SCHOOL OF ECONOMICS AND FINANCE

Discussion Paper 2008-08

Yet Another Autoregressive Duration Model: The ACDD Model

Nagaratnam Jeyasreedharan
David E Allen
Joey Wenling Yang

ISSN 1443-8593
ISBN 978-1-86295-512-7

Yet Another Autoregressive Duration Model: The ACDD Model

by

Nagaratnam Jeyasreedharan, University of Tasmania*

David E Allen, Edith Cowan University and

Joey Wenling Yang, University of Western Australia

Abstract

This paper contains three novelties. First, duration (intra-trade intervals) is assigned positive and negative values, based on whether it was a bid-trade or ask-trade. Second, as the transformed durations are no longer "asymmetric", a more GARCH-like alternative to the ACD model is proposed for modelling durations. The alternative model is called the ACDD model, where the DD stands for Directional Duration. Third, using the alternative ACD formulation, persistence in durations is addressed both in the mean and the variance equations of the ACDD model.

* Corresponding author.

1 Introduction

High-frequency financial time series have become widely available during the past decade or so. Records of all transactions and quoted prices are readily available in pre-determined formats from many stock exchanges. An inherent feature is that such data are irregularly spaced in time. Several approaches have been taken to address this feature of the data.

The seminal work originated with Engle and Russell (1998), where the time between events (trades, quotes, price changes etc.) or durations are the quantities being modeled. These authors proposed a class of models called the Autoregressive Conditional Duration, or ACD, models, where conditional (expected) durations are modeled in a fashion similar to the way conditional variances are modeled using ARCH and GARCH models of Engle (1982) and Bollerslev (1986).

ACD models and GARCH models share several common features, ACD models being commonly viewed as the counterpart of GARCH models for duration data. Both models rely on a similar economic motivation following from the clustering of news and financial events in the markets. The autoregressive ACD model allows for capturing the duration clustering observed in high frequency data, i.e., small (large) durations being followed by other small (large) durations in a way similar to the GARCH model accounting for the volatility clustering. Just as a low order GARCH model is often found to suffice for removing the dependence in squared returns, a low order ACD model is often successful in removing the temporal dependence in durations.

Following the GARCH literature, a number of extensions to the original linear ACD model by Engle and Russell (1998) have been suggested. These include the logarithmic ACD model of Bauwens and Giot (2000), and the threshold ACD model of Zhang, Russell and Tsay (2001). The distribution associated with the conditional durations has also been suggested to have several different shapes. Examples include the the exponential and Weibull distributions as in Engle and Russell (1998), and the Burr and generalized gamma distributions suggested by Grammig and Maurer (2000) respectively.

However, a crucial assumption for obtaining QML consistent estimates of the ACD model and its extensions is that the conditional expectation of durations is correctly specified. The QML estimation yields consistent estimates and the inference procedures in this case are straightforward to implement, but this comes at the cost of efficiency. *In practice, fully efficient ML estimates might be preferred.*

The inherent limitations in the ACD model and its extensions to date have been a direct consequence of the positive asymmetric density assumed for ε_i in all these models as time between successive trades are positive (see Hautsch (2004)). Furthermore, the assumption of iid innovations may be too strong and inappropriate for describing the behaviour of some financial durations. Empirical studies based on the linear ACD model often reveal persistence in durations as the estimated coefficients on lagged variables add up nearly to one. Moreover, many financial duration series show a hyperbolic decay, i.e., significant autocorrelations up to long lags. This suggests that a better fit might be obtained by accounting for longer term dependence in durations. Indeed, the standard ACD model imposes an exponential decay pattern on the autocorrelation function typical for stationary and invertible ARMA processes. This may be completely inappropriate in the presence of long memory processes.

In this paper we provide and slightly different approach to work originated by Engle and Russell (1998). We propose an alternative definition of durations, where positive durations depict “ask-durations” and negative durations depict “bid-durations”. This approach enables the resultant error density to be symmetrical. The resultant model is called the Autoregressive Conditional Directional Duration (ACDD) model.

