Studies in Nonlinear Dynamics and Econometrics

Quarterly Journal
April 2001, Volume 5, Number 1
The MIT Press

Nonlinearity in High-Frequency Financial Data and Hierarchical Models

Robert E. McCulloch
Graduate School of Business
University of Chicago
robert.mcculloch@gsb.uchicago.edu

Ruey S. Tsay
Graduate School of Business
University of Chicago
ruey.tsay@gsb.uchicago.edu

Abstract. This paper studies nonlinear behavior of high-frequency financial data and employs nonlinear hierarchical models for analyzing such data. We illustrate the analysis by modeling the transaction-by-transaction data of IBM stock on the New York Stock Exchange for a period of 3 months. The variables considered include time durations between trades and price changes. For a short time span of 5 trading days, a simple threshold model is found adequate for modeling time durations between trades after adjusting for the diurnal pattern of the data. When price change and time duration between price changes are considered jointly, we use a hierarchical model that consists of 6 simple conditional models to handle the dynamic structure within a trading day and the variation between trading days for the whole sample. The model shows that dynamic structure exists in the high-frequency data, but there are some special days on which the behavior of the stock seems different from the others. We use Markov chain Monte Carlo methods to estimate the hierarchical model.*

1 Introduction

Transaction-by-transaction data of individual stocks in the United States are now available in the Trades and Quotes (TAQ) database of the New York Stock Exchange. These data are commonly referred to as high-frequency financial data and are useful in studying market microstructure, such as the determinants of bid and ask quotes and the refinements in daily volatility estimation of an asset return (see Hasbrouck 1999 and Zhang, Russell, and Tsay 2000 for the dynamics of bid and ask quotes and Bai, Russell, and Tiao 2000 for

volatility estimation). Analysis of high-frequency data, however, is not simple, because such data exhibit some characteristics that are not found when the data frequency is low. Examples of the special characteristics include (a) extremely high kurtosis, (b) diurnal pattern, (c) discrete-valued observations as the price change in consecutive trades is in a multiple of tick size, (d) irregular time intervals between trades, and (e) large sample size. Wood (2000) provides a brief history of transaction data in finance.

As an illustration, there were more than 134,000 intraday trades of IBM stock in December 1999. The trading intensity exhibits a U-shaped diurnal pattern, with heavier activities during the beginning and closing of trading hours. Associated with each trade are time of the trade, measured in seconds starting from midnight; transaction price; transaction volume; and the prevailing bid and ask quotes. A thorough analysis of the transaction-by-transaction data of IBM stock in December 1999 alone would require techniques of multivariate analysis, time-series methods, and generalized linear models. The purpose of this paper is twofold. First, we consider time duration between trades and find that a threshold duration model fits better than a conditional autoregressive duration model. In other words, nonstationarity is likely to become a major problem in analyzing such data. Secondly, we jointly model durations and price levels. Here we define duration to be the time between transactions associated with a price change. We also consider the transaction data as a panel of time series in which each series corresponds to a particular day. We use a parametric model to describe the structure of the data within a day, then use a hierarchical model to analyze data from all days in the sample.

We demonstrate the proposed analysis by analyzing intraday trades of IBM stock from November 1, 1990, to January 31, 1991. The data were obtained from the Trades, Orders Reports and Quotes (TORQ) database (see Hasbrouck 1992). We use this data set because it has been widely used in the literature (see Engle and Russell 1998).

The paper is organized as follows. Section 2 provides some summary statistics of the data used and shows evidence of diurnal pattern by considering the time series of number of trades in 5-min time intervals. A linear-regression method consisting of four quadratic time functions and three indicator variables is used to remove the diurnal pattern of the time durations between trades. Section 3 focuses on the nonlinearity in time durations and shows that a simple threshold autoregressive model can adequately model the adjusted time duration process when the data span is not too long. In Section 4, we introduce variables for modeling price changes and time durations between price changes. The price change is decomposed into two components consisting of the direction and size of the change. The size is measured in multiples of tick size, which was one eighth of a dollar during the sample period. Within each time duration between price changes, we employ a counting process that enumerates the number of trades with no price change. These trades are important in measuring trading intensity but provide no information on price change. To allow for different dynamics when the price is going up or coming down, we postulate two different models for the size of a price change given the direction of price change. In short, we use six simple generalized linear models to describe the dynamic of price change and time duration between price changes in a trading day. A nonlinear hierarchical model is then proposed in Section 5 to model the day-to-day variation of the two variables of interest. Results of the hierarchical model are given and discussed.

2 Data and Their Characteristics

The data used in this study are the transaction data of IBM stock from November 1, 1990, to January 31, 1991, from the TORQ data set. This data set was used in Engle and Russell 1998 to demonstrate autoregressive conditional duration (ACD) models. There were 63 trading days and 60,328 transactions. For simplicity, we focus on the intraday transactions that occurred in the normal trading hours from 9:30 A.M. to 4:00 P.M. Eastern time. Each transaction contains a time stamp measured in seconds starting from midnight, the transaction price and volume, and the prevailing bid and ask quotes. The price change occurred in a multiple of tick size, which was one eighth of a dollar.

Table 1
Frequencies of price change in multiples of tick size for IBM stock from November 1, 1990, to January 31, 1991

Number (tick)	≤ −3	−2	−1	0	1	2	≥ 3
Percentage	0.66	1.33	14.53	67.06	14.53	1.27	0.63

To better understand transaction data, we consider some empirical characteristics of the data (see Tsay 2000, chap. 5, for further details). Table 1 shows the frequencies of price change in intraday transactions. From the table, we observe the following:

1. About two thirds of the intraday transactions are without price change.

2. The price changed in 1 tick approximately 29% of the intraday transactions.

3. Only 2.6% of the transactions are associated with 2-tick price changes.

4. Only about 1.3% of the transactions resulted in price changes of 3 ticks or more.

5. The distribution of positive and negative price changes is approximately symmetric.

Consider next the number of transactions in a 5-min time interval. Denote the series by x_t. That is, x_1 is the number of IBM transactions from 9:30 A.M. to 9:35 A.M. on November 1, 1990, Eastern time, x_2 is the number of transactions from 9:35 A.M. to 9:40 A.M., and so on. The time gaps between trading days are ignored. Figure 1(a) shows the time plot of x_t and Figure 1(b) the sample ACF of x_t for lags 1 to 260. Of particular interest is the cyclical pattern of the ACF with a periodicity of 78, which is the number of 5-min intervals in a trading day. The number of transactions thus exhibits a diurnal pattern. Figure 2 shows the average number of transactions within 5-min time intervals over the 63 days. There are 78 such averages. The plot exhibits a "smiling" shape, indicating heavier trading at the openings and closings of the market and thinner trading during the lunch hours.

