VII.2

Tracing the Source of Long Memory in Volatility

Rohit Deo* Mengchen Hsieh Clifford M. Hurvich

New York University

April 28, 2005

Abstract

We study the effects of trade duration properties on dependence in counts (number of transactions in equally-spaced intervals of clock time) and thus on dependence in volatility of returns. Long memory in counts may arise either from infinite variance in durations or from long memory in durations. The Autoregressive Conditional Duration (ACD) model can produce infinite variance in durations under certain parameter configurations, but such configurations did not arise in the actual trade duration data we analyzed. We introduce a long-memory stochastic duration (LMSD) model, which is closely related to the long-memory stochastic volatility model. The LMSD model seems to describe the actual data better than the ACD model, thereby providing some support for the idea that long memory in counts (and therefore in volatility) arises from long memory in durations.

JEL: C13, C14, C41

*8-57 KMC, 44 West 4th St. New York, NY 10012, TEL: 212-998-0469, email: rdeo@stern.nyu.edu.

It is widely accepted that volatility of financial returns has a high level of persistence, and some have argued that it has long memory (see, for example, Andersen, Bollerslev, Diebold and Labys 2001). There is also a strand of literature that argues that long memory may be an illusion caused by breaks in the model parameters (see, for example, Mikosch and Stărică 2003). In general, distinguishing between breaks and long memory is hard, since either of these properties can masquerade as the other. Nevertheless, the types of breaks considered by Mikosch and Stărică (2003) induce a unit root rather than a fractional unit root in volatility. Furthermore, the relationship between breaks and long memory may actually be a duality rather than a dichotomy since there are processes, useful for modeling volatility, that genuinely and naturally have both features. See Liu (2000), Parke (1999). For these processes, the breaks and long memory are inextricably intertwined. In this paper, we will thus assume that volatility does indeed possess long memory.

Most studies have considered measures of volatility computed using data observed at equally spaced time intervals, i.e., *clock time*. With the availability of transaction-level data in *tick time*, it may be possible to obtain a better understanding of volatility beyond that obtained by existing econometric models for clock-time data. A seminal paper in this regard is Engle and Russell (1998) which presents a model for durations, i.e., the time elapsed between events. Other papers on dependence in durations include Bauwens and Veredas (2003), Ghysels and Jasiak (1997), Grammig and Wellner (2002), and Engle and Sun (2003).

It seems plausible that the dependence in durations is intimately connected to the dependence in volatility. The dependence in durations will affect the dependence in the number of transactions within any fixed clock-time interval (the *counts*). Furthermore, dependence in the counts has long been understood to have a connection to dependence in the volatility of returns. This idea builds on the seminal work of Clark (1976), who modeled asset prices as Brownian motion subordinated to a non-decreasing positive stochastic process which could be considered as a proxy for the cumulative counts (the *counting process*), or for the volume. Clark's model implies that any autocorrelation in the increments of the subordinating process yields autocorrelation in the volatility of returns. Thus, either counts or volume may be thought of as proxies for volatility, though the empirical literature seems to suggest that the counts are preferable. See Jones, Koul and Lipson (1994) and Ané and Geman (2000).

In this paper, we study the effects of duration properties on dependence in counts and thus on dependence in volatility. More specifically, we will focus on determining which properties of durations result in long memory in volatility. Through a combination of theoretical and empirical analysis we find that either long tails or long memory in durations is sufficient for generating long memory in counts, and therefore in volatility. We will present a data and simulation analysis of durations, counts and volatility in an attempt to reconcile the theoretical results currently available with the stylized facts for volatility.

An explicit model for returns is needed for establishing a link between counts and volatility, though not for the link between durations and counts. In Section I, we present a return model, and briefly investigate its appropriateness. In Section II, we present existing theorems showing that infinite variance in durations results in long memory in counts. We then provide a new result which shows that under the Long-Memory Stochastic Duration Model which we introduce here, long memory in durations induces long memory in counts, with at least the same memory parameter. In Section III, we describe our data on durations between transactions on various stocks, and carry out some preliminary exploratory data analysis. In Section IV, we investigate the presence of long memory in durations, counts and volatility. In Section V, we fit some popular duration models to the data, investigate their adequacy, and state their implications for volatility by appealing to the theory from Section II.

I Return Model

A model that explicitly incorporates the impact of counts on volatility, inspired by the work of Clark (1976) is

$$\log P(t) = B(N(t))$$

where $P(t)$ is the price process, $B(\cdot)$ is Brownian motion and $N(\cdot)$ is a counting process with stationary increments, representing the number of events in $(0, t]$, which is independent of B. This model implies that the returns $r_{t'}$ at equally-spaced clock time intervals of width $\Delta t > 0$ may be expressed as

$$r_{t'} = \sqrt{\Delta N_{t'}}\, e_{t'} \quad , \quad t' = 1, 2, 3, \ldots \quad , \tag{1}$$

where the $\{e_{t'}\}$ are *i.i.d.* normal with zero mean, and are independent of $\{\Delta N_{t'}\}$ where $\Delta N_{t'} = N(t'\Delta t) - N[(t'-1)\Delta t]$. In this model, the squared returns $\{r_{t'}^2\}$ and the counts $\{\Delta N_{t'}\}$ have the same autocorrelations. Thus, in order to capture persistence in squared returns it is necessary in this model to allow for autocorrelation in the counts.

We can easily check the adequacy of the model in Equation (1) since data on trade counts and returns are available. One can simply examine the autocorrelations of the normalized returns $r_{t'}/\sqrt{\Delta N_{t'}}$ and their squares as well as a normal probability plot of the normalized returns. In Figures 3 and 4, we present the results for data on IBM at five-minute and thirty-minute intervals, for the year 2002. Following Ané and Geman (2000), the returns we used here were actually residuals from an $AR(10)$ process fitted to the raw returns, and the counts include transactions that had no associated price change. In the normal probability plots of Figure 3 we removed two extreme outliers to enhance the readability of the plots. Our results, which also generalize to other stocks that we considered, clearly indicate violation of both normality and independence of the normalized returns. The violation of normality evident in Figure 3 is somewhat surprising, since it seems to contradict the findings of Ané and Geman (2000). The normalized returns appear to be uncorrelated, while the squared normalized returns indicate that the $\{e_{t'}\}$ possess conditional heteroscedasticity.

We therefore modify the assumptions on $\{e_{t'}\}$ in (1), and assume now the $GARCH(1,1)$ model

$$e_{t'}|\Psi_{t'-1} \sim N(0, h_{t'}) \tag{2}$$

$$h_{t'} = \tilde{\omega} + \tilde{\alpha} e_{t'-1}^2 + \tilde{\beta} h_{t'-1} \tag{3}$$

where $\Psi_{t'-1}$ is the information set at time $t'-1$, $\tilde{\omega} > 0$, $\tilde{\alpha} > 0$, $\tilde{\beta} \geq 0$ and $\tilde{\alpha} + \tilde{\beta} < 1$.

Under this modified return model with GARCH innovations, we prove in the Appendix that the lag-k autocorrelation of the squared returns is given by

$$\text{corr}(r_{t'}^2, r_{t'+k}^2) = [m_1 + m_2(\tilde{\alpha} + \tilde{\beta})^k]\, \rho_{\Delta N,k} + m_3(\tilde{\alpha} + \tilde{\beta})^k \tag{4}$$

where $\rho_{\Delta N,k}$ is the lag-k autocorrelation of the counts and $m_1 \in (0,1)$, m_2 and m_3 are constants for any fixed value of Δt. Thus, if the autocorrelations of counts decay exponentially, then so do the autocorrelations of the squared returns. Furthermore, if the autocorrelations of counts have power-law decay, i.e., $\rho_{\Delta N,k} \sim c_1 k^{-\theta}$ where $0 < \theta < 1$ and $c_1 \neq 0$, then

$$\text{corr}(r_{t'}^2, r_{t'+k}^2) \sim m_1 \rho_{\Delta N,k} \qquad \text{as } k \to \infty, \tag{5}$$

so the squared returns have the same long-term persistence as the counts do.

In the Appendix, we show that $0 < m_1 < 1$ and hence when the counts have long memory the autocorrelations of the squared returns are ultimately proportional to those of the counts, with a proportionality constant that is less than 1. Thus, squared returns are at best a noisy proxy for volatility. This is in line with existing literature. See, e.g. Andersen and Bollerslev (1998).

It is important to note that if the squared returns have long memory then the realized volatility series constructed from squared returns at any fixed interval Δt will inherit long memory with the same memory parameter.

Since the counting process is constructed from the duration process, the statistical properties of the duration process completely determine those of the counting process, which, together with a return model such as our modified version of (1), completely determines the statistical properties of return volatility in clock time. In the remainder of the paper, we investigate the question as to which statistical properties of the durations will generate counts with long memory. We start by presenting some existing relevant results.

II Conditions on Durations to Give Long Memory in Counts

The following theorems are from Daley (1999) and Daley, Rolski and Vesilo (2000). Relevant definitions are given in the Appendix. In the theorems, C denotes a generic positive constant.

Theorem 1 *If the durations $\{\tau_k\}$ are i.i.d. with tail index $\kappa \in (1,2)$ then $VarN(t) \sim Ct^{1+2d}$ as $t \to \infty$, where $d = 1 - \kappa/2$.*

Theorem 2 *If the durations $\{\tau_k\}$ are stationary with tail index $\kappa \in (1,2)$ under the Palm probability measure (see Appendix), then $VarN(t) \sim Ct^{1+2d}$ as $t \to \infty$ where $1/2 > d \geq 1-\kappa/2$.*

These two results establish that infinite variance of the durations is sufficient to guarantee that the counting process has a variance which scales as a power of t, where the power is greater than 1, as happens for partial sums of a long memory process. Under the assumptions of the theorems, counts cannot have exponentially decaying autocorrelations. Furthermore, if the counts are assumed to have power law decay in their autocorrelations, then the theorems imply that the counts must have long memory, with memory parameter d, *i.e.*, they have lag-h autocorrelation $\rho_h \sim Ch^{2d-1}$ as $h \to \infty$.

Note that in the theorems above it is the tail behavior of the durations, and not their dependence structure, that results in long memory in the counts. This prompts the question as to whether long memory in the counts can be generated solely by dependence in finite-variance durations. An answer in the affirmative was given by Daley et. al. (2000), who provide an example outside of the framework of the popular econometric models we will consider here. We next state a theorem of more direct relevance for a duration model used in Econometrics:

Theorem 3 *Suppose the durations are given by* $\tau_k = e^{\psi_k}\epsilon_k$ *where* $\{\epsilon_k\}$ *is a positive i.i.d. process with all moments finite and let* $\{\psi_k\}$ *be a stationary Gaussian process independent of* $\{\epsilon_k\}$*. If* $\{\psi_k\}$ *is a short-memory process, then* $VarN(t) \sim Ct$ *as* $t \to \infty$*. If* $\{\psi_k\}$ *is a long-memory process with memory parameter* $d_\tau \in (0, 1/2)$*, and if the short-memory component of the count spectral density* $f_{\Delta N}$ *is continuous, then* $VarN(t) \sim Ct^{1+2d_{\Delta N}}$ *as* $t \to \infty$*, and*

$$f_{\Delta N}(\lambda) \sim C\lambda^{-2d_{\Delta N}} \text{ as } \lambda \to 0,$$

where $d_{\Delta N} \geq d_\tau$*.*

We provide some relevant simulations in Part B of Section V, which suggest that in fact $d_{\Delta N} = d_\tau$, i.e., the memory parameter of the durations propagates unchanged to the memory parameter of the counts. Note that Theorem 3 provides a relevant example where the absence or presence of long memory in durations implies the absence or presence of long memory in counts.

In the sequel, we study empirical durations and counts in the light of the theorems presented above. We start with a description of the data and an exploratory analysis.

III Data Description

We study tick-by-tick trade data from the NYSE TAQ (Trade And Quote) database. Standard filtering rules (See Saar, Yu and Boehmer) were applied to the trade data. [1] The trade duration here is defined as the waiting time between two consecutive trades. We also performed

[1] We used NYSE trades exclusively for which the TAQ field CORR ("Correction Indicator") is 0 or 1. We also used trades for which the TAQ field "Sales Condition" is blank or 0. We exclude trades with negative prices and trades which are less than 50% or greater than 150% of the previous prices. We also exclude trading days which had fewer than 6.5 regular trading hours. For the year 2002, we removed four trading days from our sample: July 5, September 11, November 29 and December 24, since the market was open for only a half day. It is not uncommon to observe multiple trades at different prices at the same time. In this case, we computed a weighted average of the trade price with weights given by the trade volume to avoid zero durations.

an analysis, not reported here, on durations defined as waiting times between changes in the mid-quote price, and obtained nearly equivalent results. We omitted all overnight returns and durations from the analysis. We considered two subsets of stocks traded on the NYSE for the year 2002: those with more than 400000 transactions ("active stocks") included in the Dow Jones Industrial Average and those with fewer than 100000 transactions ("less active stocks"). We randomly selected five active stocks: American International Group (AIG), American Express (AXP), Boeing (BA), Coca-Cola (KO) and IBM and five less active stocks: Beverly Enterprize (BEV), Bally Total Fitness (BFT), CBL Associates Properties (CBL), Commercial Federal Corporation (CFB) and SONY (SNE).

Usually one observes active trading during opening and closing hours and dormant trading around noon. This is reflected by short durations during active hours and longer durations around noon in a trading day. To remove the seasonal effects present in the duration process in a manner that is robust to outliers, we applied a semi-logarithmic version of the method suggested by Engle and Russell (1998) to get the diurnally-adjusted durations. Let τ_k denote the trade duration and $\hat{\tau_k}$ the diurnally-adjusted durations. Then we construct

$$\hat{\tau_k} = \exp[\log(\tau_k) - \hat{\phi}(t_k)] \tag{6}$$

where $\hat{\phi}(t_k)$ is the fitted value from the regression

$$\begin{aligned} \log(\tau_k) &= \sum_{j=1}^{13} \beta_j x_{j,k} + u_k \\ x_{j,k} &= \max(t_k - q_j, 0) \end{aligned}$$

where t_k denotes the event time such that $\tau_k = t_k - t_{k-1}$ and q_j denotes the jth partition endpoint. Since there are 6.5 trading hours per trading day, we partition the total number of trading hours into 13 intervals, each with a length of 30 minutes. For example, q_1 is the first endpoint, 9:30 AM, and q_{13} is the last endpoint, 3:30 PM. From now on, we use the term *duration* to denote the diurnally-adjusted durations and all of the subsequent data analysis is based on the diurnally-adjusted durations unless stated otherwise.

