_{t,i+1} = a_{t,i} + K_{t,i} F_{t,i}^{-1} v_{t,i}, \qquad P_{t,i+1} = P_{t,i} - K_{t,i} F_{t,i}^{-1} K_{t,i}', \tag{10}$$

where

$$v_{t,i} = y_{t,i} - Z_{t,i} a_{t,i}, \qquad F_{t,i} = Z_{t,i} P_{t,i} Z_{t,i}' + \sigma_{t,i}^2, \qquad K_{t,i} = P_{t,i} Z_{t,i}', \tag{11}$$

for $i = 1, \ldots, p_t$ and $t = 1, \ldots, n$. This formulation has $v_{t,i}$ and $F_{t,i}$ as scalars and $K_{t,i}$ as a column vector. The transition from time t to time $t+1$ is achieved by the relations

$$a_{t+1,1} = T_t a_{t,p_t+1}, \qquad P_{t+1,1} = T_t P_{t,p_t+1} T_t' + R_t Q_t R_t'. \tag{12}$$

These forward recursions are initialised by $a_{1,1} = a$ and $P_{1,1} = P$ as given by (8).

Minimum mean squared linear estimators using all observations Y_n are evaluated by a smoothing algorithm which require output of the Kalman filter. The basic smoothing recursions operate backwards and the equations are given by

$$\begin{array}{rclrcl} r_{t,i-1} &=& Z_{t,i}' F_{t,i}^{-1} v_{t,i} + L_{t,i}' r_{t,i}, & N_{t,i-1} &=& Z_{t,i}' F_{t,i}^{-1} Z_{t,i} + L_{t,i}' N_{t,i} L_{t,i}, \\ r_{t-1,p_t} &=& T_{t-1}' r_{t,0}, & N_{t-1,p_t} &=& T_{t-1}' N_{t,0} T_{t-1}, \end{array} \tag{13}$$

where $L_{t,i} = I - K_{t,i} Z_{t,i} F_{t,i}^{-1}$, for $i = p_t, \ldots, 1$ and $t = n, \ldots, 1$. The initialisations are $r_{n,p_n} = 0$ and $N_{n,p_n} = 0$. The equations for r_{t-1,p_t} and N_{t-1,p_t} do not apply for $t = 1$.

The output of recursions (13) can be used to construct the smoothed estimator of the disturbances, that is, for example, $\widehat{\varepsilon}_t = \mathrm{E}\left(\varepsilon_t | Y_n\right)$, together with their corresponding variances. The smoothed disturbances are computed by

$$\begin{array}{rclrcl} \widehat{\varepsilon}_{t,i} &=& \sigma_{t,i}^2 F_{t,i}^{-1} \left(v_{t,i} - K_{t,i}' r_{t,i} \right), & \mathrm{var}\left(\widehat{\varepsilon}_{t,i}\right) &=& \sigma_{t,i}^4 F_{t,i}^{-2} \left(F_{t,i} + K_{t,i}' N_{t,i} K_{t,i} \right). \\ \widehat{\eta}_t &=& Q_t R_t' r_{t,0}, & \mathrm{var}\left(\widehat{\eta}_t\right) &=& Q_t R_t' N_{t,0} R_t Q_t, \end{array} \tag{14}$$

for $t = n, \ldots, 1$. The proofs and more general results for smoothed disturbances are given by Koopman (1993).

The smoothed state vector $\hat{\alpha}_t = \mathrm{E}\left(\alpha_t | Y_n\right)$ and variance matrix $V_t = \mathrm{var}\left(\alpha_t | Y_n\right)$ also use (13) and they can be evaluated by

$$\hat{\alpha}_t = a_t + P_t r_{t-1}, \qquad V_t = P_t - P_t N_{t-1} P_t, \tag{15}$$

for $t = n, \ldots, 1$. A substantial amount of additional storage space is required for a_t and P_t. Proofs of (13) and (15) are given by de Jong (1988) and Kohn and Ansley (1989). A more efficient algorithm for calculating the smoothed estimator of the state vector only is given by

$$\widehat{\alpha}_{t+1} = T_t \widehat{\alpha}_t + R_t \widehat{\eta}_t, \qquad t = 1, \ldots, n, \tag{16}$$

with $\widehat{\alpha}_1 = a + P r_0$ and $\widehat{\eta}_t$ is given by (14); see Koopman (1993) for a discussion.

The Kalman filter smoother also provides a general procedure to handle missing observations in time series. When no observations are available for a certain time period τ, or a sequence of time periods, the dimension $p_\tau = 0$ and the updating equation (12) is applied. The smoothing recursions adjust naturally to this situation. Compared to other treatments of missing observations in statistics, this approach is very simple.

C Maximum likelihood estimation

Consider the state space model (6) and (7) with system matrices and vectors depending on the parameter vector ψ. For a given vector ψ, the output of the Kalman filter is used to construct the likelihood function. Harvey (1993) shows how the likelihood function of the state space model with normal distributed disturbances can be calculated via the prediction error decomposition. The log-likelihood function is given by

$$\log L(\psi) = \text{constant} - \frac{1}{2}\sum_{t=1}^{n}\sum_{i=1}^{p_t} \log F_{t,i} + v_{t,i}^2/F_{t,i},$$

where $v_{t,i}$ and $F_{t,i}$ are obtained from the Kalman filter which depend on parameter vector ψ. In the context of state space models, maximum likelihood estimation refers to numerically optimising the log-likelihood function with respect to ψ.

References

Admati, A. and P. Pfleiderer, 1988, "A Theory of Intraday Trading Patterns: Volume and Price Variability", *Review of Financial Studies* 1, 3-40.

Affleck-Graves, J., S. Hedge and R. Miller, 1994, "Trading Mechanisms and the Components of the Bid-Ask Spread", *Journal of Finance* 49, 1471-1488.

Amihud, Y. and H. Mendelson, 1980, "Dealership Market: Market Making with Inventory", *Journal of Financial Economics* 8, 31-53.

Anderson, B. and J. More, 1979, *Optimal Filtering*, Prentice Hall, Englewood Cliffs, New Jersey.

Bollerslev, T. and I. Domowitz, 1993, "Trading Patterns and Prices In the Interbank Foreign-Exchange Market", *Journal of Finance* 48, 1421-1443.

Bollerslev, T., I. Domowitz and J. Wang, 1997, "Order Flow and The Bid-Ask Spread: an Inpirical Probability Model of Screen-Based Trading", *Journal of Economic Dynamics and Control* 21, 1471-1491.

Bollerslev, T. and M. Melvin, 1994, "Bid-Ask Spreads and Volatility in the Foreign Exchange Market: an Empirical Analysis", *Journal of International Economics* 36, 355–372.

Board, J. and C. Sutcliff, 1995, "The Effects of Trade Transparency in the London Stock Exchange", Project Report Jointly Commissioned by London International Financial Futures and Options Exchange and the London Stock Exchange.

Box, G. and G. Jenkins, 1976, *Time Series Analysis: Forecasting And Control* (Revised ed.), San Francisco, CA: Holden Pay, .

Brock, W. and A. Kleidon, 1992, "Periodic Market Closure and Trading Volume: a Model of Intraday Bids and Asks", *Journal of Economic Dynamics and Control* 16, 451-489.

Chan, K., Y. Chung and H. Johnson, 1995, "The Intraday Behavior of Bid-Ask Spreads for NYSE Stocks and CBOE Options", *Journal of Financial and Quantitative Analysis* 30, 329–346.

Chan, K.C., W. Christie and P. Schultz, 1995, "Market Structure and the Intraday Pattern of Bid-Ask Spreads for NASDAQ Securities", *Journal of Business* 68, 35-60.

Chordia, T. and A. Subrahmanyam, 1995, "Market Making, the Tick Size, and Payment-for-Order Flow: Theory and Evidence", *Journal of Business* 68, 534-575.

Chung, Y., 1991, "A Transactions Data Test of Stock Index Futures Market: Efficiency and Index Arbitrage Profitability", *Journal of Finance* 46, 1791-1809.

Demsetz, H., 1968, "The Costs of Transacting", *Quarterly Journal of Economics* 82, 33-53.

Dutta, P. and A. Madhavan, 1997, "Competition and Collusion In Dealer Markets", *Journal of Finance* 52, 245-276.

Easley, D. and M. O'Hara, 1987, "Price, Trade Size, and Information in Securities Markets", *Journal of Financial Economics* 19, 69-90.

Easley, D. and M. O'Hara, 1992, "Time and the Process of Security Price Adjustment", *Journal of Finance* 47, 577–605.

Foster, F.D. and S. Viswanathan, 1990, "A Theory of the Interday Variations in Volume, Variance, and Trading Costs in Securities Markets", *Review of Financial Studies* 3, 593-624.

French, K. and R. Roll, 1986, "Stock-Return Variances: the Arrival of Information and the Reaction of Trades", *Journal of Financial Economics* 17, 5-26.

George, T., G. Kaul and M. Nimalendran, 1991, "Estimation of the Bid-Ask Spread and its Components: a New Approach", *Review of Financial Studies* 4, 623-656.

Glosten, L. and L. Harris, 1988, "Estimating the Components of the Bid-Ask Spread", *Journal of Financial Economics* 21, 123-142.

