

A Semiparametric Conditional Duration Model*

Mardi Dungey[†] Xiangdong Long[‡] Aman Ullah[§] Yun Wang[¶]

April, 2014

ABSTRACT

We propose a new semiparametric autoregressive duration (SACD) model, which incorporates the parametric and nonparametric estimators of the conditional duration in a multiplicative way. Asymptotic properties for this combined estimator are presented. Empirical applications to the transaction duration of the U.S. 2-Year Treasury note show the outperformance of our SACD models over parametric ACD models.

Key Words: Duration, Nonparametric Estimator, Semiparametric Model.

JEL Classification: C3, C5, G0.

*We sincerely thank Badi H. Baltagi and the referee for their many insightful comments and suggestions that lead to a substantial improvement of the presentation. Mardi Dungey acknowledges funding from ARC DP0984994. Yun Wang acknowledges "the Fundamental Research Funds for the Central Universities" in UIBE (13QNJJX03) and "Program for Innovative Research Team" in UIBE (Grant CXTD4-01).

[†]Tasmanian School of Business and Economics, University of Tasmania, Australia; Centre for Financial Analysis and Policy, University of Cambridge, UK; Centre for Applied Macroeconomic Analysis, Australian National University, Australia; Email: mardi.dungey@utas.edu.au.

[‡]Bank Of Communications Schroder Fund Management Co. Ltd; Email: xiangdong.long2008@googlemail.com.

[§]Department of Economics, University of California, Riverside, CA, 92521, USA; Email: aman.ullah@ucr.edu.

[¶]Corresponding author: School of International Trade and Economics, University of International Business and Economics, Beijing, China, 100029; Email: wyuncolor@gmail.com.

1 Introduction

The increased availability of high frequency data has motivated a surge of research interest in theoretical and empirical financial market microstructure. Information on the waiting time between consecutive transactions, quote updates, price changes and order arrival across a variety of asset types is an important element in developing an understanding of the spread of public and private information (Goodhart and O'Hara, 1997; Madhavan, 2000).

Financial data reveal strong evidence for clustering in time of both trades and quote changes, in a similar way to the clustering of volatility. This observation led Engle and Russell (1998) to combine transition analysis and ARCH models to develop the autoregressive conditional duration (ACD) model. The ACD model treats transaction arrival time as a random variable of a dependent point process, whose conditional intensity depends on past durations. This model has been generalized in a number of ways, including the log-ACD model of Bauwens and Giot (2000), the Threshold ACD model of Zhang et al. (2001), and the mixture ACD model by Hujer and Vuletic (2007). Alternative distribution processes considered for the transactions arrivals include the generalized gamma and the Burr distribution used in Lunde (2000), and Weibull and exponential distributions nested in Grammig and Maurer (2000). In parallel with the ACD model, there are a closely related but econometrically distinct class of stochastic duration models. Bauwens and Hautsch (2006) review the literature on dynamic duration and intensity process, also see an extensive survey on ACD models by Pacurar (2008).

The common shortcomings of parametric ACD models are potential misspecifications of the functional form and distribution function. It is well known that in these situations the parametric estimators of ACD models will become inconsistent and/or inefficient. Therefore, in this paper we propose a new semiparametric ACD (SACD) model inspired by the univariate and multivariate volatility works of Mishra et al. (2010) and Long et al. (2011), respectively. The SACD is consistent, and it combines parametric and nonparametric estimators of conditional duration in a multiplicative way: first modelling ACD parametrically, and then modelling the conditional mean of the standardized parametric residuals nonparametrically. The estimate of the latter serves as a nonparametric correction factor for the parametric ACD estimator. To our knowledge, this is the first combined estimator of conditional duration.

To illustrate the consequences of using the SACD we apply parametric and the SACD duration estimators to the transaction data in the 2 year U.S. Treasury note market. Model evaluation is carried out in terms of Ljung-Box Q-statistic and the mean squared error. The results show the clear advantages of the SACD approach.

The paper is organized as follows. Section 2 discusses a number of parametric duration models. In section 3 we propose the semiparametric conditional duration model. Section 4

provides the empirical application to the U.S. Treasury data. Finally, section 5 concludes.