IV Preliminary Data Analysis for Durations and Counts

In Table 1 we report the summary statistics for the logged durations. Our decision to take logs here is based on the extreme right-skewness of the non-logged durations. It is seen from the table that all of the active stocks except IBM have right-skewed log durations, while all of the less active stocks have left-skewed log durations.

The histograms of the logged durations for the active and less active stocks are presented in Figures 5 and 6 respectively. The histograms for the active stocks look different from those for the less active ones. Both sets of histograms show some evidence of bimodality.

We now investigate the serial dependence in durations and counts, where for counts we consider three values of Δt: 5 minutes, 30 minutes and 1 day. The counts were constructed from the non diurnally adjusted durations. In Figure 7 we present the ACF plots for durations, logged durations and counts of the AIG data. The ACF for both durations and counts indicates considerable persistence, and the ACF of the counts suggests seasonality. In light of the evidence of persistence, we estimate the memory parameter d for log durations and (non-logged) counts. We use log durations here to mitigate the effects of outliers and also since our parametric long memory models in Section V will be estimated directly from log durations.

A semiparametric estimator $\hat{d}_{GPH}$ was proposed by Geweke and Porter-Hudak (1983). To motivate this estimator, we note that for a long-memory time series $\{X_t\}$ the spectral density $f(\omega)$ has power-law decay, $f(\omega) \sim A\omega^{-2d}$ as $\omega \to 0$. The GPH estimator $\hat{d}_{GPH}$ is based on an OLS regression of the log periodogram $\{\log I_j\}_{j=L}^{m}$ on $\{\log \omega_j\}_{j=L}^{m}$, where $\omega_j = 2\pi j/n$, $j = 0, \ldots, n-1$ are the Fourier frequencies,

$$I(\omega_j) = \frac{1}{2\pi n} |\sum_{t=1}^{n} X_t \exp(-i\omega_j t)|^2$$

is the periodogram, L and m are positive integers, and n is the number of observations. The GPH estimator $\hat{d}_{GPH}$ then is given by $-1/2$ times the least-squares slope estimate in the OLS regression.

The statistical properties of $\hat{d}_{GPH}$ and the choice of L and m for Gaussian long-memory time series have been discussed in recent literature. Robinson (1995) showed that for Gaussian processes the GPH estimator is $m^{1/2}$-consistent and asymptotically normal if $m^{-1} + m/n \to 0$, and a suitably increasing number of low frequencies L is trimmed. Hurvich, Deo and Brodsky (1998) showed that trimming can be avoided for Gaussian processes, and under some additional conditions the optimal m is $O(n^{.8})$. In our analysis, we choose $L = 1$, and since we are working with only ten series, which are possibly non-Gaussian, we prefer to choose m by inspecting $\log-\log$ periodogram plots rather than by using an automatic procedure, such as the one proposed by Hurvich and Deo (1999). Figure 8 presents the log-log periodogram plots of the counts for our three choices of Δt, for the AIG series. Based on visual inspection of the plots, we chose the values of j which yielded m for the three values of Δt to be $m = [n^{.576}]$, $m = [n^{.69}]$ and $m = [n^{.72}]$ for Δt equal to 5 minutes, 30 minutes and 1 day, respectively. Corresponding plots for the other series were sufficiently similar to what we found for AIG that we have used these same choices of m for all of the stocks. In addition to our visually selected m, we also report estimates using $m = [n^{.8}]$. Note that the GPH estimator is unaffected by seasonality in counts since the estimator uses only Fourier frequencies in a neighborhood of zero.

The values of $\hat{d}_{GPH}$ for the counts are presented in Table 3. The estimated memory parameter in all cases is significantly greater than zero. For the visually chosen values of m, the estimated memory parameter for Δt of 5 minutes and 30 minutes is quite similar for all stocks, and the estimate for Δt of 1 day is almost invariably higher, though not significantly so since the standard error is more than doubled. Thus, our results here suggest that the memory parameter for counts is invariant to Δt, as would be implied by the mechanisms we described in Section 3 for generating long memory in counts. At this stage, then, it seems that either long tails or long memory in durations may be the source of long memory in the observed counts. Next, we investigate whether durations have long memory.

In Table 2, we present GPH estimates of d for the log durations, using $m = [n^{.4}]$ and $m = [n^{.5}]$, as suggested by visual inspection of the log − log periodogram plots. From Deo and Hurvich (2001), the GPH estimator will be consistent and asymptotically normal if the durations follow a long-memory stochastic duration model as defined in Section V. As Table 2 shows, all the estimated values of d are significantly greater than zero, and are not significantly different from the corresponding estimates for the counts. This is consistent with the paradigm implied by Theorem 3, under which long memory in durations leads to long memory in counts with at least the same memory parameter.

In Table 4 we report GPH estimates of d for the squared returns on intervals Δt of 5 minutes, 30 minutes and 1 day. The values of m used were the same as those used for the counts. In general, the estimates of d for the squared returns tend to be smaller than the corresponding estimates for the counts. This apparent phenomenon may be illusory, in keeping with the general observation that squared returns are a noisy proxy for volatility. We feel that our findings here do not contradict the hypothesis that the memory parameter for squared returns is the same as that for durations and counts.

In Table 5, for each of the ten stocks, we report GPH estimates of d for the 1 day log realized volatility series, computed as the log of the sum of squares of the intra-day five-minute returns, omitting the overnight return. In general, the estimates of d based on realized volatility are larger than those based on either five-minute squared returns or five-minute counts. However, the standard errors for the estimates based on realized volatility are very large due to the reduced sample size. Overall, there seems to be no statistically significant difference between the estimates based on the three volatility proxies. Nevertheless, the fact that realized volatility yields larger estimates of d than squared returns is consistent with the idea that realized volatility is the less noisy volatility proxy.

Our semiparametric analysis indicates evidence of long memory in durations, logged durations, counts, squared returns and realized volatility. Note, though, that from the theorems in Section II, infinite variance in durations can also lead to long memory in counts. It would thus be of interest to nonparametrically estimate the tail index of the durations. However, it is generally accepted (see Resnick (1997)) that the standard method, the Hill (1975) estimator performs very poorly even when the data are independent. Presumably, the performance of the Hill estimator will deteriorate even further under strong dependence of durations, as we indeed found in simulations not shown here. We thus prefer to investigate tail behavior of durations through fully parametric models.

We now proceed to estimate a set of parametric models for durations and investigate the implications of the fitted models for the propagation of long memory from durations to volatility through counts.

V Parametric Models for Durations

A seminal econometric model for durations is the Autoregressive Conditional Duration (ACD) model of Engle and Russell (1998). The simplest version is the $ACD(1,1)$ model, defined by

$$\tau_k = \psi_k \epsilon_k \tag{7}$$

$$\psi_k = \omega + \alpha \tau_{k-1} + \beta \psi_{k-1} \quad , \tag{8}$$

where $\omega > 0$, $\alpha \in (0,1)$, $\beta \in [0,1)$, $\alpha + \beta < 1$, and $\{\epsilon_k\}$ are *iid* random variables with unit mean and positive support. This model is closely related to the $GARCH(1,1)$ model (Bollerslev (1986)), widely used for clock-time returns. When the $\{\epsilon_k\}$ are unit exponential, the model is called the Exponential ACD, denoted by EACD(1,1). When the $\{\epsilon_k\}$ are Weibull, the model is called the Weibull ACD, denoted by WACD(1,1). Note that in the ACD model, ψ_k is the conditional mean of the durations, explicitly modeled through the past durations. Thus, the ACD model is observation driven.

An alternative duration model is Bauwens and Veredas' (2004) Stochastic Duration (SCD), which is driven by a latent series. The model is given by

$$\tau_k = e^{\psi_k} \epsilon_k \tag{9}$$

$$\psi_k = \omega + \beta \psi_{k-1} + u_k \quad , \tag{10}$$

where $\omega \in \mathbb{R}$, $|\beta| < 1$, the $\{u_k\}$ are *iid* $N(0, \sigma^2)$ and the $\{\epsilon_k\}$ are *iid* with unit mean and positive support. Note that in the SCD model, e^{ψ_k} is not the conditional mean of the durations, even though dependence in $\{\psi_k\}$ produces dependence in the durations. Thus, in spite of its acronym, the model is not a model for conditional durations.

The SCD model is closely related to the $AR(1)-$Stochastic Volatility (SV) model (see Harvey, Ruiz and Shephard (1994)) proposed for clock-time returns. Just as the $AR(1)$-SV model has been extended to the long-memory case (see Harvey 1998, Breidt, Crato, de Lima 1998) we can similarly extend the SCD model to allow for long memory. This is done by defining

$$\tau_k = e^{\psi_k} \epsilon_k \tag{11}$$

$$\psi_k = \omega + (1-L)^{-d} u_k \quad , \tag{12}$$

where the $\{\epsilon_k\}$ are as specified above, $\{u_k\}$ is a zero-mean Gaussian stationary short-memory series, L is the lag operator, and $d \in [0, 1/2)$. We will refer to this as a long-memory stochastic duration model, denoted by LMSD.

Jasiak (1998) proposed a long-memory version of the ACD model for durations, which may be viewed as an analog to the FIGARCH model for volatility. However, Davidson (2004) explains that the FIGARCH model does not actually yield long memory in volatility. For the same reasons, the model proposed by Jasiak (1998) does not yield long memory in durations, so we do not pursue this model further. Bauwens and Giot (2000) have introduced a log-ACD model where the log conditional mean duration is a linear function of past durations. It is not known whether a long memory version of the log-ACD model exists. In the sequel, we therefore focus on the ACD and LMSD models.

It is interesting to compare the dependence structure and the tail properties of the ACD and LMSD models. The tail index of durations under an $ACD(1,1)$ model can take any value κ in $(1, \infty)$, as determined by the solution to the equation

$$E\left[(\beta + \alpha\epsilon)^{\kappa}\right] = 1 \quad . \tag{13}$$

(See Nelson, 1990). If $\kappa \geq 2$, the durations are weakly stationary with exponentially decaying autocorrelations, and hence have short memory. On the other hand, if $\kappa \in (1, 2)$, then the durations are not weakly stationary, and it follows from Theorem 2 that the corresponding counting process will have long memory with memory parameter $d \geq 1 - \kappa/2$.

Note that in the ACD model, the tail index κ is inextricably connected with the dependence of the process, as both are completely determined by the parameters α and β. Increasing the persistence of the ACD durations forces the tail thickness to also increase. Specifically, as $\alpha + \beta$ approaches its maximal value, 1, κ must approach its lowest allowable value, 1.

In contrast, for the SCD and LMSD models, the tail index κ of $\tau_,$ is equivalent to the tail index of ϵ_k (See Breiman, 1965), irrespective of the dependence structure of the $\{\psi_k\}$. Here, the tail index and the dependence structure of the duration process can be separately controlled. In particular, the persistence can be increased from short memory to long memory without affecting the tail index in any way. For the LMSD models, Theorem 2 implies that the counts will have long memory if the ϵ_k have a tail index $\kappa \in (1, 2)$, irrespective of the dependence structure in the $\{\psi_k\}$. If the durations follow an LMSD model with finite variance, we cannot apply Theorem 2 to infer that the counts have long memory. Nevertheless, Theorem 3 implies that the counting process will indeed have long memory in this case, with the same memory parameter as that of the durations.

We next estimate ACD and LMSD models for our data.

A Empirical Analysis with ACD Models

We estimate the ACD(1,1) model by maximum likelihood. Since the ACD model is observation driven, we can write the log likelihood function for the parameter vector θ as

$$\log L(\theta) = \sum_{k=1}^{n} \log\{\frac{1}{\psi_k} \exp(\frac{-\tau_k}{\psi_k})\} = -\sum_{k=1}^{n} \log(\psi_k) - \sum_{k=1}^{n} \frac{\tau_k}{\psi_k}$$

if the ϵ_k are standard exponential, and as

$$\log L(\theta) = \sum_{k=1}^{n} \left\{ \log(\frac{\gamma}{\tau_k}) + \gamma \log\left(\frac{\tau_k \Gamma(1 + 1/\gamma)}{\psi_k}\right) - \left[\frac{\tau_k \Gamma(1 + 1/\gamma)}{\psi_k}\right]^{\gamma} \right\}$$

if the $\{\epsilon_k\}$ are Weibull (δ, γ) with unit mean, so that $\delta = 1/\Gamma(1+\frac{1}{\gamma})$. The MLE for the ACD(1,1) model is computed using the Berndt, Hall, Hall and Hausmann (1974) algorithm. The estimation results for the EACD(1,1) and WACD(1,1) models are presented in Table 6 and Table 7. The estimated values of α and β are quite similar for the EACD(1,1) and WACD(1,1) models, with the sum of $\hat{\alpha}$ and $\hat{\beta}$ close to one, suggesting a nearly integrated duration process. All of the less active stocks have $\hat{\gamma}$ less than one in the fitted WACD(1,1) model, indicating that the shocks have a longer tail than a standard exponential distribution. The results for the actively traded stocks are mixed since some have $\hat{\gamma}$ greater than one while others have $\hat{\gamma}$ less than one. However, the $\hat{\gamma}$ for all the active stocks are all greater than those for the less active stocks.

From (13), it can be shown that as γ decreases while holding the other parameters fixed, the tail index κ of the durations also decreases. This suggests that the durations of the active stocks tend to have shorter tails than those of the less active stocks. In Tables 6 and 7 we also present the tail indices of the durations as implied by (13) based on the estimated parameters. The estimated values of κ for the less active stocks are typically between 4 and 10, while those for the active stocks are typically between 13 and 30.

The resulting estimated values of κ are all greater than two. Thus the estimated EACD and WACD models yield finite variance in durations. Consequently, we are unable to use Theorems 1 and 2 to conclude that the corresponding counts have long memory. Note, however, that these estimates of κ will be accurate estimates of the true tail indices only if the model is correctly specified. We therefore perform diagnostic tests for the fitted ACD and WACD models.

Define the residuals from the ACD model as

$$\hat{\epsilon_k} = \frac{\hat{\tau_k}}{\psi_k} \tag{14}$$

where $\hat{\tau_k}$ are the diurnally-adjusted durations.

