Time-Deformation Modeling Of Stock Returns Directed By Duration Processes

by

Dingan Feng[1], Peter X.-K. Song[2], and Tony S. Wirjanto[3]

This Version: December 2008

Abstract

This paper presents a new class of time-deformation (or stochastic volatility) models for stock returns sampled in transaction time and directed by a generalized duration process. Stochastic volatility in this model is driven by an observed duration process and a latent autoregressive process. Parameter estimation in the model is carried out by using the method of simulated moments (MSM) due to its analytical feasibility and numerical stability for the proposed model. Simulations are conducted to validate the choices of the moments used in the formulation of the MSM. Both the simulation and empirical results obtained in this paper indicate that this approach works well for the proposed model. The main empirical findings for the IBM transaction return data can be summarized as follows: (i) the return distribution conditional on the duration process is not Gaussian, even though the duration process itself can marginally function as a directing process; (ii) the return process is highly leveraged; (iii) a longer trade duration tends to be associated with a higher return volatility; and (iv) the proposed model is capable of reproducing return whose marginal density function is close to that of the empirical return.

Keywords: Duration process; Ergodicity; Method of simulated moments; Return process; Stationarity.

JEL classification: G10, C51, C32.

[1]CIBC, 161 Bay Street, Toronto, Ontario, CANADA M5J 2S8, E-mail Address: Dingan.Feng@cibc.ca

[2]Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, Michigan, USA 48109-2029, E-mail Address: pxsong@umich.edu

[3]Department of Economics, University of Waterloo, Waterloo, Ontario, CANADA N2L 3G1, E-mail Address: twirjant@uwaterloo.ca

1 Introduction

Since the seminal paper by Clark (1973), a great deal of research has been devoted to studying the relationship between the volatility of returns and the measures of market activity such as volume and the number of trades. The relationship between the stock-price movement and the stock-trade volume is further investigated by Gallant, Rossi and Tauchen (1992), who showed that the conditional volatility of returns and the volume are positively correlated. More recently, Ane and Geman (2000) revisited Clark's method of dealing with the non-normality of observed returns by considering a general time change process. They conclude that in order to recover the normality of returns, the transactions clock is better represented by the number of trades than the trading volume.

All the above studies appear to have been motivated by the concept that a stock market is primarily driven by the information flow, which is either private or public, and with traders, either informed or uninformed (see, for example, Easley and O'Hara, 1992). If the information flow could be completely retrieved from the market, studying the mechanism of price process would be relatively straightforward. Therefore, if price process is indexed by information flow in the stock market, instead of calendar time, the resulting model would be much simpler to analyze. This motivates the introduction of a so-called time-deformation model.

Recently, with the availability of high-frequency time series, researchers have come to recognize that a trade duration (or simply a duration) process, which is defined as the time interval between two consecutive trades, conveys useful information. This results in the emergence of new statistical models designed specifically to elicit such information. Most notably, Engle and Russell (1998) introduce an Autoregressive Conditional Duration (ACD) model, which can be viewed as an ARMA process with non-Gaussian innovations characterized by deterministic GARCH process. This ACD model is recently extended by Bauwens and Veredas (2004) by allowing the evolution of the conditional duration process to be driven by a latent variable, giving rise to a whole class of Stochastic Conditional Duration (SCD) models. This SCD model, in turn, is further studied using alternative estimation methods by Feng, Jiang and Song (2004), and more recently by Knight and Ning (2008) and Xu, Knight and Wirjanto (2008). Another extension of the ACD model focuses on the interdependence between the duration process and the conditional return volatility characterized by a GARCH process. See, for instance, Ghysels and Jasiak (1998) and Grammig and Wellner (2002).

In this paper, we follow Engle (2000) and Renault and Weker (2004) in considering the use of a duration process to capture the information flow in the stock market and formulate a time-deformation model via a duration process. Due to the complexity of the stock market, a single stochastic process is unlikely to be able to embrace all the information flow in the market; as a result, a latent process is introduced in this paper to carry on the remaining information flow. In other words, we use the duration process and the latent process jointly to describe the information flow in the stock market. An advantage of this approach is that we can statistically test whether the duration process itself (or any other process) is able to share a significant amount of the information flow in the stock market.

It is important to stress that we are not the first to go down this route. Our proposed model is similar to the model studied by Chernov, Gallant, Ghysels, and Tauchen (2003) and Huang and Tauchen (2005). Both in our discrete-time model and in their continuous-time specifications, volatility is driven by two components with one component being highly persistent, and the other being not. In the above papers both volatility components are latent, whereas in this paper the persistent component is observed and captured by a duration process.

The remaining part of this paper is structured as follows. Section 2 proposes a new time-deformation model and discusses some of its statistical properties. A Monte-Carlo study on moment selections is reported in Section 3. Section 4 presents both preliminary and model-based analyses of the IBM stock return data. Finally we conclude the paper in Section 5. Proofs of theorems and propositions in this paper are collected in the Appendix.

2 Model Formulation

2.1 Model Specification

A time-deformation model generally consists of two processes: a parent process X_t, which is usually assumed to be Gaussian $N(\mu s, \sigma_0^2 s)$, and a directing process $s = g(t)$, which maps the calendar time to the operational time, so that the observed return process $Y(t)$ can be expressed as $Y(t) = X(g(t))$.

In both Clark's (1973) and Ane and Geman's (2000) models, parameters μ, and σ_0 are assumed to be constant, and $g(t)$ is specified to be the trading volume and the number of trades respectively. While in both Stock's (1988) and Ghysels, Gouriéroux and Jasiak's (1998) models, $g(t)$ is specified as a logistic function of variables with lags. One potential drawback of Stock's (1988) operational-time scale function, $g(t)$, is that it is only a deterministic function of a small numbers of variables; as such it may not be ble to fully capture the information flow in the market and the trading volume in Clark's (1973) model or the number of trades in Ane and Geman's (2000) model. Another possible limitation of the above models is that the observed sequence of prices is assumed to be equally spaced. Consequently, an irregular spaced time series will be forced to aggregate in order to be equally spaced (see Ane and Geman, 2000). This aggregation typically results in information loss, and more seriously, in a possible change in the underlying stochastic structure of the original data. To overcome some of these limitations, we present a new time-deformation model below via a duration process for irregularly spaced, high frequency time series of stock return.

The cumulative duration process may embrace a good deal of information flow in the stock market. As for high-frequency financial data, the observed series is typically irregularly spaced; so it is necessary to index the series sequentially by trades. By convention, trades that occur at the same time are treated as a single trade. Let N be the total number of trades occurred within time interval $[0, T]$. In order to establish a time-deformation framework, we first define a correspondence between the return at trade k and the duration at trade k, $k = 1, 2, \cdots, N$, following Ane and Geman (2000). By definition, the duration d_k at trade k is $d_k = t_k - t_{k-1}$. Then, the cumulative duration up to the time of trade k is $D_k = \sum_{j \leq k} d_k$. Let p_k be the asset price at trade k. The return process R_k at trade k, is defined as the difference between the logarithmic prices between two adjacent trades, k and $k-1$,

$$R_k = \ln(p_k) - \ln(p_{k-1}),\ \ k = 1, \ldots, N.$$

The primary objective of this paper is to study the time-deformation model of the return process with D_k being the directing process. That is, conditional on the cumulative duration D_k, the logarithm of asset-price process is assumed to be Gaussian and distributed as $N(\mu D_k, \sigma_0^2 D_k)$. It follows that

$$\ln(p_k) = \mu D_k + z_k \sqrt{\sigma_0^2 D_k},$$

with parameters μ, and σ_0, and z_k is the standard Gaussian, $N(0, 1)$, random variable. So, the return can be expressed as

$$R_k = \mu(D_k - D_{k-1}) + z_k \sqrt{\sigma_0^2 D_k} - z_{k-1} \sqrt{\sigma_0^2 D_{k-1}}.$$

Next we postulate another standard Gaussian random variable ϵ_k, such that

$$\epsilon_k \sqrt{D_k - D_{k-1}} = z_k \sqrt{D_k} - z_{k-1} \sqrt{D_{k-1}}$$

and $\{\epsilon_k\}$ are mutually independent, (actually z_k can be constructed recursively from the sequence $\{\epsilon_k\}$ and an initial state z_0), we obtain

$$R_k = \mu d_k + \epsilon_k \sqrt{\sigma_0^2 d_k}. \tag{1}$$

In the real market, generally D_k or d_k is not expected to be able to embrace the entire information flow. Thus, we let an additional term $\exp(V_k)$ to carry on the remaining information flow in the market. Note that at least, in the short term, it is more important to incorporate the remaining information flow in the volatility process rather than the drift process. Therefore, we allow this remaining information flow to enter the volatility component of the model. As a result, the model in (1) is extended to be of the following form:

$$R_k = \mu d_k + \epsilon_k \sqrt{\sigma_0^2 \exp(V_k) d_k}. \tag{2}$$

To capture the dynamics of the information flow, the V_k is assumed to follow a first-order Markov process,

$$V_k = \beta_0 + \beta_1 V_{k-1} + \eta_k, \tag{3}$$

where the innovation η_k is assumed to be Gaussian $N(0, \sigma^2)$, and β_0, β_1 and σ are the unknown parameters of the model with a stationarity restriction, $|\beta_1| < 1$.

Lastly, extending the model in (2) to allow for a power transformation on d_k, together with (3), we obtain a time-deformation return model directed by a duration process:

$$\begin{aligned} R_k &= \alpha_0 + \alpha_1 d_k + \exp[\alpha_2 + \alpha_3 \ln(d_k) + V_k]\epsilon_k, \\ V_k &= \beta_1 V_{k-1} + \eta_k, \end{aligned} \tag{4}$$

where $\alpha_0, \alpha_1, \alpha_2, \alpha_3$, and β_1 are the unknown parameters with a restriction on the volatility process, $|\beta_1| < 1$. The innovations (ϵ_k, η_k) are assumed to follow a bivariate Gaussian, $BVN(0, 0; 1, \sigma^2; \rho)$; that is, $corr(\epsilon_k, \eta_k) = \rho$. In addition, we also assume that the process $\{d_k\}$ is independent of the innovations, ϵ_k and η_k.

The salient feature of our model in (4) is the presence of correlation (ρ) between the return innovation process (ϵ_k) and the volatility innovation process (η_k). It is designed to capture the so-called leverage effect, that has come to characterize many of the empirical return processes. Moreover, when $\beta_1 = 0$ and $\sigma = 0$, V_k will degenerate to a constant zero. This means that the remaining information flow is not present in addition to what has been captured by the duration process. Consequently, R_k is distributed as Gaussian conditional on d_k, which effectively reduces our model to Clark's (1973) and Ane and Geman's (2000) models. When $\alpha_0 = \alpha_1 = \alpha_3 = 0$, the model in (4) becomes a simple stochastic-volatility model with correlated errors,

$$\begin{aligned} R_k &= \exp(\alpha_2 + V_k)\epsilon_k, \\ V_k &= \beta_1 V_{k-1} + \eta_k. \end{aligned} \tag{5}$$

2.2 Statistical Properties

In this section, we present some statistical properties of the time-deformation model of stock return in (4). Denote the complex unit $i = \sqrt{-1}$.

Theorem 1. *If $|\beta_1| < 1$, and V_0 has the distribution $N(0, \frac{\sigma^2}{1-\beta_1^2})$, the process R_k satisfying the model in (4) is stationary and geometrically ergodic, provided that the duration process d_k is stationary and geometrically ergodic.*

Proof: See Appendix.

Initially,we investigate the statistical properties of the model in (4) by setting $\alpha_0 = \alpha_1 = 0$. Denoting $X_k = \ln|R_k|$, and taking the logarithmic transformation on both sides of the return equation in (4) yields,

$$\begin{aligned} X_k &= \alpha_2 + V_k + \alpha_3 \ln(d_k) + \zeta_k, \\ V_k &= \beta_1 V_{k-1} + \eta_k, \end{aligned}$$

where $\zeta_k = \ln \mid \epsilon_k \mid$.

It is well-known that the characteristic function and the distribution function are a one-to-one correspondence. Specifically the characteristic function of X_k is given by

Proposition 1.

$$\begin{aligned}\Psi_k(u) &\equiv E\{\exp(iuX_k)\}\\ &= \exp\left\{iu[\alpha_2+\frac{1}{2}\ln(2-2\rho^2)]\right\}\Psi_{ld}(\alpha_3 u)\exp\left[-\frac{\sigma^2}{2(1-\beta_1^2)}\beta_1^2u^2\right]\\ &\quad\times\sum_{h=0}^{\infty}\left[\frac{1}{h!}\frac{\Gamma(\frac{1}{2}+i\frac{u}{2}+h)}{\Gamma(\frac{1}{2}+h)}I_h(u)\right],\end{aligned}$$

where $I_h(u) = \sqrt{\frac{2}{\pi}}\frac{1}{\sigma}\int\limits_0^\infty[\exp(-a\xi^2)(c\xi)^{2h}\cos(u\xi)]d\xi$ *with* $c = \frac{|\rho|}{\sigma\sqrt{2(1-\rho^2)}}$, *and* $a = \frac{1}{2\sigma^2(1-\rho^2)}$, $\Psi_{ld}(\cdot)$ *is the characteristic function of* $\ln(d_k)$, *and* $\Gamma(\cdot)$ *is the generalized gamma function.*

Proof: See Appendix.

