Journal of Financial Econometrics, 2004, Vol. 2, No. 3, 390–421

Stochastic Conditional Duration Models with "Leverage Effect" for Financial Transaction Data

DINGAN FENG
York University

GEORGE J. JIANG
University of Arizona

PETER X.-K. SONG
York University

ABSTRACT

This article proposes stochastic conditional duration (SCD) models with "leverage effect" for financial transaction data, which extends both the autoregressive conditional duration (ACD) model (Engle and Russell, 1998, *Econometrica*, 66, 1127–1162) and the existing SCD model (Bauwens and Veredas, 2004, *Journal of Econometrics*, 119, 381–412). The proposed models belong to a class of linear nongaussian state-space models, where the observation equation for the duration process takes an additive form of a latent process and a noise term. The latent process is driven by an autoregressive component to characterize the transition property and a term associated with the observed duration. The inclusion of such a term allows the model to capture the asymmetric behavior or "leverage effect" of the expected duration. The Monte Carlo maximum-likelihood (MCML) method is employed for consistent and efficient parameter estimation with applications to the transaction data of IBM and other stocks. Our analysis suggests that trade intensity is correlated with stock return volatility and modeling the duration process with "leverage effect" can enhance the forecasting performance of intraday volatility.

KEYWORDS: autoregressive conditional duration (ACD) model, ergodicity, financial transaction data, leverage effect, Monte Carlo maximum-likelihood (MCML) estimation, stationarity, stochastic conditional duration (SCD) model

We wish to thank Professor Eric Renault (the editor), an associate editor, and two referees for very helpful comments and suggestions. We also wish to thank Luc Bauwens, Christian Gouriéroux, Joann Jasiak, Tom Salisbury, Neil Shephard, and David Veredas for helpful discussion and suggestions. The usual disclaimer applies. The second and third authors acknowledge financial support from NSERC, Canada. Address correspondence to George J. Jiang, Department of Finance, Eller College of Business, University of Arizona, Tucson, AZ 85721-0108, or e-mail: gjiang@eller.arizona.edu.

doi:10.1093/jjfinec/nbh016
Journal of Financial Econometrics, Vol. 2, No. 3, © Oxford University Press 2004; all rights reserved.

Modeling the trade duration of the financial market has recently drawn a great deal of attention in the statistical and financial econometric literature. Due to the rapid development of technologies for data collection and the growing capacity of data storage, massive transaction records in the financial market are available. Such voluminous data provide a wealth of information about the activities and microstructures of the financial market, yet they also give rise to the challenge of developing appropriate dynamic models. One major complicating factor with transaction data is that they are typically irregularly spaced. Modeling of an irregularly spaced "marked point process" involves the complex dynamic structure of random arrival times. It presents a great challenge to statisticians and econometricians, as most standard econometric techniques are developed to deal with fixed-interval random processes. In their seminal article, Engle and Russell (1998) propose an autoregressive conditional duration (ACD) model, which is essentially an ARMA process with nongaussian innovations and in the line of the well-known autoregressive conditionally heteroskedastic (ARCH) [Engle (1982)] and generalized ARCH (GARCH) [Bollerslev (1986)] models for asset returns. The major advantage of the ACD model is the availability of maximum-likelihood (ML) inference, which furnishes a great deal of ease and efficiency in parameter estimation both conceptually and computationally.

Various extensions have been proposed in the literature either generalizing the distribution of the disturbance term or incorporating other state variables in the duration process. For instance, Grammig and Maurer (2000) extend the Weibull distribution in the ACD model by Engle and Russell (1998) to the Burr distribution in order to have a more flexible shape for the conditional hazard function. Veredas, Rodriguez-Poo, and Espasa (2001) further extend the model to a semiparametric framework for the joint analysis of trade duration dynamics and intraday seasonality. Bauwens and Giot (2003) model the conditional duration process based on the state of the asset price process in an asymmetric ACD model. Meddahi, Renault, and Werker (1998) extend the ACD model to the continuous-time modeling of stochastic volatility using irregularly spaced data. Other studies have extended the modeling of duration with a GARCH model for the conditional volatility of asset returns [see Ghysels and Jasiak (1998), Engle (2000), and Grammig and Wellner (2002)]. Finally, Bauwens and Giot (2000) propose a log ACD model so that the positivity constraint on the state variables can be relaxed. For a survey of the literature and a comparison of various models, see Bauwens et al. (2000).

In recent literature, the ACD model is also extended to the latent-factor models, such as the stochastic volatility duration (SVD) model of Ghysels, Gouriéroux, and Jasiak (2004) and the stochastic conditional duration (SCD) model of Bauwens and Veredas (2004). In the SVD model of Ghysels, Gouriéroux, and Jasiak (2004), the volatility of the trade duration is assumed to be stochastic and the duration is driven by a mixture of distributions, namely the combination of gamma and exponential distributions. The authors believe that it is not sufficient to model the duration process by only incorporating randomness into the conditional mean. Compared to the ACD models of Engle and Russell (1998), the

SCD models proposed by Bauwens and Veredas (2004) are based on the assumption that the evolution of the conditional duration is driven by a latent variable. By incorporating a latent variable in the conditional duration process, the SCD model in Bauwens and Veredas (2004) offers a flexible structure for the dynamics of the duration process. Extension of the SCD model over the ACD model is similar to that of the SV model over the GARCH model in the asset return literature. Similar to asset return in the SV model, trade duration under the SCD model is also modeled as a mixture of distributions. As pointed out by Ghysels, Harvey, and Renault (1996), there are various advantages of modeling asset return dynamics in an SV model framework relative to the ARCH/GARCH model framework. We shall discuss the advantages of modeling the duration process in an SCD model framework. The main challenge with the SCD model is, however, its statistical inference, as it involves unobserved latent variables in the likelihood function.

In this article we propose a further extension to Bauwens and Veredas's (2004) SCD model in order to capture the asymmetric behavior or "leverage effect" in the duration process. To reflect the asymmetric behavior, our model includes an intertemporal term associated with the observed duration in the latent process. The inclusion of such a term gives further flexibility to capture the local movements and random spikes of the trade duration. For statistical inference of our models, we adopt the Monte Carlo maximum-likelihood (MCML) approach proposed by Durbin and Koopman (1997), which produces not only consistent but also efficient parameter estimators. The MCML procedure is a powerful inference tool to deal with parameter estimation of nonnormal parametric families. With a selected importance distribution (e.g., normal distribution) under which certain standard estimation procedures apply, the original intricate estimation problem involving non-normal distributions can be reformulated. In the present article, as part of the MCML procedure, the normal distribution is used as the importance distribution under which it is possible to perform the standard Kalman filter procedure as the key to estimation.

The article is structured as follows. Section 1 presents SCD models with "leverage effect" and then discusses some analytical properties of the proposed models. Section 2 introduces the MCML estimation procedure, and its application is illustrated in Section 3 using the transaction data of IBM and other stocks. Section 4 concerns the diagnostics of model specification, with special attention to the implications of "leverage effect" in the SCD models. We conclude in Section 5. Proofs of all propositions are collected in the appendix.

1 MODELS

1.1 Formulation

In modeling the arrival times of a "marked point process," a common approach in the literature is to model the conditional intensity process. For example, Cox's doubly stochastic model assumes that there is a latent independent process that governs the arrival rate, and such a process is itself a self-exciting process. Engle and Russell (1998) introduce a new family of self-exciting processes for the irregularly spaced transaction data where the duration process at the current time is assumed to follow a multiplicative model conditional on the past. To be specific, let $d_i = t_i - t_{i-1}, i = 1, 2, \ldots$, be the length of the interval between two trade times, termed the trade duration, and let ψ_i be the conditional expectation of the ith duration given all the past durations,

$$E(d_i|d_{i-1}, \ldots, d_1) = \psi_i(d_{i-1}, \ldots, d_1; \theta) \equiv \psi_i,$$

where ψ_i may be dependent on a parameter vector θ.

The ACD(m, q) model specified in Engle and Russell (1998) takes the following form:

$$d_i = \psi_i \epsilon_i, \tag{1}$$

$$\psi_i = \omega + \sum_{j=0}^{m} \alpha_j d_{i-j} + \sum_{j=0}^{q} \beta_j \psi_{i-j},$$

where ϵ_i is an i.i.d. innovation with a given parametric density $p(\epsilon; \phi)$. That is, the conditional duration ψ_i is assumed to follow an autoregressive process with a GARCH structure. Under such a parametric specification, the ML estimation can be applied for inference.

It is noted that when log transformation is taken on both sides of the ACD model in Equation (1), it results in an additive form of the logarithmic conditional duration and the error term. This relaxes the positivity restriction on the variable ψ_i and motivates some other developments using log duration other than the duration itself [see, e.g., Bauwens and Giot (2000)].

In the present article we consider the stochastic process for the log duration and propose SCD models which, in a general setting, are in the following state-space form,

$$\begin{array}{ll} \text{the observation equation:} & \log(d_i) = g(\psi_i, \epsilon_i) \\ \text{the latent equation:} & \psi_i = h(\psi_{i-1}, \ldots, \psi_{i-p}, \epsilon_{i-1}, \ldots, \epsilon_{i-r}, \eta_i, \ldots, \eta_{i-q}), \end{array}$$

where $g(\cdot)$ and $h(\cdot)$ are known continuous functions, and error distributions for ϵ_i and η_i may be nongaussian. For example, when g is chosen such that the observation equation reduces to Equation (1) and the latent equation takes an ARMA structure with absence of η_i, this reduces to the Engle and Russell (1998) ACD model. Also, g and h can be chosen so that the model reduces to the Bauwens and Veredas (2004) SCD model with the following specification:

$$\log(d_i) = \mu + \psi_i + \epsilon_i,$$

$$\psi_i = \beta\psi_{i-1} + \eta_i, \quad |\beta| < 1.$$

It is clear that the SCD models proposed by Bauwens and Veredas (2004) are extensions of the ACD models of Engle and Russell (1998). As pointed out above, by incorporating a latent variable in the conditional duration process, the SCD models in Bauwens and Veredas (2004) offer a flexible structure for the dynamics of the duration process. We also noted that extension of the SCD model over the ACD model is similar to that of the SV model over the GARCH model in the asset return literature. To further appreciate the SCD models and, more importantly, to motivate the model specification proposed in this article, here we summarize the analysis of the SV model versus the GARCH model in Ghysels, Harvey, and Renault (1996). As noted in Ghysels, Harvey, and Renault (1996), there are various advantages of modeling asset return dynamics in an SV model framework in comparison to the ARCH/GARCH model framework. The GARCH model, proposed by Bollerslev (1986) by extending the ARCH model of Engle (1982) and applied extensively to financial time series, assumes the conditional volatility to be a deterministic function of observed variables. The appeal of the GARCH model is its straightforward application of the ML estimation. The SV model extends the GARCH model by allowing the conditional volatility to be stochastic with its own disturbance term. The SV model has been shown to have a better fit to the autocorrelation functions (ACFs) of squared asset returns [see Jacquier, Polson, and Rossi (1994)].

In terms of capturing the stylized facts of financial asset returns, namely the asset return distribution with negative skewness and excess kurtosis or fat tails, the SV model has certain advantages over the GARCH model. The SV model displays excess kurtosis even if the conditional volatility is not autoregressive. This is because the asset return under the SV model framework is modeled as a mixture of distributions. It is, however, not the case with a GARCH model, where the degree of kurtosis depends directly on the roots of the variance equation. Thus, very often a nongaussian GARCH model has to be employed to capture the excess kurtosis typically found in a financial time series. In addition, in the SV model, when the disturbance terms in the asset return process and the conditional volatility process are allowed to be correlated to each other, the model can pick up the kind of asymmetric behavior that is often found in stock prices. In particular, when the correlation between the return and conditional volatility is negative, the model induces the so-called leverage effect [see Black (1976)]. In other words, higher volatility tends to be associated with a negative return in equity or an increase of a firm's leverage (debt/equity ratio). The basic GARCH model, however, does not allow for the kind of asymmetric behavior as captured easily by the SV model. The extension to correlate asset return and conditional volatility in a GARCH model framework is less straightforward. For instance, the EGARCH model proposed by Nelson (1991) handles the asymmetry by specifying the log volatility as a function of past squared and absolute return observations.