We present Q-Q plots of the residuals for the AIG and BEV data in Figure 9 to check the adequacy of the exponential and Weibull distribution for the shocks in the EACD(1,1) and WACD(1,1) models, respectively. The plots indicate that the actual distribution has a longer right tail than that implied by the model. Results for the other stocks were similar, and are not shown here.

Following Engle and Russell (1998), we also assess the goodness of fit of the Weibull distribution by examining the survivor function of the residuals. The survivor function for the WACD(1,1) model is

$$\bar{F}(\tau_k|\psi_k) = \exp\left(-\left(\frac{\tau_k}{\phi_k\psi_k}\right)^{\gamma}\right),$$

and hence

$$-\log(\bar{F}(\tau_k|\psi_k)) = \left(\frac{\tau_k}{\phi_k\psi_k}\right)^{\gamma}.$$

Thus, the negative log of the empirical survivor function should be linearly related to the transformed standardized residuals with a slope of unity. Figure 10 presents this plot. Except for the case of Sony, we observe curvature in the empirical survivor functions, which is once again indicative that the distribution has a longer right tail than that implied by the Weibull.

We conclude that both the WACD(1,1) and EACD(1,1) models fitted to our data are misspecified in terms of the conditional distribution of the shocks. One implication of this is that the estimated tail indices of the durations based on (13) cannot be trusted. Thus the question as to whether durations have infinite variance remains open.

Next, we explore the adequacy of the ACD(1,1) model in describing the dependence among the durations. This is done by investigating the dependence in the standardized residuals and the squared standardized residuals from both the EACD(1,1) and WACD(1,1) model. Since both models assume that the shocks are independent, any dependence in either the standardized residuals or their squares indicates misspecification in the time series structure of the model. For the AIG data, the Ljung-Box statistics for both of the residuals and the squared residuals under the EACD(1,1) model as well as the WACD(1,1) model are all significant up to lag 100. We found similar results for the other stocks. We conclude that the residuals under the EACD and WACD models are not independent. Thus the dependence structure implied by the fitted ACD models is unable to adequately capture the dependence structure of the durations.

Here, we examine the extent of the misspecification of the dependence structure of the fitted ACD(1,1) models by estimating the memory parameter from the residuals. Figure 1 is the log-log periodogram plot of the AIG residuals under the EACD(1,1) model. The vertical line is drawn at $\log(j) = \log(\sqrt{n})$, where $n = 461,346$ is the number of residuals.

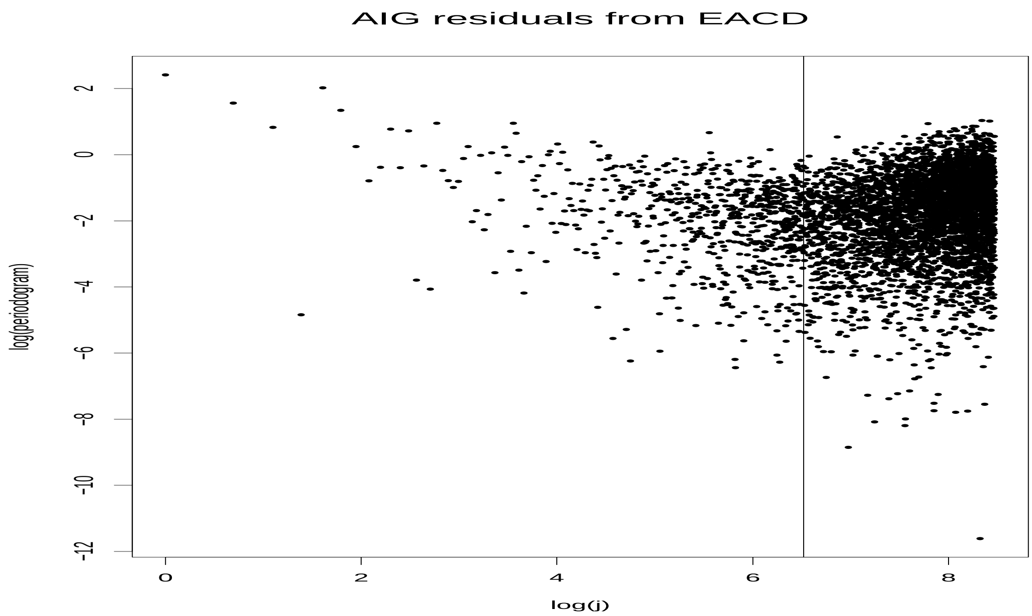
The resulting GPH estimates for the memory parameter using $m = \sqrt{n}$ frequencies were 0.2062 and 0.2185 for the AIG residuals under the EACD(1,1) and WACD(1,1) models respectively. Results were similar for the other stocks. The fitted ACD models all had short memory, whereas it appears that long memory is needed in order to adequately model the dependence structure of the durations.

Finally, we examine the adequacy of the ACD models from the point of view of the behavior of the corresponding counts. If the ACD model were adequate, then the statistical properties of the counts simulated from the ACD model should be consistent with those of the empirical counts. Table 9 and Table 10 present the lag-1 and lag-2 autocorrelations of the counts for $\Delta t = 5$ minutes, 1 hour, and 1 day generated from the simulated durations under the estimated EACD(1,1) model and WACD(1,1) models. As Δt increases, the autocorrelations of the counts become extremely small. This strongly contradicts the findings of much larger autocorrelations for counts of the empirical data (see Figure 7).

B Empirical Analysis with LMSD Models

In view of the substantial long-range dependence in the durations, we fit the LMSD model introduced in the previous section to the duration data, assuming an $AR(1)$ process for the short-memory component $\{u_k\}$. Unlike the observation-driven ACD model, the latent process $\{\psi_k\}$ in the LMSD model is not observable and thus it is difficult to implement the MLE for estimating

Figure 1: $\log - \log$ periodogram plots for the AIG residuals from the EACD(1,1)Model, $n = 461,346$ $m = \sqrt{n}$.



the parameters of the LMSD model. However, a Quasi Maximum Likelihood Estimator (QMLE) using the Whittle approximation (1962) can be used for parameter estimation. Under the LMSD model, the log durations $\{\log \tau_k\}$ in equation (11) can be written as

$$\log \tau_k = E\left[\log \epsilon_k\right] + \psi_k + \xi_k$$

where $\{\epsilon_k\}$ is the shock process and $\{\xi_k\} = \{\log \epsilon_k - E\left[\log \epsilon_k\right]\}$ with variance σ_ξ^2 and is independent of ψ_k. We assume that the latent process ψ_k follows a Gaussian $ARFIMA(1, d, 0)$ process with innovations $w_k \overset{iid}{\sim} N(0, \sigma_w^2)$, given by

$$(1 - \alpha L)(1 - L)^d \psi_k = w_k$$

with $|\alpha| < 1$. The spectral density for $\log \tau_k$ is given by

$$f_\theta(\omega_j) = \frac{\sigma_w^2}{2\pi}|1 - \alpha \exp(-i\omega_j)|^{-2}|1 - \exp(-i\omega_j)|^{-2d} + \frac{\sigma_\xi^2}{2\pi}, \tag{15}$$

where $\theta = (d, \alpha, \sigma_w^2, \sigma_\xi^2)$, and $\omega_j = 2\pi j/n$. The Whittle negative log likelihood is

$$Ł_n(\theta) = \sum_{j=1}^{[\frac{n-1}{2}]} \{\log f_\theta(\omega_j) + \frac{I(\omega_j)}{f_\theta(\omega_j)}\} \tag{16}$$

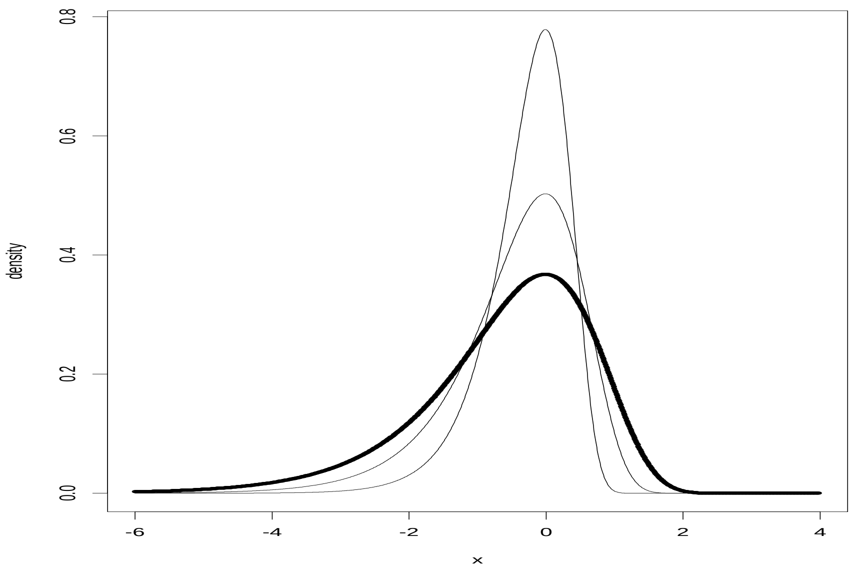
where $f_\theta(\omega_j)$ is the spectral density of $\{\log \tau_k\}$ under the LMSD model. This expression is an approximation to the negative log Gaussian quasi likelihood.

Even though $\log \tau_k$ is not Gaussian, it is known that the resulting estimator is $\sqrt{n}$ consistent and asymptotically normal. See Hosoya (1997). The assumptions on the distribution of ϵ_k determine the value of σ_ξ^2. If ϵ_k is standard exponential, then $\sigma_\xi^2 = \pi^2/6$. If ϵ_k is Weibull(γ, 1), then $\sigma_\xi^2 = \pi^2/(6\gamma^2)$, and in this case σ_ξ^2 needs to be estimated.

We fit the LMSD model with both Weibull and standard exponential shocks. The estimation algorithm failed to converge in the case of the exponential shocks for the active stocks. To try to understand this, we note first that the empirical log durations are right-skewed for the active stocks (see Figure 5), while the log Weibull and log exponential distributions are both left-skewed. However, the left skewness is less extreme for the log Weibull distribution, particularly as γ increases. This can be seen in the density plots in Figure 2.

Since the Weibull distribution is apparently less misspecified than the exponential, we focus on fitting an LMSD model with Weibull shocks. The resulting Whittle estimates are presented in Table 8. The estimates of d generally vary between .35 and .42. These parametric estimates are quite similar to the corresponding semiparametric GPH estimates of d for log durations presented in Table 2. The $AR(1)$ parameter estimates vary between $-.3$ and $-.5$. It is interesting to note that the estimates of γ for all the less active stocks are very close to 1, which corresponds to the exponential distribution, while the estimates for the active stocks are much larger than 1. This underscores the misspecification of the exponential model for these stocks.

Figure 2: Densities for the log standard exponential, log Weibull (1,1.3672) and log Weibull (1, 2.1169). The solid line represents the density for log standard exponentials. The curve with the highest peak represents the density of logged Weibull(1, 2.1169)). The values of $\hat{\gamma}$ presented here are estimated from AIG and IBM durations.



To perform diagnostics for the LMSD model, one would ideally want to compute the residuals. Unfortunately, unlike the observation-driven ACD model, the LMSD model has no explicit expression for the residuals since the latent process $\{\psi_k\}$ is not observable. Harvey (1998) proposed a smoothing methodology based on the MMSLE (Minimum Mean Square Linear Estimator) for smoothing the latent process in the Long Memory Stochastic Volatility model. It may thus be tempting to consider the sample ACF of the pseudo-residuals defined as $\log \tau_k - \hat{\psi}_k$ where $\{\hat{\psi}_k\}$ is given by

$$\hat{\psi} = \Sigma_\psi \Sigma_\tau^{-1}(\tau - \mu 1), \tag{17}$$

$\mu = E[\log \tau_k]$, $\tau = \{\log(\tau_1), \ldots, \log(\tau_n)\}$, Σ_τ is the autocovariance matrix of τ, Σ_ψ is the autocovariance matrix of $\{\psi_1, \ldots, \psi_n\}$ and 1 is an $n \times 1$ vector of ones.

Unfortunately, even if the true parameter values of the model for ψ_k are known, it can be shown that the autocorrelation of the resulting pseudo-residuals would still be nonzero at fixed lags, though slowly decaying for higher lags. More specifically, Bhansali and Karavellas (1983) point out that the spectral density of the pseudo-residuals (based on the true parameter values) is given by

$$f_\psi(\lambda) - \frac{f_\psi^2(\lambda)}{f_\psi(\lambda) + \sigma_\xi^2/(2\pi)}$$

where f_ψ is the spectral density of $\{\psi_k\}$. It is thus clear that the pseudo-residuals will not behave like a white noise process, even when the model is correctly specified. Unfortunately, there seems at present to be no satisfactory method for model diagnostics for latent variable models such as SCD or LMSD.

We therefore chose to examine the adequacy of the LMSD model by simulating durations from the estimated models and then comparing the properties of the resulting counts to the stylized facts observed in the data. In Table 11 we present the sample ACF of the simulated counts for $\Delta t = 5$ minutes, 30 minutes, and 1 hour. It is seen that the sample correlations do not die out quickly to zero for any of the choices of Δt, unlike what was found for the simulated ACD models. In Table 12, we also present GPH estimates of the long memory parameter of the simulated counts. It is seen that these estimates are significantly larger than zero and very close to the memory parameter ($d = .3545$) of the durations of the model, in keeping with Theorem 3.

VI Conclusion

In this paper, we have explored some statistical properties of durations that lead to long memory in volatility. Existing theorems show that infinite variance in durations can produce long memory in counts, though our empirical analysis did not support this route to long memory. We also presented a theorem, illustrated by simulations, that under an LMSD model for durations, the resulting memory parameter in counts and volatility would be greater than or equal to that in durations. Moreover, we found that an LMSD model adequately accounted for the long-range dependence in durations.