2 The basic ACDD model

The time series of arrival times or durations between successive occurrences of certain events associated with the trading process can be defined in a number of ways. Examples include the time between successive trades, the time until a price change occurs or until a pre-specified number of shares or level of turnover has been traded. We define directional duration as the time between successive trades where the trades at the ask-quotes are positive and at the bid-quotes are negative. In doing so we are able to differentiate between the arrival times of bid and ask-quotes.

The basic ACDD model relies on a linear parameterization of the conditional duration, ψ_i, which depends on p past absolute directional durations, $\left|\delta_{i-j}\right|$ and q past conditional durations, ψ_{i-j} as defined by:

$$\psi_i = \omega + \sum_{j=1}^{p} \alpha_j \left|\delta_{i-j}\right| + \sum_{j=1}^{q} \beta_j \psi_{i-j} \tag{1.1}$$

where $\delta_i = \gamma_i(t_i - t_{i-1})$ are the directional durations and $\gamma_i = 1$ for an ask-durations and $\gamma_i = -1$ for bid-durations with t being the trade times. To ensure positive conditional durations for all possible realizations, sufficient but not necessary conditions are $\omega > 0, \alpha \geq 0, \beta \geq 0$. The main assumption behind ACDD model is that the standardized directional durations,

$$\varepsilon_i = \frac{\delta_i}{\psi_i}, \tag{1.2}$$

are independent and identically distributed (iid) with $E(\varepsilon_i) = 0$ and $E(\varepsilon_i^2) = 1$. Directional durations thus defined enable symmetrically distributed innovation errors to be assumed as shown in Figure 1 below. It can be seen that the directional durations are both positive and negative, whereas standard durations have positive support.

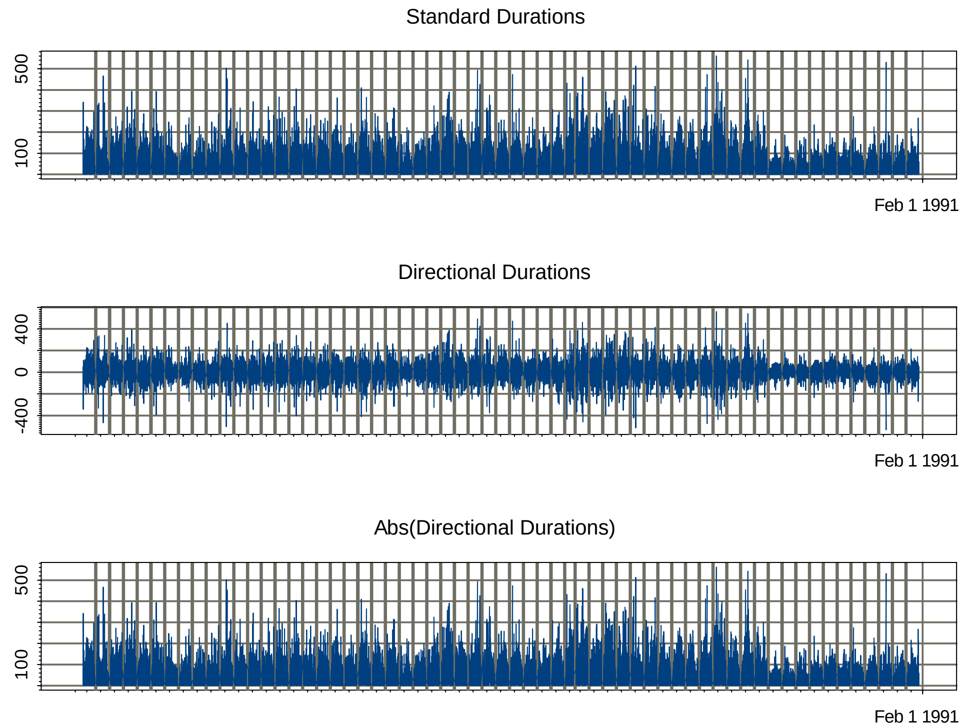

Figure 1: Standard, Directional and |Directional| Durations in seconds (IBM data)