Similar to asset return in the SV model, trade duration under the SCD model framework is also modeled as a mixture of distributions. In particular, the SCD models combine a lognormal distribution with another one of positive support. For instance, Bauwens and Veredas (2004) specify SCD models with log-Weibull (LW) and log-gamma (LG) errors for the conditional duration process. The main challenge with the SCD model is its statistical inference, as the likelihood function becomes difficult to evaluate because of the need to integrate the unobserved latent variables. The quasi-maximum-likelihood (QML) estimation method is implemented in Bauwens and Veredas (2004) for parameter estimation with application of the Kalman filter after transforming the model into a linear state-space representation.

The motivation for further extension of the Bauwens and Veredas (2004) SCD models is the asymmetric structure or the "leverage effect" incorporated in the SV model. Note that in the representation of Bauwens and Veredas's (2004) latent equation, the errors (ϵ_i) associated with the observed process are not present. Such a state-space model can capture the dynamic features of the duration process as to be driven by the Markov component. However, it may oversimplify the behavior of local movements, as this process tends to oversmooth the expected duration. Since the observed duration series may have local asymmetric changes, it seems desirable to include ϵ_i in the latent equation, which models the variation beyond what the variable η_i can describe.

The term "leverage effect," as we have noted, has specific meaning in finance. Here we borrow this term simply because of the similarity in model structure, not because of the financial interpretations. In fact, in this article we actually find a positive intertemporal correlation between observed duration and expected conditional duration, which is equivalent to a negative relationship between trade intensity and observed duration.

The extended model is specified as follows with an intertemporal term in the latent process to capture the "leverage effect,"

$$\begin{aligned} &\log(d_i) = \mu + \psi_i + \epsilon_i, \\ &\psi_i = \beta\psi_{i-1} + \gamma\epsilon_{i-1} + \eta_i, \quad |\beta| < 1, \end{aligned} \tag{2}$$

where ϵ_i and η_i are i.i.d. innovations and ϵ_i and η_i are mutually independent. Note that because of the presence of ϵ_i, the latent process is effectively intertemporally correlated with the duration process.

To parameterize the distributions of noise terms, we assume that η_i follows gaussian $N(0, \sigma_\eta^2)$. For the distribution of ϵ_i, we consider three cases, namely log-Weibull$(\nu, 1)$ (hereafter LW$(\nu, 1)$), log-gamma$(\nu, 1)$ (hereafter LG$(\nu, 1)$), and log standard exponential (LE) which is LW$(1, 1)$ or LG$(1, 1)$. We now summarize some basic properties of these three distributions in Table 1, which are useful in our later development of model estimation.

Table 1 Summary of three density functions.

Distribution	Scale parameter	Density function	Mean	Variance
LW(ν, 1)	$\nu > 0$	$f(\epsilon) = \nu\exp(\nu\epsilon - e^{\nu\epsilon})$	$-\dfrac{C}{\nu}$	$\dfrac{\pi^2}{6\nu^2}$
LG(ν, 1)	$\nu > 0$	$f(\epsilon) = \dfrac{1}{\Gamma(\nu)}\exp(\nu\epsilon - e^{\epsilon})$	$\xi(\nu)$	$\xi'(\nu)$
LE	1	$f(\epsilon) = \exp(\epsilon - e^{\epsilon})$	$-$C	$\frac{\pi^2}{6}$

The table reports the mean and variance of relevant density functions for ϵ_i, where $\xi(\nu)$ is the logarithmic derivative of the gamma function, or the so-called digamma function, namely $\xi(\nu) = \frac{d\ln\Gamma(\nu)}{d\nu}$, and the constant $C = -\xi(1) = \int_0^\infty e^{-x}\ln x dx$ is the Euler constant, which is known to be approximately equal to 0.5772157.

1.2 Statistical Properties

In this section we study statistical properties of the processes $\{y_i\}$, where $y_i = \log d_i - \mu$, and $\{d_i\}$, $i = 1, 2, \ldots,$ both specified in Equation (2). In particular, some moments of these processes are derived which will be used in the development of the MCML estimation in Section 2. Proofs of these results are given in the appendix.

Proposition 1 *The process $\{y_i = \log d_i - \mu\}$ as defined in Equation (2) is weakly stationary and geometrically ergodic if $|\beta| < 1$, so is the duration process $\{d_i\}$.*

It is noted that for the model specified in Equation (2), the condition for ergodicity is the same for stationarity. It is known that for an ergodic (or geometrically ergodic) Markov process, there exists a limiting distribution. In other words, the distribution of the process converges to the limiting distribution. Moreover, a single trajectory represents the whole probability law of the process.

Proposition 2 *For the process $\{y_i\}$ as defined above, we have the following unconditional and intertemporal moments*:

$$
\begin{aligned}
E[y_i^2] &= (1-\beta^2)^{-1}\{(1+\gamma^2-\beta^2)m_2^\epsilon + m_2^\eta\} \\
E[y_i^3] &= (1-\beta^3)^{-1}\{(1+\gamma^3-\beta^3)m_3^\epsilon + m_3^\eta\} \\
E[y_i^4] &= \left(1+\frac{\gamma^4}{1-\beta^4}\right)m_4^\epsilon + 12\frac{\gamma^2}{1-\beta^2}\left(1+\gamma^2\frac{\beta^2}{1-\beta^4}\right)(m_2^\epsilon)^2 \\
&\quad + 6\left(1+\frac{\gamma^2}{1-\beta^2}\right)m_2^\epsilon\frac{1}{1-\beta^2}m_2^\eta + \frac{m_4^\eta}{1-\beta^4} + 12\frac{\beta^2}{1-\beta^2}\frac{1}{1-\beta^4}(m_2^\eta)^2 \\
\operatorname{cov}(y_i, y_{i-s}) &= \left(\gamma\beta^{s-1} + \frac{\gamma^2\beta^s}{1-\beta^2}\right)m_2^\epsilon + \frac{\beta^s}{1-\beta^2}m_2^\eta, \quad s \geq 1,
\end{aligned}
$$

where m_j^ϵ and *m_j^η are, respectively, jth moments of ϵ and η, respectively, $j = 2, 3, 4$.*

Proposition 3 *For the process $\{d_i\}$ as defined in Equation (2), the rth moment is*

$$
Ed_i^r = \exp(r\mu)\prod_{j=0}^{\infty} m(r\alpha_j)\exp\left\{\frac{r^2\sigma^2}{2(1-\beta^2)}\right\}.
$$

In particular, the first and second moments are

$$
\begin{aligned}
\mathrm{E}d_i &= \exp(\mu)\prod_{j=0}^{\infty} m(\alpha_j)\exp\left\{\frac{\sigma^2}{2(1-\beta^2)}\right\} \\
Ed_i^2 &= \exp(2\mu)\prod_{j=0}^{\infty} m(2\alpha_j)\exp\left(\frac{2\sigma^2}{1-\beta^2}\right),
\end{aligned}
$$

where

$$
m(\alpha) = \begin{cases} \Gamma\left(\dfrac{\alpha}{\nu}+1\right), & \text{when } \epsilon_j \text{ is } LW(\nu, 1) \\ \dfrac{\Gamma(\nu+\alpha-1)}{\Gamma(\nu)}, & \text{when } \epsilon_j \text{ is } LG(\nu, 1). \end{cases}
$$

When the intertemporal term is $\gamma = 0$, the mean and variance of the duration are the same as those given in Bauwens and Veredas (2004).

Proposition 4 *For the process $\{y_i\}$ as defined in Equation (2), the following three important properties can be derived immediately when $|\beta| < 1$.*

(1) *The first lag autocorrelation function is given by*

$$\rho_1 = \frac{E[y_i y_{i-1}]}{E[y_i^2]} = \frac{\dfrac{\beta\gamma^2}{1-\beta^2}\sigma_1^2 + \gamma\sigma_1^2 + \dfrac{\beta}{1-\beta^2}\sigma_2^2}{\dfrac{\gamma^2}{1-\beta^2}\sigma_1^2 + \sigma_1^2 + \dfrac{1}{1-\beta^2}\sigma_2^2}.$$

(2) *Let $\rho_s (s \geq 1)$ be the sth ACF, then we have that $\rho_s/\rho_{s-1} = \beta$. It is obvious that the process is highly persistent if β is close to one.*

(3) *The kurtosis of the process y_i is larger than three. In other words, compared to the normal distribution, the process y_i has a leptokurtic distribution with fat tails.*

The role of the parameter γ on the unconditioned moments can be easily seen in Proposition 2. The presence of the "leverage effect," that is, $\gamma \neq 0$, inflates the variance and the fourth moment. The sign of γ determines whether or not the third moment as a function of γ increases or decreases. In addition, the kurtosis varies in terms of both the sign and magnitude of γ, which adds additional flexibility in modeling financial data.

2 ESTIMATION

The difficulty of parameter estimation for nongaussian state-space models as specified in Equation (2) arises from the fact that the conditional density of the duration involves the latent or unobserved variable. Unlike the ACD model, in which the likelihood function can be expressed in an explicit form, the likelihood function for the SCD model is very complex due to the curse of high dimensionality. Mostly the high-dimensional integral in the likelihood function cannot be expressed as the form of one-dimensional (or substantially lower) integrals. For related issues, see Danielsson (1994), Duffie and Singleton (1993) and Durbin and Koopman (1997). In the context of estimation, issues such as evaluating high-dimensional integration encountered here are very similar to those for the SV model. In fact, the SV model is proposed ahead of ARCH or GARCH models in the statistics literature, but it did not become popular until powerful computers became available to attack intensive computation [see Danielsson (1994)].

Various estimation methods have been proposed for state-space models. One of the earliest methods is the Kalman filter algorithm, first studied by Kalman and Bucy (1961). This method is developed for linear and gaussian state-space models to compute the ML estimates of the state variables recursively, where both error terms in the observation equation and the transition equation are normally distributed [Harvey (1989)]. When parameters other than the state variables are involved in a state-space model, a typical approach to parameter estimation is the expectation maximization (EM) algorithm. In the E step, the Kalman filter technique sequentially produces the estimates of the state variables via conditional expectations, and the M step then maximizes the resulting likelihood function, which can be explicitly evaluated with the given estimates of the state variables. As shown in West and Harrison (1989), the resulting estimates from the EM algorithm are the ML estimates. However, when at least one of the disturbances is nongaussian, analogy to the EM algorithm using the Kalman filter in the E step is, in general, not efficient, and even inconsistent for some cases [see Jørgensen et al. (1999)]. Due largely to the fact that the Kalman filter is conceptually simple and computationally tractable, researchers are still willing to adopt it into some estimation procedures. For instance, the quasi-ML estimation used in Bauwens and Veredas (2004) for their SCD model is based on an adopted Kalman filter technique, initially considered in Harvey, Ruiz, and Shephard (1994).

In recent years, more estimation methods have been proposed in the literature for dynamic models with latent variables. The first type of estimation method requires knowledge of the distribution function to implement the ML estimate of different variations. For example, Jacquier, Polson, and Rossi (1994) proposed the Bayesian Markov chain Monte Carlo (MCMC) method, Danielsson and Richard (1993) proposed the simulated maximum-likelihood (SML) method using the accelerated gaussian importance sampler (AGIS), and Shephard and Pitt (1997) and Durbin and Koopman (1997), as well as Sandmann and Koopman (1998), proposed the MCML method. The second type of estimation method requires only moment conditions to form estimating equations. Examples of this type include the simple method of moments by Taylor (1986), the generalized method of moments (GMM) by Hansen (1982), the simulated method of moments (SMM) by Duffie and Singleton (1993), the indirect inference developed by Gouriéroux, Monfort, and Renault (1993) and Smith (1993), and the efficient method of moments (EMM) by Gallant and Tauchen (1996).

2.1 MCML Estimation

In the present article we adopt the MCML estimation for the SCD model with "leverage effect." The main idea behind the MCML estimation is to convert the intractable likelihood function associated with nongaussian distribution into a setting where related computations become feasible. In the context of the SCD models, we first approximate the nongaussian distribution of ϵ_i by a gaussian distribution, resulting in an approximate model for which the EM algorithm, with the E step being the classical Kalman filter, is applicable. Then the Monte Carlo method is employed to evaluate the difference of the likelihood functions between the original model and the approximate model. Such an evaluation of the difference in the likelihood functions is necessary in the EM algorithm to correct the bias in estimation, which effectively leads to consistent estimators. In order to achieve efficiency, both antithetic variables and control variables in the Monte Carlo simulation [Campbell, Lo, and MacKinlay (1997)] are also used.