Therefore, the r-th original moments of X_k can be obtained as

$$E(X_k^r) = i^{-r}\left.\frac{d^r[\Psi_k(u)]}{du^r}\right|_{u=0}.$$

Now we turn to the model in (4) and derive the unconditional moments of the return process R_k.

Theorem 2. *The r-th unconditional moment of R_k is given by:*

$$\begin{aligned}m_{R_k}(r) &\equiv ER_k^r\\ &= \sum_{\substack{j+h+l=r,\\ 0\le j,h,l\le r}}\tfrac{r!}{j!h!l!}(\alpha_0)^j(\alpha_1)^h\exp(l\alpha_2)m_d(h+l\alpha_3)\exp\left(\tfrac{1}{2}\tfrac{\sigma^2}{1-\beta_1^2}l^2\right)\\ &\quad\times\left[\sum_{j_1=0}^{l}\tfrac{l!}{j_1!(l-j_1)!}(l\rho\sigma)^{l-j_1}m_N(j_1)\right].\end{aligned}$$

Especially, the unconditional mean of R_k is given by:

$$m_{R_k}(1) = \alpha_0+\alpha_1 m_d(1)+\exp(\alpha_2)m_d(\alpha_3)\rho\sigma\exp\left[\frac{\sigma^2}{2(1-\beta_1^2)}\right],$$

where $m_d(j)$ and $m_N(j)$ are j-th moments of the duration process d_k and of the standard normal variable, respectively.

Proof: See Appendix.

Proposition 2. *For the model in (5), when $\alpha_2 = 0$, the first four moments are given by:*

$$\begin{array}{rcl} m_{R_k}(1) & = & \rho\sigma A, \\ m_{R_k}(2) & = & (1+4\rho^2\sigma^2)A^4, \\ m_{R_k}(3) & = & 9\rho\sigma(1+3\rho^2\sigma^2)A^9, \\ m_{R_k}(4) & = & [3+6(4\rho\sigma)^2+(4\rho\sigma)^4]A^{16}, \end{array}$$

respectively, where $A = \exp[\frac{\sigma^2}{2(1-\beta_1^2)}]$. Moreover, the third and fourth unconditional central moments of R_k, denoted by $M_{R_k}(3)$ and $M_R(4)$, are given by:

$$M_{R_k}(3) = \rho\sigma A^3[3A^2(3A^4-1)+3(\rho\sigma A)^2(9A^4-4)+2\rho^2\sigma^2],$$

$$\begin{array}{rcl} M_{R_k}(4) & = & 3A^{16}+(\rho\sigma)^2A^8(96A^8-36A^2)+(\rho\sigma)^4A^8(4^4A^8-108A^2) \\ & & +6(\rho\sigma)^2A^6+(\rho\sigma)^4A^4(24A^2-3), \end{array}$$

respectively, where $A = \exp[\frac{\sigma^2}{2(1-\beta_1^2)}]$.

Proof: See Appendix.

For the simple volatility model in (5), we can draw on Proposition 2, to make the following remarks: (1) both the mean and the third moment of R_k are zero if and only if the correlation ρ is zero. Therefore it is not possible to fit the simple volatility model with correlated errors for the zero-mean series. Moreover, this model implies that the signs of the mean and the third moment of R_k are determined by the sign of the correlation ρ; (2) the sign of the skewness coefficient is the same as the sign of the third central moment. Because $A > 1$, R_k has negative skewness (being skewed to the left) if the correlation is negative. In particular, the R_k has zero skewness (being symmetric) if the correlation is zero; and (3) the kurtosis of R_k is larger than three regardless of whether the correlation is positive, negative or zero.[4]

3 Monte-Carlo Based Moment Selection

As mentioned in the introduction, we propose the MSM as our preferred method for the parameter estimation of the model in (4) because its log-likelihood function proves intractable.[5]

[4]This simply re-affirms the fact that the volatility model is useful to describe the leptokurtic (or heavy tailed) feature of the return data.

[5]The MSM is proposed by McFadden (1989), and has its genesis in generalized method of moments introduced by Hansen (1982). See also Gouriéroux et al. (1993) and Gouriéroux and Monfort (1996).

A critical element in implementing the MSM is to decide which and how many moment conditions are used in the estimation. The key to the selection of the moments in effect boils down to a sensible trade-off between estimation efficiency and computational feasibility.

In this section we propose a simulation-based technique for a parsimonious selection of the moments. In many applications, selecting moment conditions is done via Akaike information criterion (AIC). However, as mentioned above, the log-likelihood of our model in (4) is hard to compute; so we will base our decision on the performance of point estimation and standard error estimation among a few candidate settings of moment conditions.

To reiterate, our model in (4) has several distinctive features: (1) the two innovation terms from the return process and the volatility process, ϵ_k and η_k, are allowed to be correlated; (2) the volatility is jointly modeled by two stochastic processes, which are the observed duration process $\{d_k\}$, and the latent process $\{V_k\}$; (3) a drift term, $\alpha_1 d_k$, is included in the model; and (4) the drift and volatility terms are also allowed to be correlated as both are dependent on the duration process. These features of our model provide a great deal of flexibility in modeling the return process, but unfortunately it comes at a cost of making statistical inference somewhat less straightforward.[6] When the drift term is absent and the volatility term is fully latent ($\alpha_1 = \alpha_3 = 0$), which is a setting postulated in many prior research, there have been several estimation methods developed to estimate the resulting model with and without the presence of the correlation between ϵ_k and η_k. When there is no correlation between the two innovation processes, Anderson and Sørensen (1996) use a generalized method of moments (GMM) to estimate the model. When there is correlation in the model, Harvey and Shephard (1996) suggest a transformation on the model that leads to uncorrelated innovations, and then use a quasi-maximum likelihood to estimate their model.

[6]Strictly speaking the additional, explicit dependence of volatility on an unobservable latent process moves us away from the traditional time-deformation literature and more toward the stochastic volatility literature. In addition,we should also point out that working with a stochastic volatility model, and hence departing from the GARCH-type specification may actually constrains our ability to address the number of economic questions compared to the ACD-type model proposed by Engle (2000).

When a drift term appears in the model, a two-stage least squares (2SLS) method can be used to estimate the drift parameters ($\boldsymbol{\theta}_1 = (\alpha_0, \alpha_1)'$) in the mean part, and the volatility parameters ($\boldsymbol{\theta}_2 = (\alpha_2, \alpha_3, \sigma, \rho))'$ in the variance part. See Hamilton (1994), and Gouriéroux and Monfort (1996). That is, the parameter vector $\boldsymbol{\theta}_2$ is first estimated consistently and then the parameter vector $\boldsymbol{\theta}_1$ is estimated by least-squares. However, in the classical setting the application of the 2SLS method requires the condition that instrumental variables (or exogenous explanatory variables) are only related to the drift term but not to the volatility term. Unfortunately, there is no such instrumental variable available for the model in (4), and consequently the 2SLS is not applicable to this model. Theoretically, if the parameter vector $\boldsymbol{\theta}_2$ is fixed, or can be separately estimated consistently, weighted least squares can still be applied to estimate the parameter vector $\boldsymbol{\theta}_1$, and *vice versa*. In this paper, we develop a full MSM estimation for the parameters of the model in (4). Our choice of the MSM method is largely driven by its analytical feasibility and numerical stability in estimating the volatility and correlation parameters of the proposed model. Unfortunately, estimating these key parameters of interest in the study of return processes by other more efficient methods (such as maximum likelihood estimation) has proven to be very intricate.

In the next three sub-sections we proceed with the discussion of the parameter estimation in three stages:

Model A: $\begin{array}{l} R_k = \exp[\alpha_2 + \alpha_3 \ln(d_k) + V_k]\epsilon_k, \\ V_k = \beta_1 V_{k-1} + \eta_k. \end{array}$

Model B: Model A with nonzero $\rho = corr(\epsilon_k, \eta_k)$.

Model C: Model given by (4).

3.1 Moment Selection in Model A

We begin the analysis by noting that Model A is very similar to the classic log-normal stochastic volatility model in (5) studied in Andersen and Sørensen (1996). The main difference between the model in (5) and Model A is the inclusion of the duration process in the volatility, which is observed, stationary and exponentially ergodic, in Model A. Andersen and Sørensen (1996) use the GMM to estimate the parameters of the model in (5). In particular they were interested in the absolute value of the return, $|R_k|$ instead of the return, R_k, itself, and perform the GMM estimation based on the first four moments of $|R_k|$, and certain auto-covariances of $|R_k|$. For small sample size, they report that the GMM estimation based on nine moments yields most satisfactory results. These nine moments are the first four moments of $|R_k|$ and the first, third, fifth and seventh auto-covariances of $|R_k|$. When the sample size is larger than 2000, they found that the GMM estimation based on fourteen moments produces better results. These fourteen moments include the first, third, fifth and seventh auto-covariances of R_k^2 in addition to the previous nine moments.

Our approach is based on the logarithm of the absolute return denoted as $Z_k = \ln(|R_k|)$. There are two major advantages to using Z_k: (1) it lends itself to a model transformation from a non-linear model to a linear model which parameters are much easier to estimate; and (2) with this specification, the GMM objective function does not accumulate due to the positivity of $|R_k|$; so the range of the function would be much more reasonable. As a result, calibration would not be required in our estimation procedure. When the return is zero, a negligibly small non-zero number, say $\exp(-40)$, is assigned to the observation to allow us to validly take the logarithm of the observation. This modification will have little effect on the estimation results. Following Andersen and Sørensen (1996), we first use nine moments to perform the MSM estimation, and these moments are $\mathrm{E}Z_k, \mathrm{E}Z_k^2, \mathrm{E}Z_k^3, \mathrm{E}Z_k^4, \mathrm{E}(Z_k Z_{k-1}), \mathrm{E}(Z_k Z_{k-3}), \mathrm{E}(Z_k Z_{k-5})$, and $\mathrm{E}(Z_k Z_{k-7})$. Since the presence of $\ln(d_k)$ in Model A is related to Z_k (since the sample correlation coefficient of the IBM stock return data considered in this paper is 0.23), we use the first four cross-covariances between Z_k and $\ln d_k$, $\mathrm{E}(Z_k \ln d_k)$, $\mathrm{E}(Z_k \ln d_{k-1})$, $\mathrm{E}(Z_k \ln d_{k-2})$, and $\mathrm{E}(Z_k \ln d_{k-3})$ to perform the MSM estimation. As a result, our initial MSM estimation is based on thirteen moments. But the

MSM estimation based on these thirteen moments did not yield more superior estimation results than that based on a subset of these thirteen moments. This means that there is still a scope for simplification from the thirteen-moment based MSM. So we conduct simulations to identify those moments which are central to the formulation of the MSM estimation.

Note that there are four parameters, $\alpha_2, \alpha_3, \beta_1$, and σ, involved in the design of the simulations. In selecting benchmark (or "true") values for the parameters, we take the following three facts into consideration: (1) as stated earlier, we want to confirm whether the duration process is a directing process, and if so, α_3 should be 0.5. Because of this, we choose the benchmark values for α_3 to be around 0.5; (2) we run a simple linear regression of $\ln |R_k|$ on $\ln d_k$. Note that if Model A fits the data well, $\ln |R_k|$ should be a linear function of $\ln d_k$. The estimates of α_2 and α_3 are found to equal 0.46,and 0.21, respectively. The estimated standard deviation of the error term in the regression model equals 0.92. Therefore, we choose values around 0.9 as the benchmark values for σ; and (3) if the duration process adequately captures the information flow in the stock market, the remaining information (V_k) should be relatively immaterial. Given that the duration process d_k is highly persistent, it is likely that V_k would have small persistence, and, for this reason, the benchmark values of β_1 is chosen to be around 0. In summary the benchmark values for the simulation exercises are chosen to be:

$$\begin{array}{rlllll} \beta_1: & -0.3, & -0.1, & 0.1, & 0.3; & \\ \sigma: & 0.9, & 1.2, & 1.6; & & \\ \alpha_2: & -0.3, & -0.1, & 0.1, & 0.3, & 0.6; \\ \alpha_3: & 0.3, & 0.6, & 0.9. & & \end{array}$$

Thus we end up with eight settings of the parameter values, as shown in Table 1. In the first row of Table 1, where $\alpha_2 = 0.3$, $\alpha_3 = 0.6$, $\beta_1 = 0.3$ and $\sigma = 1.2$, we first simulate a return series based on Model A, then carry out the MSM estimation of the parameters by using the simulated returns. We find that the MSM estimation based on the following seven moments produces a reasonable result. These moments are $\mathrm{E}Z_k$, $\mathrm{E}Z_k^2$, $\mathrm{E}Z_k^3$, $\mathrm{E}Z_k^4$, $\mathrm{E}(Z_k Z_{k-1})$, $\mathrm{E}[Z_k \ln(d_k)]$, and $\mathrm{E}[Z_k \ln(d_{k-1})]$.