Let $f(\varepsilon,\theta_\varepsilon)$ be the density function for ε with parameters θ_ε. A natural choice convenient for estimation will be any family of suitable symmetrical distributions. We adopt the generalized error distribution (GED) family proposed by Nelson (1991) to capture the fat tails, if any, in the error terms. If a random variable ε_i has a GED with mean zero and unit variance, the PDF of ε_i is given by:

$$f(\varepsilon_i)=\frac{\upsilon\, exp\left[-(1/2)\left|\varepsilon_i/\lambda\right|^{\upsilon}\right]}{\lambda\bullet 2^{(\upsilon+1)/\upsilon}\Gamma(1/\upsilon)} \tag{1.3}$$

where

$$\lambda=\left[\frac{2^{-2/\upsilon}\Gamma(1/\upsilon)}{\Gamma(3/\upsilon)}\right]^{1/2} \tag{1.4}$$

and υ is a positive parameter governing the thickness of the tail behaviour of the distribution. When $\upsilon = 2$ the above PDF reduces to the standard normal PDF; when $\upsilon < 2$, the density has thicker tails than the normal density; when $\upsilon > 2$, the density has thinner tails than the normal density. When the tail thickness parameter $\upsilon = 1$, the PDF of the GED reduces to the PDF of a double exponential distribution (analogous to the exponential distribution in the basic ACD model of Engle and Russell (1998)).

Based on the above PDF, the log-likelihood function of ACDD model with GED errors can be constructed as such maximum likelihood (ML) estimators for the ACDD parameters can be obtained as opposed to the quasi-maximum likelihood (QML) estimators used in ACD models. Furthermore, the re-definition of durations to bid- and ask-based durations enables us to fully adopt GARCH formulations, meaning both the mean equation and the variance equation in the standard GARCH model and its extensions can be utilitsed.

The autocorrelation properties of standard durations led to the development of the original ACD model where the error terms were assumed to be iid. Distributions defined on positive support typically imply a strict relationship between the first moment and higher order moments and do not disentangle the conditional mean and variance function. For example, under the exponential distribution, all higher order moments directly depend on the first moment. Hence the inherent restrictiveness or inflexibility encountered with ACD models.

In fact, a battery of tests on the IBM data (not reported in this paper) reveal that the durations, apart from being autocorrelated and having arch effects, also exhibit long range dependence (long memory) and non-stationarity. Thus, whist crucial for the ACD model and its extensions the “assumptions of iid innovations may be too strong and inappropriate for describing the behaviour of trade durations” (see Pacurar (2006)).

In the above ACDD formulation, the directional durations still exhibit long range dependence (long memory) and non-stationarity in addition to being autocorrelated and arched. In addition, diurnal and day-of-the-week (DoW) components have also been observed in duration data (see …). To address these stylised characteristics and as several ‘trend-generating’ mechanisms may occur simultaneously we include a SEMIFAR-type mean equation into the ACDD model.

3 The SEMIFAR-ACDD model

Semiparametric fractional autoregressive (SEMIFAR) models (see Beran and Feng (2002)) have been introduced for modelling different components in the mean function of a financial time series simultaneously, such as nonparametric trends, stochastic nonstationarity, short- and long-range dependence as well as antipersistence. SEMIFAR includes ARIMA and FARIMA processes (see Hosking (1981); Granger and Joyeux (1980)).

Let $d = (-0.5, 0.5)$ be the fractional differencing parameter, $m \in \{0,1\}$ be the integer differencing parameter, L be the lag or backshift operator, $\phi(L)$ and $\theta(L)$ be the lag polynomials in L with no common factors and all roots outside the unit circle and ε_i be white noise, then the SEMIFAR model can be defined as (see Feng, Beran and Yu (2007)):

$$\phi(L)(1-L)^d [(1-L)^m y_i - g(\tau_i) = \theta(L)\varepsilon_i \qquad (1.5)$$

where $\tau_i = t_i / n$ is the rescaled time.