To present the estimation procedure for the SCD model, we first give a brief summary of MCML estimation proposed by Durbin and Koopman (1997). Let us first consider a more general setup than Equation (2) with $y_i = \log(d_i) - \mu$ defined by the following data-generating process,

$$y_i|\psi_i = F(\psi_i, \epsilon_i; \boldsymbol{\theta}) \quad \text{and} \quad \psi_i|I_{i-1} = G(\psi_{i-1}, \epsilon_{1-1}, \eta_i; \boldsymbol{\theta}), \tag{3}$$

where ϵ_i follows a distribution with density $p(\epsilon_i)$, η_i is normally distributed with $\eta_i \sim N(0, \Sigma_i)$, θ is the set of parameters to be estimated, and F and G are two given functions. Let $y_1, \ldots, y_n$ be the observations of trade durations, and

$$\boldsymbol{\psi} = (\psi_1, \ldots, \psi_n)', \quad \text{and} \quad \mathbf{y} = (y_1, \ldots, y_n)'.$$

In the following, $p(\cdot|\cdot)$ is a generic notation denoting a conditional density function. Then the likelihood function for the parameter $\boldsymbol{\theta}$ is

$$L(\boldsymbol{\theta}) = p(\mathbf{y}|\boldsymbol{\theta}) = \int p(\mathbf{y}, \boldsymbol{\psi}|\boldsymbol{\theta}) d\boldsymbol{\psi} = \int p(\mathbf{y}|\boldsymbol{\psi}, \boldsymbol{\theta}) p(\boldsymbol{\psi}|\boldsymbol{\theta}) d\psi. \tag{4}$$

We now approximate $p(\epsilon_i)$ by a normal density of $N(\mu_i, H_i)$ with both μ_i and H_i matching the first two moments of $p(\epsilon_i)$, respectively.

Under the proposed approximate $N(\mu_i, H_i)$ for ϵ_i in Equation (3), the resulting likelihood function becomes

$$L_g(\boldsymbol{\theta}) = g(\mathbf{y}|\boldsymbol{\theta}) = \frac{g(\mathbf{y}, \boldsymbol{\psi}|\boldsymbol{\theta})}{g(\boldsymbol{\psi}|\mathbf{y}, \boldsymbol{\theta})} = \frac{g(\mathbf{y}|\boldsymbol{\psi}, \boldsymbol{\theta}) p(\boldsymbol{\psi}|\boldsymbol{\theta})}{g(\boldsymbol{\psi}|\mathbf{y}, \boldsymbol{\theta})}, \tag{5}$$

where $g(\cdot|\cdot)$ is a generic notation for a conditional density function corresponding to the approximate model with error terms ϵ_i being normally distributed.

It follows from Equation (5) that

$$p(\boldsymbol{\psi}|\boldsymbol{\theta}) = L_g(\boldsymbol{\theta}) \frac{g(\boldsymbol{\psi}|\mathbf{y}, \boldsymbol{\theta})}{g(\mathbf{y}|\boldsymbol{\psi}, \boldsymbol{\theta})}. \tag{6}$$

Plugging Equation (6) into Equation (4), we get

$$L(\boldsymbol{\theta}) = L_g(\boldsymbol{\theta}) \int \frac{p(\mathbf{y}|\boldsymbol{\psi}, \boldsymbol{\theta})}{g(\mathbf{y}|\boldsymbol{\psi}, \boldsymbol{\theta})} g(\boldsymbol{\psi}|\boldsymbol{\theta}, \mathbf{y}) d\boldsymbol{\psi} = L_g(\boldsymbol{\theta}) \mathrm{E}_g \left[\frac{p(\mathbf{y}|\boldsymbol{\psi}, \boldsymbol{\theta})}{g(\mathbf{y}|\boldsymbol{\psi}, \boldsymbol{\theta})} \right] = L_g(\boldsymbol{\theta}) \mathrm{E}_g [w(\mathbf{y}, \boldsymbol{\psi}|\boldsymbol{\theta})], \tag{7}$$

where $w(\mathbf{y}, \boldsymbol{\psi}|\boldsymbol{\theta}) = \frac{p(\mathbf{y}, \boldsymbol{\psi}|\boldsymbol{\theta})}{g(\mathbf{y}, \boldsymbol{\psi}|\boldsymbol{\theta})}$. Taking the logarithm on both sides of Equation (7) leads to

$$\log\{L(\boldsymbol{\theta})\} = \log\{L_g(\boldsymbol{\theta})\} + \log\{\mathrm{E}_g[w(\mathbf{y}, \boldsymbol{\psi}|\boldsymbol{\theta})]\}. \tag{8}$$

Thus, finding the ML estimator of θ from $\log\{L(\boldsymbol{\theta})\}$ can be done via the following two steps: step 1 maximizes $\log\{L_g(\boldsymbol{\theta})\}$ with respect to $\boldsymbol{\theta}$ through the EM algorithm where the E step is computed by the Kalman filter, and step 2 evaluates the log expectation of $w(y, \psi|\theta)$ for bias correction in step 1. Finally, the Monte Carlo method with the antithetic variable and control variable is used to evaluate the log expectation. An unbiased estimator of the log expectation is given by $\log \bar{w} + \frac{s_w^2}{2N\bar{w}^2}$, where $\bar{w}$ and s_w^2 are the sample mean and variance of a Monte Carlo sample of size N for variable w, respectively [see Durbin and Koopman (1997)].

Under certain mild regularity conditions, the MCML estimators are consistent and asymptotically normally distributed, with the asymptotic variance-covariance matrix being the inverse of the Fisher information matrix. That is, for the MCML estimator, $\hat{\boldsymbol{\theta}}$, obtained by maximizing the log likelihood function in Equation (8), we have $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \overset{a}{\sim} N(0, I^{-1}(\boldsymbol{\theta}))$, where $I(\boldsymbol{\theta})$ is the observed Fisher information matrix, which can be computed based on the second derivative of Equation (8) with respect to the parameter vector.

2.2 Estimation of the SCD Model

Now we consider estimation of the SCD model with "leverage effect" as specified in Equation (2) with ϵ_i following LW or LG, respectively. Either case with the scale parameter equal to one leads to LE. Both LW and LG distributions are commonly used for modeling financial variables with positive supports. Here we describe the gaussian approximation in the MCML estimation, that is, how to approximate a nongaussian conditional density $p(y_i|\psi_i)$ by a gaussian conditional density $g(y_i|\psi_i)$. The idea is to simply choose μ_i and H_i such that the first two derivatives of $p(y_i|\psi_i)$ and $g(y_i|\psi_i)$ (or $\ln p(y_i|\psi_i)$ and $\ln g(y_i|\psi_i)$) with respect to ψ_i are equal.

Based on the model specification in Equation (2), $\eta_i = y_i - \psi_i$, thus $p(y_i|\psi_i) = f(\eta_i)$. For the LW($\nu$, 1) distribution, at a given i, we have

$$\frac{\partial \log f(\epsilon_i)}{\partial \epsilon_i} = -\nu + \nu e^{\nu\epsilon_i} \quad \text{and} \quad \frac{\partial^2 \log f(\epsilon_i)}{\partial \epsilon_i^2} = -\nu^2 e^{\nu\epsilon_i}.$$

Using $N(\mu_i, H_i)$ to approximate LW(ν, 1), we match their first two moments by the following equations:

$$-\nu + \nu e^{\nu\epsilon_i} + H_i^{-1}(\xi_i - \mu_i) = 0 \quad \text{and} \quad -\nu^2 e^{\nu\epsilon_i} + H_i^{-1} = 0, \tag{9}$$

where ξ_i is the variable that follows $N(\mu_i, H_i)$. The solutions to the equations, at a given iteration, are functions of $\hat{\psi}$, which are obtained by the Kalman filter or smoother.

Similarly the LG(ν, 1) distribution leads to

$$\frac{\partial \log f(\epsilon_i)}{\partial \epsilon_i} = -\nu + e^{\epsilon_i} \quad \text{and} \quad \frac{\partial^2 \log f(\epsilon_i)}{\partial \epsilon_i^2} = -e^{\epsilon_i}.$$

Thus the moments matching requires

$$-\nu + e^{\epsilon_i} + H_i^{-1}(\xi_i - \mu_i) = 0 \quad \text{and} \quad -e^{\epsilon_i} + H_i^{-1} = 0. \tag{10}$$

Solutions to the equations allow us to proceed with the MCML procedure in which, at a given iteration, μ_i and H_i are evaluated with given $\hat{\boldsymbol{\psi}}$.

3 EMPIRICAL RESULTS

3.1 The Data

We now apply the SCD model with "leverage effect" as proposed in Equation (2) to the transaction data of IBM and other stocks. The IBM transaction data is downloaded from Professor Robert Engle's website. The data contain various trade records, such as transaction time, price, and volume. All trades occurred from November 1, 1990, to January 31, 1991. Instead of using the whole sample, we only use the data from November 1, 1990, to December 21, 1990, to avoid any holiday effects. There are a total of 35 trading days in these two months. As in Engle and Russell (1998), we delete the trades that occurred before 9:50 A.M. and after 4:00 P.M. to eliminate the irregularities during the open and close period of the trading day. On the other hand, we initialize the duration process for each trading day following the procedure in Engle and Russell (1998). That is, the first duration for each day is calculated as the average duration from 9:50 A.M. to 10:00 A.M. After all deletions, the total number of transactions is 24,765. The trade time is recorded in seconds and the trade duration is defined as the time difference between two consecutive trades. Of all the durations, the largest is 502 seconds and the smallest is 1 second (trade time unit). Most of the durations are less than 100 seconds (more than 94%), and the mean and median durations are 30 and 17 seconds, respectively. As a robustness check of our empirical results, transaction data of other stocks are also used in our application. We focus on those stocks that have been used in existing empirical studies such as Bauwens and Veredas (2004) and Bauwens et al. (2000). For brevity, we only report the results for Boeing and Coca Cola. The transaction data for both stocks are extracted from the TAQ database over the period of February to March 2002. Similar to the IBM data, the trades that occurred before 9:50 A.M. or after 4:00 P.M. were deleted and the first duration for each day was calculated as the average duration from 9:50 A.M. to 10:00 A.M. This results in 41,482 and 44,042 total transactions for Boeing and Coca Cola, respectively. The mean, median, minimum, and maximum of trade durations for these two stocks are 20.86 and 19.67, 26.78 and 24.48, 1 and 1, and 494 and 703, respectively.

Table 2 reports the first 15 autocorrelations and partial autocorrelations of the IBM duration series. From Table 2, we can see that the AC coefficients decay very slowly, while the PAC coefficients at lag 1 is clearly larger than the other coefficients. The first-order autocorrelation is 0.127, and the ratio of the consecutive autocorrelations is about 0.9. In other words, the data present a very strong autoregressive (AR) and moving average (MA) or ARMA structure. Similar dynamic properties are found for the trade durations of Boeing and Coca Cola.

3.2 Seasonal Adjustment

To remove seasonality from the data, the technique with piecewise cubic spline (available in S-Plus software with the function smooth.spline(·)) is employed.

Table 2 Dynamic Properties of the IBM Trading Durations.

	Raw data		Seasonally adjusted data	
Lag	Autocorrelation	Partial autocorrelation	Autocorrelation	Partial autocorrelation
1	0.12690	0.12690	0.12574	0.12574
2	0.10693	0.09232	0.10738	0.09304
3	0.09187	0.06954	0.09035	0.06806
4	0.09092	0.06499	0.09072	0.06515
5	0.08221	0.05231	0.08106	0.05144
6	0.08340	0.05174	0.08133	0.05003
7	0.08716	0.05333	0.08526	0.05211
8	0.09850	0.06210	0.09552	0.05988
9	0.08481	0.04317	0.08416	0.04371
10	0.07259	0.02929	0.06930	0.02682
11	0.08289	0.04106	0.07969	0.03906
12	0.07682	0.03266	0.07490	0.03229
13	0.06791	0.02280	0.06595	0.02227
14	0.05986	0.01517	0.05775	0.01461
15	0.06180	0.01845	0.05922	0.01742

The table reports the first 15 autocorrelations and partial autocorrelations of the IBM trading duration for both the raw and seasonally adjusted data.