Hansch, O. and A. Neuberger, 1993, "Block Trading on the London Stock Exchange", Working Paper 182,Institute of Finance and Accounting, London Business School.

Hansch, O., N. Naik and S. Viswanathan, 1996, "Does Inventory Matter in Dealership Markets: Evidence From the London Stock Exchange", Working Paper 225, Institute of Finance and Accounting, London Business School.

Harris, L., 1991, "Stock Price Clustering and Discreteness", *Review Of Financial Studies* 4, 389–415.

Harris, L., 1994, "Minimum Price Variations, Discrete Bid-Ask Spreads, and Quotation Sizes", *Review of Financial Studies* 7, 149–178.

Harvey, A.C., 1989, *Forecasting, Structural Time Series Models and the Kalman Filter*,Cambridge University Press, Cambridge.

Harvey, A.C., 1993, *Time Series Models* (2nd ed.), Harvester Wheatsheaf, London.

Harvey, A.C. and S.J. Koopman, 1993, "Forecasting Hourly Electricity Demand Using Time-Varying Splines", *Journal of the American Statistical Association* 88, P.1228-1236.

Hasbrouck, J., 1988, "Trades, Quotes, Inventories and Information", *Journal of Financial Economics* 22, 229-252.

Hasbrouck, J., 1991a, "Measuring the Information Content of Stock Trades", *Journal of Finance* 46, 179-207.

Hasbrouck, J., 1991b, "The Summary Informativeness of Stock Trades: An Econometric Analysis", *Review of Financial Studies* 4, 571-595.

Hasbrouck, J., 1996, "The Dynamics of Discrete Bid and Ask Quotes", working paper, Stern School of Business, New York University.

Hasbrouck, J., G. Sofianos and D. Sosebee, 1993, "New York Stock Exchange Systems and Trading Procedures", Working Paper 93-01, New York Stock Exchange.

Hausman, J., A. Lo and A.C. Mackinlay, 1991, "An Ordered Probit Analysis of Transaction Stock Prices", *Journal of Financial Economics* 31, 319–379.

Ho, T. and H. Stoll, 1981, "Optimal Dealer Pricing Under Transactions and Return Uncertainty", *Journal of Financial Economics* 9, 47-73.

Ho, T. and H. Stoll, 1981, 1983, "The Dynamics of Dealer Markets Under Competition", *Journal of Finance* 47, 247-270.

Huang, R. and H. Stoll, 1997, "The Components of Bid-Ask Spread: A General Approach", *Review of Financial Studies* 10, 995-1034.

Jain, P. and G. Joh, 1988, "The Depencence Between Hourly Prices And Trading Volume", *Journal of Financial and Quantitative Analysis* 23, 269–283.

Johnston, J., 1984, *Econometric Methods* (3rd ed.), Mcgraw-Hill, Singapore.

Jones, C.M., G. Kaul and M.L. Lipson, 1994, "Transactions, Volume and Volatility", *Review of Financial Studies* 7, 631-651.

Jong de, F., T. Nijman and A. Roëll, 1995, "A Comparison of The Costs of Trading French Shares on the Paris Bourse and on SEAQ International", *European Economic Review* 39, 1277-1301.

Jong de, P., 1988, "A Cross Validation Filter for Time Series Models", *Biometrika*, 75, 594-600.

Jong de, P., 1991, "The Diffuse Kalman Filter", *Annals of Statistics* 19, 1073-1083.

Kohn, R., and Ansley, C.F. 1989, "A Fast Algorithm for Signal Extraction, Influence and Cross-Validation in State Space Models", *Biometrika*, 76, 65-79.

Koopman, S.J., 1993, "Disturbance Smoother for State Space Models", *Biometrika*, 80, 117-126.

Koopman, S.J., 1997, "Exact Initial Kalman Filtering and Smoothing for Nonstationary Time Series Models", *Journal of the American Statistical Association*, 92, 1630-1638.

Koopman, S.J. and J. Durbin, 1998, "Fast Filtering and Smoothing for Multivariate State Space Models", working paper, Tilburg University.

Kyle, A., 1985, "Continuous Auctions and Insider Trading", *Econometrica* 53, 1315-1335.

Lee, C., B. Mucklow and M. Ready, 1993, "Spreads Depths and the Impact of Earnings Information: an Intraday Analysis", *Review of Financial Studies* 6, 345–374.

Lee, C. and M. Ready, 1991, "Inferring Trade Direction from Intraday Data", *Journal of Finance* 46, 733–746.

Lehmann, B. and D. Modest, 1994, "Trading Liquidity On the Tokyo Stock Exchange: a Bird's-Eye View", *Journal of Finance* 49, 951-984.

Locke, P. and P. Venkatesh, 1997, "Futures Market Transaction Costs", *Journal of Futures Markets* 17, 229–245.

London Stock Exchange, 1997, *The Stock Exchange Fact Book* 1996.

Madhavan, A. and S. Smidt, 1991, "A Bayesian Model of Intraday Specialist Trading", *Journal of Financial Economics* 30, 99-134.

Manrique, A. and N. Shephard, 1997, "Likelihood Analysis of a Discrete Bid/Ask Price Model for a Common Stock", working paper, Oxford University.

Naik, N., A. Neuberger and S. Viswanathan, 1994, "Disclosure Regulation in Competitive Dealership Markets: Analysis of the London Stock Exchange", Working Paper 183, Institute of Finance and Accounting, London Business School.

Neuberger, A. and O. Hansch, 1996, "Strategic Trading By Market Makers On the London Stock Exchange", Working Paper 224, Institute of Finance and Accounting, London Business School.

O'Hara, M., 1995, *Market Microstructure Theory*, Blackwell, Cambridge, Mass.

Pagano, M. and A. Roëll, 1992, "Auction and Dealership Markets - What Is the Difference", *European Economic Review* 36, 613-623.

Park, H., 1993, "Trading Mechanisms and Price Volatility: Spot Versus Futures", *Review of Economics and Statistics* 75, 175–179.

Poirier, D., 1973, "Piecewise Regression Using Cubic Splines", *Journal of American Statistical Association* 68, 515-524.

Poirier, D., 1976, *The Econometrics of Structural Change with Special Emphasis On Spline Functions*, North Holland, Amsterdam.

Ravenswaaij Van, M., 1997 "Intraday Patterns and the Bid-Ask Spread On the Paris Bourse", working paper, Tilburg University.

Reiss, P. and I. Werner, 1996, "Transaction Costs in Dealer Markets: Evidence From the London Stock Exchange", in A. Lo (ed.), *The Industrial Organisation and Regulation of the Securities Industries*, University of Chicago Press, London.

Reiss, P. and I. Werner, 1997, "Interdealer Trading: Evidence from London", working paper, Stanford University.

Roll, R., 1984, "A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market", *Journal of Finance* 39, 1127-1139.

Sheikh, A. and E. Ronn, 1994, "A Characterization of the Daily And Intraday Behavior of Returns On Options", *Journal of Finance* 49, 557-579.

Slezak, S., 1994, "A Theory of the Dynamics of Security Returns Around Market Closures", *Journal of Finance* 49, 1163-1211.

Smith, M., C.-M. Wong and R. Kohn, 1998, "Additive Nonparametric Regression with Autocorrelated Errors", *Journal of the Royal Statistical Society Series B*, 60, 311-332.

Stoll, H., 1978a, "The Supply of Dealer Services in Securities Markets", *Journal of Finance* 33, 1133-1151.

Stoll, H., 1978b, "The Pricing of Security Dealer Services: an Empirical Study of NASDAQ Stocks", *Journal of Finance* 38, 1153-1172.

Stoll, H., 1989, "Inferring the Components of the Bid-Ask Spread: Theory and Empirical Tests", *Journal of Finance* 44, 115-134.

Wang, G., R. Michalski, J. Jordon and E. Moriority, 1994, "An Intraday Analysis of Bid-Ask Spreads and Price Volatility in the S&P 500 Index Futures", *Journal of Futures Markets* 14, 837–859.

Werner, I. and A. Kleidon, 1996, "U.K. and U.S. Trading of British Cross-Listed Stocks: an Intraday Analysis of Market Integration", *Review of Financial Studies* 9, 619–664.

Wood, R. and T. Mcinish, 1992, "An Analysis of Intraday Patterns in Bid/Ask Spreads for NYSE Stocks", *Journal of Finance* 47, 753-764.

Wood, R., T. Mcinish and J. Ord, 1985 "An Investigation Of Transactions Data for NYSE", *Journal of Finance* 40, 723-739.

Table 1: **Descriptive statistics of sample stocks**

Company name		Glaxo	BT	Shell
Industry		Pharmaceutical	Telecommunications	Oil and Gas
Original sample				
number of observations		61,111	124,694	48,660
total trading volumes		474,829	912,699	297,926
max number of trades in 1 minute		38	89	56
Constructed sample				
number of trades	buy	31,734	29,198	13,301
	sell	28,476	91,797	34,273
	total	60,210	120,995	47,584
price	mean	852.00	360.09	873.84
	max	970.00	386.50	820.00
	min	765.00	326.00	949.00
	std	50.99	14.34	33.41
size of trade	mean	7,790	7,395	6,203
	max	3,200,000	4,484,000	1,798,800
	min	1	1	1
total trading volumes	buy	241,965	447,720	132,179
	sell	227,113	460,036	162,986
	total	469,078	907,756	295,165
number of market makers		19	19	19

The constructed sample consists of the first ten largest trades with the same time stamp. The directions of the trades are identified by the capacity flags which are described in Section 3.2. The unit of volumes is 1,000 shares, the unit of the size of trade is one share, and the price unit is one penny.