Similarly, in the SEMIFAR–ACDD model, the mean equation in is defined as follows:

$$\phi(L)(1-L)^d [(1-L)^m \delta_i - g(\tau_i) = \theta(L)\zeta_i \qquad (1.6)$$

with the duration equation defined by:

$$\psi_i = \omega + \sum_{j=1}^{p} \alpha_j \left|\zeta_{i-j}\right| + \sum_{j=1}^{q} \beta_j \psi_{i-j} \qquad (1.7)$$

where ζ_i is then the adjusted directional duration. To ensure positive conditional durations for all possible realizations, sufficient but not necessary conditions are that $\omega > 0, \alpha \geq 0, \beta \geq 0$. The main assumption behind SEMIFAR-ACDD model is that the standardized directional durations,

$$\varepsilon_i = \frac{\zeta_i}{\psi_i}, \tag{1.8}$$

are independent and identically distributed (iid) with $E(\varepsilon_i) = 0$ and $E(\varepsilon_i^2) = 1$.

4 Methodology

Based on the SEMIFAR-ACDD model above and the asymptotic results for the SEMIFAR-GARCH formulation obtained by Feng, Beran and Yu (2007), the following algorithm in S-PLUS is proposed for the practical implementation of the SEMIFAR–ACDD model:

(a) Carry out data-driven SEMIFAR fitting using algorithm, AlgB in Beran and Feng (2002) to the square-root of observations to obtain $g(\tau), \phi(L)$ and $\theta(L)$;

(b) Calculate the residuals $\xi_i = \delta_i - g(\tau_i)$ and invert ξ_i using $\phi(L)$ and $\theta(L)$ into $\hat{\zeta}_i$, the estimates of ζ_i;

(c) Estimate the ACDD model using S-PLUS/GARCH software taking the residuals of the SEMIFAR model in (b) above.

The best SEMIFAR-ACDD model is then determined as follows:

(a) For $p = 1, p_{max}$ and $q = 1, q_{max}$ estimate ACDD(p,q) and calculate BIC(p, q);

(b) Choose the ACDD(p,q) model that minimizes the BIC. We obtain the fitted ACDD model, where the BIC in S-PLUS will be used, which is given by:

$$BIC(p, q) = -2 \bullet \log(\text{maximized likelihood}) + (\log n)(p + q + 2) \tag{1.9}$$

The estimated parameter vectors for the SEMIFAR and the ACDD models are asymptotically independent (see Feng, Beran and Yu (2007)). With the trend function in the SEMIFAR–ACDD model, it is inconvenient to select the two equations (1.6 and 1.7) at the same time. Thus they are selected separately. The best-fit SEMIFAR model is chosen from $r=0,1,2$ and $s=0$ and the best-fit ACDD model selected from $p = 0,1,2$ and $q = 0, 1, 2$, by means of the minimal BIC.

5 The Data

The dataset in this paper is the IBM data used in the seminal paper titled "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data" by Engle and Russell (1998) and downloaded from http://weber.ucsd.edu/~mbacci/engle. This is to enable direct comparisons to be made with the basic ACD model and data.

A total of 60328 transactions were recorded for IBM over the 3 months of trading on the consolidated market from November 1990 through January 1991. As per the seminal paper, two days from the three months were deleted. A halt occurred on 23rd November and a more than one hour opening delay occurred on 27th December. Following Engle and Russell (1998) the first half hour of the trading day (i.e. trades before 10.00am) is omitted. This is to avoid modeling the opening of the market which is characterized by a call auction followed by heavy activity. The dynamics are likely to be quite different over this period. Furthermore, the call auction transactions are not recorded at the same time each morning.

In addition, all trades after 4.00pm were also omitted. After omitting these two days and times, of the original 60328 transactions there were 51356 observations left. Of the transactions occurring at non-unique trading times, nearly all of them corresponded with zero price movements. Engle and Russell (1998) suggest that these transactions may reflect large orders that were broken up into smaller pieces. As it is not clear that each piece should be considered a separate transaction, the zero-second durations were considered to be a single transaction and were deleted from the data set as in Engle and Russell (1998). After all the adjustments to the data, 46052 observations were collated.