Since we focus on transactions that occurred in the normal trading hours of a trading day, there are 59,838 time intervals in the data. These intervals are called the intraday *durations* between trades. For the IBM stock,
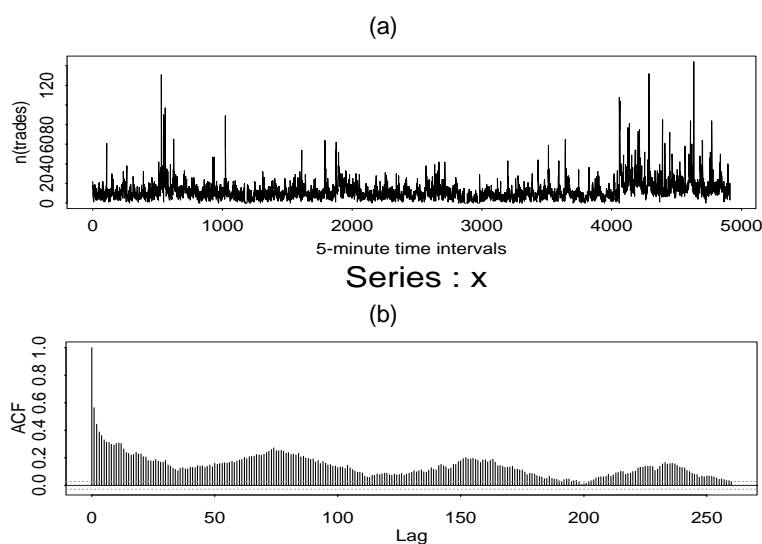
Figure 1
IBM transactions data from November 1, 1990, to January 31, 1991: (a) number of transactions in 5-min time intervals; (b) sample ACF of the series in part (a).
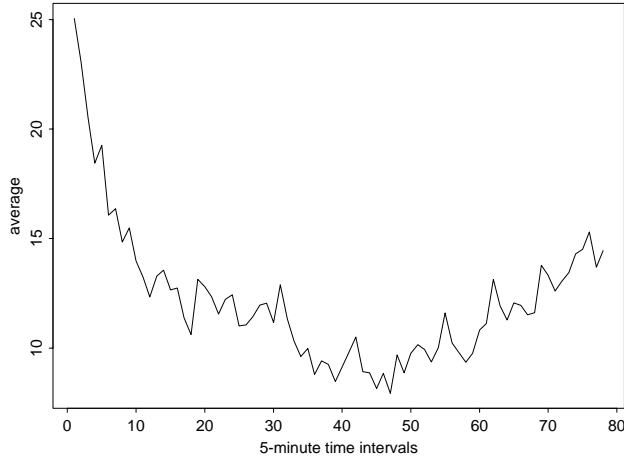
Figure 2
Time plot of the average number of transactions in 5-min time intervals. There are 78 observations, averaging over the 63 trading days from November 1, 1990, to January 31, 1991, for IBM stock.

there were 6,531 zero time intervals. That is, during the normal trading hours of the 63 trading days from November 1, 1990, to January 31, 1991, multiple transactions in a second occurred 6,531 times, which is about 10.91%. Among these multiple transactions, 1,002 had different prices, which is about 1.67% of the total number of intraday transactions. Therefore, multiple transactions, that is, zero durations, may become an issue in statistical modeling of the time durations between trades. For simplicity, we follow Engle and Russell (1998) by focusing on nonzero time durations.

A simple approach to modeling the diurnal pattern of transaction durations is to assume that the diurnal pattern is deterministic and follows a smooth quadratic function. Let Δt_i be the time duration between trades. The assumption implies that the adjusted duration

$$\Delta t_i^* = \Delta t_i / f(t_i) \tag{2.1}$$

where $f(t_i)$ is a smooth deterministic function, has no diurnal pattern. For the IBM data, we assume

$$f(t_i) = \exp[d(t_i)], \quad d(t_i) = \beta_0 + \sum_{j=1}^{7} \beta_j f_j(t_i) \tag{2.2}$$

where

$$f_1(t_i) = -\left(\frac{t_i - 43200}{14400}\right)^2, \quad f_3(t_i) = \begin{cases} -(\frac{t_i - 38700}{7500})^2 & \text{if } t_i < 43{,}200 \\ 0 & \text{otherwise} \end{cases}$$

$$f_2(t_i) = -\left(\frac{t_i - 48300}{9300}\right)^2, \quad f_4(t_i) = \begin{cases} -(\frac{t_i - 48600}{9000})^2 & \text{if } t_i \geq 43{,}200 \\ 0 & \text{otherwise} \end{cases}$$

$f_5(t_i)$ and $f_6(t_i)$ are indicator variables for the first and second 5 min of market opening, that is, $f_5(.) = 1$ if and only if t_i is between 9:30 A.M. and 9:35 A.M. Eastern time, and $f_7(t_i)$ is the indicator for the last 30 min of daily trading, that is, $f_7(t_i) = 1$ if and only if the trade occurred between 3:30 P.M. and 4:00 P.M. Eastern time. Figure 3 shows the plot of $f_i(.)$ for $i = 1, \cdots, 4$, where the time scale in the x-axis is in minutes. Note that $f_3(43{,}200) = f_4(43{,}200)$, where 43,200 corresponds to 12:00 noon.
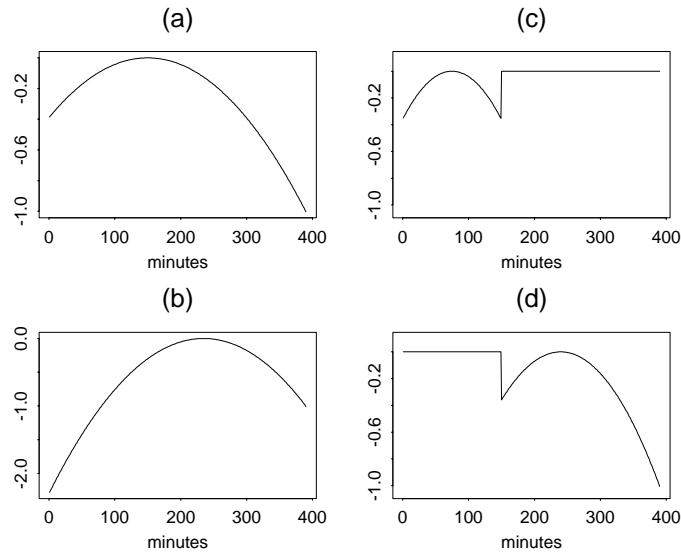
Figure 3
Quadratic functions used to remove the deterministic component of IBM intraday trading durations: (a)–(d) are the functions $f_1(.)$ to $f_4(.)$, respectively, of Equation (2.2).

The coefficients β_j of Equation (2.2) are obtained by the least squares method of the linear regression

$$\ln(\Delta t_i) = \beta_0 + \sum_{i=j}^{7} \beta_j f_j(t_i) + \epsilon_i$$

The fitted model is

$$\ln(\widehat{\Delta t_i}) = 2.555 + .159 f_1(t_i) + .270 f_2(t_i) + .384 f_3(t_i) + .061 f_4(t_i) - .611 f_5(t_i) - .157 f_6(t_i) + .073 f_7(t_i)$$

Figure 4 shows the time plot of average durations in 5-min time intervals over the 63 trading days before and after adjusting for the deterministic component. Panel (a) shows the average durations of Δt_i and as expected, it exhibits a diurnal pattern. Panel (b) shows the average durations of Δt_i^*, that is, after the adjustment, and the diurnal pattern is largely removed.