Next a Monte-Carlo simulation is conducted to check the performance of the MSM with the chosen seven moments. Specifically, $1,000$ simulations are carried out under each of eight parameter settings. The simulation results are reported in Table 1, including the averages of estimates over 1000 simulations and the sample standard deviations reported in the parentheses.[7] We note that the 95% confidence intervals easily include the benchmark value of each parameter. This indicates that the MSM based on the selected seven moments worked reasonably well, at least for the eight selected parameter settings. All of the estimates of σ are very close to their benchmark values. In general the bias associated with the MSM estimate tends to increase with the benchmark value of σ. But the estimates of β have little bias except for $\beta = -0.1$, and the estimates of α_2 and α_3 are reasonably close to their benchmark values, except in case #6 in which there is some evidence of biases. In summary, the MSM estimation procedure formed by the selected seven moments appears to work reasonably well for all of the parameters of the model.

3.2 Moment Selection in Model B

When the correlation between the return innovation process and the volatility innovation process, ρ, is treated as a free parameter in the model, the mean of the return process is no longer zero. In this case, the logarithmic transformation on the absolute return $|R_k|$, (Z_k) or the innovation, $|\epsilon_k|$, would lead to loss of information on the sign of the correlation between the original, untransformed innovations ϵ_k and η_k (See Harvey and Shephard, 1996). In order to estimate the sign of the correlation, some odd-order moments of the return process (and not Z_k) would need to be included in the MSM estimation, in addition to the previously chosen seven moments. Specifically the moments used are selected from two different (but related) sources; one is from the return process, R_k, and another is from the process Z_k.

[7] Note that after the first simulation, the parameter estimates are used as new starting values in the next simulation, so the remaining simulations will take less time than the previous ones.

The odd-order moments of R_i is known to contain information about the correlation between the return innovation process and the volatility innovation process, ρ. Importantly, these moments can be used to determine the sign of R_k . In effect, we add the following additional moments: the second, third, and fourth moments of R_k, as well as the first-order auto-covariance of R_k, namely , $\mathrm{E}R_k^2$, $\mathrm{E}R_k^3$, $\mathrm{E}R_k^4$, and $\mathrm{E}(R_k R_{k-1})$. Note that the first moment of R_k is reserved for estimation of the constant drift, α_0, in the model in (4). To specify the benchmark values in the Monte-Carlo simulation, due to our focus on the correlation ρ, we choose six different values of ρ in the range of $(-1, 1)$: $-0.9, -0.4, -0.1, 0, 0.1, 0.4$, and 0.9. The benchmark values for $\alpha_2, \alpha_3, \beta_1$, and σ remain the same as in the previous simulation exercise (see Table 1). Again, there are 8 cases under investigation, as shown in Table 2.

At each parameter setting, the MSM is applied to obtain the parameter estimates of the model. Interestingly, we find that the second and third moments of R_k seem to be the key contributors to the estimation of ρ. So we decide to use the following nine moments: $\mathrm{E}Z_k$, $\mathrm{E}Z_k^2$, $\mathrm{E}Z_k^3$, $\mathrm{E}Z_k^4$, and $\mathrm{E}(Z_k Z_{k-1})$, $\mathrm{E}(Z_k \ln(d_k))$, $\mathrm{E}(Z_k \ln(d_{k-1}))$, $\mathrm{E}R_k^2$, $\mathrm{E}R_k^3$ as a basis for the MSM estimation. As in the previous section, a simulation is used to examine the appropriateness of the chosen nine moments. For each parameter setting, the simulation is replicated 1,000 times.

Table 2 presents the simulation results. Once again, the MSM formed from the selected nine moments seems to provide satisfactory estimates of the model parameters, as every benchmark parameter falls within the corresponding 95% confidence interval. In particular the estimates of the correlation ρ are very close to their benchmark values with the correct expected signs. When $\rho = 0$, the estimate of ρ is 0.03 with only 1% bias and the corresponding 95% confidence interval contains 0. The estimates of $\alpha_2, \alpha_3, \beta_1$, and σ are reasonably close to their benchmark values, especially when the value of β_1 is not too small at, say, -0.1 (case #5), the value of σ is around 1, and the value of α_3 ranges from 0.3 to 0.9.

3.3 Moment Selection in Model C

When a drift term is included in the model, as discussed earlier, the 2SLS method is no longer applicable. From a regression model's point of view, the drift term, $\alpha_0 + \alpha_1 d_k$, of the model in (4) can be treated as a linear function of the duration d_k, while the volatility term, $\exp(\alpha_2 + \alpha_3 \ln d_k + V_k)\epsilon_k$, can be regarded as a non-Gaussian error term. In this case we can use two moments to estimate the two drift parameters α_0, and α_1: one is the mean of the returns, $\mathrm{E}R_k$, and another is the cross-covariance between R_k and d_k, $\mathrm{E}(R_k d_k)$. Note that these two moments have not been used so far in the MSM estimation. Therefore, a total of eleven moments are now used in estimation.

Again we employ a Monte-Carlo simulation to assess the appropriateness of the selected eleven moments. As before, we begin with one parameter setting, where $\alpha_0 = 0.5$, $\alpha_1 = 1$, $\alpha_2 = 0.6$, $\alpha_3 = 1$, $\beta_1 = 0.2$, $\sigma = 0.9$, $\rho = -0.3$. For estimation, we first remove the drift term from R_k, and the residuals $R_k - \hat{\alpha}_0 - \hat{\alpha}_1 d_k$ are used to fit Model A considered in the previous section. We find that the following ten moments yield reasonable estimation results: Z_k, $\mathrm{E}Z_k$, $\mathrm{E}Z_k^2$, $\mathrm{E}Z_k^3$, $\mathrm{E}Z_k^4$, and $\mathrm{E}(Z_k Z_{k-1})$, $\mathrm{E}[Z_k \ln(d_k)]$,$\mathrm{E}[Z_k \ln(d_{k-1})]$, $\mathrm{E}R_k^2$, $\mathrm{E}R_k$, $\mathrm{E}(R_k d_k)$. In other words the third moment of R_k, $\mathrm{E}R_k^3$, is found to have little contribution for the estimation and is, thus, excluded from the above eleven moments.

Next a Monte-Carlo simulation is carried out to examine the performance of the selected ten moments in the MSM. Initially we attempt to run $1,000$ replications as in the previous two simulation studies. However, the computation involved turns out to be too time intensive; so we are forced to consider only four parameter settings, as shown in Table 3, and only 100 replications are run at each parameter setting. The simulation results are reported in Table 3. The estimates are reasonably close to their benchmark values, except β_1 which is marginally insignificant at the 95% level (case#3).

Clearly, the means of the estimates of α_3, β_1, σ and ρ are close to their benchmark values, and each parameter (except β_1 in case#3) is included within the 95% confidence interval. But there is a relatively large bias in case #3, especially for α_3. The mean of the estimates of α_3 is 0.25. Fortunately, even in this worst case, the signs of all of the parameters are correctly estimated, especially the sign of β_1 which reflects the interpretation of the Markovian nature of the latent process V_k. For each of of the parameters, α_3, β_1, σ and ρ, we plot the histograms of the simulation estimates in order to assess how far the estimate of each simulation is from its benchmark value. These are shown in Figure 1 (with the term "true" value used in place of "benchmark" value in each histogram).

The first column in Figure 1 shows the histograms for the parameter α_3. The histograms appear to be skewed, but the benchmark values are situated around the middle points of the histograms, and the range in each histogram is narrow, except for case #3. Therefore, we are reasonably confident in the estimate of α_3. For the parameter β_1, the histograms shown in the second column in Figure 1 appear to resemble that of a Gaussian, but the benchmark parameters are a little farther away from the middle points. This indicates that the MSM has more chance of overestimating the parameters of the model. But the sign is consistent with that of the benchmark value in each case. For the parameter ρ, the histograms, shown in the third column of Figure 1, seems more skewed. But the benchmark value of the parameter resides in the area with the largest frequency in the histogram. The histogram for the estimate of σ is shown in the last column in Figure 1. These four histograms are slightly centralized and symmetric around the middle range. The benchmark value is very close to the middle range of each histogram, and the spread is small as well, suggesting that the MSM produces a reasonable estimate for the parameter ρ, which is one of the key parameters in the model.

The estimates of α_2 are very close to the benchmark value, and the sample standard deviations are also relatively small. The estimates of α_0, and α_1 seem to have a larger bias than those of the other five parameters. Nevertheless the 95% confidence intervals still include the benchmark value of each parameter. In case #1, the estimate of α_0 is 1.12, which is quite far from the benchmark value of 0.5, but the sample standard deviation is as large as 0.40. The same situation arises in the estimate of α_1. Overall the evidence conform to our prior expectation and suggest that the MSM estimates of the model's parameters are not fully efficient.

In conclusion, we have arrived at ten moments based upon which the MSM estimation is performed in the next section. These selected ten moments are: Z_k, $\mathrm{E}Z_k$, $\mathrm{E}Z_k^2$, $\mathrm{E}Z_k^3$, $\mathrm{E}Z_k^4$, and $\mathrm{E}(Z_kZ_{k-1})$, $\mathrm{E}(Z_k\ln(d_k))$,$\mathrm{E}(Z_k\ln(d_{k-1}))$, $\mathrm{E}R_k^2$, $\mathrm{E}R_k$, $\mathrm{E}(R_kd_k)$.

4 Application: IBM Stock Return Data Analysis

4.1 Preliminary Analysis

We are now in the position to analyze the IBM stock return data that consist of the original trade data and the original quote data from March 1, 2001 to March 31 of 2001. For the analysis in this paper, we rely on the marginal moments of the data. The data contain tick-by-tick trading records including the day, time of a trade, a trade price, bid price and ask price. Table 4 shows a selected fragment of the data. To avoid both the market opening and closing effects, all records occurred before 10:00 am and after 3:45 pm are excluded from our analysis. If multiple trades occur at the same time, for simplicity, only the last one is used as a representative observation at that time. When one record happens to be incomplete, we use the nearest-neighbor record to replace it. For example, in Table 4, the four trading records all occur at 10:57:24, on March 1. Only the last one is kept in our data set, and the other three are discarded. There is a missing value of 'ask price' in this one, which is imputed by the nearest-neighbor one as 98.61.

After the data cleaning-up process, we end up with 66,678 trading records. To avoid the bouncing problem, we take the average of bid price and ask price as the trading price.

Moreover the return is defined by[8]

$$R_k = \ln\left(\frac{\text{Ask}_k + \text{Bid}_k}{2}\right) - \ln\left(\frac{\text{Ask}_{k-1} + \text{Bid}_{k-1}}{2}\right).$$

We note that for all observations, we multiply the values of the return, R_k, by 10,000 times as they are very close to 0.

Duration data typically display two seasonality effects: the day-of-week effect and the time-of-day effect. As to the first effect, duration usually remains high between Monday and Wednesday, then decreases steadily afterwards and eventually reaches the shortest on Friday. This effect is caused by trades that appear relatively inactive during the early part of the week and become substantially more active at the end of the week, In principle this effect can be removed by taking the average sample duration for a week day. However, in this paper, we will ignore the day-of-week effect due to the short-time period of our data set. As to the time-of-day effect, duration first appears short in the morning, increases substantially around noon, and then decreases toward the closing of the market. As this second effect appears quite visible on our duration data,[9] as shown in Figure 2, we remove it by using a nonparametric method described in Engle and Russell (1998). The resulting adjusted duration data are then used for model estimation.

In the analysis, we use the last 10,000 observations of the total of 66,678 observations.[10] Figure 3 and Figure 4 show the duration data and the return data respectively.

[8] An alternative definition of the return is given by

$$R_k = \frac{1}{2}[\ln(\text{Ask}_k) - \ln(\text{Ask}_{k-1}) + \ln(\text{Bid}_k) - \ln(\text{Bid}_{k-1})].$$

This definition is particularly useful for the analysis of currency exchange data (see for example Dacorogna et al., 1993, and Ghysels, Gouriéroux and Jasiak (1998). However these two definitions are equivalent if the ratio of Ask_k to Ask_{k-1} is the same as (or close to) the ratio of Bid_k to Bid_{k-1}.