In their seminal paper, Engle and Russell (1998) reported 46091 final IBM observations. This is probably a typo (it should have been 46051) as their other reported summary statistics for the same dataset was identical with a mean duration of 28.38 seconds, maximum duration of 561 seconds and standard deviation of 38.41 seconds. We ended up with 46052 observations, the extra 1 observation is due to the way we computed the durations.

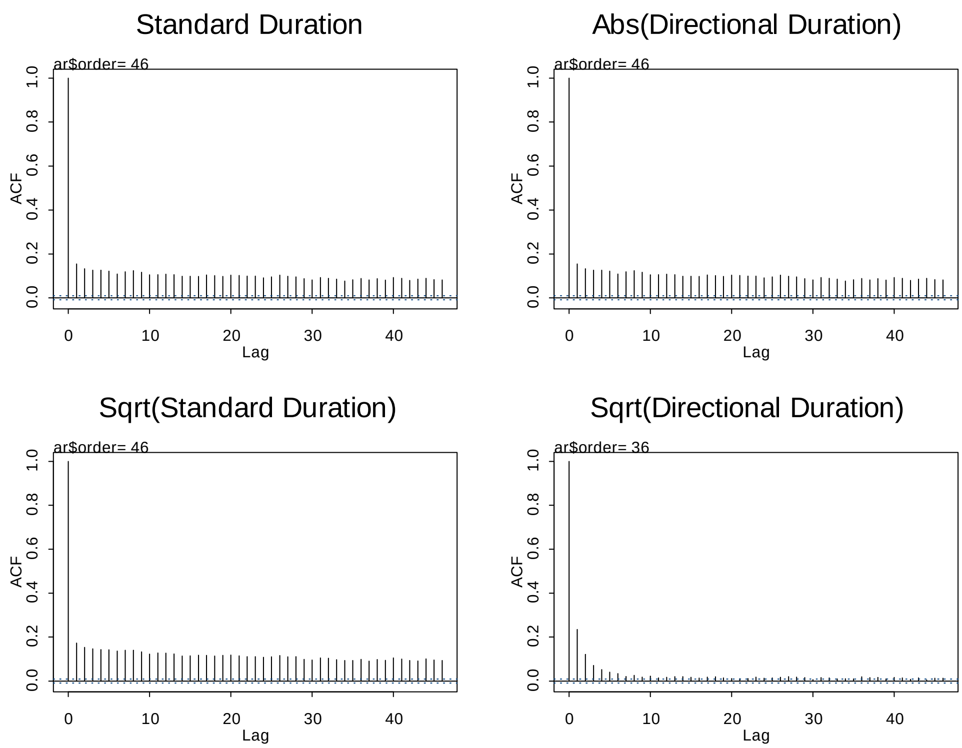

Figure 2 ACFs of functions of Standard and Directional Durations

It can be seen from Figure 2 that the autocorrelation properties of the standard duration and the absolute directional durations are identical by definition. However, the ACF plot of the directional durations exhibit AR(MA) effects.

Herein the difference between standard and directional durations: first order dependencies are fundamentally different. The inclusion of the SEMIFAR equation to ACDD model ensures that the error residuals are IID. However, the second order dependencies are identical. Consequently, the conditional duration equation remains unchanged.

6 Results

Figure 3 exhibits the adjusted durations (square-root) for both the standard and directional durations. The standard durations were adjusted using the trend function using the SEMIFAR(0,0) model. Though this adjustment is different to that done by Engle and Russell (1998) it is similar.