2 Parametric ACD Models

Denote t_i as the transaction time for i th transaction so that, $y_i = t_i - t_{i-1}$ is the duration, or waiting time, between i th and $(i-1)$ th transactions. Engle and Russell (1998) propose the ACD model for the diurnally adjusted y_i , which has a similar structure to the familiar GARCH(p, q) model,

$$y_i = \psi_{\theta,i} \varepsilon_i, \quad (2.1)$$

$$\psi_{\theta,i} = \omega + \sum_{j=1}^p \alpha_j y_{i-j} + \sum_{j=1}^q \beta_j \psi_{\theta,i-j} \quad (2.2)$$

where ε_i represents the *i.i.d.* arrival process, which is assumed to have a constant expectation, $E(\varepsilon_i) = 1$, with probability distribution $P(\varepsilon_i \leq \varepsilon) = F(\varepsilon)$, and $\theta = (\alpha, \beta)$. The conditional duration for a transaction, given available information, is given by the parameter of interest, $\psi_{\theta,i}$, that is $E(y_i | \mathcal{F}_{i-1}) = \psi_{\theta,i}$, \mathcal{F}_{i-1} represents all the information up to period $i-1$. Hence, the expected value of the duration is given as

$$E(y_i) = E(\psi_{\theta,i}) E(\varepsilon_i) = \frac{\omega}{1 - \sum_{j=1}^p \alpha_j - \sum_{j=1}^q \beta_j}. \quad (2.3)$$

The positiveness of the conditional duration is assured by imposing restrictions on the parameters ω, α_j and β_j .

Rather than modeling the intensity function, the ACD model considers the dynamics of $\psi_{\theta,i}$. An ACD(p, q) model can be equivalently expressed as an ARMA($\max(p, q), q$) model,

$$y_i = \omega + \sum_{j=1}^{\max(p,q)} (\alpha_j + \beta_j) y_{i-j} - \sum_{j=1}^q \beta_j \eta_{i-j} + \eta_i \quad (2.4)$$

where $\eta_i = y_i - \psi_{\theta,i}$.

The parametric ACD model has been extended in two directions; by changing the dynamic functional forms for $\psi_{\theta,i}$ or by assuming more flexible distributions for ε_i , or a combination of both. Bauwens and Giot (2000) propose the Logarithmic ACD (Log-ACD) model, which modifies equations (2.1) and (2.2) to be

$$y_i = e^{\phi_i} \varepsilon_i, \text{ where } P(\varepsilon_i \leq \varepsilon) = F(\varepsilon) \text{ and } E(\varepsilon_i) = 1 \quad (2.5)$$

$$\psi_{\theta,i} = \ln E(y_i | \mathcal{F}_{i-1}) \quad (2.6)$$

$$\psi_{\theta,i} = \omega + \sum_{j=1}^p \alpha_j \ln y_{i-j} + \sum_{j=1}^q \beta_j \psi_{\theta,i-j} \quad (2.7)$$

with no sign constraints on the parameters ω , α_j and β_j , while $|\sum_{j=1}^p \alpha_j + \sum_{j=1}^q \beta_j| < 1$ guarantees the covariance stationarity of $\ln y_i$. The advantage of this formulation is in retaining a positive $\psi_{\theta,i}$ in the presence of covariates. Neither exponential GARCH models nor the Log-ACD model have an explicit form for unconditional moments.

3 A Semiparametric ACD Model

Mishra et al. (2010) construct a two-step semiparametric estimator for conditional variance by multiplicatively combining a parametric estimator for conditional variance and nonparametric conditional variance estimator of standardized residuals. This is extended in Long et al. (2011) to a conditional covariance matrix modelling framework. Motivated by their work, we propose a two-stage SACD model $\psi_i = \psi_{\theta,i} \psi_{np,i}$, which has both parametric $\psi_{\theta,i}$ and nonparametric $\psi_{np,i}$ parts for conditional duration, where $\psi_{\theta,i} \in \mathcal{F}_{i-1}$ is specified by a parametric conditional duration model and $\psi_{np,i} \equiv E((y_i/\psi_{\theta,i}) | \mathcal{F}_{i-1})$ leads to the identity $E(y_i | \mathcal{F}_{i-1}) = \psi_{\theta,i} E((y_i/\psi_{\theta,i}) | \mathcal{F}_{i-1})$ analogous to Glad (1998), see also Martins-Filho et al. (2008). The misspecifications in the parametric model may be corrected by the nonparametric model where the correction factor $\psi_{np,i}$ degenerates to a constant if the parametric model is correctly specified. The conditional variable for the first stage parametric estimator $\hat{\psi}_\theta$ is a $d_1 \times 1$ vector $X_{1,i} \equiv X_{1,i}(\theta^0)$, and the $d_2 \times 1$ vector, $X_{2,i} \equiv X_{2,i}(\theta^0)$ is the state variable in the second stage of the nonparametric estimation. The true value of the finite dimensional parameter θ^0 is unknown, and replaced in estimation with its estimator $\hat{\theta}$, where the corresponding state variable is $\hat{X}_{j,i} \equiv X_{j,i}(\hat{\theta})$ for $j = 1, 2$. Mishra et al. (2010) discuss the properties of the $d \times 1$ disjoint union vector of $X_{1,i}$ and $X_{2,i}$, given by $X_i \equiv \{X_{1,i}^T(\theta^0), X_{2,i}^T(\theta^0)\}^T$ belonging to \mathcal{F}_{i-1} , emphasizing the minimal reducible dimension.