Table 2: **The timing and magnitude of changes in variances**

stock	date	time	a_t	b_t	events
Glaxo	31/01	08:03		100	Optimism that a cocktail of its AZT and 3TC drugs could be a weapon against HIV.
	06/03	09:13		40	ABN Amro HG downgrades the share.
		10:48-10:55	6.67	800	Galxo may give up US patent fight. Disappointment on the full-year figures.
	26/03	10:05-10:20	2	20	Glaxo is about to link up with Pfizer.
	16/05	14:10-14:50	4	1000	A better-than-expected report on sales
		15:46-15:56		2	growth.
BT	06/03	11:45		50	The government plans to liberalise the overseas telephone call market
	08/03	13:35-14:25		8	
	28/03	15:59-16:06	2.5	200	Dealers on bid alert as speculation builds up
	29/03	07:46-08:07	50	600	BT and Cable &Wireless re-open talks on merger
	02/04	11:50-12:00		5	France Telecom and Deutsch Telecom indicate they are not interested in bidding for Mercury
	18/04	08:40-09:00		5	BT and C&W may announce deal
	02/05	17:46-	20	1000	Shock news that BT had terminated merger talks with Cable and Wireless came after market hours.
	03/05	-8:02	20		BT marked down sharply from the outset on news of the groups' doomed merger.
	17/05	09:10		50	BT and BSkyB plan joint venture.
	10/06	09:12-09:45		6	Markets welcome ambitious venture with MCI.
Shell	15/02	10:16		80	23% rise in dividends
		13:15-14:50		5	Bad result for the 4th quarter 1995 Downgraded by Goldman Sachs and ABN Amro in New York opening
	08/03	13:55-14:35	4	150	The stock could not resist the pressure from Wall Street
	09/05	10:00-10:08	6.67	200	Recorded quarterly result of net income. Recommended by Goldman Sachs

This table shows the time when a_t and b_t, defined in equation (9), are adjusted. The magnitude of the adjustments is also shown. Events are news items of the companies reported in the *Financial Times* on the same day or the next day of the adjustments.

Table 3: **LR statistics for size, time and volume effects**

		Glaxo		BT		Shell		df
Proposed model	$\log L$	-9897.07		19941.92		4054.34		
			LR		LR		LR	
No size effect	$\log L_0$	-9933.33	72.52	19778.89	326.06	3932.50	243.68	2
No time effect	$\log L_0$	-9903.70	13.26	19940.31	3.22	4040.84	27.00	3
No volume effect	$\log L_0$	-9987.14	180.14	19838.05	207.74	3952.20	204.28	2

The proposed model is equation (1) with size and time spline as in (3) and with adverse selection effect q_t defined in (5). The model without size effect is the same as the proposed model except that $w_x'\delta$ of (3) is dropped and replaced by a constant term for the spread. The model without time effect is the proposed model without $z_t'\gamma$ of (3). The model without volume effect is the proposed model with no q_t in (1). The log-likelihood is defined in Appendix C, the likelihood ratio is defined as LR= $-2 * \log(L_0/L) = 2 * (\log L - \log L_0)$, and df is the difference of the number of parameters between the proposed model and the model under the null hypothesis. The critical value for $\chi^2_{(2,0.99)}$ is 9.21 and $\chi^2_{(3,0.99)}$ is 11.341.

Table 4: **AIC statistics for size, time and volume effects**

AIC	Glaxo	BT	Shell	k
Proposed model	19814.1	-39863.8	-8089.7	10
No size effect	19882.7	-39541.8	-7849.0	8
No time effect	19821.4	-39866.6	-8067.7	7
No volume effect	19990.3	-39660.1	-7889.4	6

The same models are considered as for Table 3. AIC is defined as AIC$= 2\,(k - \log L)$, where k is the number of parameters, including the signal-to-noise ratio ω.

Table 5: **Parameter estimates**

	Glaxo		BT		Shell	
ω	0.115461		0.009990		0.116872	
σ_η^2	0.784730		0.219089		0.341186	
σ_ε^2	0.090606		0.014243		0.039875	
		t-value		*t-value*		*t-value*
δ_1	1.2759	342.81	0.6939	549.71	0.9664	307.48
δ_2	0.7822	3.81	0.4670	10.65	0.6670	8.62
δ_3	5.3844	4.41	4.6262	17.12	5.3851	12.09
β_1	6.7621e-05	8.93	9.1462e-06	7.02	7.7344e-05	8.90
β_2	1.7988e-05	2.36	1.2122e-05	9.25	2.4020e-05	2.79

This table presents the parameters of the final model, that is without time spline. The signal-to-noise ratio ω is $\sigma_\eta^2/\sigma_\varepsilon^2$, where σ_ε^2 and σ_η^2 are respectively the variance of $\varepsilon_{t,i}$ and η_t, as defined in (1). The estimates are reported of the elements δ_1, δ_2 and δ_3 of the size spline vector δ which is defined in (3). The estimates for β_1 and β_2, the elements of vector β of (5), are also reported. The usual t-values are also reported.

Table 6: **Summary statistics for standardized residuals**

	Glaxo	BT	Shell
mean	0	0	0
variance	1	1	1
skewness	0.0819	0.1787	0.0869
kurtosis	6.1759	7.7223	6.0047
$\rho(1)$	0.1896	0.2659	0.1639

This Table presents the summary statistics of the average standardised residuals $u_{t,i}$ which are defined in equation (17).

Description of Figures

Figure 1: Example of regression, piece-wise regression and cubic spline. Three regression methods are used to fit the random points in the diagram: the solid line is of the OLS regression, the dash line is of the piecewise regression, and the dotted line is of the cubic spline regression.

Figure 2: An illustration of the LSE data set. This is a simplified example of settlement records retrieved from *Transaction Data Service.* Each trade record consists of the time stamp of the trade (to the nearest minute), the codes of the buyer and the seller (Buy/Sell firm), the capacity flag of the traders (Buy/Sell cap), of which market makers are marked with an "M", the price of the trade (in pence), and the size of the trade (in number of shares).

Figure 3: Means of absolute standardised errors of ten-minute intervals. The panels (a), (b) and (c) are respectively the bar charts of means of absolute standardised errors of Glaxo, BT and Shell. The standardised error $u_{t,i}$ is defined as

$$u_{t,i} = v_{t,i}/F_{t,i}^{\frac{1}{2}}, \tag{17}$$

where prediction error $v_{t,i}$ and its variance $F_{t,i}$ are computed as in equation (11). The absolute values of $u_{t,i}$ are grouped by the time of the trade in ten-minute intervals for which the means are calculated.

Figure 4: Estimated time splines of the spreads. The three panels on the left-hand side are the estimated time splines of Glaxo, BT and Shell, respectively. The estimated time spline is defined according to equation (3) as $z_j'\hat{\gamma}$ for $j = 1, \ldots, 1440$. The splines are measured in pences. The panels on the right-hand side are the estimated time splines $z_j'\hat{\gamma}$ for $j = 510, \ldots, 990$. This range of index j reflects the mandatory quote period from 8:30 to 16:30.

Figure 5: Estimated size splines of the spreads. The three panels are the estimated size splines of Glaxo, BT and Shell, respectively. The estimated size spline is defined according to equation (3) as $w_x'\hat{\delta}$ and is measured in pence. The quantity x is measured in number of shares.

Figure 6: Average of prediction residuals for each minute. The three panels are plots of the average of residuals of Glaxo, BT and Shell, respectively, against time. The average residual is defined as

$$\bar{u}_t = \sum_{s=1}^{N_t} u_{t,i} \tag{18}$$

where $u_{t,i}$ is defined in equation (17).

Figure 7: Correlogram for prediction residuals. Three panels are respectively plots of the autocorrelation functions of Glaxo, BT and Shell against the lag $\tau = 1, 2, ..., 100$. The vertical axis represents the function, which is defind as

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)} = \frac{\sum_{t=\tau+1}^{T} \bar{u}_t \bar{u}_{t-\tau}/(T-\tau)}{\sum_{t=1}^{T} \bar{u}_t^2/T},$$

where $\bar{u}_t$ is defined in equation (18).

Figure 8: Observed and smoothed fundamental price of Glaxo Wellcome. This figure presents the observed prices of Glaxo, $y_{t,i}$, indicated by dots, and the estimated fundamental price μ_t, represented by the solid line.

Figure 9: Observed and smoothed fundamental price of British Telecommunications. This figure presents the observed prices of BT, $y_{t,i}$, indicated by dots, and the estimated fundamental price μ_t, represented by the solid line.

Figure 10: Observed and smoothed fundamental price of Shell Transport. This figure presents the observed prices of Shell, $y_{t,i}$, indicated by dots, and the estimated fundamental price μ_t, represented by the solid line.