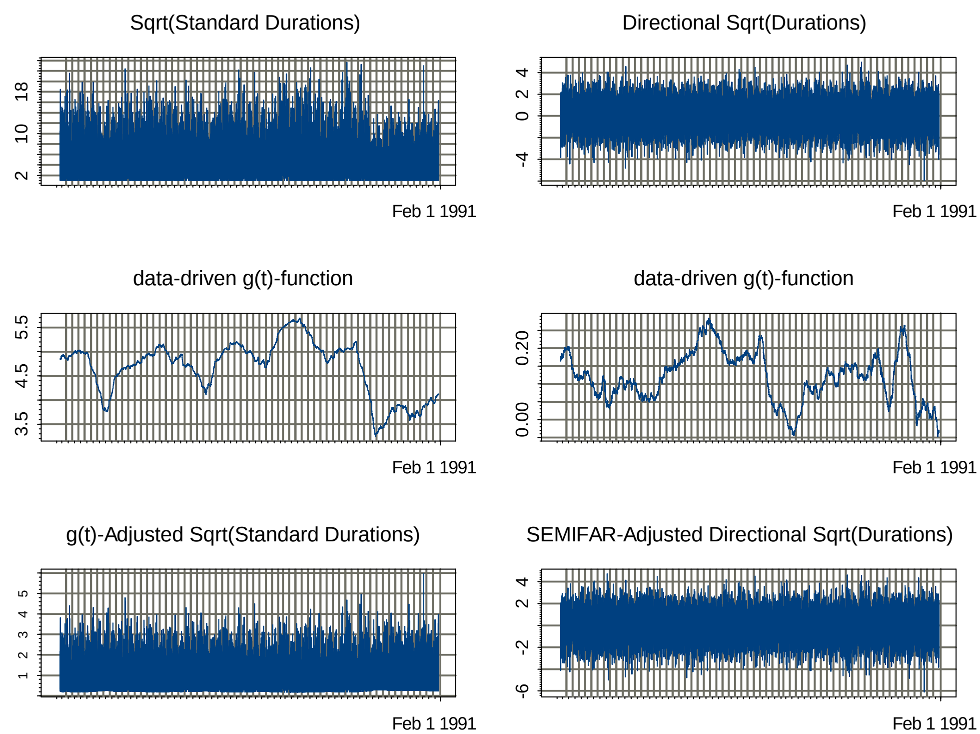

Figure 3: Adjusted Durations (square-root)

As mentioned earlier, the AlgB in Beran and Feng (2002) was used for estimating the SEMIFAR portion of the model. The trend was estimated by local linear regression using a kernel as the weight function. For the shortmemory effects, only an AR component was considered. The SEMIFAR model is chosen from r = 0,1,2.

	lb-stat	**lm-stat**	**rs-stat**	**kpss-stat**
AdjsqSD	10725.0039	2333.2642	1.2841	0.0444
p.value	<0.01	<0.01	>0.1	>0.1
AdjsqDD	30.7122	1733.0241	0.6662	0.0117
p.value	>0.1	<0.01	>0.1	>0.1

Table 1: Adjusted Duration Statistics (square-root)

The adjusted standard durations (ASD) are still contain autocorrelations and arch effects. The directional durations are based on the adjusted durations and further adjusted using the SEMIFAR model. As can be seen from Table 1 the ADD series have no autocorrelation and long memory, but still have arch effects. All the ADD statistics are however significantly lower than that of the ASD statistics. The ACD, ACDD and SEMIFAR-ACDD models are then fitted and tested. The ACD and ACDD coefficients for the duration equations are listed in Table 2. The duration parameters are of the same order. The GED parameter values for all models are greater than the value of 2 for a normal distribution.

	"ACD"	**"ACDD"**	**"SEMIFAR-ACDD"**
"A"	0.0195	0.0174	0.018
"ARCH(1)"	0.066	0.0622	0.057
"GARCH(1)"	0.9191	0.9242	0.9297
"dist.par"	2.9635	2.9423	2.1682

Table 2: ACDD coefficients and ged-parameter

The ACD and ACDD statistics are displayed in Table 3. The ACD and ACDD models have dependent residuals, arch effects and non-normal distributions. The SEMIFAR-ACDD MODEL has no dependency, still has some arch effects and non-normality in the standardized residuals.

	lb-stat	**lm-stat**	**jb-stat**
ACD	42.1238	35.2505	10643.7801
p.value	0.0002	0.0023	0
ACDD	4112.4265	33.0222	780.0414
p.value	0	0.0047	0
SEMIFAR-ACDD	7.2107	37.4595	255.0674
p.value	0.9515	0.0011	0

Table 3: ACD/ACDD statistics

The Ljung-Box statistic for the ACD model was 42.12 is of the same order of the reported statistic (lb-stat=38.08) in Engle and Russell (1998).