In recent articles by Engle and Russell (1998) and Veredas, Rodriguez-Poo, and Espasa (2001), the spline or nonparametric functions capturing diurnal variations are estimated simultaneously along with the duration process. In particular, Veredas, Rodriguez-Poo, and Espasa (2001) propose an integrated method to estimate the deterministic seasonality jointly with the stochastic duration process. Their model is semiparametric: nonparametric for the seasonality and parametric (of the log-ACD type) for the duration process. As shown in Veredas, Rodriguez-Poo, and Espasa (2001), however, preadjusting the data has no important consequences for the estimation of the autoregressive parameters since the seasonal component does not carry a lot of information about intertemporal dynamics. Since the estimation method employed in this article for the duration process involves a great deal of simulation, we rely on the simple cubic spline technique to preadjust the seasonality of the data. As in Engle and Russell (1998) and Bauwens and Giot (2000), two different effects are considered. One is the day-of-week effect, the other is the time-of-day effect. Typically the duration remains constantly high between Monday and Wednesday, then decreases continuously afterward, and finally becomes the shortest on Friday. This reflects the fact that trades appear relatively inactive during the early part of the week and become a lot more active at the end of the week. To eliminate this day-of-week effect, the average sample duration is calculated for a weekday, denoted by F_w, $w = 1, 2, 3, 4, 5$, see Figure 1a for the IBM data. The duration after removing the day-of-week effect is given as

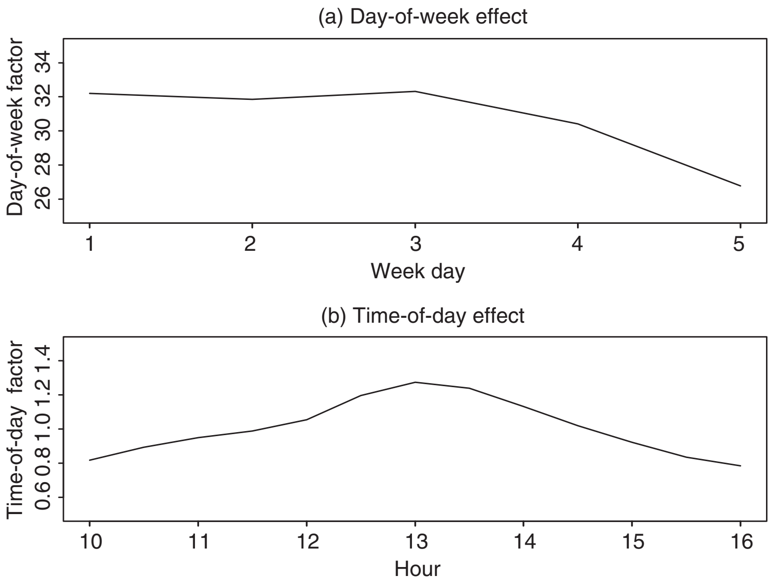

Figure 1 Seasonal effects of the IBM trading durations.

$\frac{d_i}{F_w}$, denoted by $\tilde{d}_i$. Extra seasonality presented in $\tilde{d}_i$ would be attributed to the well-known time-of-day effect. The duration first appears short in the morning, rises up dramatically around noon, and drops toward the close of the market. Again, we use the spline method to remove the time-of-day effect. First, 13 knots are chosen over each trading day, with the first one being at 10:00 A.M., the last one at 4:00 P.M., and the remaining knots 30 minutes apart. Second, the value (duration) at each knot is calculated by averaging the durations around the knot. We use the 30-minute window (15 minutes for both the left side and right side of the knot). The average duration in the interval for 35 days is regarded as the duration at the knot. Finally, the daily seasonal factor is calculated, denoted by F_t (t is the time in seconds from 10:00 A.M. to 4:00 P.M.). Then the adjusted duration data are calculated as $\frac{\tilde{d}_i}{F_t}$. The time-of-day pattern, as in Figure 1b for the IBM data, clearly shows that the duration increases in the morning and reaches a maximum at around 1 P.M., then decreases toward the close of the market in an average trading day. As shown in Table 2 for the AC and PAC coefficients, the seasonally adjusted duration process remains highly persistent, which, again, provides evidence of the ARMA-type structure in the data-generating process. The seasonally adjusted duration data of the stocks considered in our study are used in the model estimation.

3.3 Estimation Results

The SCD models specified in Equation (2) are fitted to the seasonally adjusted duration series with three error distributions—LW(ν, 1), LG(ν, 1) and LE—for the disturbance of the observation equation. The parameter vector to be estimated is

Table 3 Estimation results of SCD models without "leverage effect."

		β	σ	μ	ν
Panel A: IBM					
LW(ν, 1) ($\gamma = 0$)	Parameter estimate	0.9707	0.1149	−0.7391 (−8.60)	0.9462
	p-value $H_0 : \boldsymbol{\theta}_j = 0$	0.000	0.000	0.000	—
	$H_0 : \boldsymbol{\theta}_j = 1$	0.000	—	—	0.000
LG(ν, 1) ($\gamma = 0$)	Parameter estimate	0.9646	0.1303	−0.7166 (−5.80)	0.9569
	p-value $H_0 : \boldsymbol{\theta}_j = 0$	0.000	0.000	0.000	—
	$H_0 : \boldsymbol{\theta}_j = 1$	0.000	—	—	0.000
LE ($\gamma = 0$)	Parameter estimate	0.9584	0.1448	−0.6940 (−5.10)	1.0000
	p-value $H_0 : \boldsymbol{\theta}_j = 0$	0.000	0.000	0.000	—
	$H_0 : \boldsymbol{\theta}_j = 1$	0.000	—	—	—
Panel B: Boeing					
LW(ν, 1) ($\gamma = 0$)	Parameter estimate	0.9420	0.1390	−0.6252 (−7.39)	0.9272
	p-value $H_0 : \boldsymbol{\theta}_j = 0$	0.000	0.000	0.000	—
	$H_0 : \boldsymbol{\theta}_j = 1$	0.000	—	—	0.000
LG(ν, 1) ($\gamma = 0$)	Parameter estimate	0.9451	0.1501	−0.6202 (−7.61)	0.9305
	p-value $H_0 : \boldsymbol{\theta}_j = 0$	0.000	0.000	0.000	—
	$H_0 : \boldsymbol{\theta}_j = 1$	0.000	—	—	0.000
LE ($\gamma = 0$)	Parameter estimate	0.9280	0.1622	−0.6098 (−6.94)	1.0000
	p-value $H_0 : \boldsymbol{\theta}_j = 0$	0.000	0.000	0.000	—
	$H_0 : \boldsymbol{\theta}_j = 1$	0.000	—	—	—
Panel C: Coca Cola					
LW(ν, 1) ($\gamma = 0$)	Parameter estimate	0.9564	0.0823	−0.5530 (−4.23)	0.9389
	p-value $H_0 : \boldsymbol{\theta}_j = 0$	0.000	0.000	0.000	—
	$H_0 : \boldsymbol{\theta}_j = 1$	0.000	—	—	0.000
LG(ν, 1) ($\gamma = 0$)	Parameter estimate	0.9500	0.0911	−0.5312 (−4.71)	0.9411
	p-value $H_0 : \boldsymbol{\theta}_j = 0$	0.000	0.000	0.000	—
	$H_0 : \boldsymbol{\theta}_j = 1$	0.000	—	—	0.000
LE ($\gamma = 0$)	Parameter estimate	0.9523	0.0867	−0.5399 (−4.05)	1.0000
	p-value $H_0 : \boldsymbol{\theta}_j = 0$	0.000	0.000	0.000	—
	$H_0 : \boldsymbol{\theta}_j = 1$	0.002	—	—	—

The table reports estimation results of the SCD models without "leverage effect," that is, $\gamma = 0$, as specified in Equation (2). The value in the brackets beside the estimate of μ is the z-value or the t-statistic defined as the ratio of parameter estimate and standard deviation. No z-values are reported for other parameter estimates (β, σ, ν) because they are estimated indirectly via certain transformations to ensure positivity. For example, σ is estimated via the transformation $\sigma = \exp(c)$, where c is estimated. Instead, p-values of relevant hypotheses are reported in the table where $\boldsymbol{\theta}_j = 0$ or 1 means the jth component of $\boldsymbol{\theta}$ is 0 or 1, j = 1, 2, 3, 4 with $\boldsymbol{\theta} = (\beta, \sigma, \mu, \nu)$.

$\theta = (\beta, \sigma, \mu, \nu, \gamma)'$. First we consider estimation of the models without the "leverage effect" or in the absence of the intertemporal term, that is, $\gamma = 0$. The results are reported in Table 3. It is noted that for all three stocks with different model specifications, the persistence parameter β is close to but significantly smaller than one, suggesting high persistence and stationarity of the duration process. All three models have similar estimates for parameters β, σ, and μ. However, for both log-Weibull and log-gamma models, the scale parameter ν is significantly different from one, suggesting that the null hypothesis $H_0 : \nu = 1$ is rejected. In other words, the log-exponential model is misspecified. The estimated variance of the error term, η_i, in the transition process is significantly different from zero for all models. Therefore the log-ACD model is strongly rejected, and it is necessary to model the conditional expectation of the duration as a latent process. Overall our results are similar to those in Bauwens and Veredas (2004) based on trade durations of other stocks. Their results also indicate the misspecification of the log-exponential model. In particular, they note that among different durations (trade, price, and volume), the trade durations tend to be more persistent. The correlations between the estimated latent variables and observed durations are smaller for the trade durations than for the other kinds of durations, indicating a poorer "fit" of the model to trade duration process. A comprehensive empirical comparison between log-ACD and SCD models is also performed in Bauwens and Veredas (2004). They find that in terms of unconditional densities, the log-ACD model cannot account for the hump in the density of the trade durations and the SCD model clearly outperforms the log-ACD model.

In this article, our focus is whether the further extension of "leverage effect" can improve modeling of the duration process. The estimation results for the SCD models with "leverage effect" as specified in Equation (2) are reported in Table 4. Again, for all three stocks with different model specifications, the persistence parameter β is very close to but significantly smaller than one, suggesting high persistence and stationarity of the duration process. The estimated variance of the error term, η_i, in the transition process remains to be significantly different from zero for all models. It further confirms the necessity of modeling the conditional expectation of the duration as a latent process. While the estimates of parameter μ are numerically similar to those in the SCD models without "leverage effect" as reported in Table 3, there is a clear increase of z-values or a decrease of standard deviations for μ. From our subsequent diagnostic analysis based on the filtered series $\hat{\epsilon}_i$ and $\hat{\eta}_i$, we note that the presence of the γ term helps to remove some large spurious noise in the observation process, which results in better estimation of the constant term μ with smaller standard deviations.

This suggests that with the addition of the intertemporal term, the model clearly provides a better structure for the disturbance term of the duration process as specified in Equation (2). It provides evidence for the necessity of further extending the standard SCD model specification. The scale parameter ν is close to but significantly different from one for both the SCD LW(ν, 1) model and the SCD LG(ν, 1) model. This suggests again that the log-exponential model is misspecified. Most importantly, the intertemporal term γ has an overall positive sign for all stocks with different model specifications and is highly significant for the SCD LW(ν, 1) model. The SCD LE model has the least significance for the intertemporal term. As we have mentioned, however, the model is clearly misspecified.

The significant positive sign suggests that there is a positive intertemporal correlation between trade duration and the conditional expected duration. That is,

Table 4 Estimation results of SCD models with "leverage effect."