3 Nonlinear Models for Duration

Duration models, such as the autoregressive conditional duration (ACD) model of Engle and Russell (1998), are concerned with the dynamic structure of the adjusted duration in Equation (2.1). For simplicity in notation, we let $x_i = \Delta t_i^*$ be the adjusted time duration between trades. Let F_{i-1} be the σ-field generated by x_{i-j} for $j > 0$ and $\psi_i = E(x_i|F_{i-1})$ be the conditional expectation of the adjusted duration. An ACD model assumes

$$x_i = \psi_i \epsilon_i \qquad (3.1)$$

where $\{\epsilon_i\}$ is a sequence of independent and identically distributed positive random variables satisfying $E(\epsilon_i) = 1$ and ψ_i follows the model

$$\psi_i = \alpha_0 + \sum_{v=1}^{r} \alpha_v x_{i-v} + \sum_{v=1}^{s} \beta_v \psi_{i-v}$$

where r and s are non-negative integers, and α_i and β_j satisfy some positiveness conditions so that ψ_i is positive. Engle and Russell (1998) use standardized exponential or Weibull distribution for e_i, and Zhang,
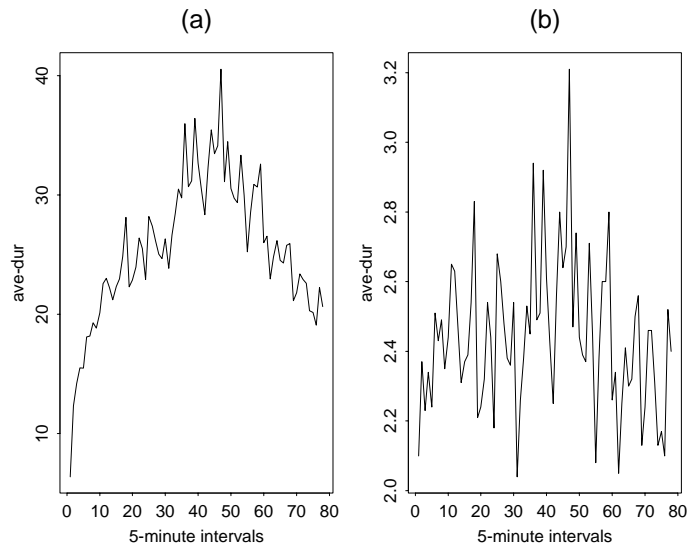
Figure 4
IBM transactions data from November 1, 1990, to January 31, 1991: (a) average durations in 5-min time intervals; (b) average durations in 5-min time intervals after adjusting for the deterministic component.

Russell, and Tsay (2001) use generalized Gamma distribution. Maximum likelihood method can be used to estimate the ACD model.

Stationarity, however, often becomes an important issue in high-frequency data analysis. For the IBM data considered in this paper, Zhang, Russell, and Tsay (2001) detect several highly significant structural changes using an ACD model with generalized Gamma distribution. For the purpose of this paper, we focus on time durations between trades for the first 5 trading days from November 1 to November 7, 1990. There were 3,534 observations (see Figure 5a). If ACD models with Weibull distribution are entertained, we obtain the model

$$x_i = \psi_i \epsilon_i, \quad \psi_i = 0.291 + 0.077 x_{i-1} + 0.836 \psi_{i-1} \tag{3.2}$$

Figure 5
Time plots of durations for IBM traded in the first five trading days of November 1990: (a) adjusted series; (b) normalized residuals of an ACD(1,1) model. There are 3,534 nonzero durations.

Table 2
Nonlinearity tests for IBM transaction durations from November 1 to November 7, 1990

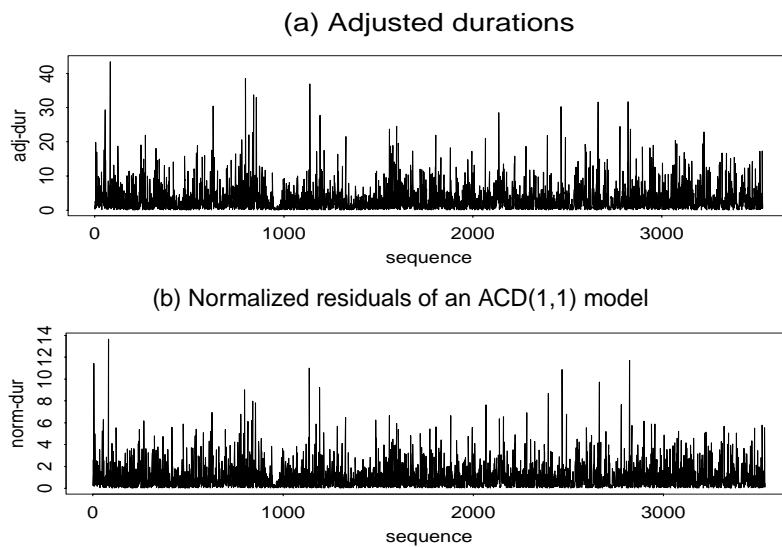| (a) Normalized residuals of an ACD(1,1) model					
Type	Ori-F	Tar-$F(1)$	Tar-$F(2)$	Tar-$F(3)$	Tar-$F(4)$
Test	0.343	3.288	3.142	3.128	0.297
p-value	0.969	0.006	0.008	0.008	0.915
(b) Normalized innovations of a threshold ACD(1,1) model					
Type	Ori-F	Tar-$F(1)$	Tar-$F(2)$	Tar-$F(3)$	Tar-$F(4)$
Test	0.163	0.746	1.899	1.752	0.270
p-value	0.998	0.589	0.091	0.119	0.929

Note: Only intraday durations are used. Ori-F denotes the nonlinearity test of Tsay 1986 and Tar-F(d) is the threshold nonlinearity test of Tsay 1989 with delay d.

where ϵ_i follows a standardized Weibull distribution with parameter $\hat{\eta}$ = 0.878, the standard error of which is 0.011. The standard errors of the three coefficients of model (3.2) are 0.076, 0.012, and 0.029, respectively.

Let $\hat{\epsilon}_t = x_i/\hat{\psi}_i$ be the normalized residual of model (3.2) (see Figure 5b). The Ljung-Box statistics of $\hat{\epsilon}_i$ give $Q(12)$ = 5.7 and $Q(24)$ = 19.9, indicating that there is no serial correlation in the normalized residuals. Furthermore, the Ljung-Box statistics of the squared residuals $\hat{\epsilon}_i^2$ give $Q(12)$ = 6.5 and $Q(24)$ = 15.1. There exists no conditional heteroscedasticity in the normalized residuals, either. These diagnostic statistics suggest that model (3.2) captures adequately the linear dynamic structure and the volatility of the data.

The model in (3.2), however, fails to pass some nonlinearity tests. For illustration, we apply the F-test of Tsay 1986 and the threshold test of Tsay 1989 to the normalized residual $\hat{\epsilon}_i$. Using an AR(4) model, the test results are given in part (a) of Table 2, where Ori-F denotes the nonlinearity test of Tsay 1986 and Tar-$F(d)$ is the threshold nonlinearity test of Tsay 1989 with delay d. As expected from the results of Ljung-Box statistics, the Ori-F test indicates no quadratic nonlinearity in the normalized residuals. However, the Tar-F test statistics suggest strong nonlinearity.

Based on the test results in Table 2, we entertain a threshold duration model with two regimes for the IBM intraday durations. The threshold variable is x_{t-1}, that is, lag-1 adjusted duration. The estimated threshold value is 3.79. The fitted threshold ACD(1,1) model is $x_i = \psi_i\epsilon_i$, where

$$\psi_i = \begin{cases} 0.020 + 0.257x_{i-1} + 0.847\psi_{i-1}, & \epsilon_i \sim w(0.901), \quad \text{if } x_{i-1} \leq 3.79 \\ 1.808 + 0.027x_{i-1} + 0.501\psi_{i-1}, & \epsilon_i \sim w(0.845), \quad \text{if } x_{i-1} > 3.79 \end{cases} \tag{3.3}$$

where $w(\alpha)$ denotes a standardized Weibull distribution with parameter α. The numbers of observations in the two regimes are 2,503 and 1,030, respectively. In Equation (3.3), standard errors of the parameters for the first regime are 0.043, 0.041, 0.024, and 0.014, whereas those for the second regime are 0.526, 0.020, 0.147, and 0.020, respectively.