[9] The time-of-day pattern in Figure 2 shows that the duration tends to increase in the morning, in particular from 11:00 onwards, and reaches a maximum at around 12:30 pm, then decreases (except during the interval [13:00, 14:00]) toward the closing of the market in an average trading day.

[10] This is based on the view that if the return process is stationary and exponentially ergodic, using many more observations need not lead to tangible efficiency gains in the estimation of the model' parameters, except, perhaps, to raise the required computing time substantially.

Table 5 presents the basic statistics for the duration data, the return data and the adjusted duration data.

From Table 5 we make the following remarks: [i] although the duration and adjusted duration show little differences in terms of their values of skewness and kurtosis, each displays large skewness and kurtosis values, suggesting that the duration's unconditional distribution is both asymmetric and leptokurtic; while [ii] the return data is approximately symmetric around zero, suggesting no strong drift occurring over the sequence of the trades. But the large value of kurtosis indicates that the return's unconditional distribution is leptokurtic.

The autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the duration series, shown respectively in Figure 5 and Figure 6, indicate that the duration process has long memory in its dependence structure. However both the ACF and PACF plots of the return series in Figure 7 and Figure 8 show that the returns resemble a process of independent increments. If the returns were truly independent, the absolute values of the returns would be independent. Figure 9 shows that the ACF of the absolute return series decays very slowly, suggesting that the volatility of the return process is highly persistent and, therefore, appropriately modeling the return series will have to account for the time varying variance of the series.

To inspect the association between the absolute return series and the duration series, we use the sample cross correlation (CC) function. The CC function between the durations and the returns at leads is shown in Figure 10 and the difference of two CC functions between leads and lags is plotted in Figure 11. The CC function decays sharply at the first order then slowly in both cases, and the difference between the CC function at leads and the CC function at lags is not significant, suggesting symmetry in leads versus lags. However, it is clear that both the durations and returns are positively correlated at both leads and lags; in other words, a longer duration tends to have a larger volatility.

4.2 Model-based Analysis

In this subsection we fit the adjusted durations and returns to the time-deformation model in (4) in the presence of both the drift term and the correlation coefficient ρ. The MSM formed from the chosen ten moments is employed in the parameter estimation, and the results are reported in Table 6. All parameters are estimated to be significantly different from zero, except for β_1, which is marginally insignificant from zero at the 95% level.

As mentioned earlier, the MSM has its advantages and disadvantages compared to other methods, in particular, compared to Monte-Carlo maximum-likelihood (MCML) proposed by Durbin and Koopman (1997).[11] One advantage of the MSM is its simplicity as only moments are required for estimation. Another advantage of the MSM is that computation time is relatively low compared to the MCML approach. However one well-known limitation of the MSM relative to the MCML is the apparent lack of efficiency, although, unlike the MCML, it affords robustness with respect to the distributional assumption relative to the MCML approach. Another disadvantage of the MSM relative to the MCML is that the estimates of the error terms are not available for analysis. Therefore, the assumptions about the error terms can not be empirically verified; so we can not directly check the empirical validity of the normality assumption of the innovations ϵ_k and η_k. However, once all of the parameters in the model in (4) have been estimated, the statistical properties of the return process R_k defined by the model can be derived. That is, given its stationarity and ergodicity, we can compare the marginal distribution of the process to the empirical distribution of the observed returns. However since this marginal distribution is a mixture of d_k, ϵ_k, and η_k, there is no closed-form expression available. So, to tackle this problem, we resort to the Monte-Carlo technique to simulate a path of the return process with the length equal to

[11] The idea of MCML estimation is to reformulate an intractable likelihood function of non-Gaussian distribution into one of Gaussian (treated as the importance) distribution for which the EM algorithm, with the E-step being the Kalman filter, can then be applied. Feng. Jiang and Song (2004) use this method to estimate stochastic duration models.

$10,000$, then compare the estimated marginal density based on the simulated data to the empirical one. To minimize the initial value effect of the simulation, the first 1000 simulated returns are discarded. Because β_1 is marginally significant at the 95% level, two different paths are generated, one with the parameter β_1 included in the model, and another with the parameter β_1 set equal to zero in the model.

The smoothing spline method is used to obtain the marginal density functions of the empirical and simulated data. Figure 12 shows the empirical and simulated-based densities for the model with the parameter β_1 included in the model. Figure 13 shows these two density functions for the model with the parameter β_1 set equal to zero in the model. A simple comparison indicates that the two density functions in Figure 12 are closer to each other than those in Figure 13. This points to the importance of the effect of β_1 on the model estimation. The two curves in Figure 12 are very similar except that the simulated return process has a relatively higher probability of having zero returns, and a relatively lower probability of having returns in intervals $(-8, -7)$ and $(6, 7)$. We also use the Kolmogorov-Smirnov statistic to formally test that the two smoothed density functions are not statistically different from each other. We find that both density functions are indeed not different from each other with a probability of 0.19. We take this as evidence that the model is adequate for the purpose of describing the marginal property of the return process.

It is important to note that the first case in Table 3 is intentionally set up, so that the benchmark parameters are close to the estimates obtained from the data analysis. This is because in doing so, we can examine whether the chosen ten moments lead to a proper MSM estimation. Apparently, the simulation results support our selection of these moments for estimation. In this case the 95% confidence intervals always include the benchmark values. So, we conclude tentatively that no additional moment is needed for the MSM estimation.

For the latent process of the model in (4), $V_k = \beta_1 V_{k-1} + \eta_k$, the variance σ of η_k is estimated at 0.788 with standard error of 0.024 (which suggests that σ is statistically significantly different from 0). This implies that the innovation η_k plays a role in explaining the return process in the model. We conclude that the return is not Gaussian conditional on the duration process.[12] Furthermore, β_1 is marginally different from 0, and there are some evidence to suggest that V_k is an AR(1) process. In other words, the duration process d_k can be used marginally as a directing process. Lastly, $\hat{\beta}_1$ equals to 0.2, which points to evidence of lack of persistency (evidence of high persistency is usually characterized by $1.0 > \hat{\beta} > 0.9$), so the informational effect represented by the latent process V_k resides in the market only for a short while, although an AR(1) structure of V_k still remains a relevant process.

The estimated parameter $\hat{\alpha}_3$ is positive at 0.998; so the volatility of the return process has a positive relationship with $\exp(\hat{\alpha}_3 \ln d_k) = d_k^{\hat{\alpha}_3} \approx d_k$. This means that a longer duration, d_i, tends to be associated with a higher volatility of the return. One explanation for this result can be offered at this time by reasoning that IBM not only is one of the most heavily traded stock, but it is a relatively large company. In such a company, it makes sense to assume that uninformed investors tend to dominate informed investors. This is because investors often like to trade large companies' stocks for reasons other than the companies own news. This suggests that investor trading intensity may be less sensitive to the large company specific news.

The estimated correlation between the two innovations ϵ_k and η_k is -0.25. This shows a pronounced leverage effect in the IBM stock returns. Recall that the leverage effect refers to an asymmetric effect between positive returns and negative returns on volatility. When there exists a leverage effect, negative return sequences are always associated with an increase in volatility, and positive returns are associated with a decrease in volatility. This phenomenon is common for equities and represents a well-known stylized fact in the literature (e.g. Black 1976 or Nelson 1991).

[12]WQe note that this evidence does not necessarily invalidate the time-deformation literature. An and Geman (2000), among others, could simply argue that we are using an insufficient economic clock in our analysis.

The estimated coefficient $\hat{\alpha}_1$ is recorded at 1.148. This means that a longer duration leads to a larger return. Following Engle (2000), we argue that during a slow trading activity, more informed trades are active in the market, and the spread is high; as a result the return R_k must be higher in order to compensate for acquiring new information and high spread.

Next the estimated intercept term, $\hat{\alpha}_0$, is calculated at 0.584. This estimate represents a long-term average return offset by the mean of the duration process and the mean of $\exp(\alpha_2 + \alpha_3 \ln d_k + V_k)\epsilon_k$. Lastly the estimated constant $\hat{\alpha}_2$ equals to 0.645, which, along with the duration process, captures the long-term volatility of the return.

5 Conclusion

The focus of the paper has been on modeling returns in the time-deformation framework. We assumed that the duration process is a possible directing process and proposed a bivariate stochastic time-deformation model. The method of simulated moments (MSM) was applied to estimate the parameters of this model. An advantage of the MSM, as shown in this paper, is that it is flexible enough to be adapted for estimation of different specifications nested within our proposed model. In addition, in comparison to other methods, such as the maximum-likelihood monte-carlo (MCML) proposed by Durbin and Koopman (2004), the MSM is likely to take much less computational time to estimate the model presented in this paper. In particular our MSM, via Monte-Carlo study, selected a relatively small number of moments, which yield satisfactory estimation results of the model although it obviously lacks efficiency vis-a-vis the MCML approach. Our main findings in this paper can be summarized as follows: (1) there was evidence that the duration process is marginally a proper directing process for the IBM stock return; (2) there was a pronounced leverage effect in the IBM stock-return process; (3) a longer duration tends to be associated with a higher volatile return; and lastly (4) the proposed model is capable of reproducing the return whose marginal density function is close to that of the empirical return.

Acknowledgements

The authors are grateful to Christian Gouriéroux, Joann Jasiak and George Jiang for their valuable discussions and suggestions on earlier versions of this paper. However we are still responsible for all of the remaining errors and omissions.

References

Admati, A., and P. Pfeiderer, A Theory of Intraday Patterns: Volume and Price Variability. *Review of Financial Studies* **1** (1988), pp. 3-40.

An, H. Z., and F. C. Huang, The Geometrical Ergodicity of Nonlinear Autoregressive Models. *Statistica Sinica* **6** (1996), pp. 943-956.

Andersen, T. G., and A. E. Sørensen, GMM Estimation of a Stochastic Volatility Model: a Monte Carlo Study *Journal of Business and Economic Statistics* **14** (1996), pp. 328-352.

Ane, T., and H. Geman, Order Flow, Transaction Clock, and Normality of Asset Returns. *The Journal of Finance* **55** (2000), pp. 2259-2284.

Bauwens, L., and D. Veredas, The Stochastic Conditional Duration Model: A Latent Variable Model for the Analysis of Financial Durations. *Journal of Econometrics* **119** (2004), pp. 381-482.

Bennett, B. M., Some Further Extensions of Fieller's Theorem. *Annals of the Institute of Statistical Mathematics* **5** (1954), pp. 103-106.

Black, F., Capital Market Equilibrium with Restricted Borrowing. *Journal of Business* **45** (172), pp. 444-454.

Chernov, M., A.R. Gallant, E. Ghysels, and G. Tauchen, Alternative Models for Stock Price Dynamics. *Journal of Econometrics* **116** (2003), pp. 225-257.

Clark, P., A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices. *Econometrica* **41** (1973), pp.135-155.

Dacorogna, M. M., U. A. Muller, R. J. Nagler, R. B. Olsen, and O. V. Pictet, (1993), A Geographical Model for the Daily and Weekly Seasonal Volatility in the Foreign Exchange Market. *Journal of International Money and Finance* **12** (1993), pp. 413-438.

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

Easley, D., and M. O'Hara, Time and the Process of Security Price Adjustment. *Journal of Financial Economics* **19** (1992), pp. 69-90.

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

Feng, D., G. J. Jiang and P. X-K. Song, Stochastic Conditional Duration Models with Leverage Effect for Financial Transaction Data. *Journal of Financial Econometrics* **3** (2004), pp. 390-421.

Gallant, A. R., P. E. Rossi and G. Tauchen, Stock Prices and Volume. *Review of Financial Studies* **5** (1992), pp. 199-242.

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

Ghysels, E., C. Gouriéroux and J. Jasiak, Market Time and Asset Price Movement: Theory and Estimation, in *Statistics in Finance* (1997), D. Hand and S. Jacka (eds.), Edward Arnold, London, Chapter 15, pp. 307-332.

Gouriéroux, C., and A. Monfort, Simulation-based Econometric Methods (1996). *Oxford University Press.*

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

Hamilton, J. D., Time Series Analysis (1994). *Princeton University Press.*

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

Harvey, A. C., and N. Shephard, Estimation of an Asymmetric Stochastic Volatility Model for Asset Returns. *Journal of Business and Economic Statistics* **14** (1996), pp. 429-434.

Huang, X., and G. Tauchen, The Relative Contribution of Jumps to Total Price Variation. *Journal of Financial Econometrics* **3** (2005), pp. 456-499.

Knight, J. L., and C. Ning, Estimation of the Stochastic Conditional Duration Model via Alternative Methods, *The Econometrics Journal* (2008), in press.

McFadden, D., A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration. *Econometrica* **57** (1989), pp. 995-1026.

Meyn, S. P., and R. L. Tweedie, Markov Chains and Stochastic Stability (1993). *Springer-Verlag.*

D.B. Nelson, D. B., Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica* **60** (1992), pp. 347-370.