Let $r_i \equiv y_i \psi_{\theta^0}(x_1) / \psi_{\theta^0}(X_{1,i})$ and $x = \{x_1^T, x_2^T\}^T$, by the law of iterative expectation

$$\begin{aligned} E(r_i | X_i = x) &= \psi_{\theta^0}(x_1) E\left[E\left(\frac{y_i}{\psi_{\theta^0}(X_{1,i})} | \mathcal{F}_{i-1}\right) | X_{2,i} = x_2\right] \\ &= \frac{\psi_{\theta^0}(x_1)}{\psi_{\theta^0}(X_{1,i})} \psi_{\theta^0}(X_{1,i}) \psi_{np}(x_2) = \psi_i(x). \end{aligned}$$

The nonparametric regression of r_i on $X_{2,i}$ could estimate $\psi(x)$ by the Nadaraya-Watson (NW) method, the local linear method, or the local exponential method. The local exponential method solves both the boundary bias problem of the NW method and the nonnegativity ensuring challenge of the local linear method.

To summarize, in the first stage we estimate $\hat{\theta}$ for θ^0 in the parametric model, obtain the standardized residual $\hat{z}_i \equiv y_i / \hat{\psi}_{\hat{\theta}}(\hat{X}_{1,i})$; and then minimize the following loss function to

obtain $\widehat{\beta}$ in the second stage:

$$\widehat{\beta} \equiv \arg \min_{\beta} n^{-1} \sum_{i=1}^n \left\{ \widehat{r}_i - \Psi \left(\beta_0 + \sum_{j=1}^{d_2} \beta_j \left(\widehat{X}_{2,ij} - x_{2,j} \right) \right) \right\}^2 K_h \left(\widehat{X}_{2,i} - x_2 \right), \quad (3.1)$$

where $\beta \equiv (\beta_0, \beta_1, \dots, \beta_{d_2})^T \in \mathcal{R}^{d_2+1}$, $\widehat{r}_i \equiv \psi_{\widehat{\theta}}(x_1) \widehat{z}_i = y_i \psi_{\widehat{\theta}}(x_1) / \widehat{\psi}_{\theta^0}(X_{1,i})$, Ψ is a monotone function with at least two continuous derivatives on its support, $h \equiv (h_0, h_1, \dots, h_{d_2})^T$ is a $(d_2 + 1)$ vector of bandwidth, $K_h(u) \equiv \prod_{j=1}^{d_2} h_j^{-1} k(u_j/h_j)$, and $k(\cdot)$ is a symmetric density with compact support on \mathcal{R} . Our semiparametric estimator for conditional duration is $\widehat{\psi}(x) = \Psi(\widehat{\beta}_0)$.

We make the following remarks.

(1) It is not necessary that the parametric specification of $\psi_{\theta,i}$ is correct. The consistency of $\widehat{\theta}$ with an \sqrt{n} convergence rate for some pseudo-true parameters θ^0 does not assure the consistency of the estimated $\widehat{\psi}_{\theta,i}$.

(2) Dropping the first stage of the parametric estimator, and applying our nonparametric technique directly to $\{x_i\}_{i=1}^n$, will lead to the nonparametric estimator of conditional duration. However, a nonparametric estimation where x_i vector contains unobserved variables, as in the case of GARCH-type specifications, is not implemented. In this sense, the SACD estimator is more useful, where these unobserved variables are already estimated.