Considering the normalized residuals $\hat{\epsilon}_i = x_i/\hat{\psi}_i$ of the threshold ACD(1,1) model in Equation (3.3), we obtain $Q(12)$ = 9.8 and $Q(24)$ = 23.9 for $\hat{\epsilon}_i$ and $Q(12)$ = 8.0 and $Q(24)$ = 16.7 for $\hat{\epsilon}_i^2$. Thus, there is no significant serial correlation in the $\hat{\epsilon}_i$ and $\hat{\epsilon}_i^2$ series. Furthermore, applying the same nonlinearity tests as before to this newly normalized residual series $\hat{\epsilon}_i$, we detect no nonlinearity (see panel (b) of Table 2). Consequently, the two-regime threshold ACD(1,1) model in Equation (3.3) is adequate.

If we classify the two regimes as heavy and thin trading periods, then the threshold model suggests that the trading dynamics measured by intraday transaction durations are different between heavy and thin trading periods for the IBM stock, even after the adjustment of diurnal pattern. This is not surprising, as market activities are often driven by arrivals of news and other information.

4 A Model for Price Change and Duration

In this section we consider jointly the process of price change and the associated time duration. As mentioned before, many intraday stock transactions result in no price change. These transactions are highly relevant to trading intensity, but they do not contain direct information on price movement. Therefore, to simplify the complexity involved in modeling price change, we shall focus on transactions that result in a price change and propose a price change and duration (PCD) model to describe the multivariate dynamics of price change and the associated time duration.

4.1 The PCD model

Let t_i be the calendar time of the ith price change of an asset. As before, t_i is measured in seconds from midnight of a trading day. Let P_{t_i} be the transaction price when the ith price change occurred and $\Delta t_i = t_i - t_{i-1}$ be the time duration between price changes. In addition, let N_i be the number of trades in the time interval (t_{i-1}, t_i) that result in no price change. This new variable is used to represent trading intensity during a period of no price change. Finally, let D_i be the direction of the ith price change, with $D_i = 1$ when price goes up and $D_i = -1$ when the price comes down, and S_i be the size of the ith price change measured in ticks. Under the new definitions, the price of a stock evolves over time by

$$P_{t_i} = P_{t_{i-1}} + D_i S_i \tag{4.1}$$

and the transaction data consist of $\{\Delta t_i, N_i, D_i, S_i\}$ for the ith price change. The proposed PCD model is concerned with the joint analysis of $(\Delta t_i, N_i, D_i, S_i)$.

Focusing on transactions that result in a price change can reduce the sample size dramatically. For example, consider the intraday transaction data of IBM stock from November 1, 1990, to January 31, 1991. There were 60,265 intraday trades, but only 19,022 of them resulted in a price change. In addition, there is no diurnal pattern in time duration associated with a price change.

The decomposition of a price change into direction and size follows that of Rydberg and Shephard (1998), who consider all intraday trades and decompose the price series as

$$P_i = P_{i-1} + A_i D_i S_i,$$

where $A_i = 1$ if the trade results in a price change and $A_i = 0$ otherwise (see also Ghysels 2000). We focus on trades associated with a price change and introduce the variable N_i to simplify the analysis.

To illustrate the relationship between price movements of all transactions and those of transactions associated with a price change, we consider the intraday tradings of IBM stock on November 21, 1990 (day 15). There were 726 transactions on that day during the normal trading hours, but only 195 trades resulted in a price change. Figure 6 shows the time plot of the price series for both cases. As expected, the price series are the same.

The proposed PCD model decomposes the conditional joint distribution of $(\Delta t_i, N_i, D_i, S_i)$ as follows:

$$f(\Delta t_i, N_i, D_i, S_i | F_{i-1}) = f(S_i | D_i, N_i, \Delta t_i, F_{i-1}) f(D_i | N_i, \Delta t_i, F_{i-1}) f(N_i | \Delta t_i, F_{i-1}) f(\Delta t_i | F_{i-1}) \tag{4.2}$$

This partition enables us to specify suitable econometric models for the conditional distributions and hence to simplify the modeling task. There are many ways to specify models for the conditional distributions. A proper specification might depend on the asset under study. Here we use some generalized linear models for the discrete-valued variables and a simple time-series model for the continuous variable $\ln(\Delta t_i)$.

For the time duration between price changes, we use the model

$$\ln(\Delta t_i) = \beta_0 + \beta_1 \ln(\Delta t_{i-1}) + \beta_2 S_{i-1} + \sigma \epsilon_i \tag{4.3}$$

(a) All transactions
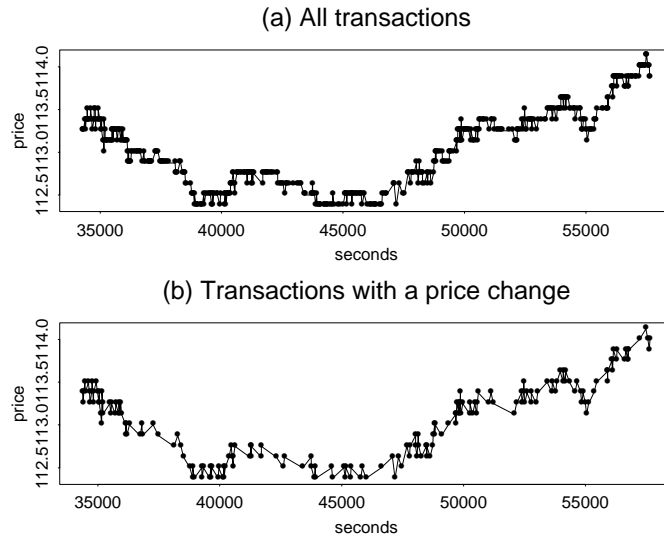
(b) Transactions with a price change

Figure 6

Time plots of the intraday transaction prices of IBM stock on November 21, 1990: (a) all transactions; (b) transactions that resulted in a price change.

where σ is a positive number, and $\{\epsilon_i\}$ is a sequence of independent $N(0, 1)$ random variables. Here $N(\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2. This is a simple regression model with lagged variables. Other explanatory variables can be added if necessary.

The conditional model for N_i is further partitioned into two parts, because empirical data suggest a concentration of N_i at 0. The first part of the model for N_i is the logit model

$$p(N_i = 0|\Delta t_i, F_{i-1}) = \text{logit}[\alpha_0 + \alpha_1 \ln(\Delta t_i)] \tag{4.4}$$

where $\text{logit}(x) = \exp(x)/[1 + \exp(x)]$, whereas the second part of the model is

$$N_i|(N_i > 0, \Delta t_i, F_{i-1}) \sim 1 + g(\lambda_i), \quad \lambda_i = \frac{\exp[\gamma_0 + \gamma_1 \ln(\Delta t_i)]}{1 + \exp[\gamma_0 + \gamma_1 \ln(\Delta t_i)]} \tag{4.5}$$

where $\sim$ means "is distributed as" and $g(\lambda)$ denotes a geometric distribution with parameter λ, which is in the interval (0,1).