Figure 11: Fundamental price for one specific day. This figure plots the observed prices of Shell, $y_{t,i}$ and the estimated fundamental price μ_t for the mandatory period of 11 June 1996. The solid line represents the price μ_t and the observation $y_{t,i}$ is represented by a circle if the trade is a buy and by a cross if the trade is a sell.

Figure 1: An example of OLS regression, piece-wise regression and cubic spline

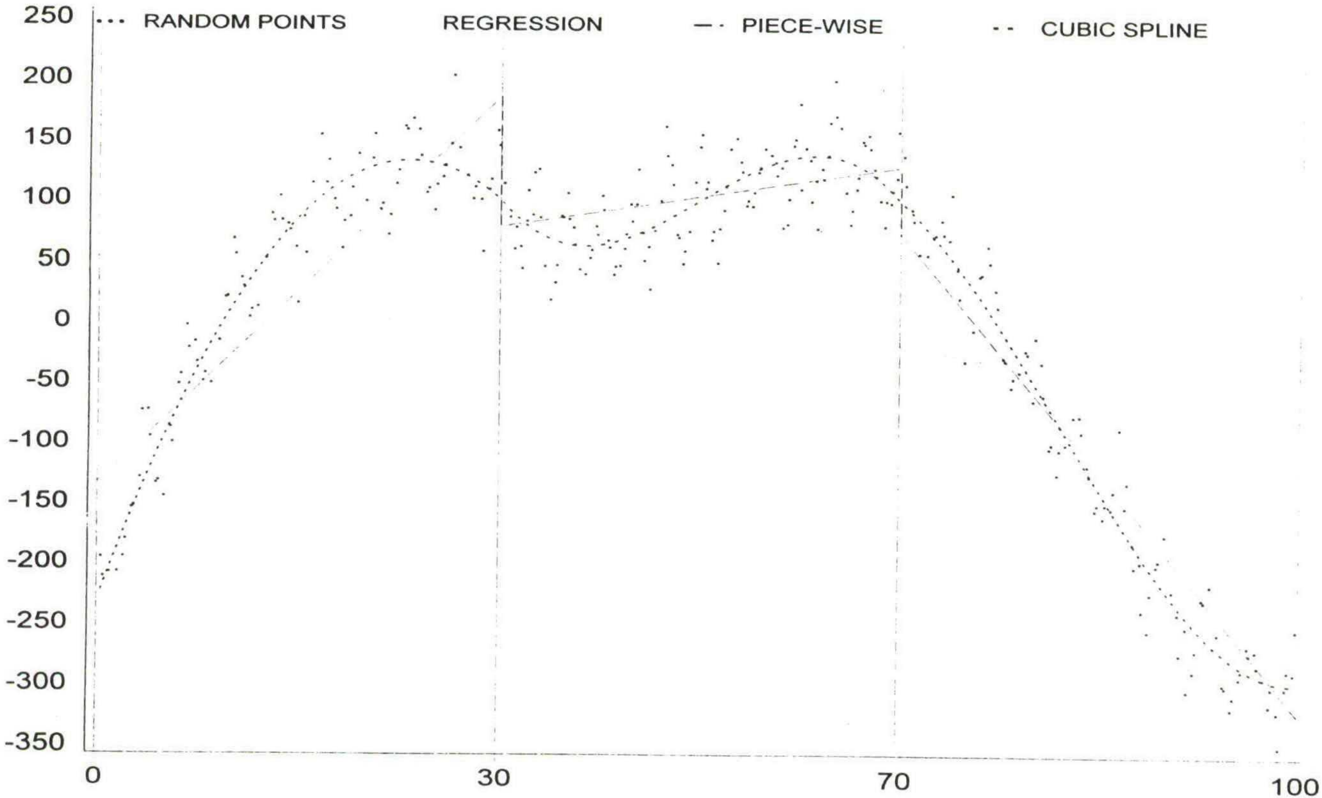


Figure 2: An illustration of the LSE data set

DATE	TIME	BUY FIRM	SELL FIRM	BUY CAP	SELL CAP	PRICE	QUAN-TITY
01/04/96	11:09	ABC	DEF	M	A	379.00	55
01/04/96	11:09	ABC	FGH	M	A	379.00	440
01/04/96	11:09	IJK	LMN	M	A	379.00	400
01/04/96	11:09	IJK	OPQ	M	A	379.00	3045
01/04/96	11:09	RST	UVW	M	A	380.00	25000
01/04/96	11:09	OPQ	IJK	A	M	381.00	783
01/04/96	11:10	IJK	OPQ	M	A	379.00	880
01/04/96	11:10	IJK	IJK	M	N	379.00	1000
01/04/96	11:10	IJK	LMN	M	A	379.00	1270
01/04/96	11:10	XYZ	CAE	M	A	379.00	550
01/04/96	11:10	XYZ	OPQ	M	A	379.00	1500
01/04/96	11:10	RST	EWP	M	A	379.50	130000
01/04/96	11:10	FEN	RST	A	M	380.00	3000
01/04/96	11:10	OPQ	IJK	A	M	380.50	2267
01/04/96	11:10	LOU	ABC	A	M	381.00	700
01/04/96	11:10	OPQ	IJK	A	M	381.00	796