We have argued that long memory in volatility may be a result of long memory in durations, but this raises a new question: What is the source of the long memory in durations? We speculate that the long memory in intertrade durations may arise from the properties of the (unobservable) news arrival process. We will not attempt to give a precise definition of news, but we suppose that relevant news events arrive at random time points, and we suppose also that the partial sums of durations between news events converge to a process with finite variance. Then by Lamperti's Theorem (see Beran, 1994, pp. 48–50), the limit process is self-similar, e.g., Brownian Motion, Fractional Brownian Motion. We can rule out Brownian Motion, since then the counts of news events would have virtually no autocorrelation if Δt is large, just as we found earlier in our analysis of the counts induced by finite-variance ACD models. If the limit process is self-similar but not Brownian motion, this suggests that the individual durations between news events should have long memory. A different argument which also leads to the conclusion that news interarrival times have long memory was provided by Bollerslev and Jubinski (1999), who invoked the arguments of Parke (1999). Thus, the ultimate source of long memory in volatility may be long memory in the news arrival process.

VII Appendix

A Proof of Equation (5)

Since $\{e_{t'}\}$ follows a GARCH(1,1) model, it is well known that we can represent $\{e_{t'}^2\}$ as an ARMA(1,1) process of the form

$$e_{t'}^2 = \tilde{\omega} + (\tilde{\alpha} + \tilde{\beta}) e_{t'-1}^2 + v_{t'} - \tilde{\beta} v_{t'-1} \quad ,$$

where $\{v_{t'}\}$ is a white noise process, provided that $E[e_{t'}^4] < \infty$. Hence,

$$\text{cov}\left(e_{t'}^2,\ e_{t'+k}^2\right) = (\tilde{\alpha} + \tilde{\beta})^k C_2$$

where C_2 is a positive constant, and thus

$$E\left[e_{t'}^2 e_{t'+k}^2\right] = \text{cov}\left(e_{t'}^2,\ e_{t'+k}^2\right) + E[e_{t'}^2] E[e_{t'+k}^2] = (\tilde{\alpha} + \tilde{\beta})^k C_2 + C_1^2 \tag{18}$$

where $C_1 = E[e_{t'}^2]$. From Equation (18), we have

$$\begin{aligned} \text{cov}(r_{t'}^2,\ r_{t'+k}^2) &= E\left[\Delta N_{t'} \Delta N_{t'+k}\right] \left[(\tilde{\alpha} + \tilde{\beta})^k C_2 + C_1^2\right] - C_1^2 E\left[\Delta N_{t'}\right] E\left[\Delta N_{t'+k}\right] \\ &= C_1^2 \text{cov}\left(\Delta N_{t'},\ \Delta N_{t'+k}\right) + C_2 (\tilde{\alpha} + \tilde{\beta})^k E\left[\Delta N_{t'} \Delta N_{t'+k}\right] \end{aligned}$$

and

$$\text{var}(r_{t'}^2) = \text{cov}(r_{t'}^2, r_{t'}^2) = C_1^2 \text{var}(\Delta N_{t'}) + C_2 E[\Delta N_{t'}^2] \quad .$$

Using this result and noting that

$$\begin{aligned} \text{cov}(r_{t'}^2, r_{t'+k}^2) &= \text{cov}(\Delta N_t' e_t'^2, \Delta N_{t'-1} e_{t'+k}^2) = E\left[\Delta N_{t'} e_{t'}^2 \Delta N_{t'+k} e_{t'+k}^2\right] - E\left[\Delta N_{t'} e_{t'}^2\right] E\left[\Delta N_{t'+k} e_{t'+k}^2\right] \\ &= E\left[\Delta N_{t'} \Delta N_{t'+k}\right] E\left[e_{t'}^2 e_{t'+k}^2\right] - E\left[\Delta N_{t'}\right] E\left[\Delta N_{t'+k}\right] E\left[e_{t'}^2\right] E\left[e_{t'+k}^2\right] \quad , \end{aligned}$$

we obtain

$$
\begin{aligned}
\mathrm{corr}(r_{t'}^2,\ r_{t'+k}^2) &= \frac{\mathrm{cov}(r_{t'}^2,\ r_{t'+k}^2)}{\mathrm{var}(r_{t'}^2)} \\
&= \frac{C_1^2 \mathrm{cov}(N_{t'},\ N_{t'+k})}{C_1^2 \mathrm{var}(\Delta N_{t'}) + C_2 E\left[\Delta N_{t'}^2\right]} + \frac{C_2(\tilde{\alpha}+\tilde{\beta})^k E\left[\Delta N_{t'+k}\Delta N_{t'}\right]}{C_1^2 \mathrm{var}(\Delta N_{t'}) + C_2 E\left[\Delta N_{t'}^2\right]} \\
&= \frac{\rho_k^{\Delta N}}{1+\frac{C_2 E[\Delta N_{t'}^2]}{C_1^2 \mathrm{var}(\Delta N_{t'})}} + \frac{\frac{(\tilde{\alpha}+\tilde{\beta})^k E[\Delta N_{t'+k}\Delta N_{t'}]}{E[\Delta N_{t'}^2]}}{1+\frac{C_1^2 \mathrm{var}(\Delta N_{t'})}{C_2 E[\Delta N_{t'}^2]}} \quad . \qquad (19)
\end{aligned}
$$

Noting that

$$
\frac{E[\Delta N_{t'}^2]}{\mathrm{var}(\Delta N_{t'})} = 1 + \frac{E[\Delta N_{t'}]^2}{\mathrm{var}(\Delta N_{t'})} \qquad (20)
$$

and

$$
\begin{aligned}
\frac{E[\Delta N_{t'+k}\Delta N_{t'}]}{E[\Delta N_{t'}^2]} &= \frac{\mathrm{cov}(\Delta N_{t'+k},\ \Delta N_{t'}) + E[\Delta N_{t'}]E[\Delta N_{t'+k}]}{\mathrm{var}(\Delta N_{t'}) + E[\Delta N_{t'}]^2} \\
&= \frac{\rho_k^{\Delta N} + E[\Delta N_{t'}]E[\Delta N_{t'+k}]/\mathrm{var}(\Delta N_{t'})}{1+E[\Delta N_{t'}]^2/\mathrm{var}(\Delta N_{t'})} \quad , \qquad (21)
\end{aligned}
$$

we obtain

$$
\mathrm{corr}(r_{t'}^2, r_{t'+k}^2) = \left[m_1 + m_2(\tilde{\alpha}+\tilde{\beta})^k\right]\rho_k^{\Delta N} + m_3(\tilde{\alpha}+\tilde{\beta})^k, \qquad (22)
$$

where m_1, m_2, m_3 are constants with

$$
m_1 = \frac{1}{1+\frac{C_2 E[\Delta N_{t'}^2]}{C_1^2 \mathrm{var}(\Delta N_{t'})}} \quad .
$$

Note that $0 < m_1 < 1$ since $C_2 > 0$. Furthermore, since $\tilde{\alpha}+\tilde{\beta} < 1$,

$$
\mathrm{corr}(r_{t'}^2, r_{t'+k}^2) \sim m_1 \rho_k^{\Delta N},
$$

as $k \to \infty$.

B Definitions

Definition: A point process $N(t) = N(0,t]$ is *stationary* if for every $r = 1, 2, \ldots$ and all bounded Borel sets $A_1, \ldots, A_r$, the joint distribution of $\{N(A_1+t), \ldots, N(A_r+t)\}$ does not depend on $t \in [0, \infty)$.

Definition: Given a stationary point process $N(t)$, there is a corresponding *Palm probability measure* P^0 given by

$$P^0(A) = P(A|t_0 = 0) \tag{23}$$

for all $A \in \sigma(\{\tau_k\}_{k=-\infty}^{\infty}, \{t_k\}_{k=-\infty}^{\infty})$, where P is the marginal probability measure for the interarrival times, t_k is the event time for the k'th event and $\tau_k = t_k - t_{k-1}$ is the waiting time between the $k-1$'st and the k'th event. The Palm probability $P^0(A)$ is the conditional probability of A given that an event occurred at time zero. Note that $P^0(t_0 = 0) = 1$. It can be shown that for a stationary renewal process (in which the $\{\tau_k\}$ are *iid*)

$$P^0(\tau_1) = P(\tau_1). \tag{24}$$

Definition: The *tail index* κ of a distribution function F is given by

$$\kappa = \sup\left\{k : \int |x|^k dF(x) < \infty\right\} \quad .$$

C Proof of Theorem 3

We give a proof by contradiction. Accordingly, assume that the counts have memory parameter $d_{\Delta N} \in [0, d_\tau)$. Suppose we have a realization of n counts. Let f_1 and I_1 denote the spectral density and periodogram of the counts at the first Fourier frequency.

We first show that I_1/n^{2d_τ} converges in distribution to a nondegenerate random variable with support on the entire positive real line. Let $S_n(t) = n^{-d_\tau-1/2}\sum_{k=1}^{[nt]}(\tau_k - E(\tau_k))$, $t \in (0,1)$. It is shown in Surgailis and Viano (2002) that $S_n(t) \overset{d}{\Rightarrow} B_{d_\tau+1/2}(t)$ in $D(0,1)$ where $B_{d_\tau+1/2}(t)$ is fractional Brownian motion with Hurst parameter $d_\tau + 1/2$. Thus, by Iglehart and Whitt (1971), it follows that $t^{-d_\tau-1/2}N(t) - E[N(t)] \overset{d}{\Rightarrow} AB_{d_\tau+1/2}(t)$ in $D(0,1)$, where A is a nonzero constant. Our claim in the first sentence of this paragraph follows by the continuous mapping theorem.

Thus, there exists a constant $C > 0$ such that for all sufficiently large n,

$$Pr\{I_1/n^{2d_\tau} > 1\} \geq C \quad . \tag{25}$$

By Hurvich and Beltrao (1993), Theorem 1, there exists a positive constant A such that for all n,

$$E[I_1] \leq An^{2d_{\Delta N}}$$

This, together with Chebyshev's inequality, implies

$$Pr\{I_1/n^{2d_\tau} > 1\} \leq n^{-2d_\tau}E[I_1] \leq An^{2(d_{\Delta N}-d_\tau)} \to 0 \quad ,$$

thereby contradicting (25). It follows that

$$f_{\Delta N}(\lambda) \sim C\lambda^{-2d_{\Delta N}} \ as \ \ \lambda \to 0,$$

where $d_{\Delta N} \geq d_\tau$. This, in turn, implies scaling of $VarN(t)$ as claimed in the statement of the theorem. See Beran (1994).□

Figure 3: Normal QQ Plots of Normalized IBM Returns.

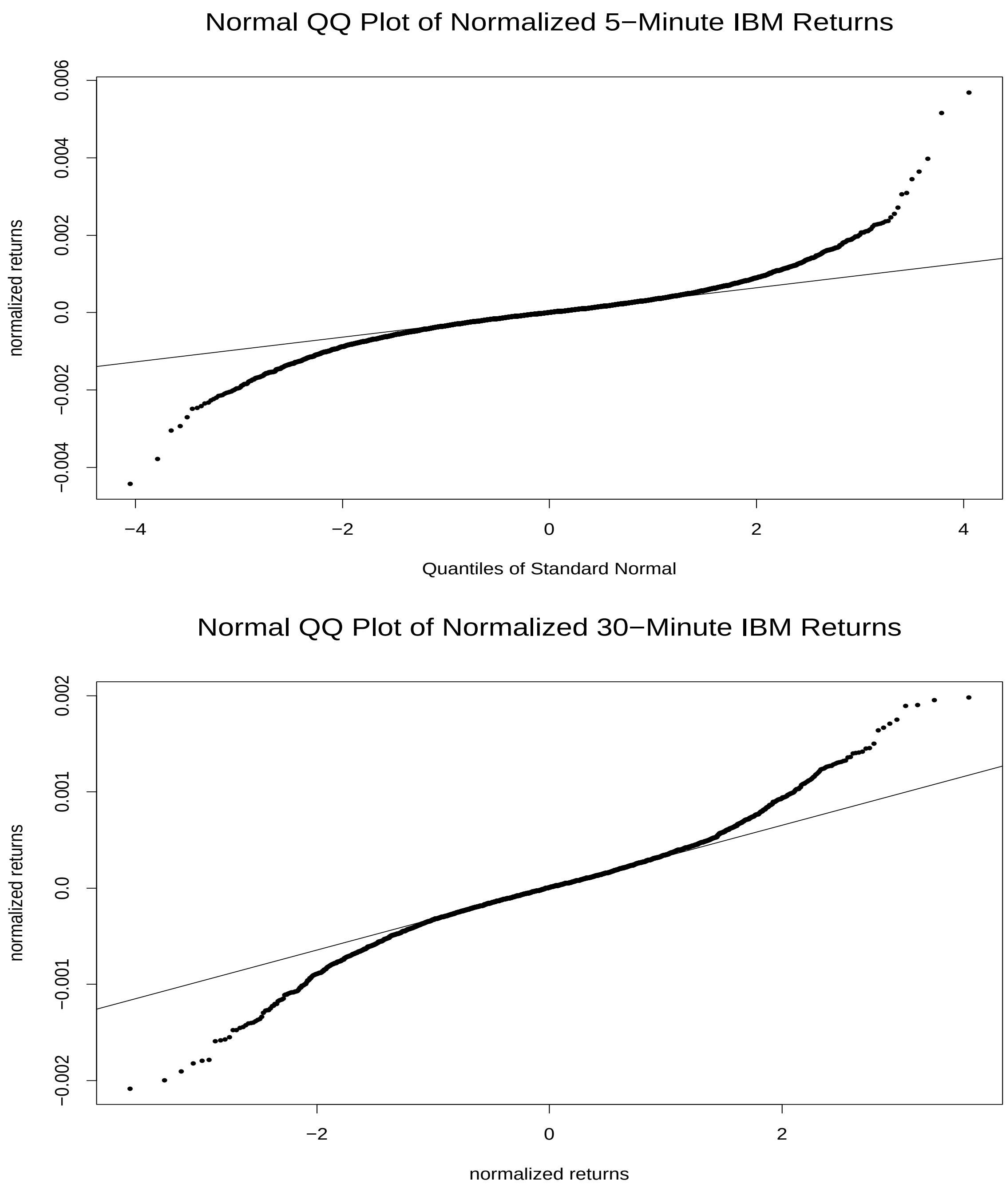


Figure 4: ACF plots of the adjusted IBM returns and squared returns.