The model for direction D_i is

$$D_i|(N_i, \Delta t_i, F_{i-1}) = \text{sign}(\mu_i + \sigma_i\epsilon) \tag{4.6}$$

where ϵ is a $N(0, 1)$ random variable, and

$$\mu_i = \omega_0 + \omega_1 D_{i-1} + \omega_2 \ln(\Delta t_i)$$

$$\ln(\sigma_i) = \beta|\sum_{j=1}^{4} D_{i-j}| = \beta|D_{i-1} + D_{i-2} + D_{i-3} + D_{i-4}|$$

In other words, D_i is governed by the sign of a normal random variable with mean μ_i and variance σ_i^2. A special characteristic of the above model is the function for $\ln(\sigma_i)$. For intraday transactions, a key feature is the *price reversal* between consecutive price changes, that is, the bid-and-ask bounce. This feature is modeled by the dependence of D_i on D_{i-1} in the mean equation. However, there exists an occasional local trend in the price movement. The above variance equation allows for such a local trend by increasing the uncertainty in
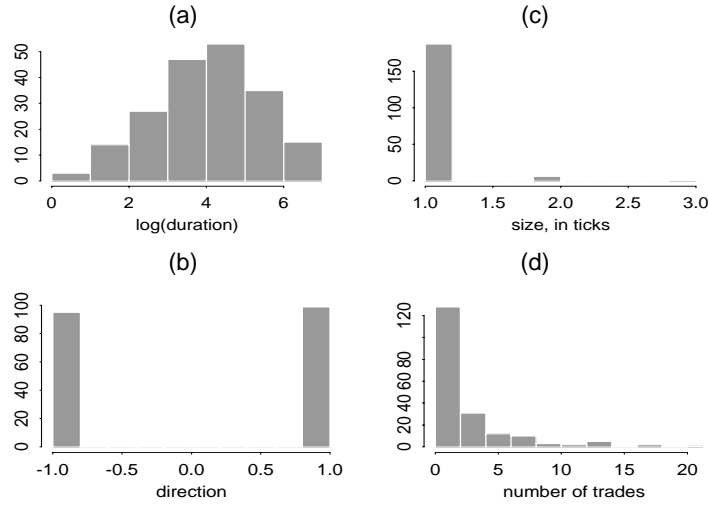
Figure 7
Histograms of intraday transactions data for IBM stock on November 21, 1990 (day 15): (a) log durations; (b) direction of price movement; (c) size of price change measured in ticks; (d) number of trades without a price change.

the direction of price movement when the past data showed evidence of a local trend. For a normal distribution with a fixed mean, increasing its variance makes a random draw have the same chance to be positive and negative. This in turn increases the chance for a sequence of all positive or all negative draws. Such a sequence creates a local trend in price movement.

To allow for different dynamics between positive and negative price movements, we use different models for the size of a price change. Specifically, we have

$$S_i | (D_i = -1, N_i, \Delta t_i, F_{i-1}) \sim p(\lambda_{d,i}) + 1, \quad \text{with}$$

$$\ln(\lambda_{d,i}) = \eta_{d,0} + \eta_{d,1} N_i + \eta_{d,2} \ln(\Delta t_i) + \eta_{d,3} S_{i-1} \tag{4.7}$$

$$S_i | (D_i = 1, N_i, \Delta t_i, F_{i-1}) \sim p(\lambda_{u,i}) + 1, \quad \text{with}$$

$$\ln(\lambda_{u,i}) = \eta_{u,0} + \eta_{u,1} N_i + \eta_{u,2} \ln(\Delta t_i) + \eta_{u,3} S_{i-1} \tag{4.8}$$

where $p(\lambda)$ denotes a Poisson distribution with parameter λ, and 1 is added to the size because the minimum size is 1 tick when there is a price change.

In this paper, to estimate the models in Equations (4.3)–(4.8), we use a Bayesian analysis with proper, but diffuse, priors. Markov chain Monte Carlo methods are used to compute the posteriors; this facilitates the hierarchical model in Section 5.

4.2 Illustration

Consider the intraday transactions of IBM stock on November 21, 1990 (day 15). There are 194 price changes within the normal trading hours. Figure 7 shows the histograms of $\ln(\Delta t_i)$, N_i, D_i and S_i. The data for D_i are about equally distributed between "upward" and "downward" movements. Only a few transactions resulted in a price change of more than 1 tick; as a matter of fact, there were 7 changes with 2 ticks and 1 change with 3 ticks. Using Markov chain Monte Carlo (MCMC) methods, we obtained the following models for the data. The reported estimates and their standard deviations are the posterior means and standard deviations of MCMC draws with 9,500 iterations. The model for the time duration between price changes is

$$\ln(\Delta t_i) = 4.023 + 0.032 \ln(\Delta t_{i-1}) - 0.025 S_{i-1} + 1.403 \epsilon_i$$

where standard deviations of the coefficients are 0.415, 0.073, 0.384, and 0.073, respectively. The fitted model indicates that there was no dynamic dependence in the time duration. For the N_i variable, we have

$$Pr(N_i > 0|\Delta t_i, F_{i-1}) = \text{logit}[-0.637 + 1.740\ln(\Delta t_i)]$$

where standard deviations of the estimates are 0.238 and 0.248, respectively. Thus, as expected, the number of trades with no price change in the time interval (t_{i-1}, t_i) depends positively on the length of the interval. The magnitude of N_i when it is positive is

$$N_i|(N_i > 0, \Delta t_i, F_{i-1}) \sim 1 + g(\lambda_i), \quad \lambda_i = \frac{\exp[0.178 - 0.910\ln(\Delta t_i)]}{1 + \exp[0.178 - 0.910\ln(\Delta t_i)]}$$

where standard deviations of the estimates are 0.246 and 0.138, respectively. The negative and significant coefficient of $\ln(\Delta t_i)$ means that N_i is positively related to the length of duration Δt_i, because a large $\ln(\Delta t_i)$ implies a small λ_i, which in turn implies higher probabilities for larger N_i.

The fitted model for D_i is

$$\mu_i = 0.049 - 0.840 D_{i-1} - 0.004\ln(\Delta t_i)$$

$$\ln(\sigma_i) = 0.244|D_{i-1} + D_{i-2} + D_{i-3} + D_{i-4}|$$

where standard deviations of the parameters in the mean equation are 0.129, 0.132, and 0.082, respectively, whereas that for the parameter in the variance equation is 0.182. The price reversal is clearly shown by the highly significant negative coefficient of D_{i-1}. The marginally significant parameter in the variance equation is exactly as expected. Finally, the fitted models for the size of a price change are

$$\ln(\lambda_{d,i}) = 1.024 - 0.327 N_i + 0.412\ln(\Delta t_i) - 4.474 S_{i-1}$$

$$\ln(\lambda_{u,i}) = -3.683 - 1.542 N_i + 0.419\ln(\Delta t_i) + 0.921 S_{i-1}$$

where standard deviations of the parameters for the "down size" are 3.350, 0.319, 0.599, and 3.188, respectively, whereas those for the "up size" are 1.734, 0.976, 0.453, and 1.459. The interesting estimates of the above two equations are the negative estimates of the coefficient of N_i. A large N_i means there were more transactions in the time interval (t_{i-1}, t_i) with no price change. This can be taken as evidence that no new information was available in the time interval (t_{i-1}, t_i). Consequently, the size for the price change at t_i should be small. A small $\lambda_{u,i}$ or $\lambda_{d,i}$ for a Poisson distribution gives precisely that.

In summary, granted that a sample of 194 observations in a given day does not contain sufficient information about the trading dynamic of IBM stock, the fitted models appear to provide some sensible results.

5 A Hierarchical Model

In Section 4 we applied the PCD model of Section 3 to data arising from a single day. In practice, we have data for many trading days, and an important question is how to analyze such data. At one extreme, we could simply apply the PCD model separately to each trading day. At the other extreme, one could concatenate the data from all days together into one long series and then fit a single PCD model. The first extreme does not combine information from all trading days. On the other hand, the second extreme assumes that the same model applies to each day, and this may not be the case.