Figure 3: Means of absolute standardised errors of ten-minute intervals

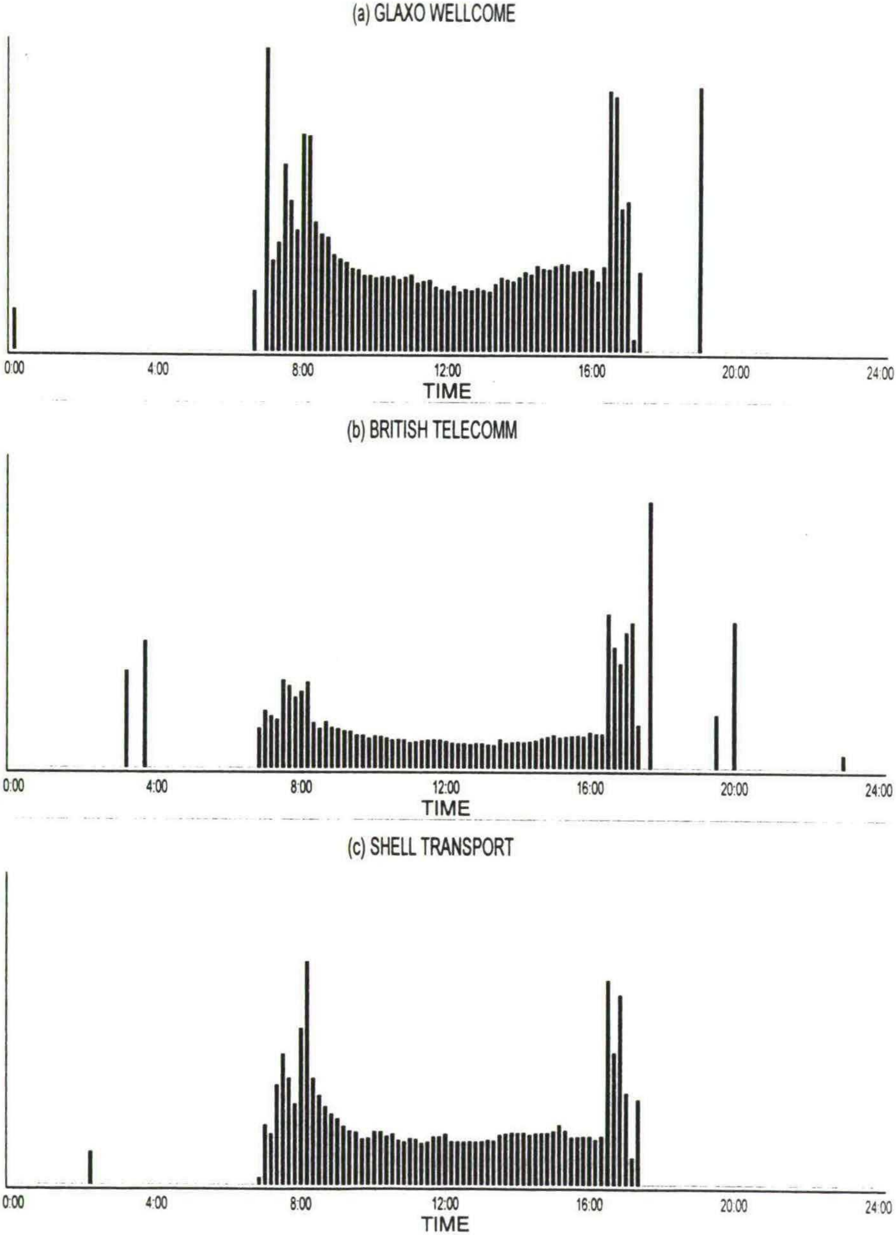


Figure 4: Estimated time splines of the spreads

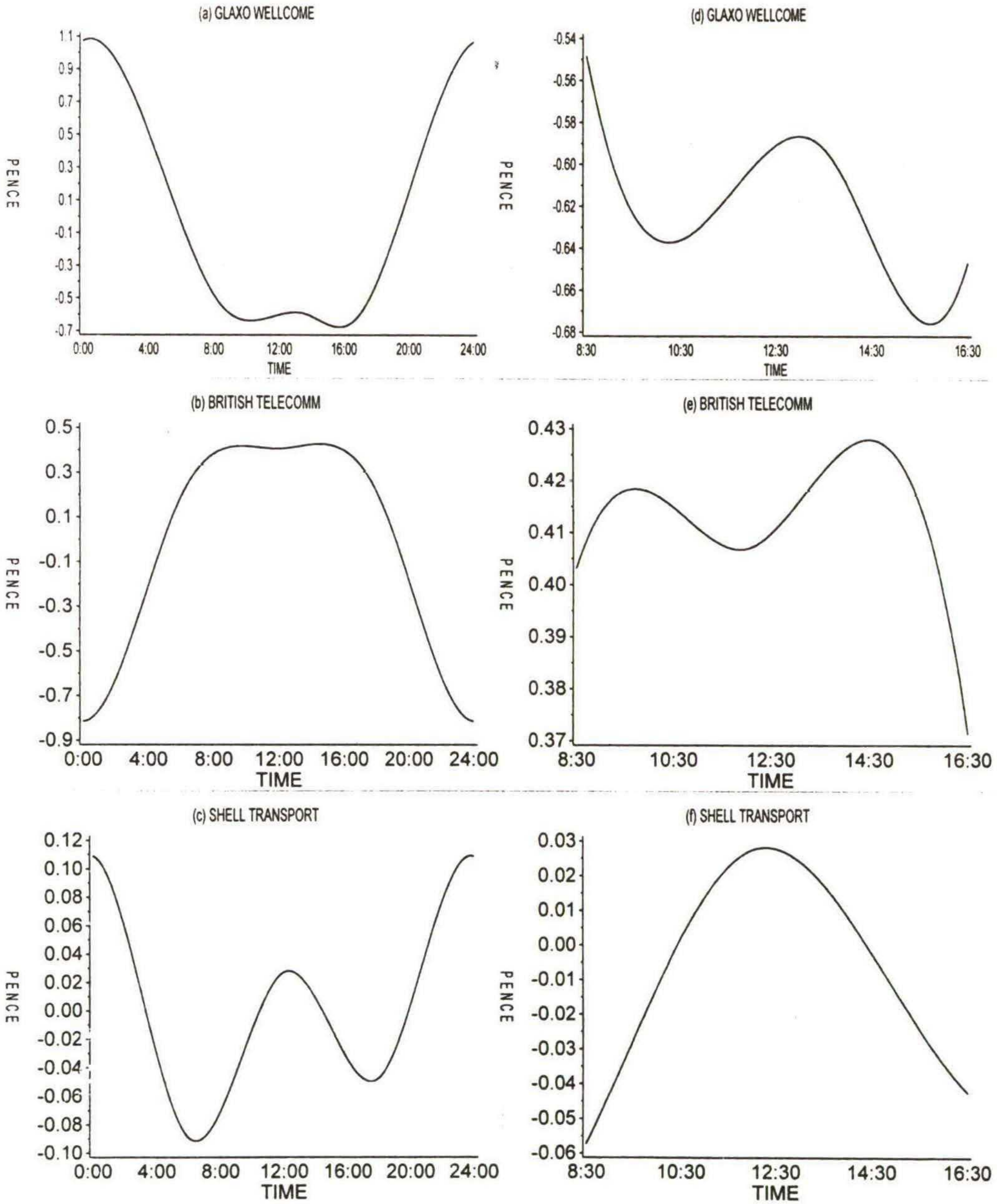


Figure 5: Estimated size splines of the spreads

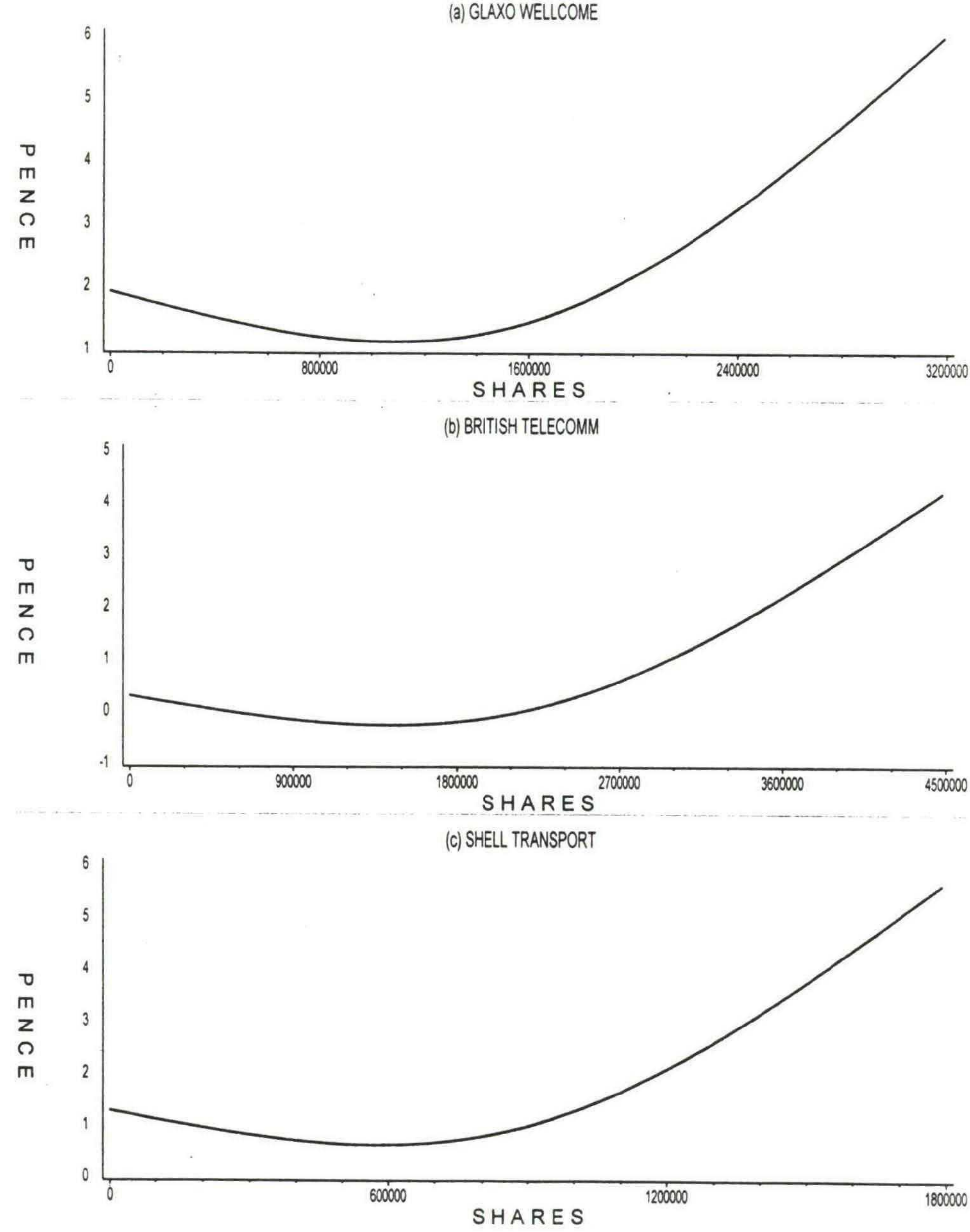


Figure 6: Average of prediction residuals for each minute

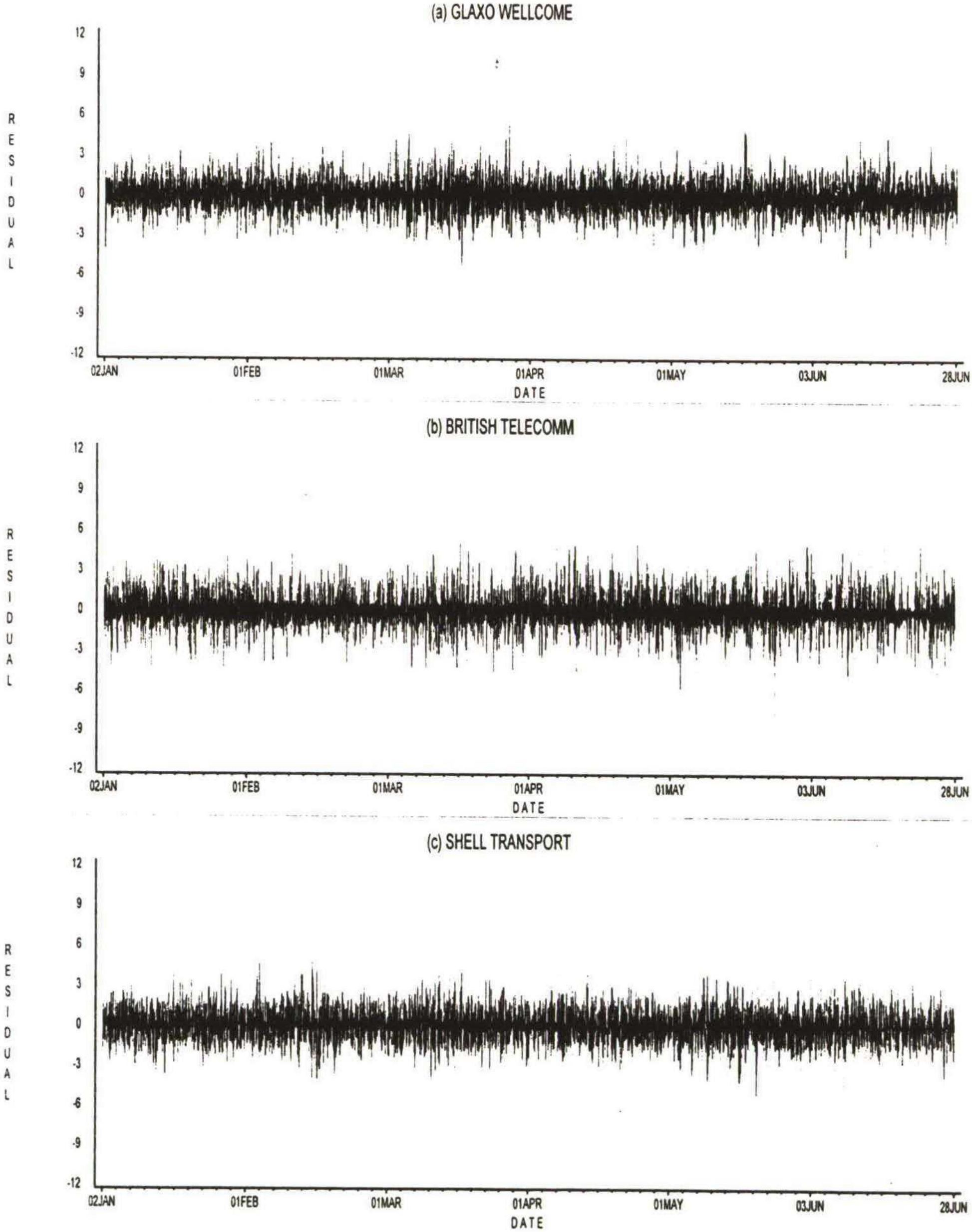


Figure 7: Correlogram for prediction residuals

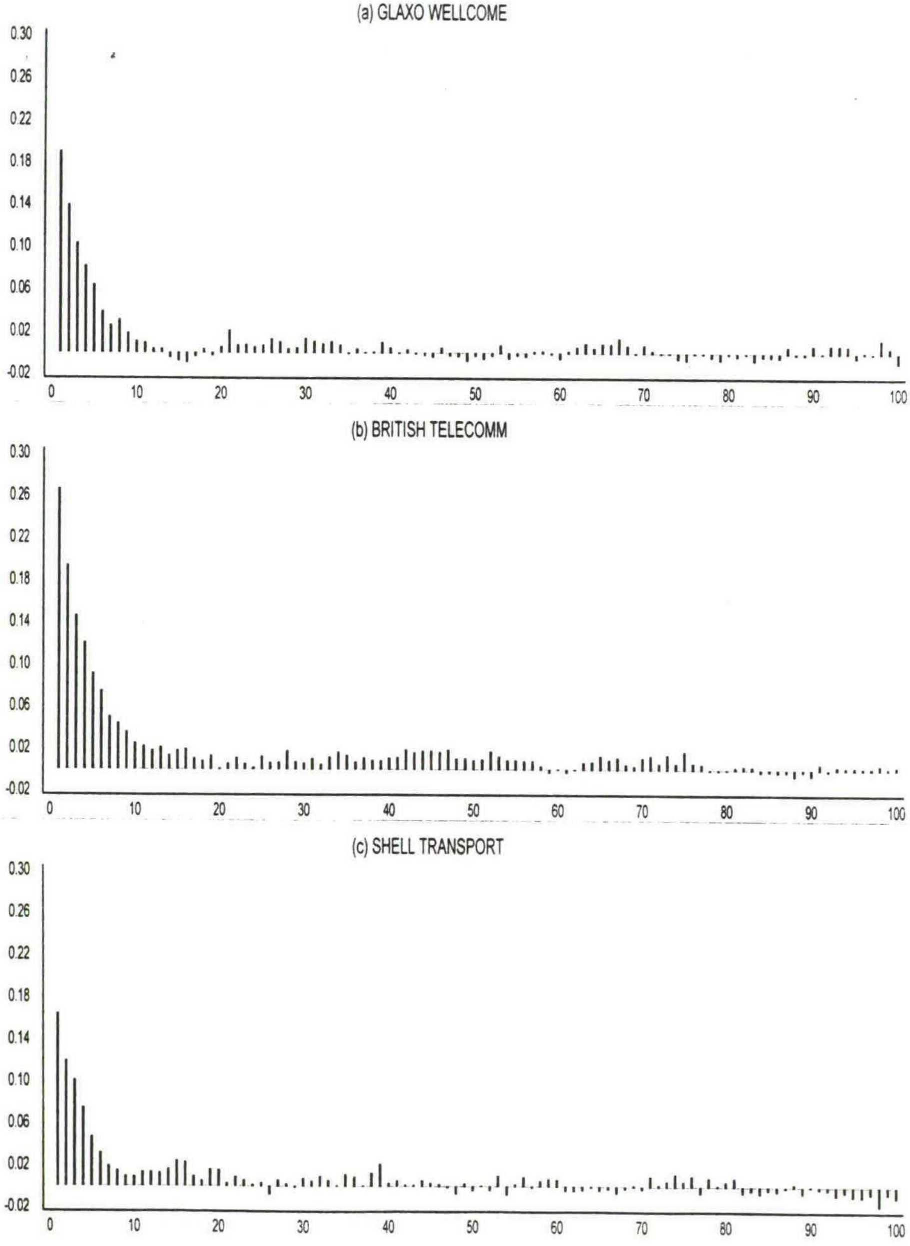


Figure 8: Observed and smoothed fundamental price of Glaxo Wellcome

Figure 9: Observed and smoothed fundamental price of British Telecommunications

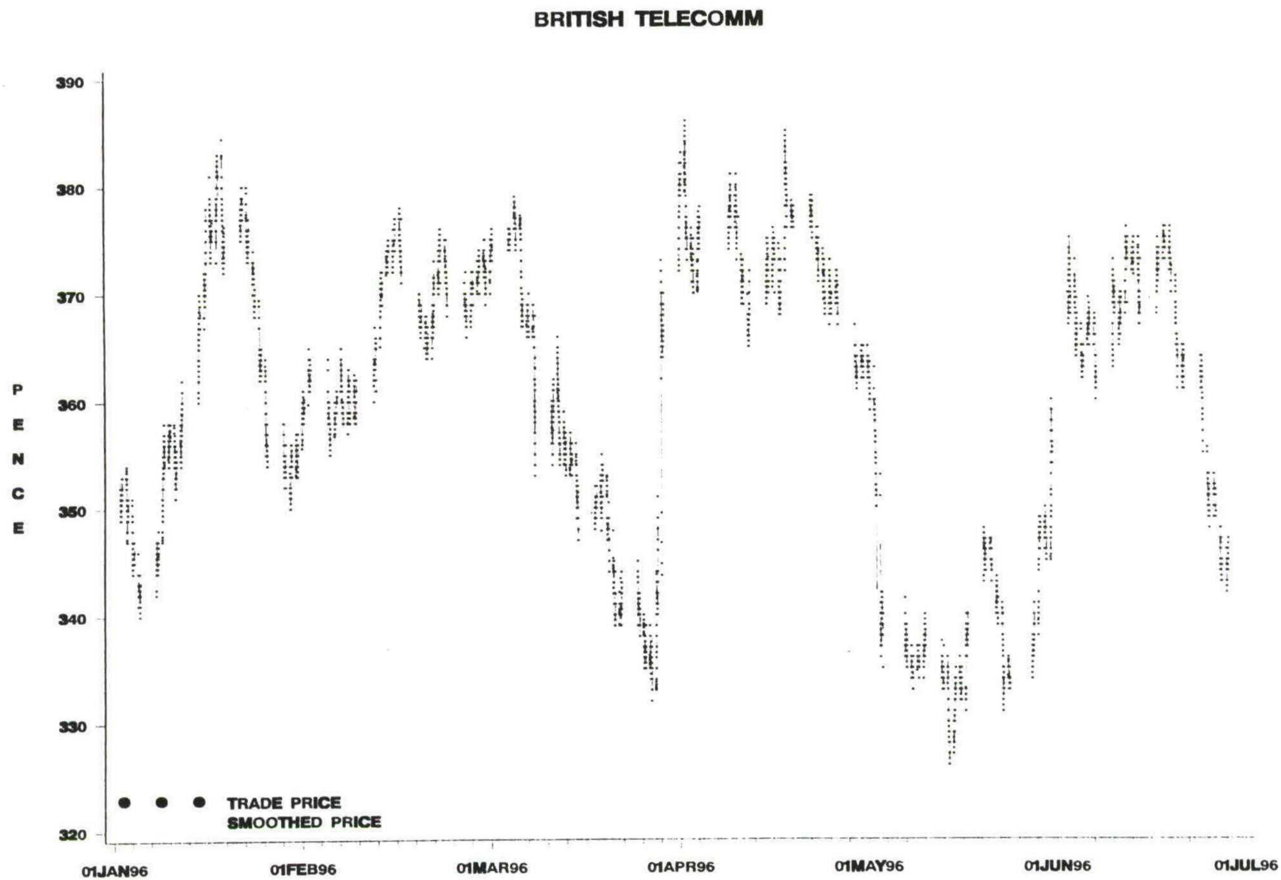


Figure 10: Observed and smoothed fundamental price of Shell Transport

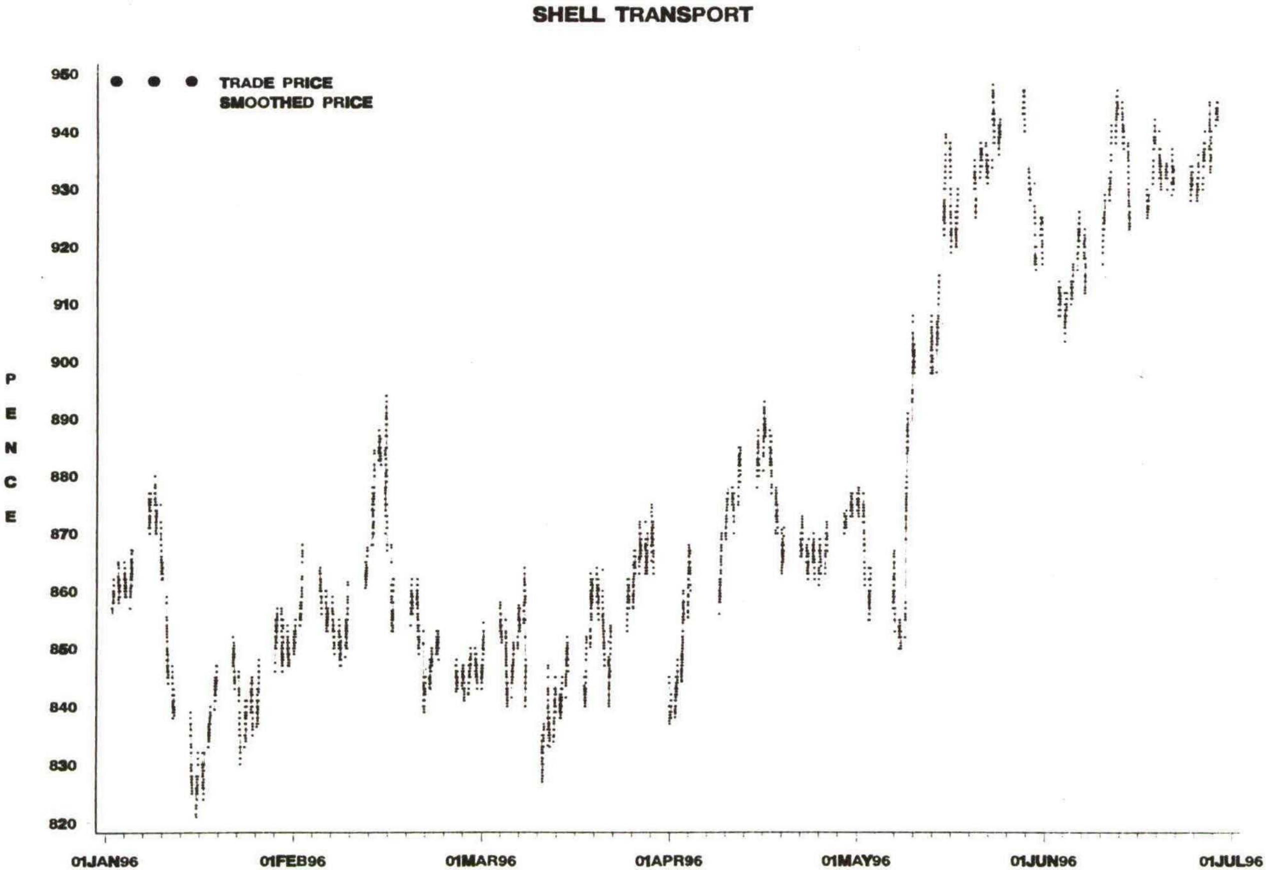
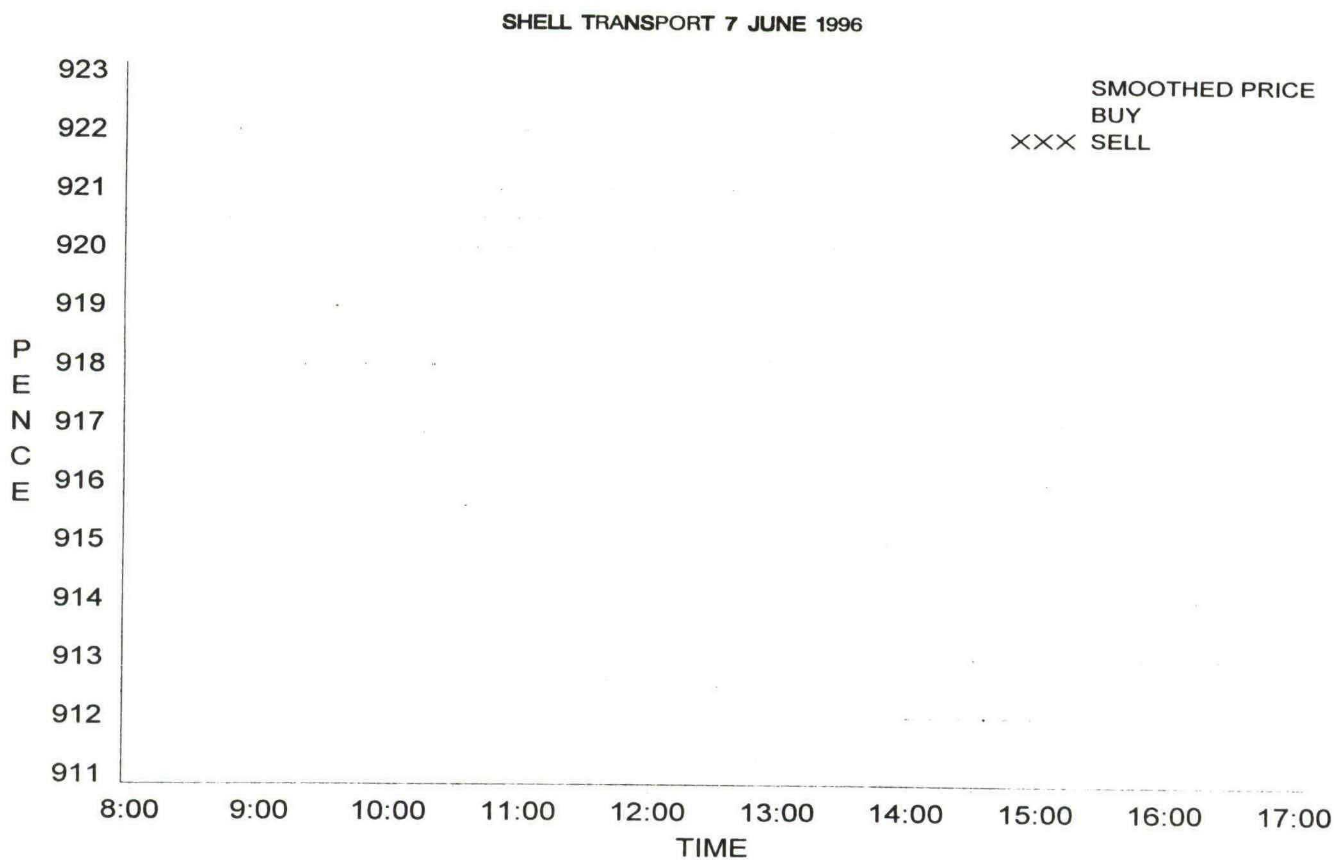


Figure 11: Fundamental price for one specific day



No.	Author(s)	Title