		β	σ	μ	ν	γ
Panel A: IBM						
LW(ν, 1)	Parameter estimate	0.9716	0.1100	−0.7488 (−10.3)	0.9404	0.0125 (2.70)
	p-value $H_0: \boldsymbol{\theta}_j = 0$	0.000	0.000	0.000	—	0.005
	$H_0: \boldsymbol{\theta}_j = 1$	0.000	—	—	0.000	—
LG(ν, 1)	Parameter estimate	0.9649	0.1293	−0.7166 (−28.8)	0.9551	0.0024 (0.50)
	p-value $H_0: \boldsymbol{\theta}_j = 0$	0.000	0.000	0.000	—	0.600
	$H_0: \boldsymbol{\theta}_j = 1$	0.000	—	—	0.000	—
LE	Parameter estimate	0.9581	0.1463	−0.7014 (−30.4)	1.0000	−0.0081 (−1.70)
	p-value $H_0: \boldsymbol{\theta}_j = 0$	0.000	0.000	0.000	—	0.090
	$H_0: \boldsymbol{\theta}_j = 1$	0.000	—	—	—	—
Panel B: Boeing						
LW(ν, 1)	Parameter estimate	0.9231	0.1302	−0.6361 (−19.5)	0.9220	0.0817 (4.30)
	p-value $H_0: \boldsymbol{\theta}_j = 0$	0.000	0.000	0.000	—	0.000
	$H_0: \boldsymbol{\theta}_j = 1$	0.000	—	—	0.000	—
LG(ν, 1)	Parameter estimate	0.9282	0.1430	−0.6271 (−15.7)	0.9245	0.0639 (2.99)
	p-value $H_0: \boldsymbol{\theta}_j = 0$	0.000	0.000	0.000	—	0.002
	$H_0: \boldsymbol{\theta}_j = 1$	0.000	—	—	0.000	—
LE	Parameter estimate	0.9167	0.1505	−0.6203 (−21.2)	1.0000	0.0552 (2.13)
	p-value $H_0: \boldsymbol{\theta}_j = 0$	0.000	0.000	0.000	—	0.033
	$H_0: \boldsymbol{\theta}_j = 1$	0.000	—	—	—	—
Panel C: Coca Cola						
LW(ν, 1)	Parameter estimate	0.9483	0.0791	−0.5643 (−9.21)	0.9371	0.0451 (5.32)
	p-value $H_0: \boldsymbol{\theta}_j = 0$	0.000	0.000	0.000	—	0.000
	$H_0: \boldsymbol{\theta}_j = 1$	0.000	—	—	0.000	—
LG(ν, 1)	Parameter estimate	0.9437	0.0855	−0.5520 (−10.2)	0.9416	0.0622 (7.13)
	p-value $H_0: \boldsymbol{\theta}_j = 0$	0.000	0.000	0.000	—	0.000
	$H_0: \boldsymbol{\theta}_j = 1$	0.000	—	—	0.000	—
LE	Parameter estimate	0.9467	0.0901	−0.5407 (−8.32)	1.0000	0.0101 (1.40)
	p-value $H_0: \boldsymbol{\theta}_j = 0$	0.000	0.000	0.000	—	0.161
	$H_0: \boldsymbol{\theta}_j = 1$	0.000	—	—	—	—

The table reports estimation results of the SCD models with "leverage effect" as specified in Equation (2). The values in the brackets beside the estimates of μ and γ are the z-value or the t-statistic defined as the ratio of parameter estimate and standard deviation. No z-values are reported for other parameter estimates (β, σ, ν) because they are estimated indirectly via certain transformations to ensure positivity. For example, σ is estimated via the transformation $\sigma = \exp(c)$, where c is estimated. Instead, p-values of relevant hypotheses are reported in the table where $\boldsymbol{\theta}_j = 0$ or 1 means the jth component of $\boldsymbol{\theta}$ is 0 or 1, $j = 1, 2, 3, 4$ with $\boldsymbol{\theta} = (\beta, \sigma, \mu, \nu, \gamma)$.

the conditional expected duration is not only highly persistent, but also responds to the informational shock in the duration process. More specifically, as a negative shock occurs to the trade duration, there tends to be a decrease in the conditional expected duration and equivalently an increase in trade intensity. In other words, the trade intensity reacts in an asymmetric manner to information shock in the duration process. This reflects a similar asymmetric behavior in the conditional volatility of asset returns, where the conditional volatility is not only highly persistent, but also reacts to information shock in the asset returns. In particular, the conditional volatility typically rises as a result of large negative returns.

4 DIAGNOSTIC ANALYSIS

In this section, diagnostic analysis is performed for the fitted models. In all subsequent analysis, we focus on IBM stock, as it has been the subject of many other empirical studies. For IBM stock, since the SCD LE model is misspecified and the SCD LG(ν,1) model turned out to have an insignificant intertemporal effect, both models are excluded in our following analysis. We focus on the SCD LW(ν,1) model, with and without "leverage effect," and the SCD LG(ν,1) model without "leverage effect." In terms of improving the goodness-of-fit, the difference between the SCD LW(ν,1) model and the SCD LG(ν,1) model, both without "leverage effect," is useful to analyze the potential impact of distributional assumption on the disturbance of the observation process, while the difference between the SCD LW(ν,1) models, with and without "leverage effect," can reflect the potential impact of including a term associated with the duration process in the latent process.

4.1 Basic Diagnostics

In all three models, both error terms ϵ_i and η_i in the observation equation and the transition equation are assumed to be i.i.d. If the models are correctly specified, the estimates of the error terms should confirm, to some extent, the independence assumption. We obtain the estimates of η_i in the transition equation, denoted by $\hat{\eta}_i$, by the Kalman smoothing filter, then the estimates of ϵ_i in the observation equation, denoted by $\hat{\epsilon}_i$, by substituting $\hat{\psi}_i$ into the observation equation.

It is well known that the estimated error terms or residuals are not uncorrelated, even in a simple regression model setting. Because of the presence of two random resources (ϵ_i and η_i) in the SCD models, the autocorrelation structure of residuals ($\hat{\epsilon}_i$ and $\hat{\eta}_i$) are too complicated to obtain analytically. Consequently this will stop us from using some classical tools such as the ACF plots (in which the asymptotic confidence limits are unknown) to draw sensible conclusions. Instead here we concentrate on the lag-1 autocorrelation, hoping to confirm whether the model has addressed part of the dependent structure of the duration process. Two different tools are applied for this purpose.

One is the linear regression technique: regressing $\hat{\epsilon}_i$ on $\hat{\epsilon}_{i-1}$ and $\hat{\eta}_i$ on $\hat{\eta}_{i-1}$, respectively. Namely $\hat{\epsilon}_i = a + b\hat{\epsilon}_{i-1} + e_i$ and $\hat{\eta}_i = a + b\hat{\eta}_{i-1} + e_i$, where slope b reflects the strength of the lag-1 autoregressive relation. So if the first-order autocorrelation of $\hat{\epsilon}_i$ or $\hat{\eta}_i$ is small, then the coefficient of $\hat{\epsilon}_{i-1}$ or $\hat{\eta}_{i-1}$ should not be significantly different from zero. Table 5 reports the results for both linear regressions. The p-values indicate that the SCD LW model with "leverage effect" has the strongest evidence (p-value = .8667) that the first autocorrelation for ϵ_i is zero.

Table 5 Results of the AR(1) $\hat{\epsilon}_i$ and $\hat{\eta}_i$ regressions.

	$\hat{\epsilon}_i$			$\hat{\eta}_i$		
	p-value for a	p-value for b	ACF(1)	p-value for a	p-value for b	ACF(1)
SCD-"LE" LW	0.0002	0.8667	0.0011	0.5364	0.0004	0.0223
SCD LW	0.0000	0.2567	0.0072	0.0069	0.0000	0.8112
SCD LG	0.0718	0.6292	0.0031	0.4345	0.0000	0.0361

The table reports the results of the AR(1) regressions of $\hat{\epsilon}_i$ and $\hat{\eta}_i$ for the IBM stock, where SCD-"LE" LW denotes the SCD LW model with "leverage effect."

The second tool is the scatter plot of $\hat{\epsilon}_i$ versus $\hat{\epsilon}_{i-1}$ and $\hat{\eta}_i$ versus $\hat{\eta}_{i-1}$. Because of the considerably large number of observations, the plots would be less indicative if the entire series of $\hat{\epsilon}_i$ or $\hat{\eta}_i$ are used. Instead, we only plot a random sample of 100 observations from the residuals $\hat{\epsilon}_i$ and $\hat{\eta}_i$. The plot of sampled $\hat{\epsilon}_i$ is shown in Figure 2, while the plot of sampled $\hat{\eta}_i$ is shown in Figure 3.

It is noted that the η_i's are highly autocorrelated for all three models. Possible explanations are as follows. First, the AR(1) structure assumed in the latent process may not be sufficient to fully capture the dynamics of the duration process. In other words, higher-order terms in the latent process may be needed to improve the fitting of the model. While the ACF(1)'s of all three models are difficult to distinguish, it is interesting to note that the SCD LW model with "leverage effect" has the lowest first-order autocorrelation. This suggests that the inclusion of an intertemporal term in the latent process can help to address the dependence structure of the duration process. Second, since all trade durations are recorded in seconds, the systematic upward measurement error for a duration of less than one second may introduce a deterministic component in the observations of duration. As the latent process has less variation than the observation process, evidenced by $\widehat{\mathrm{var}[\eta_i]} = \hat{\sigma}^2 < 0.2$ and $\widehat{\mathrm{var}[\epsilon_i]} > 1$, the systematic effect is more pronounced for the residuals of latent variable process. Finally, the distribution of error term ϵ_i may not be optimal to address the duration process, and a more flexible distribution of ϵ_i, for example, a generalized gamma distribution [Lunde (1999)], may be employed. We attempt to address these issues in our future research.

4.2 Assessment of Density Functions

As we also assume normality for the distribution of η_i, the QQ plots are also reported in Figure 4 for all three models. Visually the $\hat{\eta}_i$'s for all three models appear to have different tail shapes than the normal distribution, but the QQ plot of $\hat{\eta}_i$ for the SCD LW model with "leverage effect" is the closest to the normal distribution.

To assess the distributional assumption of duration, we compare the marginal density of duration derived from the models to the empirical marginal density

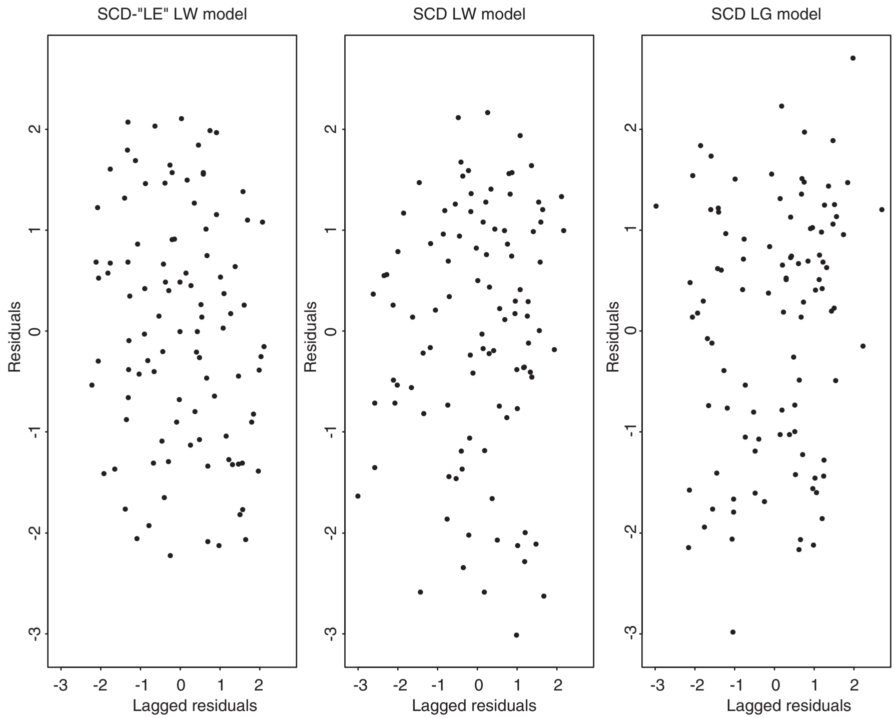

Figure 2 Plot of $\hat{\epsilon}_i$ versus $\hat{\epsilon}_{i-1}$. The panels plot the filtered $\hat{\epsilon}_i$ against $\hat{\epsilon}_{i-1}$ under different model specifications for the trading duration process of the IBM stock, where SCD-"LE" LW denotes the SCD LW model with "leverage effect."

directly obtained from the observed durations. Since the marginal distribution of logarithmic duration is a mixture of two different distributions in the three models, the marginal distribution of duration is not a closed-form expression. To overcome this problem, we simulate a large sample of durations from each fitted model, then obtain the density function. Since the models are of geometric ergodicity, we employ a Markov chain simulation method to generate a large sample (with sample size 30,000), and the first 10,000 simulated values are discarded to eliminate initial value effect. Because of the discreteness of raw data, we also round simulated durations to the unit of a second.