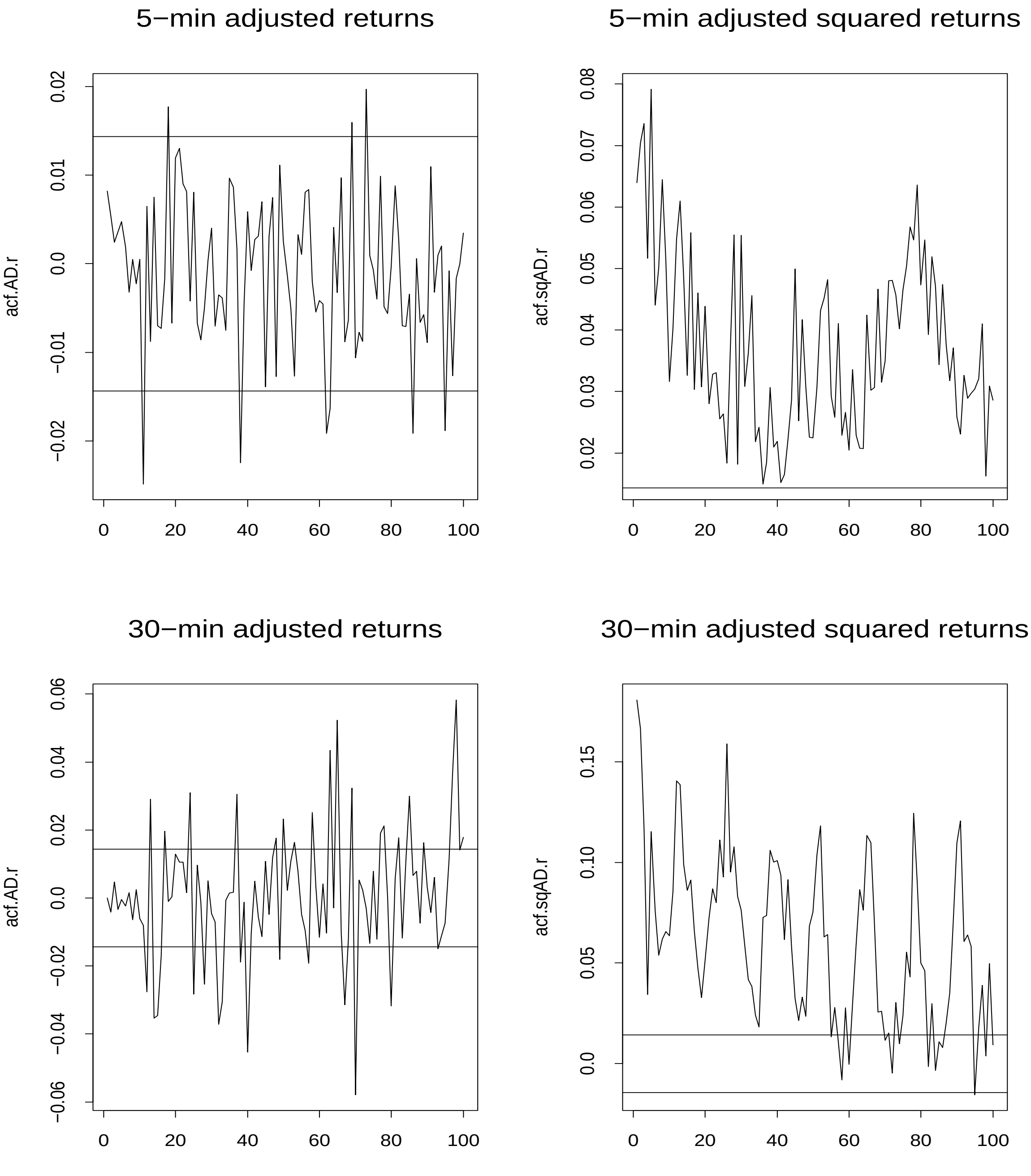


Figure 5: Histograms of the logged diurnally-adjusted durations for Active Stocks.

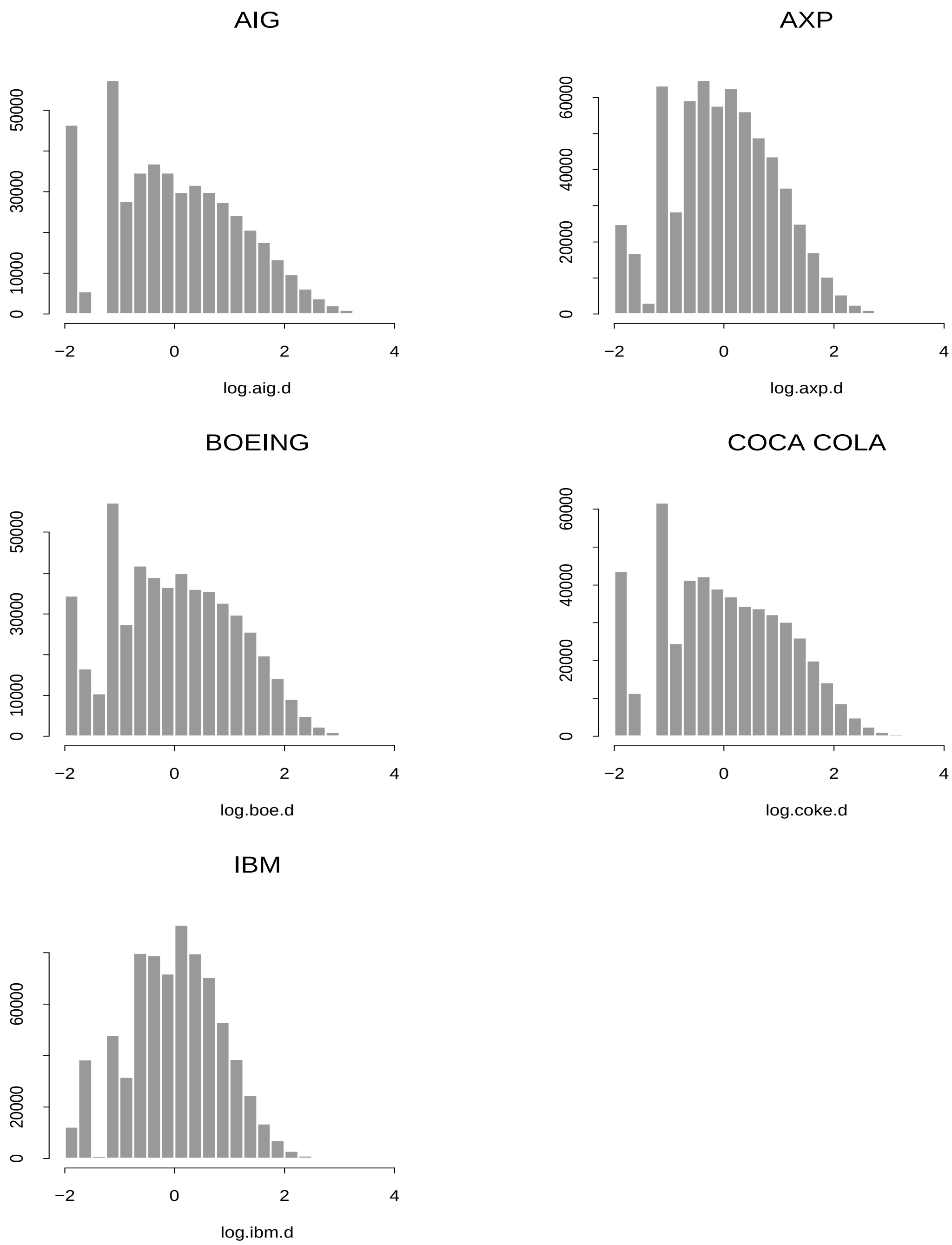


Figure 6: Histograms of the logged diurnally-adjusted durations for Less Active Stocks.

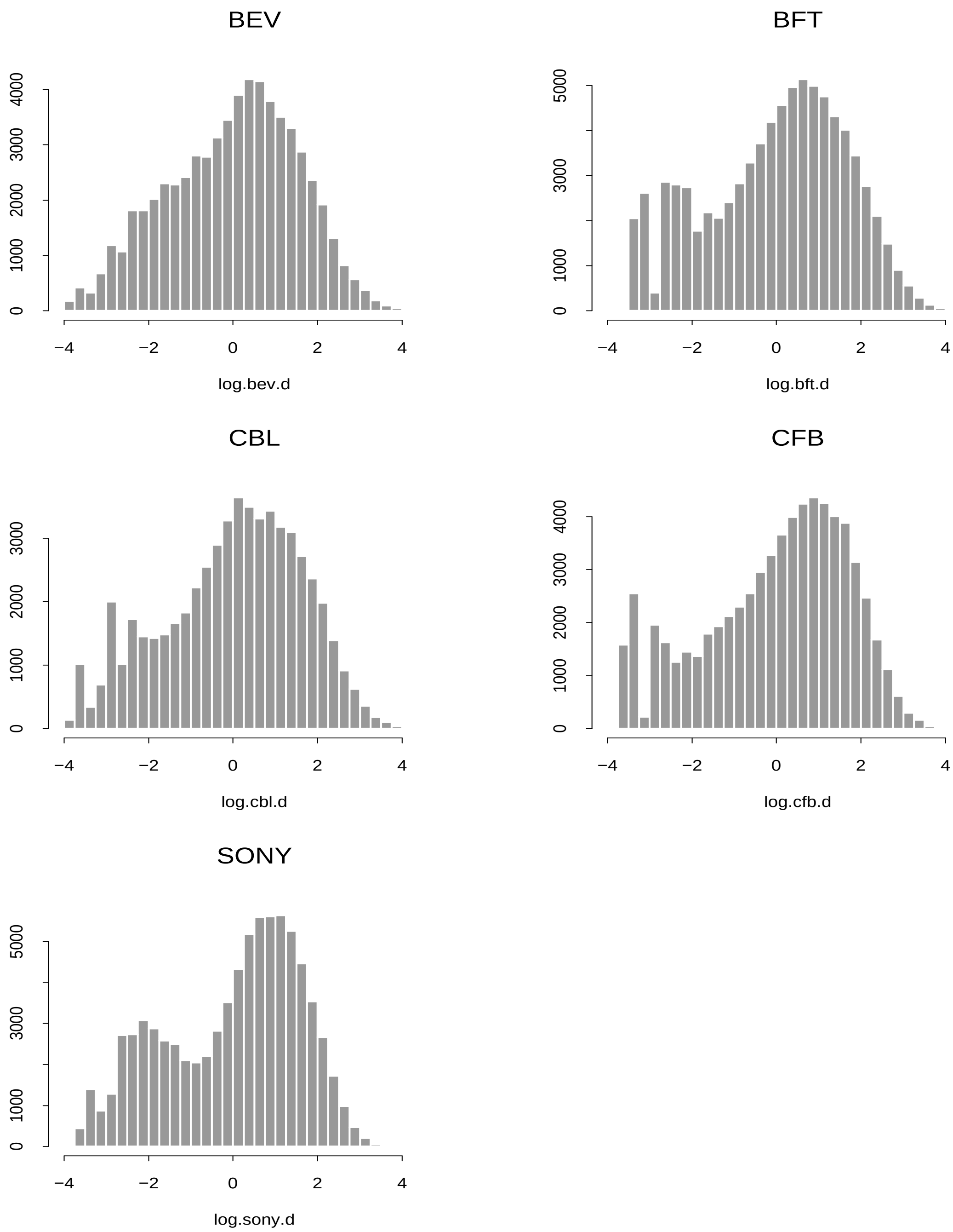


Figure 7: ACF plots for the AIG durations and counts.

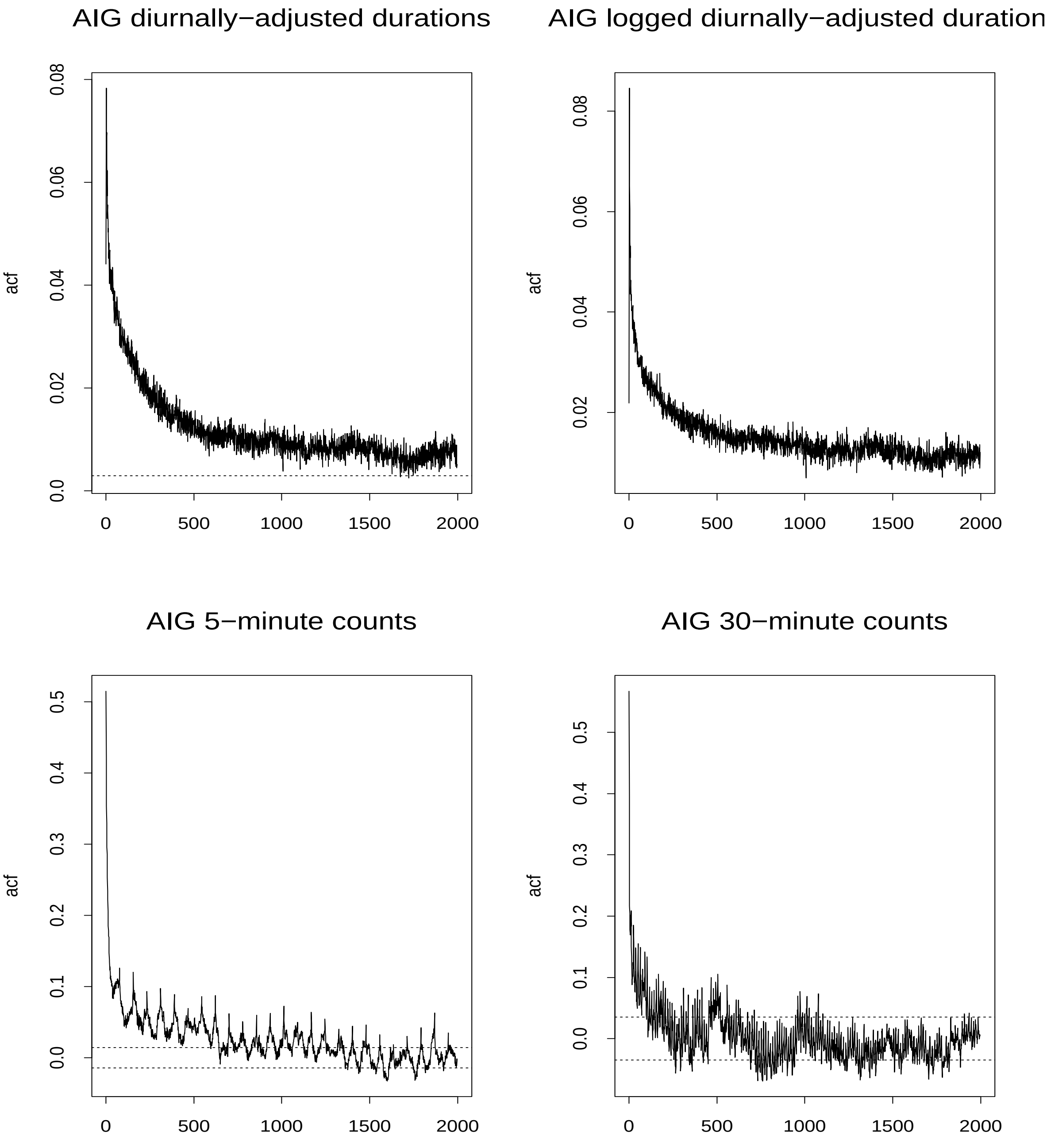


Figure 8: log-log periodogram plots for the AIG counts. The first vertical line in each plot is the visually chosen m, and the second is $m = n^{.8}$.