Hierarchical models have been extensively used in recent years to deal with this situation (see Gelman et al. 1995, chap. 5). We employ a PCD model for each trading day and model the variation in parameters from day to day. The PCD model actually consists of six model components. There are the four basic models for Δt, N, D, and S, with the models for S and N each having two components. We apply the hierarchical modeling strategy separately to each of these six components.

We first provide detailed discussion of hierarchical modeling of the simple logit component of the model for N in order to illustrate the approach. We then present results for all six model components.
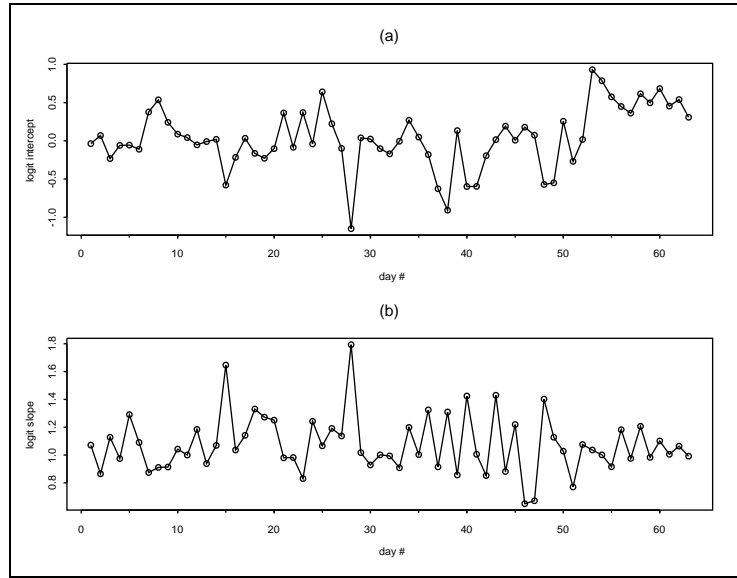
Figure 8
Time series plots of posterior means for the logit component of N by applying the PCD model to each trading day: (a) intercept α_{0j}; (b) slope α_{1j}.

5.1 The hierarchical modeling for the logit component of N

Let $\boldsymbol{\theta}_j = (\alpha_{0j}, \alpha_{1j})$ from Equation (4.4), in which j indexes the day. Thus, $\boldsymbol{\theta}_j$ represents the parameters for the logit component of the N model on day j. Note that in order to roughly orthogonalize the intercept and slope parameters, the overall (using all the days) mean of $\ln(\Delta t) = 3.5$ has been subtracted from all of the $\ln(\Delta t)$ values (on all days).

Figure 8 displays the time series of estimates (posterior means) of α_{0j} and α_{1j} obtained by applying the PCD to each day. The average intercept estimate is about .04, and the values range from -1.2 to .93. The average slope estimate is about 1.08, and the slopes range from .65 to 1.8. The average intercept is small $(\text{logit}(.04) = \exp(.04)/(1 + \exp(.04)) = .51)$. However, $\text{logit}(-1.2) = .22$ and $\text{logit}(.93) = .74$, so that the day-to-day variation in the intercepts is substantial. Since the .1 and .9 quantiles of $\ln(\Delta t)$ are 1.4 and 5.3 (see Figure 7a), we see that the slopes suggest dependence of the event $N > 0$ on $\ln(\Delta t)$ and the day-to-day variation in the slope estimates is substantial.

From Figure 8 we might wonder if day 28 is unusual because of the relatively large slope and small intercept. There is also the suggestion that the intercepts are larger after day 53 (inclusive). Neither of these features, however, is clearly distinguishable from the overall variation. In addition, we must remember that the quantities plotted are only estimates and that no attempt is made in the figure to represent our uncertainty.

We would like to elaborate our model to include a description of the variation in parameters from day to day. The way in which we elaborate our model depends on the goal of our study. If our goal were prediction for subsequent days, capturing any temporal pattern (structural shift, etc.) would be quite important. Instead, we ask the pair of somewhat simpler questions: (1) overall (that is, over all the days in our sample), what are the PCD parameters like, and (2) for which days is there strong evidence that the parameter values are "different"? For this goal the standard i.i.d. shrinkage model is suitable and convenient: we let $\boldsymbol{\theta}_j \sim N(\theta_*, \Sigma_*)$ i.i.d. Given choices for the prior distribution of (θ_*, Σ_*), we can compute the posterior distribution of these quantities and each $\boldsymbol{\theta}_j$. The parameter $\boldsymbol{\theta}_*$ can be regarded as the "overall mean" of the parameter $\boldsymbol{\theta}_j$ across all trading days, and the parameter Σ_* describes the day-to-day variation. This model also allows for adaptive shrinkage of the $\boldsymbol{\theta}_j$ toward the overall mean $\boldsymbol{\theta}_*$. If the data for a day j suggest a vector $\boldsymbol{\theta}_j$ different from the rest relative to the overall variation, but the strength of the evidence in the data is weak, then the support of
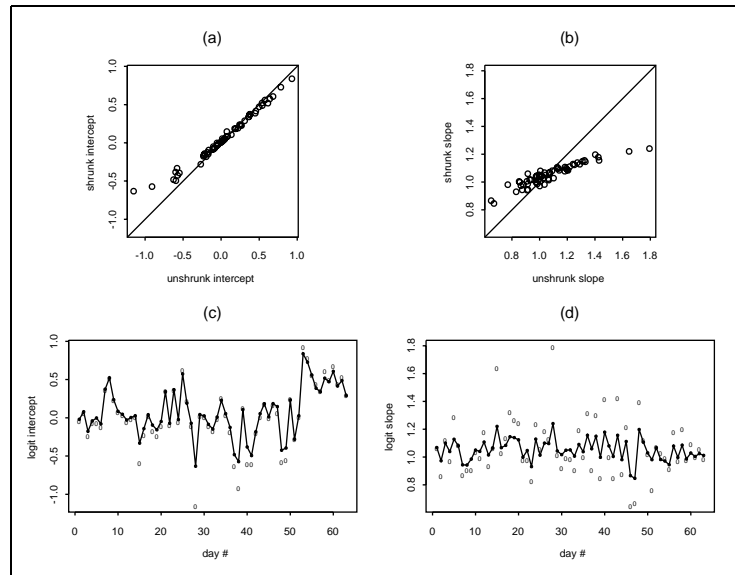
Figure 9

Comparison between hierarchical and nonhierarchical models: (a) Scatterplots of posterior means of the intercept; (b) scatterplots of posterior means of the slope; (c) time-series plots of intercept posterior means, with solid line denoting estimates of the hierarchical model; (d) time-series plots of slope posterior means, with solid line denoting estimates of the hierarchical model.

the posterior will be shrunk toward the overall mean. If the evidence is strong, however, the shrinkage will be negligible. In this way our model strikes an adaptive compromise between the extremes of treating each day separately and lumping them all together. Thus, the posterior distribution of $\boldsymbol{\theta}_*$ answers question (1) and days for which the posterior of $\boldsymbol{\theta}_j$ is not shrunk to $\boldsymbol{\theta}_*$ are the days corresponding to question (2).