The histograms of the simulated durations and the observed durations with frequencies in seconds are plotted in Figures 5 for all three models. It is noted that all three models fit the right-hand tail of the distributions very well, but not so for the low durations. Visually it is difficult to distinguish among three models. Numerically we calculate the sum of squared errors between the two marginal density functions for each model. The sums of squared errors are, respectively, 8.707638×10^{-4} for the SCD LW model with "leverage effect," 1.264881×10^{-3} for the SCD LW model without "leverage effect," and 8.734705×10^{-4} for the SCD LG

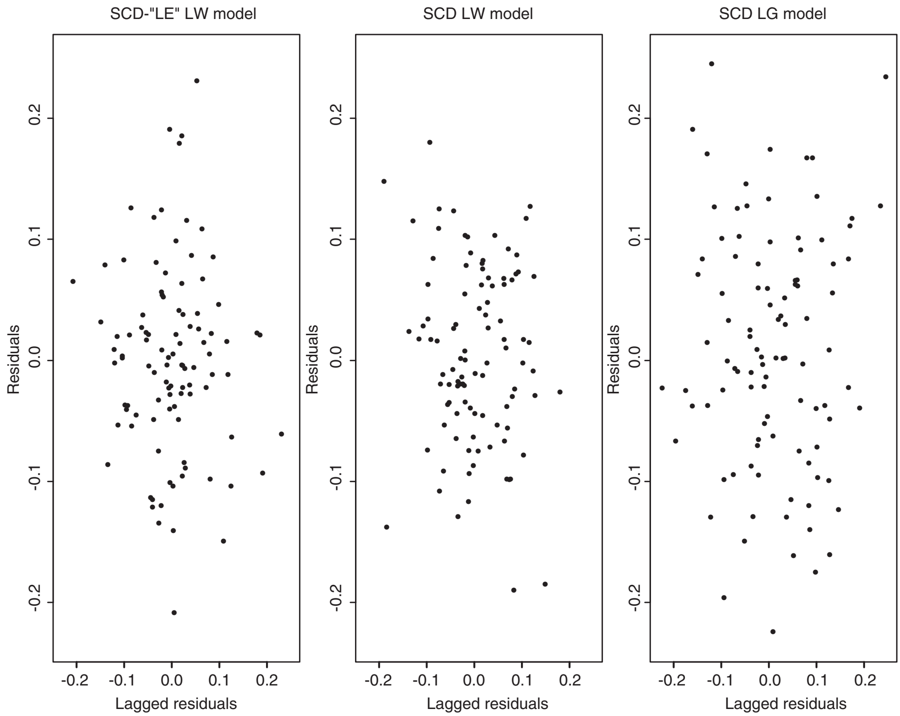

Figure 3 Plot of $\hat{\eta}_i$ versus $\hat{\eta}_{i-1}$. The panels plot the filtered $\hat{\eta}_i$ against $\hat{\eta}_{i-1}$ under different model specifications for the trading duration process of the IBM stock, where SCD-"LE" LW denotes the SCD LW model with "leverage effect."

model without "leverage effect." The overall difference among models is marginal, indicating the choice for each of them gives a similar conclusion. However, the SCD LW model with "leverage effect" works slightly better.

4.3 In-Sample Forecasting Performance

As the last diagnostic check, we investigate the goodness-of-fit of all three models based on the in-sample forecasting performance. In-sample forecasts are the fitted values for the response variable in the sample space and can be used to measure the dynamic properties of the model. The in-sample forecasting performances are investigated for the last day of the IBM dataset, December 21, 1990, with a sample size of 692. First, the Kalman smoother was used to obtain the estimates of conditional mean of the latent variable ψ_i at trade i given all observed log durations, then the logarithm of conditional mean of seasonally adjusted duration is estimated by $\log(\hat{d}_i) = \hat{\mu} + \hat{\psi}_i$ for each model. Finally, multiplying the above values by both the day-of-week and time-of-day seasonal factors leads to the in-sample forecasts of durations.

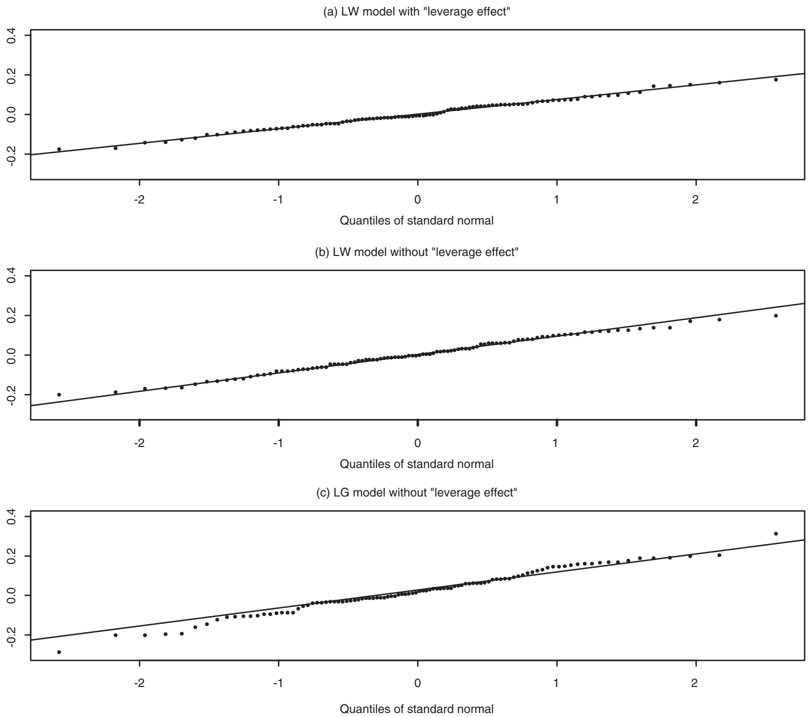

Figure 4 QQ plots of $\hat{\eta}_i$. The panels are the QQ plots of filtered $\hat{\eta}_i$ under different model specifications for the trading duration process of the IBM stock.

To quantitatively evaluate the forecasting performance of the three models, we run the following regression:

$$d_j = a + b\hat{d}_j + u_j, \quad j = 1, \ldots, 692, \tag{11}$$

where d_j and $\hat{d}_j$ are the observed durations and in-sample forecasts, respectively, a and b are regression coefficients, and u_j is white noise with variance σ_u^2. The least-squares estimates with their estimated standard errors in the parentheses are reported in Table 6. The differences of R^2 among models are marginal. However, the SCD LW model with "leverage effect" appears to be slightly better than the other two.

The out-of-sample forecasting performance of the SCD models is also investigated in an earlier version of the article. We note that the out-of-sample forecast based on the SCD LW(ν, 1) model with "leverage effect" has more fluctuations and its movement is much closer to the observed duration. Compared to those of the SCD LW(ν, 1) and SCD LG(ν, 1) models without "leverage effect," the out-of-sample forecast of the SCD LW(ν, 1) model with "leverage effect" reflects better the local dynamic behavior of the duration process. However, according to Bauwens et al. (2000), model comparison based on density forecasts suggests that the latent factor models (such as SCD and SVD) are not really superior to

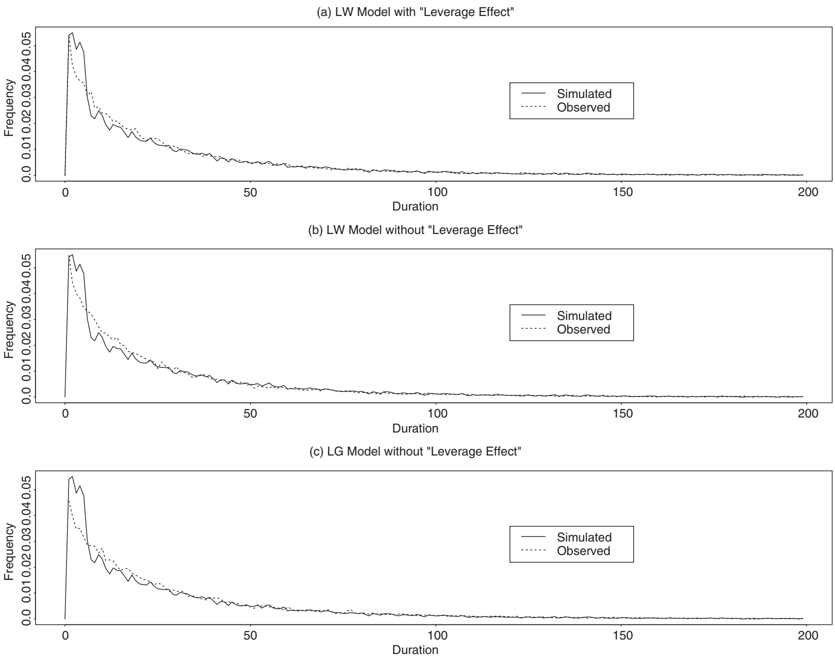

Figure 5 Histograms of model-simulated durations and observed durations. The panels plot the histogram of the simulated durations under different model specifications together with that of the observed trading durations of the IBM stock.

Table 6 In-sample forecasting performance of the models.

Model	a	b	R^2
SCD LW(ν, 1) model with "leverage effect"	−12.5678 (4.1)	2.8650 (0.25)	0.1547
SCD LW(ν, 1) model without "leverage effect"	−7.6796 (3.8)	2.5024 (0.22)	0.1491
SCD LG(ν, 1) model without "leverage effect"	−8.5385 (4.1)	2.6145 (0.25)	0.1369

The table reports the in-sample forecasting performance of three different model specifications for IBM stock.

the standard ACD models with given innovation distribution and specification of the expected conditional duration process.[1] As shown in Bauwens et al. (2000), and illustrated in Figure 5, the drawback associated with single-factor latent model is that it predicts poorly the left tail of the unconditional duration distribution, which happens to be the area with a heavy mass of probability. To achieve better out-of-sample forecasts of very small trade durations, Bauwens et al. (2000), identify some models, such as the threshold ACD model by Zhang, Russell, and Tsay (2001) and a simple log-ACD model in the Bauwens and Giot (2000) framework with the generalized gamma innovation distribution as proposed in Lunde (1999).

4.4 Transaction Intensity and Intraday Volatility

As we have mentioned, the main objective of including an intertemporal term in the latent process is to capture the asymmetric behavior or "leverage effect" of the conditional expected duration. Our results suggest that the conditional expected duration is not only highly persistent with an autoregressive structure, but also reacts to the information shock in the duration process. In particular, with a negative shock to the duration, the conditional expected duration will subsequently decrease and equivalently the trade activity will subsequently intensify. This reflects a similar behavior in the conditional volatility of asset returns, where the conditional volatility is not only highly persistent, but also reacts to information shock in the asset returns. In particular, the conditional volatility typically rises as a result of negative shock to the returns. It is generally believed that trading activity and asset return volatility are both highly correlated with the intensity of market information flow. For instance, trading typically becomes more active as information flow intensifies. As a result, trade durations tend to be shorter. On the other hand, the market adjusts the valuation of an asset according to the arrival of new information. As a result, asset return volatility tends to rise. Therefore it would be interesting to investigate whether these two variables share common information content. More interestingly, whether a better modeling of duration process can enhance the forecasting performance of intraday volatility. In the theoretical market microstructure literature, however, conflicting results have been derived on the relationship between transaction intensity and price volatility. Namely, the Easley and O'Hara (1992) model predicts that the number of transactions would influence the price process through information-based clustering of transactions, while the Admati and Pfleiderer (1988) model predicts that the number of transactions would have no impact on the price intensity. As pointed out by Engle and Russell (1998), with a continuous record of market trading activities, these theoretical hypotheses can be empirically tested.

Engle and Russell (1998) derive the relationship between the price intensity and the instantaneous volatility of asset returns. In particular, the expected conditional volatility over an infinitesimal time interval can be expressed as a function of the price intensity. Let the instantaneous volatility at time t be defined as $\sigma^2(t) = \lim_{\Delta t \to 0} E\left\{\frac{1}{\Delta t}\left[\frac{P(t+\Delta t)-P(t)}{P(t)}\right]^2\right\}$ and suppose the stock price follows a binomial process. The probability that stock price changes by c over a time interval Δt is $\lambda(t|t_{N(t)}, \ldots, t_1)\Delta t + o(\Delta t)$, and otherwise there is no change. Then, by taking the limit, the conditional volatility in the instant after t can be written as $\sigma^2(t|t_{N(t)}, \ldots, t_1) = \left(\frac{c}{P(t)}\right)^2 \lambda(t|t_{N(t)}, \ldots, t_1)$, where $\lambda(t|t_{N(t)}, \ldots, t_1)$ is the price intensity. Thus, with an estimate of the price intensity $\hat{\lambda}(t|t_{N(t)}, \ldots, t_1)$ based on their estimated model, a forecast of the instantaneous volatility can be obtained. Using the price intensity as a volatility forecast, they find that the instantaneous volatility has a significantly negative relationship with the transaction intensity. The measure of transaction intensity is constructed using the number of transactions over each price duration, as in general the price duration is longer than the trade duration. The negative relation suggests that following periods of high trade intensity, the expected price durations, and equivalently the instantaneous volatility, is higher. The results are consistent with the Easley and O'Hara model.