The SEMIFAR-ACDD model is chosen from p = 0,1,2 and q = 0,1,2 by means of BIC/AIC/LL. The best ACDD model fitted is the SEMIFAR-ACDD(1,2) as shown by the lowest values in Table 4.

	acdd01	**acdd02**	**acdd10**	**acdd11**	**acdd12**	**acdd20**	***acdd21***	**acdd22**
AIC	141633	143134	140967	139181	139173	140722	***139169***	140768
BIC	141659	143169	140993	139216	139217	140757	***139213***	140820
LL	-70814	-71563	-70481	-69587	-69582	-70357	***-69580***	-70378

Table 4: BIC, AIC and LL-values for ACDD models

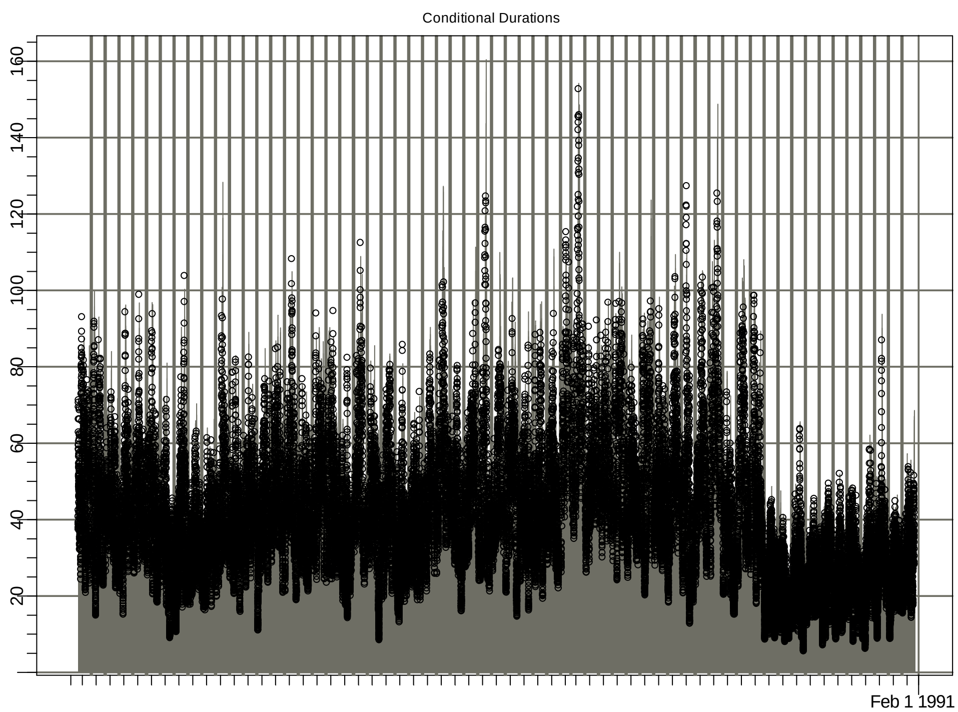

Figure 4: ACD and SEMIFAR-ACDD conditional durations

In Figure 4 above, the points in black depict the conditional durations from the SEMIFAR-ACDD model, whereas the grey lines depict the conditional durations from the ACD model. The results are similar but not identical.

7 Conclusion

This paper modifies the ACD approach to a SEMIFAR–ACDD model so that trends and long memory in durations can be modelled using a more parsimonious parameterisation. A semi-parametric estimation procedure is proposed for the trend function. Asymptotic results on the SEMIFAR-GARCH model as reported by Feng, Beran and Yu (2007) are extended to the SEMIFAR-ACDD model.

The important property that the estimates of the FARIMA and ACDD parameter vectors are independent of each other, allow us to apply the data-driven SEMIFAR algorithms to estimate the trend and the FARIMA parameters in the SEMIFAR–ACDD model. The ACDD parameters from the approximated ACDD innovations are calculated by inverting the final residuals.

The results show that the proposed algorithm and model can be used to model dependencies in durations data. Further possible extensions to the model include leverage effects and the range of GARCH-type extensions.