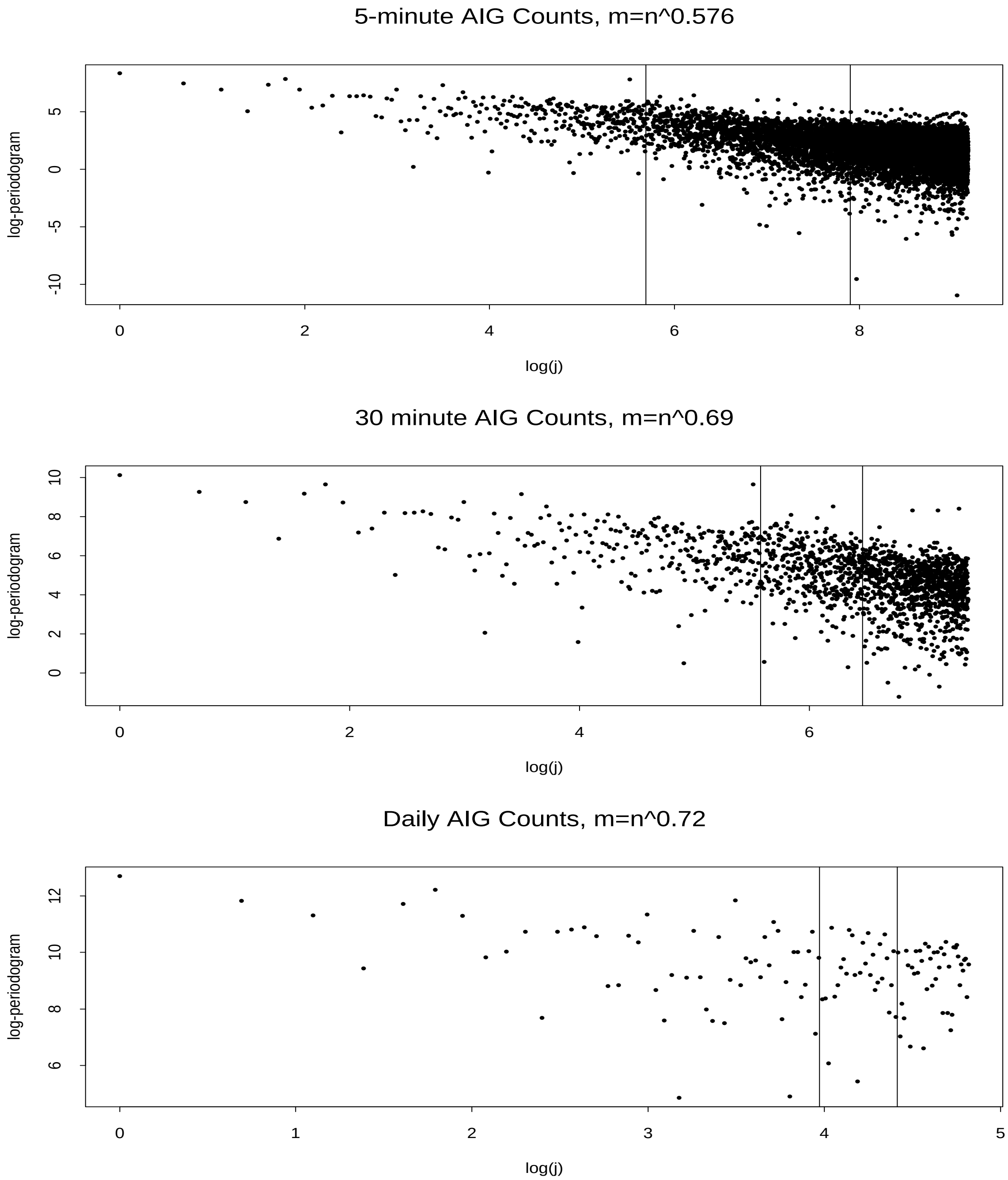


Figure 9: Q-Q plots for the residuals of the EACD(1,1) and WACD(1,1)model.

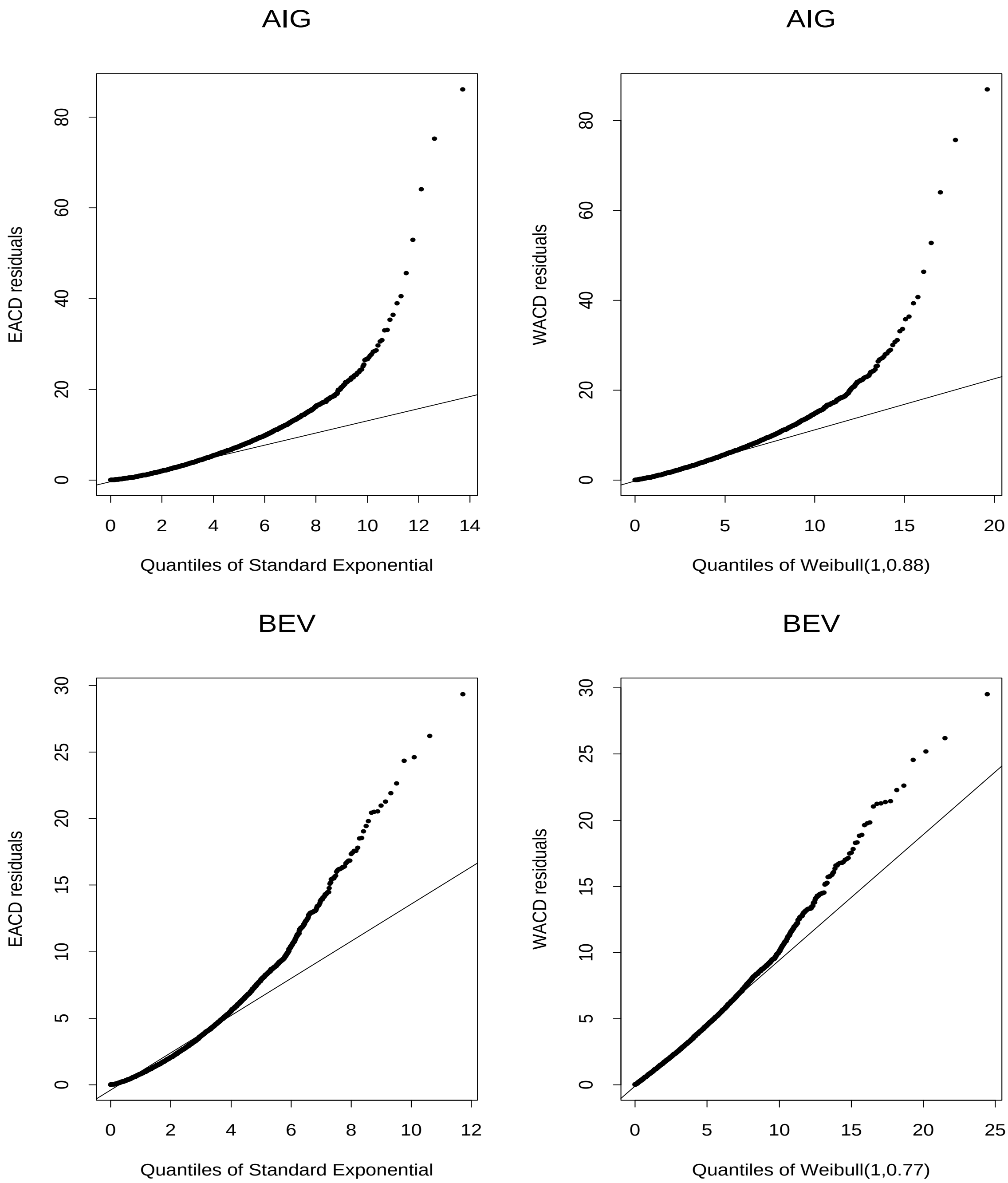


Figure 10: Plots of $-\log$ empirical survivor function vs. transformed WACD residuals.

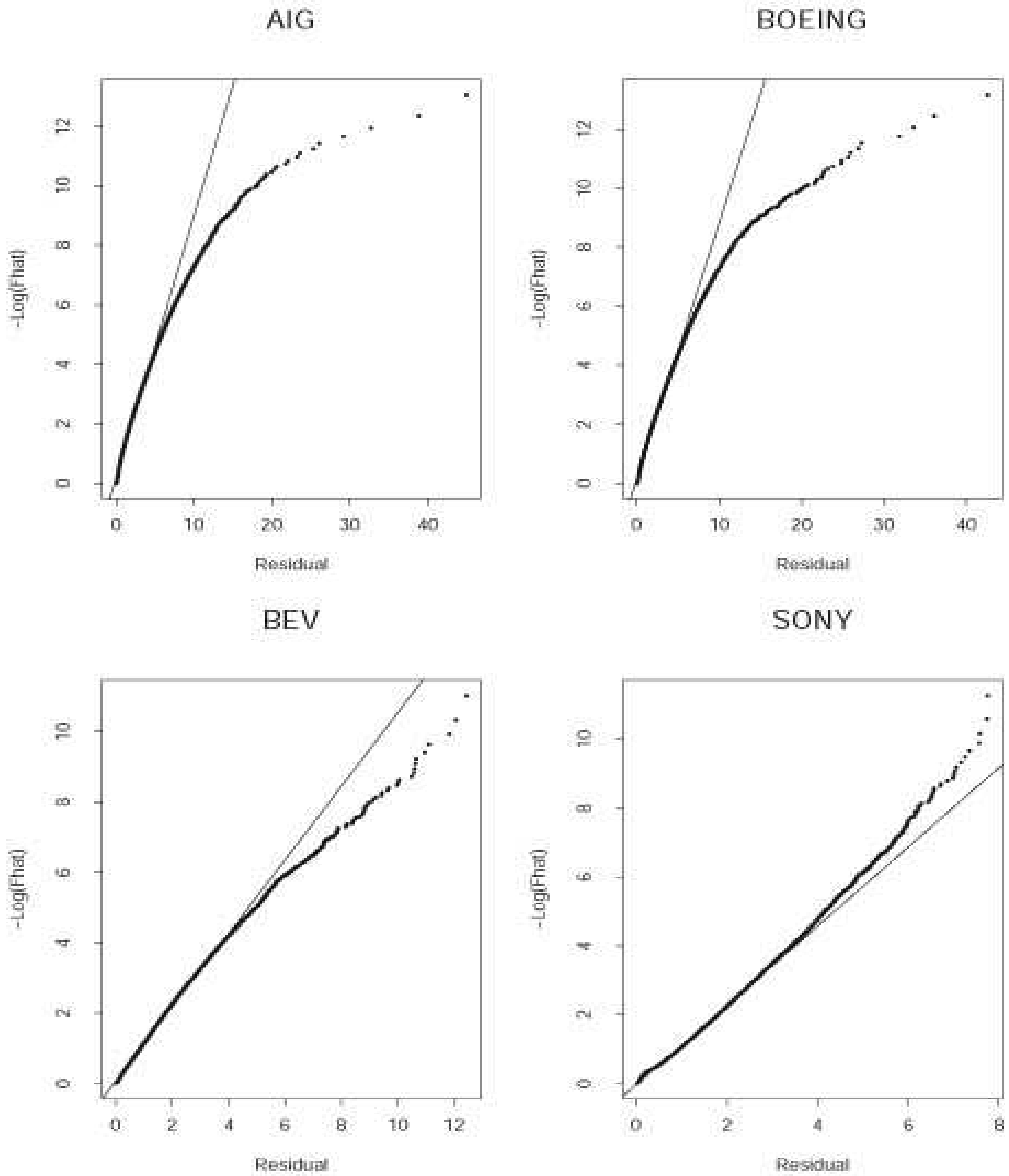


Figure 11: Histograms of the logged residuals from the EACD(1,1) and WACD(1,1) model.