9759	R.M.W.J. Beetsma and H. Uhlig	An Analysis of the "Stability Pact"
9760	M. Lettau and H. Uhlig	Preferences, Consumption Smoothing, and Risk Premia
9761	F. Janssen and T. de Kok	The Optimal Number of Suppliers in an (s,Q) Inventory System with Order Splitting
9762	F. Janssen and T. de Kok	The Fill Rate Service Measure in an (s,Q) Inventory System with Order Splitting
9763	E. Canton	Fiscal Policy in a Stochastic Model of Endogenous Growth
9764	R. Euwals	Hours Constraints within and between Jobs
9765	A. Blume	Fast Learning in Organizations
9766	A. Blume	Information Transmission and Preference Similarity
9767	B. van der Genugten	Canonical Partitions in the Restricted Linear Model
9768	W. Güth and B. Peleg	When Will the Fittest Survive? -An Indirect Evolutionary Analysis-
9769	E. Rebers, R. Beetsma and H. Peters	When to Fire Bad Managers: The Role of Collusion Between Management and Board of Directors
9770	B. Donkers and A. van Soest	Subjective Measures of Household Preferences and Financial Decisions
9771	K. Kultti	Scale Returns of a Random Matching Model
9772	H. Huizinga and S.B. Nielsen	A Welfare Comparison of International Tax Regimes with Cross-Ownership of Firms
9773	H. Huizinga and S.B. Nielsen	The Taxation of Interest in Europe: A Minimum Withholding Tax?
9774	E. Charlier	Equivalence Scales for the Former West Germany
9775	M. Berliant and T. ten Raa	Increasing Returns and Perfect Competition: The Role of Land
9776	A. Kalwij, R. Alessie and P. Fontein	Household Commodity Demand and Demographics in the Netherlands: a Microeconometric Analysis
9777	P.J.J. Herings	Two Simple Proofs of the Feasibility of the Linear Tracing Procedure
9778	G. Gürkan, A.Y. Özge and S.M. Robinson	Sample-Path Solutions for Simulation Optimization Problems and Stochastic Variational Inequalities
9779	S. Smulders	Should Environmental Standards be Tighter if Technological Change is Endogenous?