Newey, W. K., and K. D. West, A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* **55** (1987), pp. 703-8.

Renault, E., and B. Werker, Stochastic Volatility Models with Transaction Time Risk. Working paper, 2004.

Stock, J. H., Estimating Continuous-Time Processes Subject to Time Deformation. *Journal of the American Statistical Association* **83** (1988), pp. 77-85.

Tong, H. Non-linear Time Series: A Dynamic System Approach (1990). *Clarendon Press Oxford.*

Xu, D., J. Knight and T. S. Wirjanto, Asymmetric Stochastic Conditional Duration Model - A Mixture of Normals Approach, Unpublished Manuscript, 2008.

Appendix A: Proofs of Theorems and Propositions

Proof of theorem 1: To proceed with the proof, we first introduce the following lemma (Tong,1990).

Lemma 1. *Assume that (1) $\mathbf{X}_k$ is aperiodic and irreducible, (2) there exists a small set C (see Meyn and Tweedie, 1993), a non-negative measurable function g, and constants $a < 1, \gamma > 0$ and $b > 0$ such that*

$$E(g(\mathbf{X}_{k+1})|\mathbf{X}_k = \mathbf{x}) < ag(\mathbf{x}) - \gamma, \ \mathbf{x} \notin C;$$

and

$$E(g(\mathbf{X}_{k+1})|\mathbf{X}_k = \mathbf{x}) < b, \ \mathbf{x} \in C.$$

Then $\mathbf{X}_k$ is geometrically ergodic.

Let $X_k = \ln|R_k - \alpha_0 - \alpha_1 d_k|$, $\zeta_k = \ln|\epsilon_k|$, and $Y_k = X_k - \alpha_3 \ln d_k$. Without the loss of generality, assume that the means of $Y_k, \ln d_k$, and V_k are all zero (because a constant mean does not have any impact on ergodicity and stationarity), and assume that the mean of ζ_k is zero (otherwise use $\hat{\zeta}_k = \zeta_k - \mathrm{E}(\zeta_k)$ to replace ζ_k). The model in (4) can be rewritten as

$$\begin{aligned} Y_k &= \beta_1 V_{k-1} + \eta_k + \zeta_k, \\ V_k &= \beta_1 V_{k-1} + \eta_k. \end{aligned}$$

Let $\mathbf{Y}_k = (Y_k, V_k)'$, $\mathbf{u}_k = (\zeta_k, \eta_k)'$, and

$$A = \begin{pmatrix} 0 & \beta_1 \\ 0 & \beta_1 \end{pmatrix}, \ B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Thus, the model takes the form of a linear VAR(1) model

$$\mathbf{Y}_k = A\mathbf{Y}_{k-1} + B\mathbf{u}_k. \tag{6}$$

Let $\mathbf{U}_k = B\mathbf{k}_k$. So $\{\mathbf{U'}_k\}$ are i.i.d..

Since (ϵ_k, η_k) is $BVN(0,0;1,\sigma^2,\rho)$ and independent of (ϵ_k, η_k), the mean and variance and covariance matrix of $\mathbf{u}_k$ are respectively given by,

$$\begin{aligned} \mathrm{E}(\zeta_k) &= 0, \ \mathrm{E}(\eta_k) = 0, \\ \mathrm{var}(\zeta_k) &= \frac{1}{4}\frac{1}{4}\psi^1(\frac{1}{2}) = \frac{1}{32}\pi^2, \ \mathrm{var}(\eta_k) = \sigma^2, \\ \mathrm{cov}(\zeta_k, \eta_k) &= 0, \end{aligned}$$

where ψ is digamma function.

It follows that the mean and the variance-covariance matrix of $\mathbf{U_k}$ are,

$$\mathrm{E}\mathbf{U}_k = (0,0)';$$

and

$$\mathrm{var}(\mathbf{U}_k) = \begin{pmatrix} \frac{1}{32}\pi^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}.$$

Given model (6), because the spectral norm of A is less than 1, i.e. $|\beta_1| < 1$, there must exist a matrix norm $||\cdot||_m$, which is induced by a vector norm $||\cdot||_v$, and a positive real number $\lambda < 1$, such that

$$||A\mathbf{y}||_v \leq ||A||_m ||\mathbf{y}||_v \leq \lambda ||\mathbf{y}||_v, \text{ for any } \mathbf{y} \in R^2.$$

(For more detail of this, see An and Huang, 1996)

Now we prove that the conditions of Lemma 1 are satisfied by the model in (6). Let the Euclidean norm $||\cdot||_e$ be the test function g. First define a small set $C \subseteq R^2$ in the following way,

$$C = \{\mathbf{y} \in R^2 : g(\mathbf{y}) \leq c\},$$

where c is a positive number and will be decided later. So,

$$\begin{aligned} \mathrm{E}[g(\mathbf{Y}_{k+1})|\mathbf{Y}_k = \mathbf{y}] &= \mathrm{E}(||\mathbf{Y}_{k+1}||_e|\mathbf{Y}_k = \mathbf{y}) \\ &= \mathrm{E}(||A\mathbf{y} + \mathbf{U_k}||_e|\mathbf{Y}_k = \mathbf{y})\} \\ &\leq \mathrm{E}(||A\mathbf{y}||_e|\mathbf{Y}_k = \mathbf{y}) + \mathrm{E}(||\mathbf{U_k}||_e|\mathbf{Y}_k = \mathbf{y}) \\ &\leq \lambda||\mathbf{y}||_e + \mathrm{E}(||\mathbf{U}_k||_e). \end{aligned}$$

Because $\mathrm{E}\{||\mathbf{U}_k||_e\}$ is a positive constant, and $\lambda < 1$, there must exist a positive number $\lambda < a < 1$, and a positive number c such that

$$a - \lambda > \frac{2\mathrm{E}(||\mathbf{U}_k||_e)}{c}.$$

Let $\gamma = \mathrm{E}(||\mathbf{U}_k||_e)$, and $b = ac + \mathrm{E}(||\mathbf{U}_k||_e)$,

Thus,

$$\begin{aligned} \lambda||\mathbf{y}||_e + \mathrm{E}(||\mathbf{U}_k||_e) &\leq [a - \tfrac{2\mathrm{E}(||\mathbf{U}_k||_e)}{c}]||\mathbf{y}||_e + \mathrm{E}(||\mathbf{U}_k||_e) \\ &\leq a||\mathbf{y}|| - \gamma, \text{ when } \mathbf{y} \notin C; \end{aligned}$$

and

$$\lambda||\mathbf{y}||_e + \mathrm{E}(||\mathbf{U}_k||_e) \leq b, \text{ when } \mathbf{y} \in C.$$

That is, for the above chosen γ, and b, we have that

$$\mathrm{E}[g(\mathbf{Y}_{k+1})|\mathbf{Y}_k = \mathbf{y}] \leq \begin{cases} a||\mathbf{y}||_e - \gamma, & \text{when } \mathbf{y} \notin C; \\ b, & \text{when } \mathbf{y} \in C. \end{cases}$$

Therefore, $\mathbf{Y}_k$ is geometrically ergodic. The stationarity is simply due to the stationarity of d_k and V_k.

If $\ln d_k$ is stationary, geometrically ergodic and independent of (ζ_k, ξ_k), then the geometric ergodicity of Y_k implies that $X_k = \ln|R_k - \alpha_0 - \alpha_1 d_k|$ is geometrically ergodic.

Next we discuss the geometric ergodicity of $\{R_k - \alpha_0 - \alpha_1 d_k\}$. Denote this process as $W_k = R_k - \alpha_0 + \alpha_1 d_k$. Therefore,

$$W_k = \begin{cases} \exp(X_k), & W_k \geq 0; \\ -\exp(X_k), & W_k < 0; \end{cases}$$

then we define a function $\tilde{g}$ as:

$$\tilde{g}(w_k, v_k) = \begin{cases} g(\exp(\frac{x_k}{2}), v_k), & w_k \geq 0; \\ g(-\exp(\frac{x_k}{2}), v_k), & w_k < 0; \end{cases}$$

where $x_k = 2\ln|w_k|$.

It is clear that $\tilde{g}$ is measurable, and because $\tilde{g}(w, v) = g(\mathbf{x})$, the two conditions listed in Lemma 1 hold for new process $\{W_k, V_k\}$, so the process $\{W_k\}$ is geometrically ergodic and stationary.

Moreover if d_k is stationary and geometrically ergodic, then the process R_k is stationary and geometrically ergodic.

Proof of Proposition 1: The following Lemma is originally due to Bennett (1954).

Lemma 2. *Let $Y_1, \cdots .Y_n$ be i.i.d. $N(\delta, 1)$. $\sum\limits_{h=1}^{n} Y_h^2$ is a non-central χ^2 with n degree freedom. Let $Y = \ln(\chi^2/n)$. Then the characteristic function (CF) of Y is given by:*

$$\begin{aligned} \Psi_Y(u) &= E[\exp(iuY)] \\ &= \exp[-\lambda - iu\ln(\frac{n}{2})]\frac{\Gamma(\frac{n}{2}+iu)}{\Gamma(\frac{1}{2}n)}F(\frac{n}{2}+iu, \frac{1}{2}n, \lambda), \end{aligned}$$

where $\lambda = \frac{1}{2}n\delta^2$, $F(\alpha, \beta, x) = \sum\limits_{h=0}^{\infty} \frac{\Gamma(\alpha+h)\Gamma(\beta)}{\Gamma(\alpha)\Gamma(\beta+h)}\frac{x^h}{h!}$, and $\Gamma(\cdot)$ is the gamma function.

Let $\Psi(u) \equiv \mathrm{E}[\exp(iuX_k)]$ be the CF of X_k. Let $\Psi_Z(u)$ denote the CF of Z_k. $\Psi_Z(u) = \mathrm{E}[\exp(iuZ_k)]$.

To proceed with the proof of Proposition 1, the following two lemmas, Lemma 3 and Lemma 4, are introduced first.

Lemma 3.

$$E[\exp(iu\zeta)|\eta] \;=\; \exp\left[i\frac{u}{2}\ln(2-2\rho^2)\right]\sum_{h=0}^{\infty}\left[\frac{1}{h!}\frac{\Gamma(\frac{1}{2}+i\frac{u}{2}+h)}{\Gamma(\frac{1}{2}+h)}\exp(-c^2\eta^2)(c^2\eta^2)^h\right]. \quad (7)$$

Proof. The following generic notation is used. For any random variable X, $\tilde{X}$ is denoted $X|\eta$, X conditional on η.

Since (ϵ,η) is $BVN(0,0,;1,\sigma^2;\rho)$, $\epsilon|\eta \sim N(\frac{\rho}{\sigma}\eta, 1-\rho^2)$, i.e. $\tilde{\epsilon} \sim N(\frac{\rho}{\sigma}\eta, 1-\rho^2)$, and furthermore, $\tilde{\epsilon}_{new} \equiv \frac{1}{\sqrt{1-\rho^2}}\tilde{\epsilon} \sim N(\frac{\rho}{\sigma\sqrt{1-\rho^2}}\eta, 1)$.

By Lemma 2, the CF of $\ln(\tilde{\epsilon}^2_{new})$ is

$$\begin{aligned}
&\mathrm{E}[\exp(iu\ln\tilde{\epsilon}^2_{new})] \\
=\;& \frac{1}{\sqrt{\pi}}\exp[-\lambda_0 + iu\ln 2]\Gamma(\frac{1}{2}+iu)F(\frac{1}{2}+iu,\frac{1}{2},\lambda_0), \qquad (8)
\end{aligned}$$

where $\lambda_0 = \frac{1}{2}\left(\frac{\rho}{\sigma\sqrt{1-\rho^2}}\eta\right)^2 \equiv c^2\eta^2$, with $c = \frac{|\rho|}{\sigma\sqrt{2(1-\rho^2)}}$.