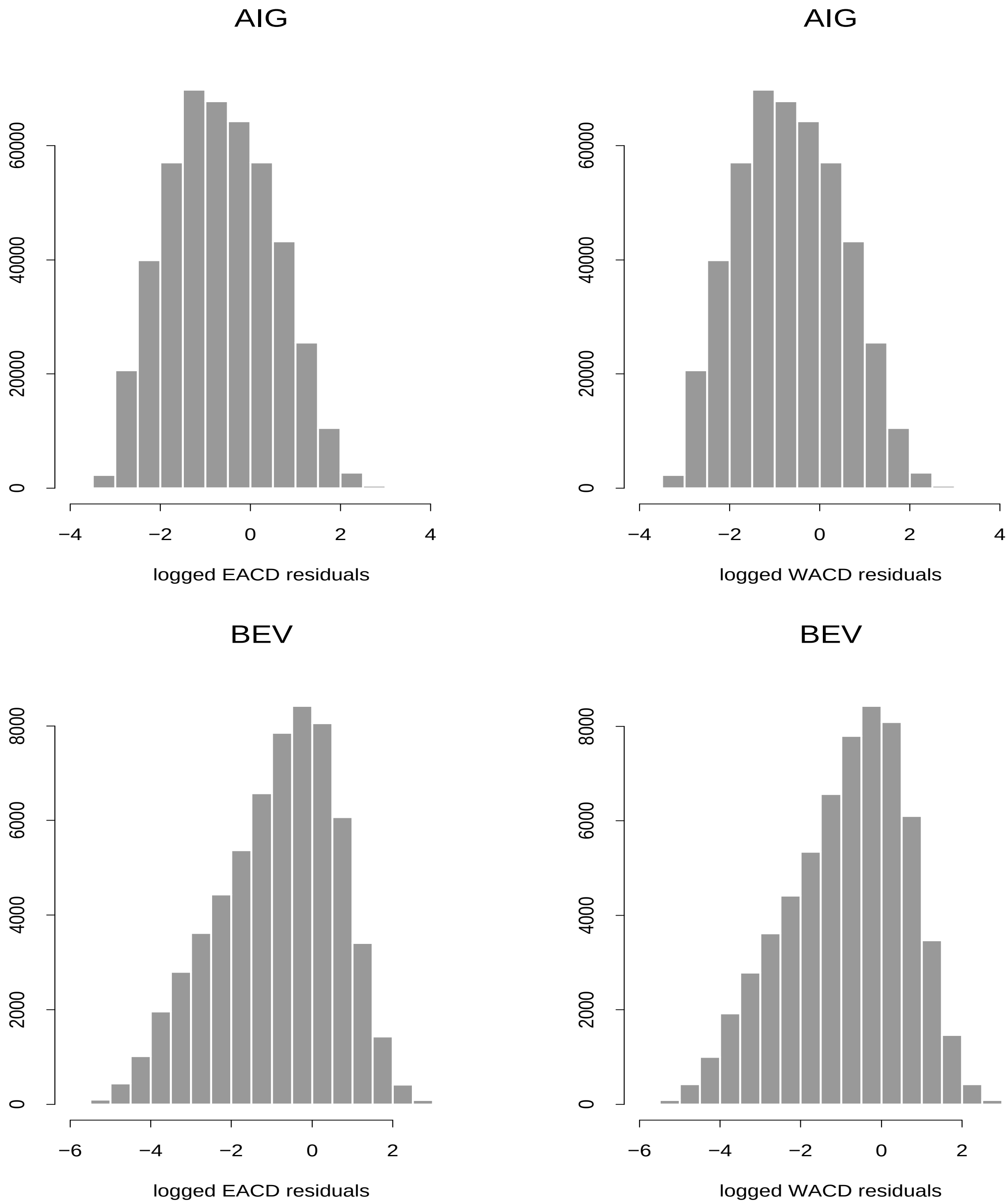


Figure 12: Densities of the durations and the logged durations simulated from the WACD(1,1) model with Weibull (1, 0.942916).

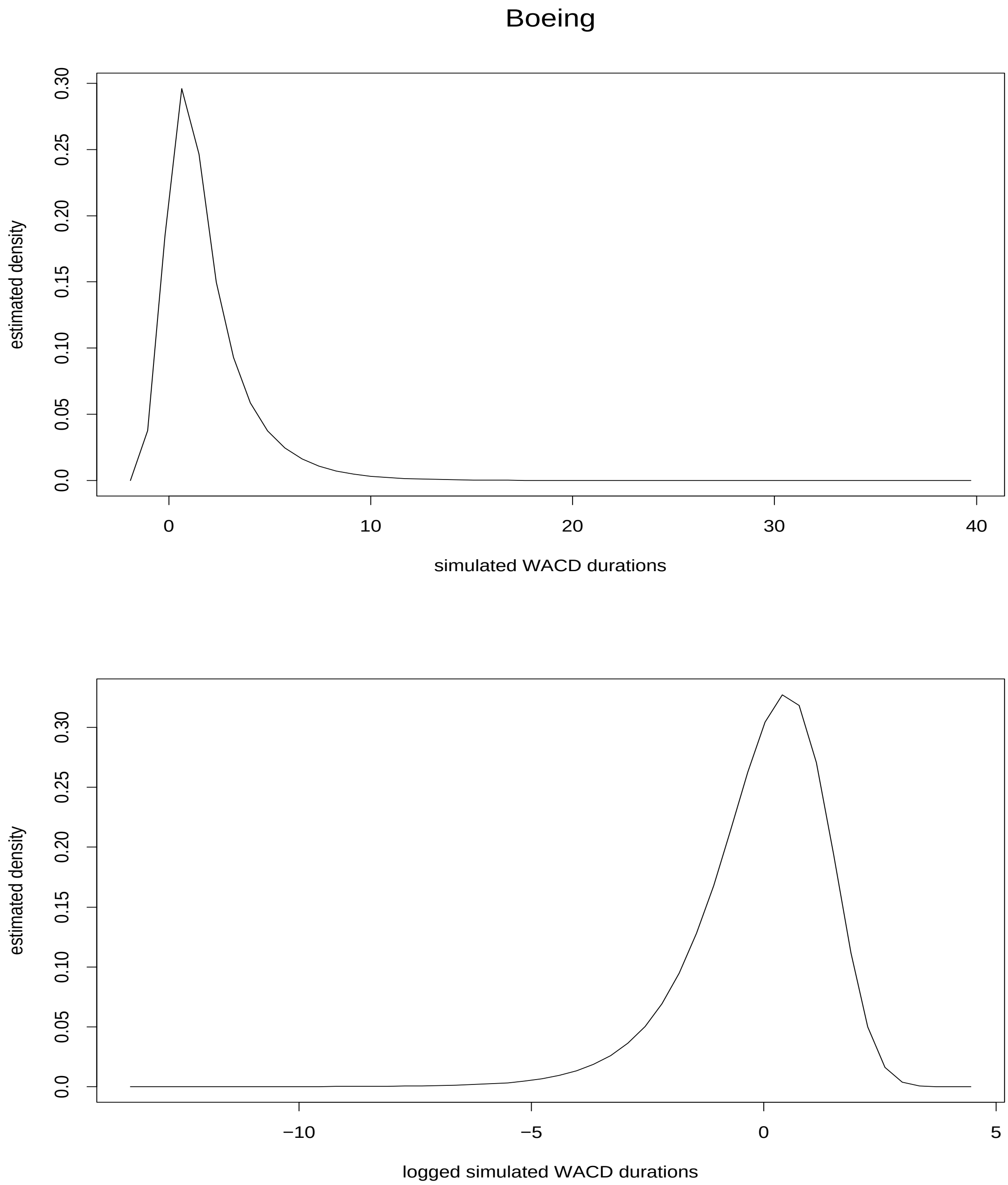


Table 1: Descriptive Statistics for logged durations of the Active Stocks and the Less-Active Stocks.

	AIG	AXP	BA	KO	IBM
Mean	-3.67×10^{-12}	-4.62×10^{-12}	-4.13×10^{-12}	-3.48×10^{-12}	$-.6.84 \times 10^{-12}$
Median	-0.092	-0.029	-0.043	-0.060	0.027
Std Deviation	1.167	0.943	1.107	1.117	0.845
Skewness	0.268	0.111	0.171	0.185	-0.057
Kurtosis	2.370	2.614	2.273	2.296	2.733
Range	6.907	6.316	6.689	6.387	6.455
n	461347	626717	515304	509897	745826

	BEV	BFT	CBL	CFB	SNE
Mean	-4.55×10^{-13}	-1.13×10^{-12}	2.13×10^{-12}	-1.85×10^{-13}	6.17×10^{-13}
Median	0.151	0.237	0.179	0.303	0.348
Std Deviation	1.522	1.656	1.622	1.689	1.590
Skewness	-0.227	-0.309	-0.333	-0.483	-0.418
Kurtosis	2.439	2.256	2.417	2.380	2.169
Range	8.454	8.102	8.108	8.209	7.619
n	62046	80364	56571	66876	78939

Table 2: GPH estimators (standard errors) for the logged diurnally-adjusted durations. For each stock, we report two GPH estimators. Here m_1 and m_2 denote the number of frequencies used in the GPH regression, where $m_1 = n^{0.4}$ and $m_2 = n^{0.5}$ (n is the sample size).

	$\hat{d}_{GPH_{m1}}$	$\hat{d}_{GPH_{m2}}$
AIG	0.2907 (0.0498)	0.3093 (0.0251)
AXP	0.3426 (0.0498)	0.2956 (0.0251)
BA	0.3557 (0.0498)	0.3489 (0.0251)
IBM	0.3711 (0.0498)	0.3446 (0.0251)
KO	0.4583 (0.0498)	0.2982 (0.0251)
BEV	0.3297 (0.0498)	0.3750 (0.0251)
BFT	0.2583 (0.0498)	0.3127 (0.0251)
CBL	0.4619 (0.0498)	0.3500 (0.0251)
CFB	0.3982 (0.0498)	0.2728 (0.0251)
SNE	0.3763 (0.0498)	0.2966 (0.0251)

Table 3: GPH estimators (standard errors) for the counts generated from the empirical durations for different values of Δt. For each Δt, we report two GPH estimators. Here m_1 and m_2 denote the number of frequencies used in the GPH regression. We chose $m_1 = n^{0.576}$, $n^{0.69}$, $n^{0.72}$ for $\Delta t = 5$ min , 30 min , 1 day respectively and $m_2 = n^{0.8}$ for all choices of Δt.

	Δt	$\hat{d}_{GPH_{m1}}$	$\hat{d}_{GPH_{m2}}$
AIG	5 min	0.2764 (0.0388)	0.3811 (0.0124)
	30 min	0.2685 (0.0411)	0.3165 (0.0258)
	1 day	0.4198 (0.0996)	0.3245 (0.0776)
AXP	5 min	0.3155 (0.0388)	0.2919 (0.0124)
	30 min	0.3141 (0.0411)	0.3369 (0.0258)
	1 day	0.4857 (0.0996)	0.3581 (0.0776)
BA	5 min	0.3548 (0.0388)	0.3559 (0.0124)
	30 min	0.3608 (0.0411)	0.3768 (0.0258)
	1 day	0.3556 (0.0996)	0.4522 (0.0776)
IBM	5 min	0.2311 (0.0388)	0.3139 (0.0124)
	30 min	0.2214 (0.0411)	0.3479 (0.0258)
	1 day	0.3775 (0.0996)	0.3306 (0.0776)
KO	5 min	0.3247 (0.0388)	0.3040 (0.0124)
	30 min	0.3173 (0.0411)	0.3760 (0.0258)
	1 day	0.3069 (0.0996)	0.2884 (0.0776)
BEV	5 min	0.3123 (0.0388)	0.3173 (0.0124)
	30 min	0.3296 (0.0411)	0.3635 (0.0258)
	1 day	0.4454 (0.0996)	0.4044 (0.0776)
BFT	5 min	0.4273 (0.0388)	0.3169 (0.0124)
	30 min	0.4547 (0.0411)	0.4013 (0.0258)
	1 day	0.3372 (0.0996)	0.3718 (0.0776)
CBL	5 min	0.3509 (0.0388)	0.2733 (0.0124)
	30 min	0.3640 (0.0411)	0.3308 (0.0258)
	1 day	0.4327 (0.0996)	0.4483 (0.0776)
CFB	5 min	0.3185 (0.0388)	0.2675 (0.0124)
	30 min	0.3128 (0.0411)	0.3412 (0.0258)
	1 day	0.5383 (0.0996)	0.5117 (0.0776)
SNE	5 min	0.3008 (0.0388)	0.2554 (0.0124)
	30 min	0.3214 (0.0411)	0.3063 (0.0258)
	1 day	0.4671 (0.0996)	0.3439 (0.0776)

Table 4: GPH estimators (standard errors) for the squared returns on different lengths of time intervals Δt. For each Δt, we report two GPH estimators. Here m_1 and m_2 denote the number of frequencies used in the GPH regression. We chose $m_1 = n^{0.576}$, $n^{0.69}$, $n^{0.72}$ for $\Delta t = 5$ min , 30 min, 1 day respectively and $m_2 = n^{0.8}$ for all choices of Δt.

	Δt	$\hat{d}_{GPH_{m1}}$	$\hat{d}_{GPH_{m2}}$
AIG	5 min	0.4364 (0.0388)	0.2385 (0.0124)
	30 min	0.3304 (0.0411)	0.2918 (0.0258)
	1 day	0.3123 (0.0996)	0.2575 (0.0776)
AXP	5 min	0.2850 (0.0388)	0.1662 (0.0124)
	30 min	0.2443 (0.0411)	0.1576 (0.0258)
	1 day	0.2033 (0.0996)	0.1266 (0.0776)
BA	5 min	0.3815 (0.0388)	0.1616 (0.0124)
	30 min	0.1663 (0.0411)	0.1456 (0.0258)
	1 day	0.1044 (0.0996)	0.0636 (0.0776)
IBM	5 min	0.3334 (0.0388)	0.1658 (0.0124)
	30 min	0.2876 (0.0411)	0.1478 (0.0258)
	1 day	0.2402 (0.0996)	0.1452 (0.0776)
KO	5 min	0.4283 (0.0388)	0.2020 (0.0124)
	30 min	0.3219 (0.0411)	0.2141 (0.0258)
	1 day	0.3194 (0.0996)	0.0860 (0.0776)
BEV	5 min	0.2553 (0.0388)	0.1739 (0.0124)
	30 min	0.1561 (0.0411)	0.1551 (0.0258)
	1 day	0.1925 (0.0996)	0.0824 (0.0776)
BFT	5 min	0.3033 (0.0388)	0.1454 (0.0124)
	30 min	0.2045 (0.0411)	0.1543 (0.0258)
	1 day	0.1416 (0.0996)	0.0544 (0.0776)
CBL	5 min	0.2717 (0.0388)	0.1383 (0.0124)
	30 min	0.2016 (0.0411)	0.0552 (0.0258)
	1 day	0.3161 (0.0996)	0.1803 (0.0776)
CFB	5 min	0.2663 (0.0388)	0.0971 (0.0124)
	30 min	0.2268 (0.0411)	0.1716 (0.0258)
	1 day	0.2119 (0.0996)	0.0919 (0.0776)
SNE	5 min	0.1851 (0.0388)	0.1247 (0.0124)
	30 min	0.2374 (0.0411)	0.1313 (0.0258)
	1 day	0.1531 (0.0996)	0.0224 (0.0776)

Table 5: GPH estimators (standard errors) of the daily log realized volatility generated from 5-minute returns. Let m_1 and m_2 denote the number of frequencies used in the OLS regression. We computed the GPH estimators based on the values of $m_1 = n^{0.6}$ and $m_2 = n^{0.8}$, where $n = 249$ for all stocks.