No.	Author(s)	Title
9780	B.J. Heijdra and L. Meijdam	Public Investment in a Small Open Economy
9781	E.G.F. Stancanelli	Do the Rich Stay Unemployed Longer? An Empirical Study for the UK
9782	J.C. Engwerda and R.C. Douven	Local Strong *d*-Monotonicity of the Kalai-Smorodinsky and Nash Bargaining Solution
9783	J.C. Engwerda	Computational Aspects of the Open-Loop Nash Equilibrium in Linear Quadratic Games
9784	J.C. Engwerda, B. van Aarle J.E.J. Plasmans	The (In)Finite Horizon Open-Loop Nash LQ-Game: An Application to EMU
9785	J. Osiewalski, G. Koop and M.F.J. Steel	A Stochastic Frontier Analysis of Output Level and Growth in Poland and Western Economies
9786	F. de Jong	Time-Series and Cross-Section Information in Affine Term Structure Models
9787	G. Gürkan, A.Y. Özge and S.M. Robinson	Sample-Path Solution of Stochastic Variational Inequalities
9788	A.N. Banerjee	Sensitivity of Univariate AR(1) Time-Series Forecasts Near the Unit Root
9789	G. Brennan, W. Güth and H. Kliemt	Trust in the Shadow of the Courts
9790	A.N. Banerjee and J.R. Magnus	On the Sensitivity of the usual *t*- and *F*-tests to AR(1) misspecification
9791	A. Cukierman and M. Tommasi	When does it take a Nixon to go to China?
9792	A. Cukierman, P. Rodriguez and S.B. Webb	Central Bank Autonomy and Exchange Rate Regimes - Their Effects on Monetary Accommodation and Activism
9793	B.G.C. Dellaert, M. Prodigalidad and J.J. Louvriere	Family Members' Projections of Each Other's Preference and Influence: A Two-Stage Conjoint Approach
9794	B. Dellaert, T. Arentze, M. Bierlaire, A. Borgers and H. Timmermans	Investigating Consumers' Tendency to Combine Multiple Shopping Purposes and Destinations
9795	A. Belke and D. Gros	Estimating the Costs and Benefits of EMU: The Impact of External Shocks on Labour Markets
9796	H. Daniëls, B. Kamp and W. Verkooijen	Application of Neural Networks to House Pricing and Bond Rating
9797	G. Gürkan	Simulation Optimization of Buffer Allocations in Production

No.	Author(s)	Title
		Lines with Unreliable Machines
9798	V. Bhaskar and E. van Damme	Moral Hazard and Private Monitoring
9799	F. Palomino	Relative Performance Equilibrium in Financial Markets
97100	G. Gürkan and A.Y. Özge	Functional Properties of Throughput in Tandem Lines with Unreliable Servers and Finite Buffers
97101	E.G.A. Gaury, J.P.C. Kleijnen and H. Pierreval	Configuring a Pull Production-Control Strategy Through a Generic Model
97102	F.A. de Roon, Th.E. Nijman and C. Veld	Analyzing Specification Errors in Models for Futures Risk Premia with Hedging Pressure
97103	M. Berg, R. Brekelmans and A. De Waegenaere	Budget Setting Strategies for the Company's Divisions
97104	C. Fernández and M.F.J. Steel	Reference Priors for Non-Normal Two-Sample Problems
97105	C. Fernández and M.F.J. Steel	Reference Priors for the General Location-Scale Model
97106	M.C.W. Janssen and E. Maasland	On the Unique D1 Equilibrium in the Stackelberg Model with asymmetric information
97107	A. Belke and M. Göcke	Multiple Equilibria in German Employment -Simultaneous Identification of Structural Breaks-
97108	D. Bergemann and U. Hege	Venture Capital Financing, Moral Hazard, and Learning
97109	U. Hege and P. Viala	Contentious Contracts
97110	P.J.-J. Herings	A Note on "Stability of Tâtonnement Processes of Short Period Equilibria with Rational Expectations"
97111	C. Fernández, E. Ley, and M.F.J. Steel	Statistical Modeling of Fishing Activities in the North Atlantic
97112	J.J.A. Moors	A Critical Evaluation of Mangat's Two-Step Procedure in Randomized Response
97113	J.J.A. Moors, B.B. van der Genugten, and L.W.G. Strijbosch	Repeated Audit Controls
97114	X. Gong and A. van Soest	Family Structure and Female Labour Supply in Mexico City
97115	A. Blume, D.V. DeJong, Y.-G. Kim and G.B. Sprinkle	Evolution of Communication with Partial Common Interest
97116	J.P.C. Kleijnen and R.G. Sargent	A Methodology for Fitting and Validating Metamodels in Simulation

No.	Author(s)	Title
97117	J. Boone	Technological Progress and Unemployment
97118	A. Prat	Campaign Advertising and Voter Welfare
9801	H. Gersbach and H. Uhlig	Debt Contracts, Collapse and Regulation as Competition Phenomena
9802	P. Peretto and S. Smulders	Specialization, Knowledge Dilution, and Scale Effects in an IO-based Growth Model
9803	K.J.M. Huisman and P.M. Kort	A Further Analysis on Strategic Timing of Adoption of New Technologies under Uncertainty
9804	P.J.-J. Herings and A. van den Elzen	Computation of the Nash Equilibrium Selected by the Tracing Procedure in *N*-Person Games
9805	P.J.-J. Herings and J.H. Drèze	Continua of Underemployment Equilibria
9806	M. Koster	Multi-Service Serial Cost Sharing: A Characterization of the Moulin-Shenker Rule
9807	F.A. de Roon, Th.E. Nijman and B.J.M. Werker	Testing for Mean-Variance Spanning with Short Sales Constraints and Transaction Costs: The Case of Emerging Markets
9808	R.M.W.J. Beetsma and P.C. Schotman	Measuring Risk Attitudes in a Natural Experiment: Data from the Television Game Show Lingo
9809	M. Bütler	The Choice between Pension Reform Options
9810	L. Bettendorf and F. Verboven	Competition on the Dutch Coffee Market
9811	E. Schaling, M. Hoeberichts and S. Eijffinger	Incentive Contracts for Central Bankers under Uncertainty: Walsh-Svensson non-Equivalence Revisited
9812	M. Slikker	Average Convexity in Communication Situations
9813	T. van de Klundert and S. Smulders	Capital Mobility and Catching Up in a Two-Country, Two-Sector Model of Endogenous Growth
9814	A.Belke and D. Gros	Evidence on the Costs of Intra-European Exchange Rate Variability
9815	J.P.C. Kleijnen and O. Pala	Maximizing the Simulation Output: a Competition
9816	C. Dustmann, N. Rajah and A. van Soest	School Quality, Exam Performance, and Career Choice
9817	H. Hamers, F. Klijn and J. Suijs	On the Balancedness of *m*-Sequencing Games
9818	S.J. Koopman and J. Durbin	Fast Filtering and Smoothing for Multivariate State Space Models

No.	Author(s)	Title
9819	E. Droste, M. Kosfeld and M. Voorneveld	Regret Equilibria in Games
9820	M. Slikker	A Note on Link Formation
9821	M. Koster, E. Molina, Y. Sprumont and S. Tijs	Core Representations of the Standard Fixed Tree Game
9822	J.P.C. Kleijnen	Validation of Simulation, With and Without Real Data
9823	M. Kosfeld	Rumours and Markets
9824	F. Karaesmen, F. van der Duyn Schouten and L.N. van Wassen-hove	Dedication versus Flexibility in Field Service Operations
9825	J. Suijs, A. De Waegenaere and P. Borm	Optimal Design of Pension Funds: A Mission Impossible
9826	U.Gneezy and W. Güth	On Competing Rewards Standards -An Experimental Study of Ultimatum Bargaining-
9827	M. Dufwenberg and U. Gneezy	Price Competition and Market Concentration: An Experimental Study
9828	A. Blume, D.V. De Jong and G.R. Neumann	Learning in Sender-Receiver Games
9829	B.G.C. Dellaert, J.D. Brazell and J.J. Louviere	Variations in Consumer Choice Consistency: The Case of Attribute-Level Driven Shifts in Consistency
9830	B.G.C. Dellaert, A.W.J. Borgers, J.J. Louviere and H.J.P. Timmermans	Consumer Choice of Modularized Products: A Conjoint choice Experiment Approach
9831	E.G.A. Gaury, H. Pierreval and J.P.C. Kleijnen	New Species of Hybrid Pull Systems
9832	S.J. Koopman and H.N. Lai	Modelling Bid-Ask Spreads in Competitive Dealership Markets

P.O. BOX 90153, 5000 LE TILBURG, THE NETHERLANDS