It follows from (8) that

$$\begin{aligned}
\mathrm{E}[\exp(iu\zeta)|\eta] &= \mathrm{E}[\exp(iu\tilde{\zeta})] \\
&= \mathrm{E}\left\{\exp\left[i\frac{u}{2}\ln(\sqrt{1-\rho^2}\tilde{\epsilon}_{new})^2\right]\right\} \\
&= \exp\left[i\frac{u}{2}\ln(1-\rho^2)\right]\frac{1}{\sqrt{\pi}}\exp(-\lambda_0 + i\frac{u}{2}\ln 2)\Gamma(\frac{1}{2}+i\frac{u}{2})F(\frac{1}{2}+i\frac{u}{2},\frac{1}{2},\lambda_0) \\
&= \frac{1}{\sqrt{\pi}}\exp\left[i\frac{u}{2}\ln(2-2\rho^2)\right]\Gamma(\frac{1}{2}+i\frac{u}{2})\exp(-c^2\eta^2) \\
&\quad \times\sum_{h=0}^{\infty}\frac{\Gamma(\frac{1}{2}+i\frac{u}{2}+h)\Gamma(\frac{1}{2})}{\Gamma(\frac{1}{2}+i\frac{u}{2})\Gamma(\frac{1}{2}+h)}\frac{(c^2\eta^2)^h}{h!} \\
&= \exp\left[i\frac{u}{2}\ln(2-2\rho^2)\right]\sum_{h=0}^{\infty}\left[\frac{1}{h!}\frac{\Gamma(\frac{1}{2}+i\frac{u}{2}+h)}{\Gamma(\frac{1}{2}+h)}\exp(-c^2\eta^2)(c^2\eta^2)^h\right].
\end{aligned}$$

□

Lemma 4.

$$E\{\exp(iu\eta)E[\exp(iu\zeta) \mid \eta]\} = \exp\left[i\frac{u}{2}\ln(2-2\rho^2)\right]\sum_{h=0}^{\infty}\frac{1}{h!}\frac{\Gamma(\frac{1}{2}+i\frac{u}{2}+h)}{\Gamma(\frac{1}{2}+h)}I_h(u), \tag{9}$$

where $I_h(u) = \mathrm{E}[(c\eta)^{2h}\exp(-c^2\eta^2)\exp(iu\eta)]$.

Proof. Applying (7), we have,

$$\begin{aligned}
&\mathrm{E}\left\{\exp(iu\eta)\mathrm{E}\left[\exp(iu\zeta \mid \eta)\right]\right\} \\
&= \mathrm{E}\left\{\exp(iu\eta)\exp\left[i\frac{u}{2}\ln(2-2\rho^2)\right]\sum_{h=0}^{\infty}\left[\frac{1}{h!}\frac{\Gamma(\frac{1}{2}+i\frac{u}{2}+h)}{\Gamma(\frac{1}{2}+h)}\exp(-c^2\eta^2)(c^2\eta^2)^h\right]\right\} \\
&= \exp\left[i\frac{u}{2}\ln(2-2\rho^2)\right]\sum_{h=0}^{\infty}\left\{\frac{1}{h!}\frac{\Gamma(\frac{1}{2}+i\frac{u}{2}+h)}{\Gamma(\frac{1}{2}+h)}\mathrm{E}\left[\exp(iu\eta)\exp(-c^2\eta^2)(c^2\eta^2)^h\right]\right\} \\
&= \exp\left[i\frac{u}{2}\ln(2-2\rho^2)\right]\sum_{h=0}^{\infty}\frac{1}{h!}\frac{\Gamma(\frac{1}{2}+i\frac{u}{2}+h)}{\Gamma(\frac{1}{2}+h)}I_h(u).
\end{aligned}$$

□

Finally we turn to the proof of Proposition 1.

Proof. By noting $V_{k-1} \sim N(0, \frac{\sigma^2}{1-\beta_1^2})$, we have

$$\begin{aligned}
\Psi(u) &= \mathrm{E}[\exp(iuX_k)] \\
&= \mathrm{E}\{\exp[iu(\alpha_2 + \alpha_3\ln d_k + \beta_1 V_{k-1} + \eta_k + \zeta_k)]\} \\
&= \exp(iu\alpha_2)\Psi_{ld}(\alpha_3 u)\mathrm{E}\{\exp[iu(\beta_1 V_{k-1})]\}\mathrm{E}\{\exp[iu(\eta_k+\zeta_k)]\} \\
&= \exp(iu\alpha_2)\Psi_{ld}(\alpha_3 u)\exp\left[-\frac{1}{2}\frac{\sigma^2}{1-\beta_1^2}(\beta_1 u)^2\right]\mathrm{E}\{\exp[iu(\eta+\zeta)]\} \\
&= \exp(iu\alpha_2)\Psi_{ld}(\alpha_3 u)\exp(-\frac{1}{2}\frac{\sigma^2}{1-\beta_1^2}\beta_1^2 u^2)\mathrm{E}\{\exp(iu\eta)\mathrm{E}[\exp(iu\zeta\ |)\eta]\}. \qquad (10)
\end{aligned}$$

Plugging (9) into (10), we obtain the marginal characteristic function of X_k as:

$$\begin{aligned}
\Psi(u) &= \mathrm{E}[\exp(iuX_k)] \\
&= \exp\left\{iu\left[\alpha_2 + \frac{1}{2}\ln(2-2\rho^2)\right]\right\}\Psi_{ld}(\alpha_3 u)\exp\left[-\frac{\sigma^2}{2(1-\beta_1^2)}\beta_1^2 u^2\right] \\
&\quad \times \sum_{h=0}^{\infty}\left[\frac{1}{h!}\frac{\Gamma(\frac{1}{2}+i\frac{u}{2}+h)}{\Gamma(\frac{1}{2}+h)}I_h(u)\right].
\end{aligned}$$

$\square$

Now we focus on the calculation of $I_h(u)$. Because $\eta \sim N(0,\sigma^2)$, we have that

$$\begin{aligned} I_h(u) &= \mathrm{E}[(c\eta)^{2h}\exp(-c^2\eta^2)\exp(iu\eta)] \\ &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}\left\{(c\eta)^{2h}\exp\left[(-c^2-\frac{1}{2\sigma^2})\eta^2\right]\cos(u\eta)\right\}d\eta \\ &= \sqrt{\frac{2}{\pi}}\frac{1}{\sigma}\int_0^{\infty}\left[\exp(-a\eta^2)(c\eta)^{2h}\cos(u\eta)\right]d\eta, \end{aligned}$$

where $a = c^2 + \frac{1}{2\sigma^2} = \frac{1}{2\sigma^2(1-\rho^2)}$. To carry the above calculation further, let

$$J_h(u) = \sqrt{\frac{2}{\pi}}\frac{1}{\sigma}\int_0^{\infty}\left[\exp(-a\eta^2)(c\eta)^{2h+1}\sin(u\eta)\right]d\eta,$$

Applying integration by part, when $h \geq 1$, we obtain

$$\begin{aligned} I_h(u) &= \sqrt{\frac{2}{\pi}}\frac{1}{\sigma}\left(-\frac{c}{2a}\right)\int_0^{\infty}[(c\eta)^{2h-1}\cos(u\eta)]d\left[\exp(-a\eta^2)\right] \\ &= \sqrt{\frac{2}{\pi}}\frac{1}{\sigma}\frac{c}{2a}\int_0^{\infty}\exp(-a\eta^2)d[(c\eta)^{2h-1}\cos(u\eta)] \\ &= \frac{(2h-1)c^2}{2a}I_{h-1}(u) - \sqrt{\frac{2}{\pi}}\frac{1}{\sigma}\frac{cu}{2a}\int_0^{\infty}\left[\exp(-a\eta^2)(c\eta)^{2h-1}\sin(u\eta)\right]d\eta \\ &= \frac{(2h-1)\rho^2}{2}I_{h-1}(u) - \frac{|\rho|u}{2\sqrt{a}}J_{h-1}(u). \end{aligned} \tag{11}$$

Similarly,

$$J_h(u) = \rho^2 J_{h-1}(u) + \frac{|\rho|u}{2\sqrt{a}}I_h(u). \tag{12}$$

It is easy to obtain the initial values for the recursive formulas in(11) and (12) with $h=0$ and $h=1$ respectively.

$$I_0(u) = \sqrt{1-\rho^2}\exp(-\frac{1}{4a}u^2), \tag{13}$$

$$J_0(u) = \frac{|\rho|u}{2\sqrt{a}}I_0(u). \tag{14}$$

By direct calculation, we can show that the mean of X_k is given by:

$$
\begin{aligned}
\mathrm{E}(X_k) &= \alpha_2 + \alpha_3 \mathrm{E}(\ln d_k) + \mathrm{E}(\eta_k + \zeta_k) \\
&= \alpha_2 + \alpha_3 \mathrm{E}(\ln d_k) - \tfrac{1}{2}\ln(2) + \tfrac{1}{2}\psi(\tfrac{1}{2}) \\
&\approx \alpha_2 + \alpha_3 \mathrm{E}(\ln d_k) - 0.6354,
\end{aligned}
$$

where $\psi(u) = \frac{d \ln \Gamma(u)}{du}$.

Proofs of Theorem 2 and Proposition 2: Let $m_X(r)$ be the r-th moment of a random variable X, namely $m_X(r) = \mathrm{E}(X^r)$. ¿From the model in (4), the r-th moment of R_k is

$$
\begin{aligned}
m_{R_k}(r) &= \mathrm{E}[\alpha_0 + \alpha_1 d_k + \exp(\alpha_2 + V_k + \alpha_3 \ln d_k)\epsilon_k]^r \\
&= \sum_{\substack{j+h+l=r,\\ 0\le j,h,l\le r}} \frac{r!}{j!h!l!}(\alpha_0)^j \mathrm{E}\{(\alpha_1 d_k)^h [\exp(\alpha_2 + V_k + \alpha_3 \ln d_k)\epsilon_k]^l\} \\
&= \sum_{\substack{j+h+l=r,\\ 0\le j,h,l\le r}} \frac{r!}{j!h!l!}(\alpha_0)^j \alpha_1^h \exp(l\alpha_2) m_{d_k}(h + l\alpha_3) \mathrm{E}[\exp(l\beta_1 V_{k-1})] \mathrm{E}[\exp(l\eta_k)\epsilon_k^l] \\
&= \sum_{\substack{j+h+l=r,\\ 0\le j,h,l\le r}} \frac{r!}{j!h!l!}(\alpha_0)^j (\alpha_1)^h \exp(l\alpha_2) m_d(h + l\alpha_3) \exp\left(\frac{1}{2}\frac{\sigma^2}{1-\beta_1^2} l^2 \beta_1^2\right) \mathrm{E}[\exp(l\eta)\epsilon^l] \qquad (15)
\end{aligned}
$$

(16)

Note that we have removed the subscript k in the last equation above due to stationarity. In the mean time, because $\eta|\epsilon \sim N(\rho\sigma\epsilon, \sigma^2(1-\rho^2))$, we have

$$
\begin{aligned}
\mathrm{E}\{\exp(l\eta)\epsilon^l\} &= \mathrm{E}\{\mathrm{E}[\exp(l\eta)\epsilon^l|\epsilon]\} \\
&= \mathrm{E}\{\epsilon^l \exp\left[l\rho\sigma\epsilon + \frac{1}{2}l^2\sigma^2(1-\rho^2)\right]\} \\
&= \exp\left[\frac{1}{2}l^2\sigma^2(1-\rho^2)\right] \int_{-\infty}^{\infty} \left[x^l \exp(l\rho\sigma x - \frac{1}{2}x^2)\right] dx/\sqrt{2\pi} \\
&= \exp\left(\frac{1}{2}l^2\sigma^2\right) \sum_{j_1=0}^{l} C_l^{j_1} (l\rho\sigma)^{l-j_1} m(j_1), \qquad (17)
\end{aligned}
$$

where $y = x - l\rho\sigma$, and $m(j_1)$ is the j_1-th moment of the standard normal distribution with

$m(1) = m(3) = 0$, $m(2) = 1$, and $m(4) = 3$. In particular,

$$\begin{aligned} & \mathrm{E}\{\exp(l\eta)\epsilon^l\} \\ &= \begin{cases} 1, & l = 0; \\ \rho\sigma \exp(\frac{1}{2}\sigma^2), & l = 1; \\ (1 + 4\rho^2\sigma^2)\exp(2\sigma^2), & l = 2; \\ 9\rho\sigma(1 + 3\rho^2\sigma^2)\exp(\frac{9}{2}\sigma^2), & l = 3; \\ (3 + 96\rho^2\sigma^2 + 4^4\rho^4\sigma^4)\exp(8\sigma^2), & l = 4. \end{cases} \end{aligned}$$

Plugging (17) into (16), unconditional moments of R_k can be expressed as:

$$\begin{aligned} m_R(r) &= \sum_{\substack{j+h+l=r, \\ 0 \le j,h,l \le r}} \frac{r!}{j!h!l!}(\alpha_0)^j(\alpha_1)^h \exp(l\alpha_2) m_d(h + l\alpha_3) \exp\left(\frac{1}{2}\frac{\sigma^2}{1-\beta_1^2}l^2\right) \\ &\quad \times \left[\sum_{j_1=0}^{l} C_l^{(j1)}(l\rho\sigma)^{l-j_1} m(j_1)\right]. \end{aligned} \tag{18}$$

The expectation of R_k is

$$m_R(1) = \alpha_0 + \alpha_1 m_d(1) + \exp(\alpha_2) m_d(\alpha_3)\rho\sigma \exp\left[\frac{\sigma^2}{2(1-\beta_1^2)}\right].$$

So, the proof of Theorem 2 is complete.