To implement this approach we must first choose a prior for $(\boldsymbol{\theta}_*, \boldsymbol{\Sigma}_*)$ and then compute the posterior. The approach of Barnard, McCulloch, and Meng (2000) is used. The chosen priors are extremely diffuse. $\boldsymbol{\theta}_*$ and $\boldsymbol{\Sigma}_*$ are independent. The components of $\boldsymbol{\theta}_*$ are i.i.d. normal with mean 0 and standard deviation 1,000. The square roots of the diagonals of $\boldsymbol{\Sigma}_*$ (the standard deviations) are i.i.d. log-normal with a mean of $-.5$ and a standard deviation of 1.5. The diagonal elements of $\boldsymbol{\Sigma}_*$ are independent of the correlation matrix, which is uniformly distributed on the set of positive definite matrices having unit diagonals.

Figure 9 compares the estimates (posterior means) of $\boldsymbol{\theta}_j$ obtained by applying the PCD separately to each day (those displayed in Figure 8) with those obtained from the hierarchical model. Panel (a) plots the daily intercept estimates obtained from the hierarchical model on the vertical axis versus the nonhierarchical estimates on the horizontal axis. Panel (b) is the corresponding plot for the slopes. The line $y = x$ is drawn through both panels (a) and (b). Panel (c) plots the time series of intercept estimates where the hierachical estimates are connected by the line and the nonhierarchical estimates are plotted with an O. Panel (d) is the slope version of panel (c). We see that whereas most intercept estimates are only slightly altered by the shrinkage, the few unusually small values are substantially shrunk toward zero so that, for example, day 28 no longer looks like a dramatic outlier. The shrinkage of the slope is quite dramatic, as the shrunk values are much more tightly clustered near one. Even with the shrinkage, there is still the suggestion that the intercepts jump to larger values for the last several days.

Figure 10 displays the posterior of $(\boldsymbol{\theta}_*, \boldsymbol{\Sigma}_*)$. Panels (a)–(e) display, respectively, draws from the marginal posteriors of the first component of $\boldsymbol{\theta}_*$ (the "overall" intercept), the second component of $\boldsymbol{\theta}_*$ (the "overall" slope), the square root of the first diagonal of $\boldsymbol{\Sigma}_*$ (the standard deviation describing the day-to-day variation in intercepts), the square root of the second diagonal of $\boldsymbol{\Sigma}_*$ (the standard deviation describing the day-to-day variation in slopes), and a single correlation from the two-by-two $\boldsymbol{\Sigma}_*$ (the day-to-day correlation between
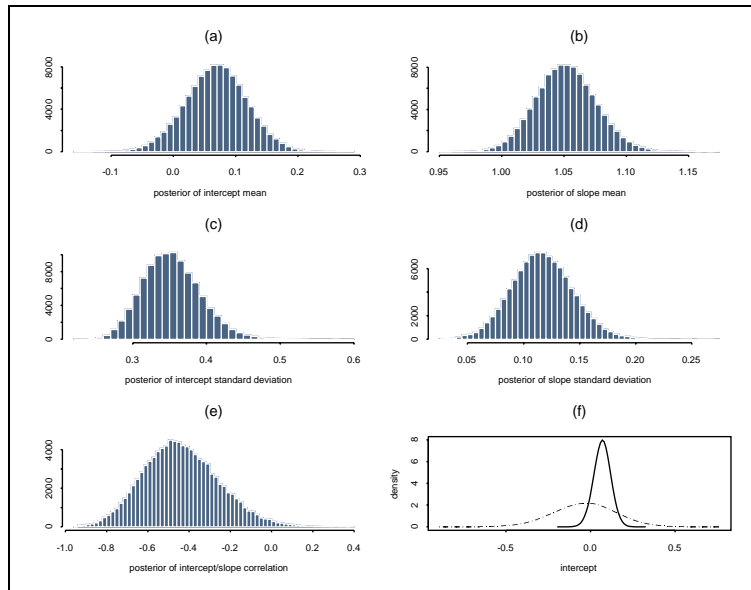
Figure 10
Posterior density functions of overall parameters of the hierarchical model.

intercepts and slopes). The solid line in panel (f) displays a kernel estimate of the posterior of the overall intercept (same quantity displayed in panel (a)), and the dashed line displays the kernel estimate of the posterior of the intercept where only the data from the first day in our sample is used. The relative tightness of the solid kernel illustrates the effect of pooling information from all 63 days.

5.2 Hierarchical results for all model components

We now present results obtained from applying the hierarchical model to each of the six components of the PCD model. For each model we present the .025, .5, and .975 posterior quantiles for the "overall mean" and the standard deviation of each parameter (components of $\boldsymbol{\theta}_*$ and square roots of diagonal elements of $\boldsymbol{\Sigma}_*$ in Section 5.1). The mean gives us an overall idea of the parameter over all 63 days, and the standard deviation describes the variation of the parameter from day to day. We do not report the posterior distributions of the day-to-day correlations between pairs of parameters.

Table 3 contains the quantiles. The table has three columns. The first column (labeled "parameter") identifies the particular parameter of the model, using the notation of Section 4.1. The second and third columns (labeled "mean" and "stan dev") give the overall mean and standard deviation quantiles. The three quantiles are listed from least to largest.

So, for example, the two rows labeled α_0 and α_1 under the heading "Logit Model for N Positive" give the intervals for the logit model discussed in detail in Section 5.1. To connect the table with our detailed discussion, the quantiles $(-.029, .069, .164)$ and $(.283, .351, .436)$ in the row labeled α_0 summarize panels (a) and (c) of Figure 10. The quantiles $(1.01, 1.05, 1.10)$ and $(.065, .116, .172)$ summarize panels (b) and (d). Note that for the time duration model (Equation (4.3)) only the βs are modeled hierarchically, so that the error standard deviation σ varies "freely" from day to day. In all other model components, all parameters are included in the hierarchical setup.

The table reveals how spread out our priors are relative to the posteriors. For all the mean parameters the prior was $N(0, 1{,}000^2)$. All of the posterior intervals in the mean column are extremely tight relative to this prior. For all the standard deviation parameters the prior was $\sigma \sim \exp(N(-.5, 1.5^2))$. This prior is again very diffuse relative to the posterior intervals. The prior is very skewed, with a 99% quantile of about 20. The 1%

Table 3

95% posterior intervals for the hierarchical parameters

parameter	mean	stan dev
	Time Duration Model	
β_0	(3.74, 3.85, 3.97)	(.319, .395, .494)
β_1	(.072, .094, .118)	(.045, .066, .090)
β_2	(−.281, −.212, −.152)	(.077, .132, .204)
	Logit Model for N Positive	
α_0	(−.029, .069, .164)	(.283, .351, .436)
α_1	(1.01, 1.05, 1.10)	(.065, .116, .172)
	Geometric Model for Positive N	
γ_0	(−.167, −.064, .044)	(.314, .384, .476)
γ_1	(−.807, −.771, −.737)	(.057, .088, .124)
	Direction Model	
ω_0	(−.011, .012, .034)	(.010, .031, .065)
ω_1	(−.914, −.879, −.844)	(.064, .099, .139)
ω_2	(−.048, −.032, −.017)	(.004, .016, .038)
β	(−.056, −.022, .012)	(.047, .082, .125)
	Poisson Model for Price Down	
$\eta_{d,0}$	(−3.04, −2.80, −2.58)	(.608, .750, .938)
$\eta_{d,1}$	(−.144, −.089, −.044)	(.096, .131, .182)
$\eta_{d,2}$	(−.232, −.175, −.114)	(.026, .106, .174)
$\eta_{d,3}$	(.526, .614, .713)	(.177, .245, .340)
	Poisson Model for Price Up	
$\eta_{u,0}$	(−3.43, −3.13, −2.86)	(.773, .973, 1.21)
$\eta_{u,1}$	(−.433, −.319, −.231)	(.179, .257, .367)
$\eta_{u,2}$	(−.177, −.125, −.067)	(.041, .101, .170)
$\eta_{u,3}$	(.817, .939, 1.07)	(.315, .409, .518)

and 99% prior quantiles of the $\sigma|\sigma < 1$ (the prior conditional on which most of posterior supports are) are .014 and .98, so that restricted to this more realistic interval, the prior is still quite spread out.