The results in Engle and Russell (1998) also suggest that trade intensity, or equivalently trade duration, has certain forecasting ability of expected asset return volatility. In this article, we investigate whether allowing for "leverage effect" in the duration process can further enhance the forecasting performance of intraday volatility. Similar to the in-sample forecasting performance, we perform the analysis on the last day of the sample, December 21, 1990. At given trade i, $i = 1, 2, \ldots,$ 692, we construct the one-step-ahead forecast of trade duration based on the estimated models. Then a measure of subsequent realized volatility is regressed against the forecast of trade duration. The purpose is to investigate how much variation in realized volatility can be explained by the forecast of trade duration based on specific duration models. The one-step-ahead forecast of the trade duration is constructed as follows. Let the smoothing value of ψ_i at trade i be $\hat{\psi}_i$ and the estimate of the error term η_i be $\hat{\eta}_i$, then the one-step-ahead forecast of the expected trade duration is $\hat{\psi}_{i+1} = \hat{\beta}\hat{\psi}_i + \hat{\gamma}\hat{\eta}_i$, where $\hat{\beta}$ and $\hat{\gamma}$ are the estimates in the respective models. Taking the logarithm of the seasonally adjusted duration and then multiplying to both the day-of-week and time-of-day seasonal factors yields the out-of-sample duration forecasts. The realized volatility is calculated over the fixed time interval following each trade i. We use 30 seconds as the time interval to measure the intraday volatility (60- and 120-second intervals are also used and the results are not significantly different). Following each trade i, we have all transaction prices and quotes (bid-ask average) of the stock over the interval $[t_i, t_i + 30)$, where t_i is the trade time in seconds. We calculate the sum of squared changes in log stock prices, and the realized volatility following trade i is given by its square root.

A linear regression of the realized volatility against the duration forecast is estimated and the estimation results with standard errors are reported in Table 7. Not surprisingly, the R^2's are all very small, as the volatility measure constructed here is a very noisy realized volatility estimator. In addition, as we are dealing with the continuous record of market trading activities, various other factors are contributing to the volatility of the market, especially the market microstructure-related noise. Similar to Engle and Russell (1998), we find a significantly negative

Table 7 The relationship between duration forecasts and intraday volatility.

Model	a	b	R^2
SCD LW(ν, 1) model with "leverage effect"	1.243 (0.133)	−0.039 (0.010)	0.024
SCD LW(ν, 1) model without "leverage effect"	1.241 (0.158)	−0.042 (0.012)	0.016
SCD LG(ν, 1) model without "leverage effect"	1.235 (0.157)	−0.041 (0.012)	0.016

The table reports the regression results of intraday volatility against the duration forecasts under three different model specifications of the IBM trading duration process.

relation between trade duration and stock price volatility, or equivalently a positive relation between trade intensity and price volatility. In other words, following periods of high trade intensity, the instantaneous asset return volatility tends to be higher. The results provide further support for the Easley and O'Hara model. It should be noted that different than Engle and Russell (1998), where the volatility forecasts are derived from the estimates of price intensity, here the realized volatility is directly measured using realized stock price changes. The realized volatility is a model-free measure of the ex post asset return volatility. More interestingly, the R^2 for the SCD LW model with "leverage effect," while small in magnitude, is about 50% higher than those of other SCD models. This suggests that the conditional expected trade duration and the instantaneous volatility not only exhibit similar asymmetric behavior, but also these asymmetric movements are to a certain extent correlated with each other. The same analysis is also performed based on the estimation results of Boeing and Coca Cola, and we find even stronger evidence of enhanced intraday volatility forecast. In other words, a better modeling of duration process with "leverage effect" can capture certain common dynamic features in the financial market and contribute to better forecasting of intra day asset price volatility.

5 CONCLUSION

This article proposes SCD models with "leverage effect" under the linear non-gaussian state-space model framework. The models are extensions of the ACD models by Engle and Russell (1998) and SCD models by Bauwens and Veredas (2004). We study the statistical properties of the models and derive certain moments that are used in the development of model estimation. The MCML method is employed in this article for consistent and efficient parameter estimation. Empirical applications to the transaction data of IBM and other stocks are also performed. Our results suggest that allowing for the intertemporal correlation between the duration process and the latent process can better reflect the local dynamic behavior of the duration process. The expected trade durations are not only highly persistent over time with an autoregressive structure, but also reacts to the information shock in the observed duration process. This reflects a similar asymmetric behavior in the conditional volatility of asset returns, where the conditional volatility is not only highly persistent, but also reacts to information shock in the asset returns. Our further analysis suggests that the conditional expected trade duration and stochastic volatility not only exhibit similar asymmetric behavior, but also share common information content. In particular, the asymmetric movements in the conditional expected trade duration and stochastic volatility are to a certain extent correlated with each other. Consequently a better modeling of duration process with "leverage effect" can capture certain common dynamic features in the financial market and contribute to better forecasting of intraday asset price volatility. Our diagnostic analysis also suggests that the AR(1) structure assumed in the latent process may not be sufficient to fully capture the dynamics of the duration process. Furthermore, the distribution of error term ϵ_i

may not be optimal to address the trade duration dynamics, and a more flexible distribution of ϵ_i, for example, a generalized gamma distribution [Lunde (1999)], may be employed. We will attempt to address these issues in our future research.

APPENDIX

Proof of Proposition 1. The stationarity of the process can be easily checked. Because $y_i = \log d_i - \mu$ and y_i is the sum of two AR(∞) processes (see the proof of Proposition 2), therefore y_i is stationary, so is $\log d_i$.

Let $X_i' = (y_i, \epsilon_i)_{1\times 2}$, $V_i' = (\epsilon_i, \eta_i)_{1\times 2}$,

$$\Psi = \begin{pmatrix} \beta & -\beta+\gamma \\ 0 & 0 \end{pmatrix}_{2\times 2} \quad \text{and} \quad \Omega = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}_{2\times 2}$$

Equation (2) can be rewritten as

$$X_i = \Psi X_{i-1} + \Omega V_i.$$

To show that X_i is geometrically ergodic, first we show that the Markov chain X_i is irreducible and aperiodic. Note that the generalized controllability matrix,

$$C_{x_0}^1 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix},$$

is a full-rank matrix; the chain X_i is forward accessible based on Proposition 7.1.4 [Meyn and Tweedie (1993)]. Moreover, it is obvious that $X^* = (0,0)$ is a global attracting state, so the chain is irreducible and aperiodic according to Theorem 7.2.6 [Meyn and Tweedie (1993)].

Now we show that the chain X_i is geometrically ergodic. Let the test function be $U(x) = \|\Sigma x\| + 1$, where Σ is a specially chosen matrix for some $\epsilon > 0$, and the test set $C = \{x \in R^2\colon U(x) \le c \text{ for some } c < \infty\}$, where $\|\cdot\|$ denotes the Euclidean norm for a vector or the spectral norm for a matrix.

Then we have

$$E[U(x_i)|x_{i-1} = x] \le (1-\epsilon)U(x) + \delta I_{c(x)}$$

for some $\delta < \infty$ and for all x, where $I_{c(x)}$ is an indicator function defined as usual. From Theorem 15.0.1 [21], we have that X_i is geometrically ergodic, and so are y_i and $\log(d_i) = y_i + \mu$.

Because the one-to-one correspondence between d_i and y_i, the $\{d_i\}$ is stationary and geometrically ergodic. ■

Proof of Proposition 2. Based on Equation (2), $\psi_i = \beta\psi_{i-1} + \gamma\epsilon_{i-1} + \eta_i$, when $\beta < 1$ holds we have that

$$\begin{aligned}\psi_i &= \sum_{j=0}^{\infty} \beta^j B^j (\gamma\epsilon_{i-1} + \eta_i) \\ &= \gamma \sum_{j=0}^{\infty} \beta^j \epsilon_{i-j-1} + \sum_{j=0}^{\infty} \beta^j \eta_{i-j},\end{aligned}$$

where B is the backward operator, that is, $B^j\epsilon_i = \epsilon_{i-j}$, $j = 0, 1, 2, \ldots$, and

$$\begin{aligned} y_i &= \epsilon_i + \gamma \sum_{j=0}^{\infty} \beta^j \epsilon_{i-j-1} + \sum_{j=0}^{\infty} \beta^j \eta_{i-j} \\ &= W_i + Z_i, \end{aligned}$$

where $W_i = \sum_{j=0}^{\infty} a_j \epsilon_{i-j}$, $Z_i = \sum_{j=0}^{\infty} \beta^j \eta_{i-j}$, and

$$a_j = \begin{cases} 1, & j = 0 \\ \gamma\beta^{j-1}, & j = 1, 2, 3, \ldots. \end{cases}$$

Given that ϵ_i and η_i are i.i.d. and mutually independent, we have that

$$\begin{aligned} \mathrm{var}(y_i) &= E(y_i^2) \\ &= E(W_i^2) + E(Z_i^2) \\ &= \left(\sum_{j=0}^{\infty} a_j^2\right) m_2^\epsilon + \left(\sum_{j=0}^{\infty} \beta^{2j}\right) m_2^\eta \\ &= \left(1 + \frac{\gamma^2}{1-\beta^2}\right) m_2^\epsilon + \frac{m_2^\eta}{1-\beta^2} \\ &= \frac{(1+\gamma^2-\beta^2)m_2^\epsilon + m_2^\eta}{1-\beta^2}, \end{aligned}$$

$$\begin{aligned} E(y_i)^3 &= E(W_i + Z_i)^3 \\ &= EW_i^3 + 3EW_i^2 Z_i + 3EW_i Z_i^2 + EZ_i^3 \\ &= \left(\sum_{j=0}^{\infty} a_j^3\right) m_3^\epsilon + \left(\sum_{j=0}^{\infty} \beta^{3j}\right) m_3^\eta \\ &= \left(1 + \frac{\gamma^3}{1-\beta^3}\right) m_3^\epsilon + \frac{m_3^\eta}{1-\beta^3} \\ &= \frac{(1+\gamma^3-\beta^3)m_\epsilon^3 + m_3^\eta}{1-\beta^3}, \end{aligned}$$

$$\begin{aligned} E(y_i)^4 &= E(W_i + Z_i)^4 \\ &= EW_i^4 + 4EW_i^3 Z_i + 6EW_i^2 Z_i^2 + 4EW_i Z_i^3 + EZ_i^4 \\ &= E\left(\sum_{j=0}^{\infty} a_j \epsilon_{i-j}\right)^4 + 6EW_i^2 EZ_i^2 + E\left(\sum_{j=0}^{\infty} \beta^j \eta_{i-j}\right)^4 \\ &= \left(\sum_{j=0}^{\infty} a_j^4\right) m_4^\epsilon + 12 \sum_{j=0}^{\infty} \sum_{k=j+1}^{\infty} a_j^2 a_k^2 (m_2^\epsilon)^2 + 6\left(1 + \frac{\gamma^2}{1-\beta^2}\right) m_2^\epsilon \frac{1}{1-\beta^2} m_2^\eta \\ &\quad + \left(\sum_{j=0}^{\infty} \beta^{4j}\right) m_4^\eta + 12 \sum_{j=0}^{\infty} \sum_{k=j+1}^{\infty} \beta^{2j}\beta^{2k} (m_2^\eta)^2 \\ &= \left(1 + \frac{\gamma^4}{1-\beta^4}\right) m_4^\epsilon + 12 \frac{\gamma^2}{1-\beta^2}\left(1 + \gamma^2 \frac{\beta^2}{1-\beta^4}\right)(m_2^\epsilon)^2 \\ &\quad + 6\left(1 + \frac{\gamma^2}{1-\beta^2}\right) m_2^\epsilon \frac{1}{1-\beta^2} m_2^\eta + \frac{m_4^\eta}{1-\beta^4} + 12 \frac{\beta^2}{1-\beta^2} \frac{1}{1-\beta^4} (m_2^\eta)^2, \end{aligned}$$

$$\begin{aligned}
\operatorname{cov}(y_i, y_{i-s}) &= E(y_i y_{i-s}) \\
&= E(W_i W_{i-s}) + E(Z_i Z_{i-s}) \\
&= \sum_{j=0}^{\infty} a_j a_{j+s} m_2^{\epsilon} + \sum_{j=0}^{\infty} \beta^j \beta^{j+s} m_2^{\eta} \\
&= \left(\gamma\beta^{s-1} + \frac{\gamma^2 \beta^s}{1-\beta^2} \right) m_2^{\epsilon} + \frac{\beta^s}{1-\beta^2} m_2^{\eta} \quad s \geq 1,
\end{aligned}$$

where $m_j^{\epsilon} = E\epsilon_i^j$ and $m_j^{\eta} = E\eta_i^j$, $j = 2, 3, 4$. ■