For the model in (5), if $\alpha_2 = 0$, (18) leads to

$$\begin{aligned} m_R(1) &= \rho\sigma A, \\ m_R(2) &= (1 + 4\rho^2\sigma^2)A^4, \\ m_R(3) &= 9\rho\sigma(1 + 3\rho^2\sigma^2)A^9, \\ m_R(4) &= [3 + 6(4\rho\sigma)^2 + (4\rho\sigma)^4]A^{16}, \end{aligned}$$

where $A = \exp\left[\frac{\sigma^2}{2(1-\beta_1^2)}\right]$. Therefore, the variance of R_k is

$$\begin{aligned} \mathrm{var}(R_k) &= m_R(2) - m_R(1)^2 \\ &= (1 + 4\rho^2\sigma^2)A^4 - (\rho\sigma A)^2 \\ &= (\rho\sigma A)^2(4A^2 - 1) + A^4. \end{aligned}$$

The third central moment, $M_R(3)$, of R_k is

$$\begin{aligned}
M_R(3) &= m_R(3) - 3m_R(1)m_R(2) + 2m_R(1)^3 \\
&= 9\rho\sigma(1+3\rho^2\sigma^2)A^9 - 3(1+4\rho^2\sigma^2)A^4\rho\sigma A + 2(\rho\sigma A)^3 \\
&= \rho\sigma A^3[3A^2(3A^4-1) + 3(\rho\sigma A)^2(9A^4-4) + 2\rho^2\sigma^2],
\end{aligned}$$

and the fourth central moment, $M_R(4)$, of R_k is

$$\begin{aligned}
M_R(4) &= m_R(4) - 4m_R(1)m_R(3) + 6m_R(1)^2m_R(2) - 3m_R(1)^4 \\
&= [3 + 6(4\rho\sigma)^2 + (4\rho\sigma)^4]A^{16} - 4\rho\sigma A[9\rho\sigma(1+3\rho^2\sigma^2)A^9] \\
&\quad +6(\rho\sigma A)^2(1+4\rho^2\sigma^2)A^4 - 3(\rho\sigma A)^4 \\
&= 3A^{16} + 96(\rho\sigma)^2A^{16} + (4\rho\sigma)^4A^{16} - 36(\rho\sigma)^2A^{10} - 108(\rho\sigma)^4A^{10} \\
&\quad +6(\rho\sigma)^2A^6 + 24(\rho\sigma)^4A^6 - 3(\rho\sigma A)^4 \\
&= 3A^{16} + (\rho\sigma)^2A^8(96A^8 - 36A^2) + (\rho\sigma)^4A^8(4^4A^8 - 108A^2) \\
&\quad +6(\rho\sigma)^2A^6 + (\rho\sigma)^4A^4(24A^2 - 3).
\end{aligned}$$

Because $A > 1$, we have that $A^8 > 1$, $96A^8 - 36A^2 > 24$, $4^4A^8 - 108A^2 > 48$, and $24A^2 - 3 > 3$. Hence,

$$\begin{aligned}
M_R(4) &\geq 3A^8 + 24(\rho\sigma)^2A^8 + 48(\rho\sigma)^4A^8 - 3(\rho\sigma A)^4 \\
&\quad +3(\rho\sigma)^4A^4 - 6(\rho\sigma)^2A^6 - 24(\rho\sigma)^4A^6 \\
&= 3[A^4 + 4\rho^2\sigma^2A^4 - \rho^2\sigma^2A^2]^2 \\
&= 3[\mathrm{var}(R)]^2,
\end{aligned}$$

which means that the kurtosis of R_k is larger than 3. It is interesting to know that when $\rho = 0$, the kurtosis equals $3A^8$, which is greater than 3. This concludes the proof of Proposition 2.

Tables and Figures

Table 1: Simulation result based on 1,000 replications for Model A with $\rho = 0$.

No.	α_2		α_3		β_1		σ	
	True	Mean	True	Mean	True	Mean	True	Mean
1	0.3	0.32	0.6	0.60	0.3	0.32	1.2	1.16
		(0.029)		(0.029)		(0.020)		(0.029)
2	−0.1	−0.09	0.6	0.58	−0.1	−0.07	1.2	1.18
		(0.022)		(0.025)		(0.019)		(0.030)
3	0.3	0.30	0.9	0.90	−0.1	−0.06	0.9	0.90
		(0.016)		(0.019)		(0.027)		(0.017)
4	0.6	0.60	0.3	0.30	−0.3	−0.27	0.9	0.91
		(0.015)		(0.018)		(0.024)		(0.015)
5	−0.3	−0.19	0.6	0.51	0.1	0.11	1.6	1.38
		(0.096)		(0.081)		(0.022)		(0.201)
6	0.1	0.21	0.9	0.71	−0.1	−0.06	1.6	1.30
		(0.100)		(0.136)		(0.020)		(0.293)
7	0.3	0.31	0.9	0.90	0.1	0.14	0.9	0.89
		(0.016)		(0.020)		(0.027)		(0.016)
8	0.1	0.11	0.6	0.59	0.1	0.12	1.2	1.17
		(0.023)		(0.026)		(0.019)		(0.027)

Table 2: Simulation results based on 1,000 replications for Model B.

No.	α_2		α_3		β_1		σ		ρ	
	True	Mean	True	Mean	True	estimate	True	Mean	True	Mean
1	0.3	0.28	0.6	0.59)	0.3	0.32	0.9	0.89	-0.1	-0.08
		(0.017)		(0.019)		(0.024)		(0.016)		(0.018)
2	0.3	0.28	0.6	0.59	0.1	0.12	1.2	1.20	-0.4	-0.36
		(0.020)		(0.022)		(0.018)		(0.015)		(0.016)
3	-0.3	-0.29	0.3	0.33	-0.3	-0.29	0.9	0.91	0.9	0.90
		(0.015)		(0.020)		(0.025)		(0.017)		(0.079)
4	0.1	0.08	0.9	0.89	0.3	0.31	1.2	1.20	0.1	0.12
		(0.021)		(0.021)		(0.018)		(0.017)		(0.019)
5	0.3	0.29	0.9	0.90	-0.1	-0.07	0.9	0.91	0.1	0.12
		(0.015)		(0.019)		(0.026)		(0.015)		(0.020)
6	-0.3	-0.31	0.6	0.61	0.1	0.11	1.2	1.20	0.4	0.41
		(0.021)		(0.022)		(0.018)		(0.016)		(0.017)
7	-0.3	-0.31	0.9	0.88	-0.1	-0.09	1.2	1.20	-0.9	-0.89
		(0.018)		(0.022)		(0.023)		(0.018)		(0.029)
8	0.3	0.28	0.9	0.90	0.1	0.12	1.2	1.20	0.0	0 .03
		(0.018)		(0.021)		(0.018)		(0.016)		(0.018)

Table 3: Simulation results based on 100 replications for the complete model in (4).

No.		α_0	α_1	α_2	α_3	β_1	σ	ρ
1	True	0.5	1	0.6	1	0.2	0.9	0.3
	Mean	1.12	0.79	0.63	1.00	0.22	0.86	-0.26
	SE	0.40	0.25	0.06	0.05	0.03	0.06	0.06
2	True	0.4	0.6	0.3	1	0.2	0.6	-0.2
	Mean	0.57	0.54	0.30	1.00	0.26	0.59	-0.18
	SE	0.19	0.12	0.02	0.02	0.05	0.02	0.09
3	True	0.0	1.5	1.0	0.5	-0.2	0.6	0.9
	Mean	-0.15	1.74	1.19	0.25	-0.11	0.61	0.79
	SE	0.11	0.18	0.05	0.07	0.05	0.08	0.14
4	True	1.2	1	0.6	0.6	0.1	1.2	-0.1
	Mean	1.36	1.02	0.60	0.62	0.12	1.18	-0.09
	SE	0.23	0.16	0.03	0.03	0.02	0.03	0.02

Note: 'SE.' stands for the estimated standard error.

Table 4: Illustrative fragment of IBM stock trading data.

Day	Hour	Min	Sec.	Price	Bid	Ask
1	10	57	24	98.60	98.51	98.60
1	10	57	24	98.60	98.50	98.60
1	10	57	24	98.61	98.50	98.61
1	10	57	24	98.61	98.50	.
	⋮		⋮		⋮	

Figure 1: Histograms of 100 simulation estimates for α_3,β_1, σ and ρ.

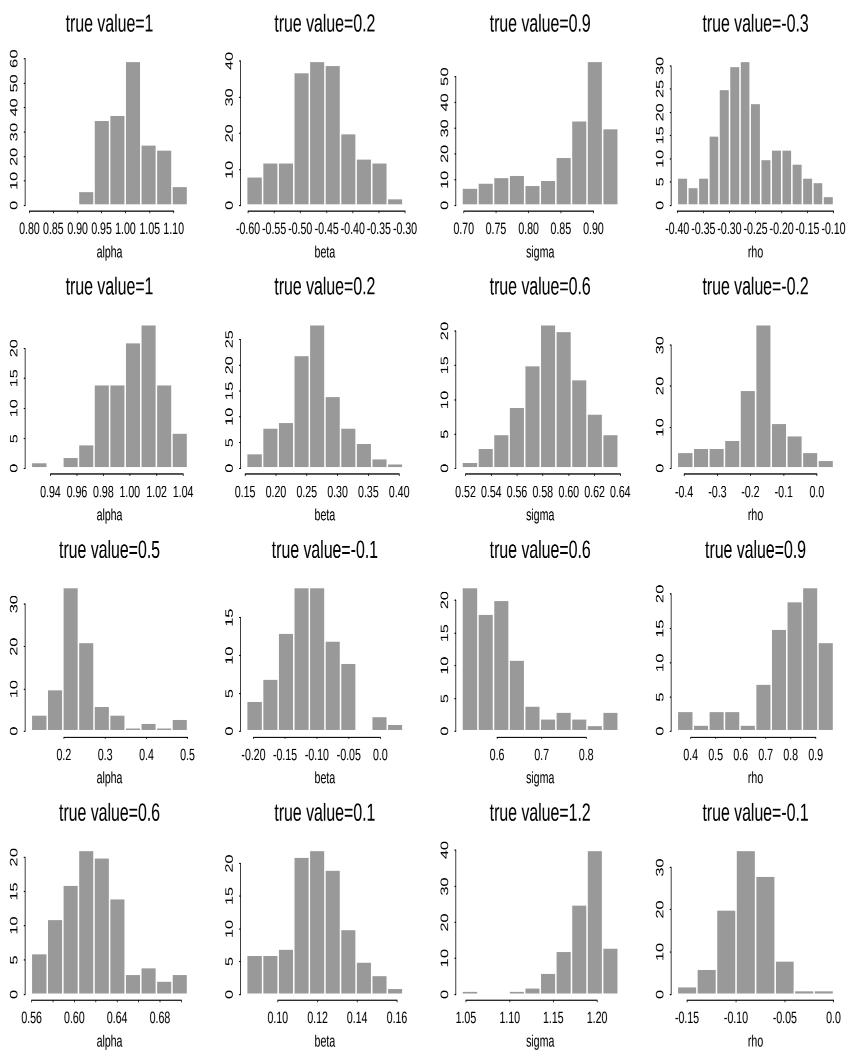

Figure 2: Time-of-day effect of trade duration.

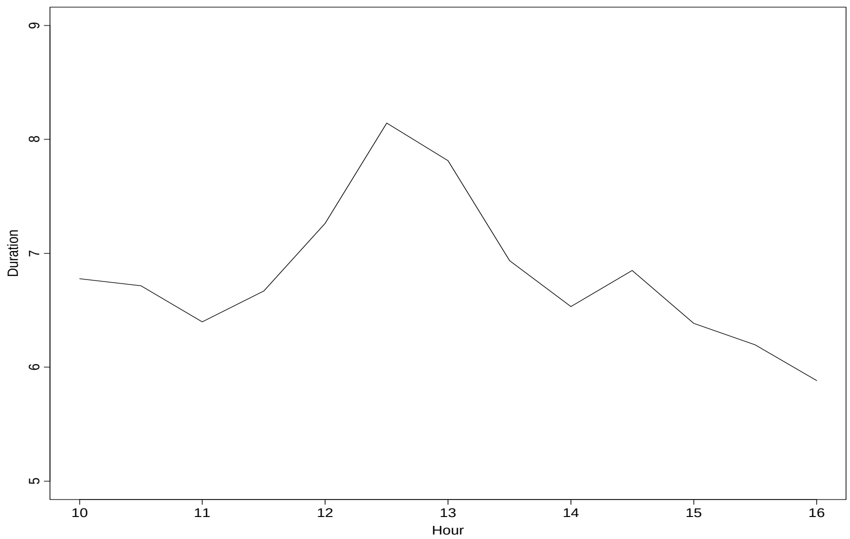

Table 5: Summary statistics of duration, adjusted duration and returns.