References

Bauwens, L, and P Giot, 2000, The Logarithmic ACD Model: an Application to the Bid-Ask Quote Process of Three NYSE Stocks, *Annales d'Économie et de Statistique* 60, 117-149.

Beran, J, and Y Feng, 2002, Iterative plug-in algorithms for SEMIFAR models—definition, convergence and asymptotic properties, *Journal of Computational Graphical Statistics* 11, 690-713.

Beran, J, and Y Feng, 2002, SEMIFAR models—a semiparametric framework for modelling trends, long-range dependence and nonstationarity, *Computational Statistics and Data Analysis* 40, 393-419.

Bollerslev, T, 1986, Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* 31, 307-327.

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

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

Feng, Y, J Beran, and K Yu, 2007, Modelling financial time series with SEMIFAR-GARCH model. , 18, 395-412., *IAM Journal of Managment Mathematics* 18, 395-412.

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

Granger, C W J, and R Joyeux, 1980, An introduction to long-memory time series models and fractional differencing, *Journal of Time Series Analysis* 1, 15-30.

Hautsch, N, 2004. *Modelling irregularly spaced financial data : theory and practice of dynamic duration models* (Springer, London).

Hosking, J R M, 1981, Fractional differencing, *Biometrika* 68, 165-176.

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

School of Economics and Finance Discussion Papers

2008-01 Calorie Intake in Female-Headed and Male Headed Households in Vietnam, **Elkana Ngwenya**

2008-02 Determinants of Calorie Intake in Widowhood in Vietnam, **Elkana Ngwenya**

2008-03 Quality Versus Quantity in Vertically Differentiated Products Under Non-Linear Pricing, **Hugh Sibly**

2008-04 A Taxonomy of Monopolistic Pricing, **Ann Marsden** and **Hugh Sibly**

2008-05 Vertical Product Differentiation with Linear Pricing, **Hugh Sibly**

2008-06 Teaching Aggregate Demand and Supply Models, **Graeme Wells**

2008-07 Demographic Demand Systems with Application to Equivalence Scales Estimation and Inequality Analysis: The Australian Evidence", **Paul Blacklow**, **Aaron Nicholas** and **Ranjan Ray**

2008-08 Yet Another Autoregressive Duration Model: The ACDD Model, **Nagaratnam Jeyasreedharan, David E Allen** and **Joey Wenling Yang**

2008-09 Substitution Between Public and Private Consumption in Australian States, **Anna Brown** and **Graeme Wells**

2007-01 Dietary Changes, Calorie Intake and Undernourishment: A Comparative Study of India and Vietnam, **Ranjan Ray**

2007-02 A Re-examination of the Real Interest Parity Condition Using Threshold Cointegration, **Arusha Cooray**

2007-03 Teaching Aggregate Demand and Supply Models, **Graeme Wells**

2007-04 Markets, Institutions and Sustainability, **Ella Reeks**

2007-05 Bringing Competition to Urban Water Supply, **Hugh Sibly** and **Richard Tooth**

2007-06 Changes in Indonesian Food Consumption Patterns and their Nutritional Implications, **Elkana Ngwenya** and **Ranjan Ray**

2007-07 The Term Spread and GDP Growth in Australia, **Jacob Poke** and **Graeme Wells**

2007-08 Moving Towards the USDA Food Guide Pyramid Food: Evidence from Household Food Group Choice in Vietnam, **Elkana Ngwenya**

2007-09 The Determinants of the Quantity-Quality Balance in Monopoly, **Hugh Sibly**

2007-10 Rationing Recreational Access to Wilderness and Other Natural Areas, **Hugh Sibly**

2007-11 Yet Another Trading Simulation: The Nonimmediacy Model, **Nagaratnam Jeyasreedharan**

Copies of the above mentioned papers and a list of previous years' papers are available on request from the Discussion Paper Coordinator, School of Economics and Finance, University of Tasmania, Private Bag 85, Hobart, Tasmania 7001, Australia. Alternatively they can be downloaded from our home site at http://www.utas.edu.au/ecofin ||