Proof of Proposition 3. From Proposition 2, we have that

$$\begin{aligned}
y_i &= \epsilon_i + \gamma \sum_{j=0}^{\infty} \beta^j \epsilon_{i-j-1} + \sum_{j=0}^{\infty} \beta^j \eta_{i-j} \\
&= W_i + Z_i,
\end{aligned}$$

where $W_i = \sum_{j=0}^{\infty} a_j \epsilon_{i-j}$, $Z_i = \sum_{j=0}^{\infty} \beta^j \eta_{i-j}$, and

$$a_j = \begin{cases} 1, & j = 0 \\ \gamma\beta^{j-1}, & j = 1, 2, 3, \ldots, \end{cases}$$

so

$$\begin{aligned}
d_i &= \exp(\mu + W_i + Z_i) \\
&= \exp(\mu) \prod_{j=0}^{\infty} \exp(a_j \epsilon_{i-j}) \prod_{j=0}^{\infty} \exp(\beta^j \eta_{i-j}).
\end{aligned}$$

The rth moment of d_i is

$$\begin{aligned}
\mathrm{E} d_i^r &= \exp(r\mu) \prod_{j=0}^{\infty} \mathrm{E}\ \exp(r\alpha_j \epsilon_{i-j-1}) \prod_{j=0}^{\infty} \mathrm{E}\ \exp(r\beta^j \eta_{i-j}) \\
&= \exp(r\mu) \prod_{j=0}^{\infty} m(r\alpha_j) \prod_{j=0}^{\infty} \exp\left(\frac{1}{2} r^2 \beta^{2j} \sigma^2 \right) \\
&= \exp(r\mu) \prod_{j=0}^{\infty} m(r\alpha_j) \exp\left(\frac{r^2 \sigma^2}{2(1-\beta^2)} \right).
\end{aligned}$$

When $r = 1$, we have that the mean of d_i is

$$\mathrm{E} d_i = \exp(\mu) \prod_{j=0}^{\infty} m(\alpha_j) \exp\left(\frac{\sigma^2}{2(1-\beta^2)} \right).$$

When $r = 2$, we have that the second moment of d_i is

$$\mathrm{E} d_i^2 = \exp(2\mu) \prod_{j=0}^{\infty} m(2\alpha_j) \exp\left(\frac{2\sigma^2}{1-\beta^2} \right),$$

where

$$m(\alpha) = \mathrm{E}\exp(\alpha\epsilon_j) = \begin{cases} \Gamma\left(\dfrac{\alpha}{\nu}+1\right), & \text{when } \epsilon_j \text{ is } LW(\nu,1) \\ \dfrac{\Gamma(\nu+\alpha-1)}{\Gamma(\nu)}, & \text{when } \epsilon_j \text{ is } LG(\nu,1). \end{cases}$$

Proof of Proposition 4 Equations (1) and (2) are direct results from Proposition 2. For the proof of Equation (3), first we prove the following lemma.

Lemma *Let X and Y be independent with mean zero and both have a kurtosis no less than three. Then kurtosis of X + Y is no less than three as well. Moreover, the kurtosis of X+Y is equal to three if and only if both the kurtosis of X and the kurtosis of Y are equal to three (i.e., X, Y are normal distributed).*

Proof Since

$$\begin{aligned}
&EX^4 \geq 3(EX^2)^2 \\
&EY^4 \geq 3(EY^2)^2 \\
&E(X+Y)^4 = E(X^4+Y^4+6X^2Y^2) \\
&V(X+Y) = V(X)+V(Y) = E(X^2)+E(Y^2) \\
&(V(X+Y))^2 = (E(X^2)+E(Y^2))^2 \\
&\qquad\qquad\quad = (EX^2)^2+(EY^2)^2+2(EX^2)(EY^2),
\end{aligned}$$

we have,

$$\begin{aligned}
\frac{E(X+Y)^4}{(V(X+Y))^2} &= \frac{E(X^4)+E(Y^4)+6E(X^2Y^2)}{(EX^2)^2+(E(Y^2)^2+2EX^2 \cdot EY^2)} \\
&\geq \frac{3((EX^2)^2+(E(Y^2))^2+6EX^2 \cdot EY^2}{(EX^2)^2+(EY^2)^2+2EX^2 \cdot EY^2} \\
&= 3
\end{aligned}$$

and $\frac{E(X+Y)^4}{(V(X+Y))^2} = 3$ if and only if $EX^4 = 3(EX^2)^2$ and $EY^4 = 3(EY^2)^2$. ■

Proof of Equation (4) Because $y_i = W_i + Z_i$, and the kurtosis of W_i and Z_i are greater than three, the kurtosis of y_i is greater than 3.

Received February 21, 2002; revised September 2, 2003; accepted April 21, 2004

REFERENCES

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

Bauwens, L., and P. Giot. (2000). "The Logarithmic ACD Model: An Application to the Bid-Ask Quote Process of Three NYSE Stocks." *Annales d'Economie et de Statistique (Special issue "Financial Market Microstructure") 60*, 117–149.

Bauwens, L., and P. Giot. (2003). "Asymmetric ACD Models: Introducing Price Information in ACD Models with a Two State Transition Model." *Empirical Economics 28(4)*, 1–23.

Bauwens, L., P. Giot, J. Grammig, and D. Veredas. (2000). "A Comparison of Financial Duration Models Through Density Forecasts." Forthcoming in *International Journal of Forecasting*.

Bauwens, L., and D. Veredas. (2004). "The Stochastic Conditional Duration Model: A Latent Factor Model for the Analysis of Financial Duration." *Journal of Econometrics 119*(2), 381–412.

Black, F. (1976). "Studies of Stock Price Volatility Changes." *Proceedings of the Business and Economic Statistics Section*, 177–181.

Bollerslev, T. (1986). "Generalized Autoregressive Conditional Heteroscedasticity." *Journal of Econometrics 31*, 307–327.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay. (1997). *The Econometrics of Financial Markets*. Princeton, NJ: Princeton University Press.

Danielsson, J. (1994). "Stochastic Volatility in Asset Prices Estimation with Simulated Maximum Likelihood." *Journal of Econometrics 64*, 375–400.

Danielsson, J., and J. F. Richard. (1993). "Accelerated Gaussian Importance Sampler with Application to Dynamic Latent Variable Models." *Journal of Applied Econometrics 8*, S153–S173.

Duffie, D., and K. J. Singleton. (1993). "Simulated Moments Estimation of Markov Models of Asset Prices." *Econometrica 61*, 929–952.

Durbin, J., and S. J. Koopman. (1997). "Monte Carlo Maximum Likelihood Estimation for Non-Gaussian State Space Models." *Biometrika 84*, 669–684.

Easley, D., and M. O'Hara. (1992). "Time and the Process of Security Price Adjustment." *Journal of Finance 19*, 69–90.

Engle, R. F. (1982). "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of the United Kingdom Inflation." *Econometrica 50*, 987–1007.

Engle, R. F. (2000). " The Econometrics of Ultra High Frequency Data." *Econometrica 68*, 1–22.

Engle, R. F., and J. R. Russell. (1998). "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data." *Econometrica 66*, 1127–1162.

Gallent, R. A. and G. Tauchen. (1996). "Which Moments to Match?" *Econometric Theory 12*, 657–681.

Ghysels, E., C. Gouriéroux, and J. Jasiak. (2004). "Stochastic Volatility Duration Models," *Journal of Econometrics 119*, 413–433.

Ghysels, E., and J. Jasiak. (1998). "GARCH for Irregularly Spaced Financial Data: The ACD-GARCH Model." *Studies in Nonlinear Economics and Econometrics 2*, 133–149.

Ghysels, E., A. C. Harvey, and E. Renault. (1996). "Stochastic Volatility." *Handbook of Statistics 14*, 119–191.

Gouriéroux, C. (1997). "*ARCH Models and Financial Application*." New York: Springer-Verlag.

Gouriéroux, C., A. Monfort, and E. Renault. (1993). "Indirect Inference," *Journal of Applied Econometrics 8*, S85–S199.

Grammig, J., and K. Maurer. (2000). "Non-Monotonic Hazard Functions and the Autoregressive Conditional Duration Model." *Econometrics Journal 3*, 16–38.

Grammig, J., and M. Wellner. (2002). "Modeling the Interdependence of Volatility and Inter-Transaction Duration Processes." *Journal of Econometrics 106*, 369–400.

Hansen, L. P. (1982). "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica 50*, 1029–1054.

Harvey, A. C. (1989). *Forecasting, Structural Models and the Kalman Filter*. Cambridge: Cambridge University Press.

Harvey, A., E. Ruiz, and N. Shephard. (1994). "Multivariate Stochastic Variance Models." *Review of Economic Studies 61*, 247–264.

Jacquier, E., N. G. Polson, and P. E. Rossi. (1994). "Bayesian Analysis of Stochastic Volatility Models." *Journal of Business & Economic Statistics*, *12(4)*, 371–389.

Jørgensen, B, S. L. Christensen, P. Song, and L. Sun. (1999). "A State Space Model for Multivariate Longitudinal Count Data." *Biometrika*, *86*, 169–181.

Kalman, R. E., and R. S. Bucy. (1961). "New Results in Linear Filtering and Prediction Theory." *Journal of Basic Engineering, Transactions, ASME, Series D 83*, 95–108.

Liesenfeld, R., and R. C. Jung. (2000). "Stochastic Volatility Models: Conditional Normality Versus Heavy–Tailed Distribution." *Journal of Applied Econometrics 15*, 137–160.

Lunde, A. (1999). "A Generalized Gamma Autoregressive Conditional Duration Model." Aalborg University.

Meddahi, N., E. Renault and B. Werker. (1998). "Continuous Time Models for High Frequency Data." *Proceedings of the Second HIGH-Frequency Data in Finance Conference*. Zurich: Olsen & Associates.

Meyn, S. P., and R. L. Tweedie. (1993). "Markov Chains and Stochastic Stability." New York: Springer-Verlag.

Nelson, D.B. (1991). "Conditional Heteroskedasticity in Asset Returns: A New Approach." *Econometrica 59*, 347–370.

Russell, J. R. (1999). "Econometric Modeling of Irregularly-Spaced Multivariate Transaction Data." Working paper. University of Chicago, Graduate School of Business.

Sandmann, G., and S. J. Koopman. (1998). "Estimation of Stochastic Volatility Models via Monte Carlo Maximum Likelihood." *Journal of Econometrics 87*, 271–301.

Shephard, N., and M. K. Pitt. (1997). "Likelihood Analysis of non-Gaussian Measurement Time Series." *Biometrika 84*, 653–667.

Smith, A. A. (1993). "Estimating Nonlinear Time Series Models Using Vector Autoregressions: Two Approaches." *Journal of Applied Econometrics 8*, S63–S84.

Taylor, S. (1986). *Modelling Financial Time Series*. Chichester, UK: Wiley.

Venables, W. N. and B. D. Ripley. (1998). *Modern Applied Statistics with S-PLUS*, 2nd ed. New York: Springer-Verlag.

Veredas, D., J. Rodriguez-Poo, and A. Espasa. (2001). "On the (Intradaily) Seasonality and Dynamics of a Financial Point Process: A Semiparametric Approach." *CORE Discussion Paper 200100. Université Catholique de Louvain*.

West, M., and J. Harrison. (1989). "Bayesian Forecasting and Dynamic Models." New York: Springer-Verlag.

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.