Data	Min.	Median	Mean	Max.	Std. Dev.	Skewness	Kurtosis
Duration	1	5	6.226	67	5.47	2.67	12.17
Adj. duration	0.12	0.65	0.9	8.98	0.78	2.56	10.91
Return	-66.63	0	0	71.26	4.13	0.26	28.47

Note: The unit of Duration and Adj. (adjusted) duration is in second, while the unit of return is 10^{-5} dollar.

Figure 3: Trade duration.

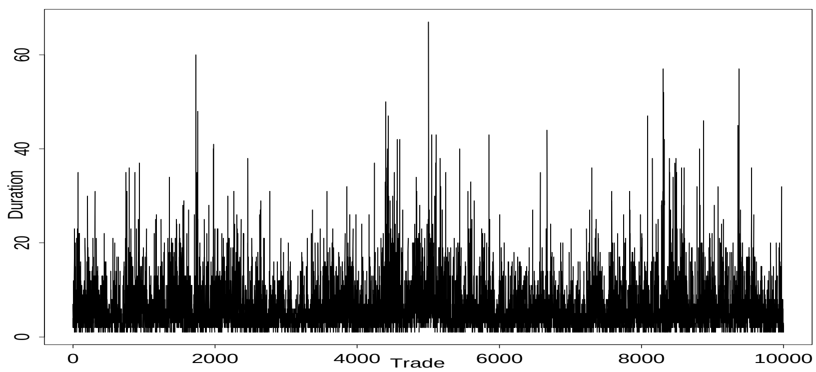

Figure 4: Trade returns.

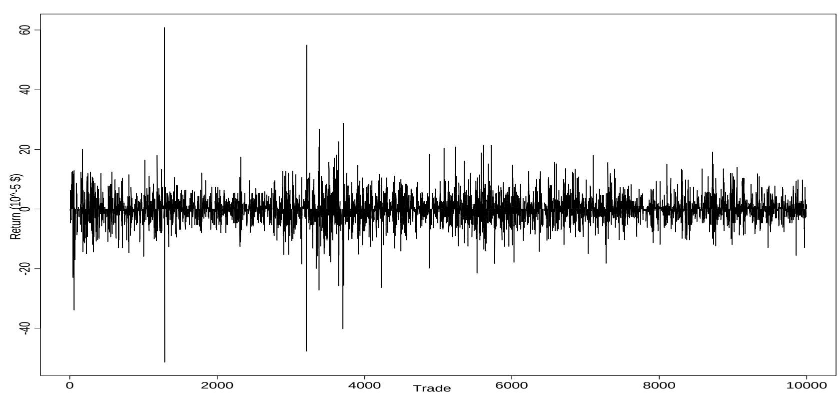

Figure 5: ACF of adjusted trade duration series.

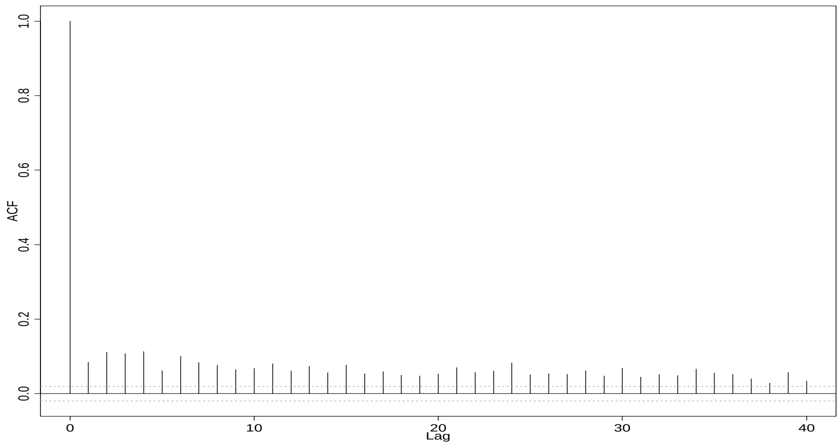

Table 6: IBM stock data analysis:
Estimates and their standard errors.

Parameter	Estimate	Std. err.	Lower 95% C.I.	Upper 95% C.I
α_0	0.584	0.002	0.580	0.588
α_1	1.148	0.019	1.110	1.186
α_2	0.645	0.024	0.597	0.693
α_3	0.998	0.030	0.938	1.058
β_1	0.207	0.117	−0.090	0.324
σ	0.788	0.024	0.740	0.836
ρ	−0.251	0.037	−0.325	−0.177

Figure 6: PACF of adjusted trade duration series.

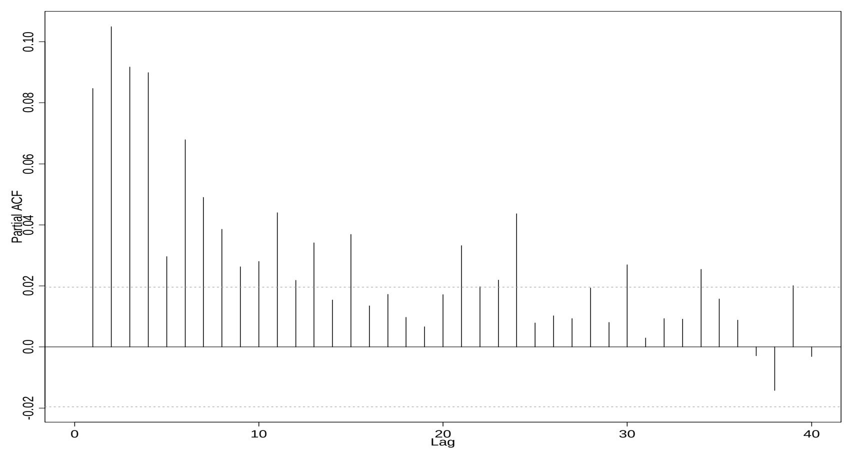

Figure 7: ACF of trade return series.

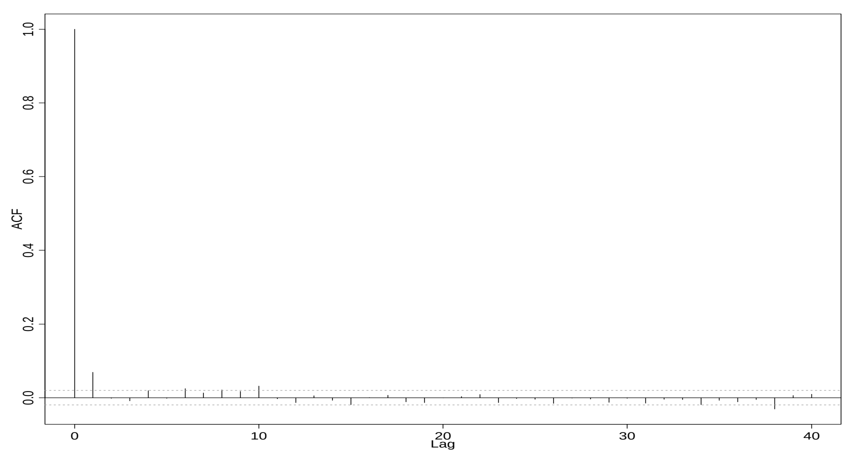

Figure 8: PACF of trade return series.

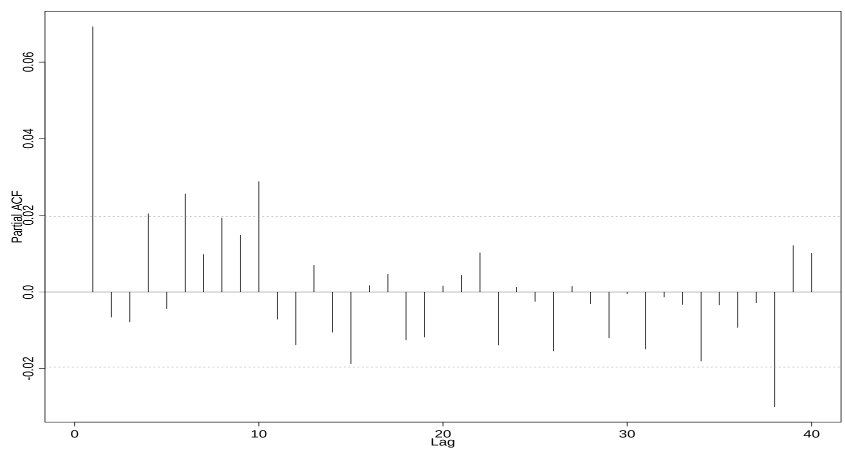

Figure 9: ACF of the absolute trade returns series.

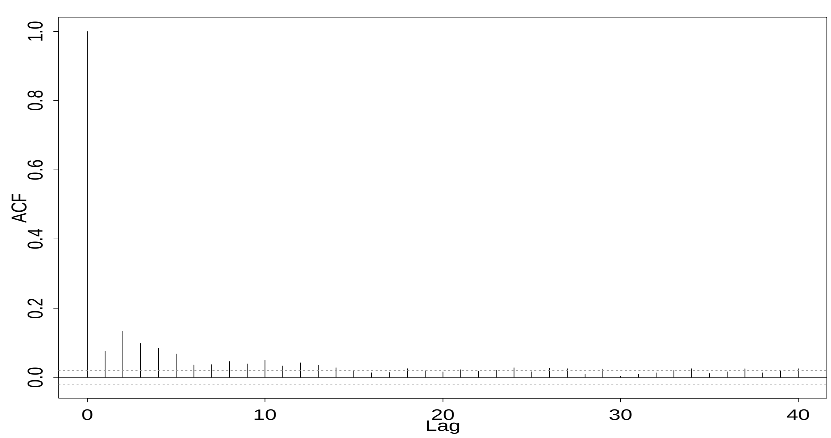

Figure 10: CC function of durations and returns at leads.

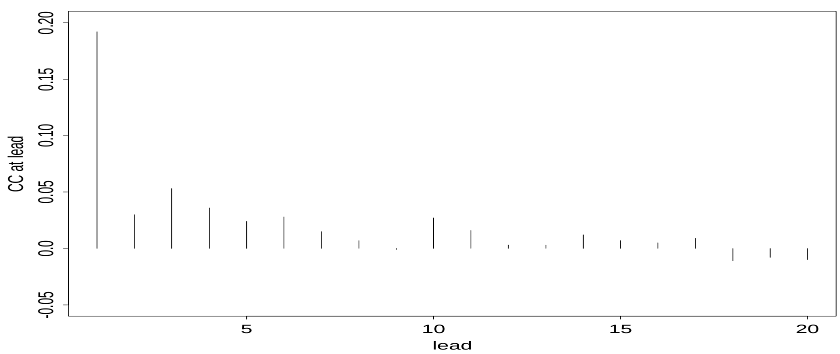

Figure 11: The difference of CC functions between leads and lags.

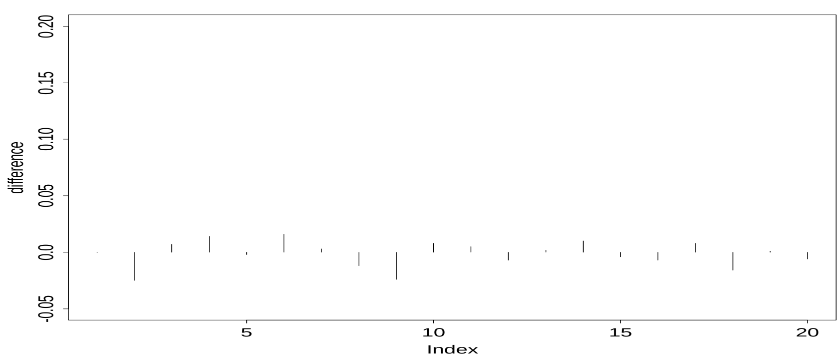

Figure 12: Empirical and implied density functions with β_1.

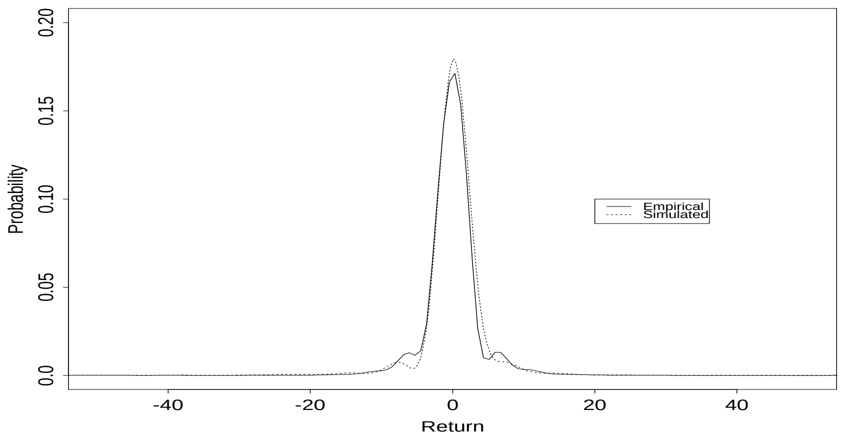

Figure 13: Empirical and implied density functions without β_1.