	$\hat{d}_{GPH_{m1}}$	$\hat{d}_{GPH_{m2}}$
AIG	0.8242 (0.1495)	0.6063 (0.0776)
AXP	0.7706 (0.1495)	0.4738 (0.0776)
BA	0.6670 (0.1495)	0.4508 (0.0776)
IBM	0.6827 (0.1495)	0.5514 (0.0776)
KO	0.8731 (0.1495)	0.6318 (0.0776)
BEV	0.5885 (0.1495)	0.4007 (0.0776)
BFT	0.3038 (0.1495)	0.2829 (0.0776)
CBL	0.8122 (0.1495)	0.4757 (0.0776)
CFB	0.5220 (0.1495)	0.3382 (0.0776)
SNE	0.7087 (0.1495)	0.3992 (0.0776)

Table 6: Parameters and tail index estimates of the EACD(1,1) model and WACD(1,1) model for the "Active Stocks". Note that κ denotes the tail index computed from the Nelson's equation. γ and δ are the parameters in the Weibull density function. The parameters marked with $*$ are statistically significant from zero at significance level of 1%. The parameter γ marked with $**$ reject the null hypothesis $\gamma = 1$ in favor of the alternative hypothesis $\gamma < 1$ at significance level of 1%.

		EACD(1,1)	WACD(1,1)
AIG	ω	0.015608*	0.016672*
$n = 461346$	α	0.029198*	0.028022*
	β	0.963152*	0.963319*
	γ		0.879661**
	δ		0.93828
	κ	14.59018	13.09937
AXP	ω	0.072136*	0.073313*
$n = 626716$	α	0.059430*	0.060287*
	β	0.894546*	0.893541*
	γ		1.089883
	δ		1.033174
	κ	14.65366	17.19794
BA	ω	0.012337*	0.012248*
$n = 515303$	α	0.023161*	0.022773*
	β	0.970158*	0.970467*
	γ		0.941954**
	δ		0.973106
	κ	19.26017	17.67065
KO	ω	0.010871*	0.010561*
$n = 509896$	α	0.015261*	0.014469*
	β	0.978925*	0.979743*
	γ		0.918299**
	δ		0.9606728
	κ	33.27753	30.84229
IBM	ω	0.049819*	0.046344*
$n = 745825$	α	0.049725*	0.049746*
	β	0.915143*	0.918233*
	γ		1.244746
	δ		1.072634
	κ	16.28203	24.58099

Table 7: Parameters and tail index estimates of the EACD(1,1) model and WACD(1,1) model for the Less Active Stocks. Note that κ denotes the tail index computed from the Nelson's equation. γ and δ are the parameters in the Weibull density function. The parameters marked with $*$ are statistically significant from zero at significance level of 1%. The parameter γ marked with $**$ reject the null hypothesis $\gamma = 1$ in favor of the alternative hypothesis $\gamma < 1$ at significance level of 1%.

		EACD(1,1)	WACD(1,1)
BEV	ω	0.022465*	0.027088*
$n = 62046$	α	0.055131*	0.055681*
	β	0.937231*	0.933935*
	γ		0.771251**
	δ		0.85886
	κ	5.421413	4.287598
BFT	ω	0.027766*	0.034344*
$n = 80363$	α	0.055813*	0.060299*
	β	0.935491*	0.928247*
	γ		0.727482**
	δ		0.8183602
	κ	5.823822	5.293579
CBL	ω	0.025468*	0.030751*
$n = 56570$	α	0.050998*	0.04760*
	β	0.940324*	0.934215*
	γ		0.745174**
	δ		0.8353984
	κ	6.809044	5.522461
CFB	ω	0.027933*	0.030879*
$n = 66875$	α	0.038498*	0.038347*
	β	0.951654*	0.950770*
	γ		0.733751**
	δ		0.8245082
	κ	11.21668	6.695557
SNE	ω	0.029339*	0.024219*
$n = 78938$	α	0.032997*	0.030115*
	β	0.955550*	0.960447*
	γ		0.768937**
	δ		0.8568542
	κ	15.64121	9.472656

Table 8: Whittle's estimators of the LMSD model for the Active Stocks and the Less Active Stocks. We assume the shock process $\{\epsilon_k\}$ follows a Weibull (1, γ) distribution. d denotes Whittle's estimator of the long memory parameter and α denotes the $AR(1)$ coefficient in an ARFIMA(1,d,0) process.

Active Stock			Less-Active Stock		
AIG	d	0.3471	BEV	d	0.4338
	σ_u^2	0.2730		σ_u^2	0.4769
	α	−0.3995		α	−0.4997
	γ	1.2720		γ	1.0075
AXP	d	0.2516	BFT	d	0.4298
	σ_u^2	0.3986		σ_u^2	0.7599
	α	−0.1814		α	−0.5588
	γ	1.9131		γ	0.9989
BA	d	0.3545	CBL	d	0.3961
	σ_u^2	0.2368		σ_u^2	0.7600
	α	−0.4212		α	−0.4360
	γ	1.3376		γ	1.0115
IBM	d	0.2616	CFB	d	0.3664
	σ_u^2	0.3139		σ_u^2	0.6050
	α	−0.2521		α	−0.4731
	γ	2.1173		γ	0.8890
KO	d	0.2744	SNE	d	0.3228
	σ_u^2	0.2771		σ_u^2	0.7001
	α	−0.3872		α	−0.4908
	γ	1.3321		γ	0.9925

Table 9: The first two sample autocorrelations of the counts generated from the estimated EACD(1,1) model for Boeing durations with $\omega = 0.012337$, $\alpha = 0.023161$, $\beta = 0.970158$ and several choices of $\triangle t$. The reported values are computed as the average of the autocorrelations from the 100 replications of the counts with sample size $n = 10,000$.

	Sample Autocorrelation	
$\triangle t$	$\hat{\rho}_1$	$\hat{\rho}_2$
5 min	0.46001	0.17361
30 min	0.10020	-0.00112
60 min	0.04625	0.00098
3 hr	0.01456	0.00094
6 hr	0.00721	0.00011

Table 10: The first two sample autocorrelations of the counts generated from the WACD model for Boeing durations with $\omega = 0.012248$, $\alpha = 0.022773$, $\beta = 0.970467$, and $\gamma = 0.941954$ and several choices of $\triangle t$. The reported values are computed as the average of the autocorrelation from the 100 replications of the counts with sample size $n = 10,000$.

	Sample Autocorrelation	
$\triangle t$	$\hat{\rho}_1$	$\hat{\rho}_2$
5 min	0.56678	0.30133
30 min	0.18167	0.00921
60 min	0.08363	-0.00201
3 hr	0.02592	0.00084
6 hr	0.01380	-0.00003

Table 11: Mean of the first two sample autocorrelations of the counts generated from the LMSD model with Weibull $(1, \gamma)$ distribution on different length of clock time intervals. We used the estimated LMSD parameters of the Boeing durations presented in Table 8 ($d = .3545$) for our simulations. We simulated 200 replications of the counts, each with sample size $n = 10,000$.

	Sample Autocorrelation	
$\triangle t$	$\hat{\rho}_1$	$\hat{\rho}_2$
5 min	0.5531	0.4186
30 min	0.5556	0.4310
60 min	0.5589	0.4756

Table 12: Mean of the GPH estimators of the counts generated from the LMSD model with Weibull (1, γ) distribution on different length of clock time intervals. We used $m = \sqrt{n}$ to compute the GPH estimators. We used the estimated LMSD parameters of the Boeing durations presented in Table 8 ($d = .3545$) for our simulations. We simulated 200 replications of the counts, each with sample size $n = 10,000$. Numbers in brackets represent standard errors of the mean of the GPH estimators.

$\triangle t$	mean($\hat{d}_{GPH}$)
5 min	0.3458 (0.004937)
30 min	0.3873 (0.009519)
60 min	0.3923 (0.009335)

References

[1] Andersen, Torben G. and Bollerslev, Tim, 1998, Answering the Skeptics: Yes, Standard Volatility Modles Do Provide Accurate Forecasts, *International Economic Review* 39, 885–905.

[2] Andersen, Torben G., Bollerslev, Tim, Diebold, Francis X. and Labys, Paul, 2001, *Journal of the American Statistical Association* 96, 453, 42–55.

[3] Andersen, Torben G., Bollerslev Tim, Diebold, Francis X. and Labys, Paul, 2003, Modelling and Froecasting Realized Volatility, *Econometrica* 71, 579–625.

[4] Ané, Thierry and Geman, H*élyette*, 2000, Order Flow, Transaction Colock, and Normality of Asset Returns, *The Journal of Finance*, 2259–2284.

[5] Bauwens, Luc and Giot, P., 2000, The Logarithmic ACD Model: An Application to the Bid/Ask Quote Process of Two NYSE Stocks, *Annales d'Economie et de Statistique* 60, 117–149.

[6] Bauwens, Luc and Veredas, David, 2004, The Stochastic Conditional Duration Model: A Latent Variable Model for the Analysis of Financial Durations, *Journal of Econometrics* 119, 381–412.

[7] Beran, Jan, 1994, *Statistics for Long-Memory Processes*, New York: Chapman and Hall.

[8] Berndt, E., Hall, B., Hall, R. and Hausmann, J., 1974, Estimation and Inference in Nonlinear Structural Models, *Annals of Economic and Social Measurement* 3, 653–665.

[9] Bertelli, Stefano and Caporin, Massimiliano, 2002, A note on calculating autocovariances of long-memory processes, *Journal of Time Series Analysis* 23, 503–508.

[10] Bahasali, R. J. and Karavellas, D., 1983, Winer Filtering (with emphasis on frequency-domain approaches), *Handbook of Statistics*, Vol. 3, 1–19.

[11] Boehmer, Ekkehart, Saar, Gideon. and Yu, Lei. Lifting the Veil: An Analysis of Pre-trade Transparancy at the NYSE, *forthcoming in the Journal of Finance.*

[12] Bollerslev, Tim and Jubinski, Dan, 1999, Equity Trading Volume and Volatility: Latent Information Arrivals and Common Long-Run Dependencies, *Journal of Business and Economic Statistics* 17, 9–21.

[13] Breiman, Leo. On Some Limit Theorems Similar to the Arc-Sin Law, 1965, *Theory of Probability and Applications* 10, 323–331.

[14] Chen,Willa W., Hurvich, Clifford M. and Lu, Yi. On the Correlation Matirx of the Discrete Fourier Transform and the Fast Solution of Large Toeplitz System For Long-Memory Time Series, *Working Paper, Stern School of Business, New York University.*

[15] Clark, Peter K., 1976, A subordinated Stochastic Process Model with finite variance for speculative prices, *Econometrica* 41, 135–155.

[16] Daley, D. J. and Vesilo, R., 1997, Long-Range Dependence of Point Processes, with Queueing Examples, *Stochastic Processes and Their Applications* 70, 265–282.

[17] Daley, D. J., 1999, The Hurst Index of Long-Range Dependent Renewal Processes, *The Annals of Probability* 27, 2035–2041.

[18] Daley, D. J., Rolski, T. and Vesilo, R., 2000, Long-Range Dependent Point Processes and their Palm-Khinchin Distributions, *Advances in Applied Probability* 32, 1051–1063.

[19] Davidson, James, 2004, Moment and Memory Properties of Linear Conditional Heteroscedasticity Models, and a New Model, *Journal of Business and Economic Statistics* 22, 16–29.

[20] Deo, Rohit S. and Hurvich, Clifford M., 2003, On the Log Periodogram Regression Estimator of the Memory Parameter in Long Memory Stochastic Volatility Models, *Econometric Theory* 17, 686–710.

[21] Engle, Robert F. and Russell, J. R., 1998, Autoregressive Conditional Duration: A New Model For Irregularly Spaced Transaction Data, *Econometrica* 66, 1127–1162.

[22] Engle, Robert F. and Sun, Zheng. Forecasting Volatility Using Tick By Tick Data, 2003, *Working Paper, Stern School of Business, New York University*

[23] Ghysels, Eric and Jasiak, Joann, 1997, GARCH for Irregular Spaced Financial Data : the ACD-GARCH Model. *Studies in Nonlinear Dynamics and Econometrics* 2, 132-149.

[24] Grammig, Joachim and Wellner, Marc, 1999, Non-monotonic Hazard Functions and the Autoregressive Conditional Duration Model. *The Journal of Econometrics* 2, 132-149.

[25] Harvey, Andrew, 1983, Long Memory in Stochastic Volatility, *Forecasting Volatility in Financial Markets.* J. Knight and S. Satchell, 307-320. Oxford: Butterworth-Heineman.

[26] Hill, Bruce M., 1975, A Simple General Approach to Inference about the Tail of A Distribution, *Annals of Statistics* 3, 1163-1174.

[27] Hosoya, Yuzo, 1997, A Limit Theory for Long-Range Dependence and Statistical Inference on Related Models, *Annals of Statistics* 25, 105–137.

[28] Hurvich, Clifford M. and Betrao, Kaizo I., 1993, Asymptotics for the Low-Frequency Ordinates of the Periodogram of a Long-Memory Time Series, *Journal of Time Series Analysis* 14, 455–472.

[29] Hurvich, Clifford M., Deo, Rohit S. and Brodsky, Julia, 1998, The Mean Squared Error of Gewek and Porter-Hudak's estimator of the Memory Parameter of a Long-Memory Time Series, *Journal of Time Series Analysis* 19, no. 1, 20–46.

[30] Hurvich, Clifford M. and Deo, Rohit S., 1999, Plug-in Selection of the Number of Frequencies in Regression Estimates of the Memory Parameter of A Long Memory Time Series. *Journal of Time Series Analysis* 20, 331–341.

[31] Jasiak, Joann, 1998, Persistence in Intertrade Durations, *Finance,* 19, 166–195.

[32] Jones, Charles M., Kaul, Gautam and Lipson, Marc L., 1994, Transactions, Volume, and Volatility, *The Review of Financial Studies* 7, 631–651.

[33] Liu, Ming, 2000, Modeling Long Memory in Stock Market Volatility, *Journal of Econometrics* 99, 139–171.

[34] Mikosch, Thomas and Stărică, Cătălin, 2003, Long-Range Dependence Effects and ARCH Modeling, *Theory and Applications of Long-Range Dependence* 439–459, Birkhauser Boston.

[35] Nelson, Daniel B., 1990, Stationarity and Persistance in the GARCH(1,1) Model, *Econometric Theory* 6, 318–334.

[36] Parke, William R., 1999, What is Fractional Integration, *Review of Economics and Statistics* 81, 632–638.

[37] Resnick, Sidney, 1997, Heavy Tail Modeling and Teletraffic Data, *Annals of Statistics* 25, 1805–1869.

[38] Whittle, Peter, 1962, Gaussian Estimation in Stationary Time Series, *Bull. Inst. Internat. Statist* 39, 105–129.