The results for α_1 and γ_1 both reflect a strong dependence of N on the duration. Detailed results for α_1 have been presented in Figure 9. The interval for the mean of γ_1 is quite tight around values that are of practical significance. If we take the median of −.77 as a point estimate and consider a change of 4 in $\ln(\Delta t)$ as possible, then an increase in duration could lead to a substantial decrease in the probability parameter of the geometric distribution (λ of Equation (4.5)). Since for the geometric distribution, smaller probabilities make larger outcomes more likely, this means that conceivable increases in $\ln(\Delta t)$ suggest larger values of N. The relatively small interval for the standard deviation (point estimate .088) tells us that the day-to-day variation in γ_1 is such that essentially the same result is obtained for all days.

Another parameter that clearly suggests an effect of practical importance for all days is ω_1 in the direction model. The posterior median of the overall mean is −.879. Since this parameter is the coefficient of lagged D, which is always ±1, the estimate suggests an important effect given the probit-type specification in Equation (4.6). The estimate (.099) and quantiles of the daily standard deviation of ω_1 clearly indicate that a comparable and important effect is likely to be present in most days. The negative sign indicates that this parameter captures the "bounce" behavior of the price: an increase is often followed by a decrease and vice versa.

The posteriors of other parameters suggest additional patterns in the data, although not as strongly: a large change in price (large S) is more likely if the previous change was large ($\eta_{d,3}$ and $\eta_{u,3}$), a longer duration
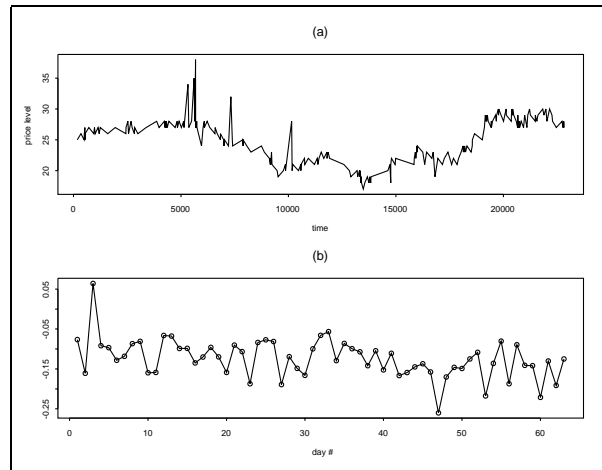
Figure 11
(a) Time plot of the price series for day 3; (b) time plot of daily shrunk estimates of the parameter $\eta_{u,2}$ in the Poisson model from price increases.
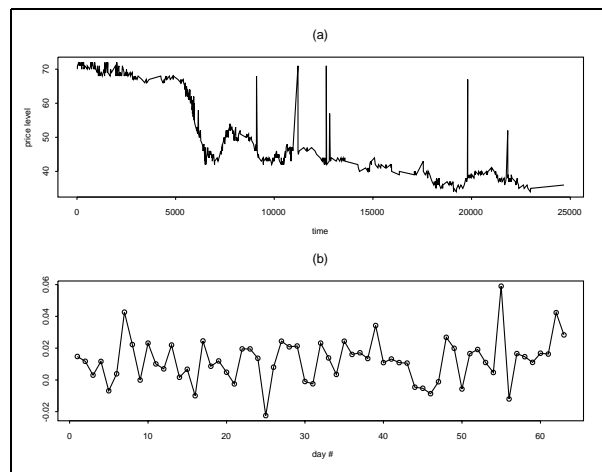

Figure 12
(a) Time plot of the price series for day 25; (b) time plot of the daily shrunk estimates of the parameter ω_0 in the direction model.

makes a price fall more likely (ω_2), and a large price change leads to a shorter duration. Overall, the only parameters whose 95% posterior intervals for the mean parameter include zero are the three intercepts α_0, γ_0, and ω_0, and β.

Some days are identified as unusual in that even estimates obtained from the shrinkage model look unusual compared to those of other days. Panel (b) of Figure 11 plots the time series of daily shrunk estimates of the parameter $\eta_{u,2}$ in the Poisson model from price increases. We see that the estimate for day 3 is unusually large. Panel (a) of Figure 11 plots the time series of prices for that day (with the first price of the first day subtracted). In the price plot there are five "spikes," indicating a sharp price increase followed by an offsetting decrease. The reader may wish to compare panel (a) to panel (b) of Figure 6, which is a more "typical" day in the judgment of the authors. Another unusual day is illustrated by Figure 12. Panel (a) shows the price series for day 25, and panel (b) shows the daily shrunk estimates of the parameter ω_0 in the direction model. The estimate for day 25 is unusually small. Actually, several parameters have unusual estimates for this day. The price series for this day has a drop followed by several spikes that seem to reach back up to the predrop

level. A plausible explanation for the sharp price changes in day 25 is the limit orders, because the transaction price of a trade associated with a sharp price increase was close to those of trades that occurred in the morning of that day.

References

Bai, X., J. R. Russell, and G. C. Tiao. (2000). "Beyond Metron's utopia: Effects of dependence and non-normality on variance estimates using high-frequency data." Working paper. Chicago: Graduate School of Business, University of Chicago.

Barnard, J., R. E. McCulloch, and X. L. Meng. (2000). "Modeling covariance matrices in terms of standard deviations and correlations with application to shrinkage." *Statistica Sinica*, 10: 1281–1311.

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

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. (1995). *Bayesian Data Analysis*. London: CRC Press.

Ghysels, E. (2000). "Some econometric recipes for high-frequency data cooking." *Journal of Business and Economic Statistics*, 18: 154–163.

Hasbrouck, J. (1992). "Using the TQRQ database." Unpublished manuscript. New York: Stern School, New York University.

Hasbrouck, J. (1999). "The dynamics of discrete bid and ask quotes." *Journal of Finance*, 54: 2109–2142.

Rydberg, T. H., and N. Shephard. (1998). "Dynamics of trade-by-trade price movements: Decomposition and models." Working paper. Oxford: Nutfield College, Oxford University.

Tsay, R. S. (1986). "Nonlinearity tests for time series." *Biometrika*, 73: 461–466.

Tsay, R. S. (1989). "Testing and modeling threshold autoregressive processes." *Journal of the American Statistical Association*, 84: 231–240.

Tsay, R. S. (2000). "Analysis of financial time series." Lecture notes. Chicago: Graduate School of Business, University of Chicago.

Wood, R. A. (2000). "Market microstructure research databases: history and projections." *Journal of Business & Economic Statistics*, 18: 140–145.

Zhang, M. Y., J. R. Russell, and R. S. Tsay. (2000). "Determinants of bid and ask quotes and implications to the cost of trading." Working paper. Chicago: Graduate School of Business, University of Chicago.

Zhang, M. Y., J. R. Russell, and R. S. Tsay. (2001). "A nonlinear autoregressive conditional duration model with applications to financial transaction data." *Journal of Econometrics*, 104: 179–207.