(3) There is an analogous relationship between ARCH models and ACD models. The theorems of QMLE properties of GARCH(1,1), such as Theorems 2 and 3 of Lumsdaine (1996) or Theorems 1 and 3 in Lee and Hansen (1994), carry over to the asymptotic properties of the EACD(1,1) model of Engle and Russell (1998). Similarly, the asymptotic properties of the semiparametric estimator of conditional variance in Mishra et al. (2010) apply to our SACD models.

To present the asymptotic properties of our model we follow Mishra et al. (2010) with appropriate changes in notation. First, let $L(X_{2,i} - x_2, \beta) \equiv \Psi \left(\beta_0 + \sum_{j=1}^{d_2} \beta_j \left(\widehat{X}_{2,ij} - x_{2,j} \right) \right)$ and $\dot{L}(x_2, \beta)$ and $\ddot{L}(x_2, \beta)$ represent the first and second derivatives of L with respect to x_2 and define $\kappa_{ij} \equiv \int u^i k(u)^j du$, for $i, j = 0, 1, 2$. We adopt assumptions A1 to A6 of Mishra et al. (2010) with the exception of three modifications. Assumption A1(i) is unnecessary to our application. We replace their assumption A2.(iii) and A3.(ii) as follows.

A2.(iii) The process $\{\varepsilon_i - 1\}$ is a stationary m.d.s such that $E(\varepsilon_i - 1 | \mathcal{F}_{i-1}) = 0$, $E(|\varepsilon_i|^{2v}) < \infty$ and $E(\|X_i\|^{2v}) < \infty$ for some $v > 1$.

A3.(ii) We replace $\sigma_p^2(\cdot)$ by $\psi_{\theta}(\cdot)$.

Consequently we can show convergence in probability of $\widehat{\beta}$ to β^0 and the asymptotic normality of our estimator.

Theorem 3.1 *Within the model structure and Assumptions A1 to A6, we have*

$$\widehat{\beta} \xrightarrow{p} \beta^0,$$

where β^0 is uniquely defined by $\psi(x) = L(0, \beta^0)$ and $\psi_\theta(x_1) \dot{\psi}_{np}(x_2) = \dot{L}(0, \beta^0)$.

Theorem 3.2 *Within the model structure and Assumptions A1 to A6, we have*

$$\begin{aligned} & \sqrt{n} \left(\prod_{i=1}^{d_2} h_i \right) \left\{ \widehat{\psi}(x) - \psi(x) - \frac{\kappa_{21}}{2} \text{tr} \left\{ D_h \left[\psi_\theta(x_1) \ddot{\psi}_{np}(x_2) - \ddot{L}(0, \beta^0) \right] \right\} \right\} \\ & \xrightarrow{d} N \left(0, \kappa_{02}^{d_2} \left(E \left(\varepsilon_i^2 | X_{2,i} = x_2 \right) - 1 \right) f^{-1}(x_2) \psi^2(x) \right), \end{aligned}$$

where $D_h \equiv \text{diag}(h_1^2, \dots, h_{d_2}^2)$, and $f(x_2)$ is the marginal density of $X_{2,i}$.

The proofs of the convergence in probability of $\widehat{\beta}$ to β^0 and the asymptotic normality of our estimator can be derived in a manner precisely analogous to the proofs in Mishra et al (2010), Theorem 2.1 and Theorem 2.2. These are available from the authors on request.

4 Empirical Examples

Here we apply the parametric models outlined in section 2 and the SACD specifications presented in Section 3 to transaction data on the 2 year constant maturity Treasury note and compare the outcomes. The parametric duration models include the Exponential ACD (EACD), Weibull ACD (WACD), Generalized Gamma ACD (GACD), and Weibull Log_ACD (Log-WACD); and the corresponding semiparametric estimators denoted as the SEACD, SWACD, SGACD, SLogWACD.

Throughout the applications, we adopt least squares cross validation method based on a grid search approach for obtaining optimal bandwidth, i.e. set $h = c\widehat{\sigma}n^{-1/(4+d_2)}$ as in Silverman (1986), where $\widehat{\sigma}$ is the sample standard deviation of X_2 , and c increments in intervals of 0.1 between 0.1 and 2.

Trading frequency data often displays diurnality. In equity markets this is well-known to have an inverted U-shape, representing high trading frequency at the beginning and end of the trading hours. Diurnality in bond markets is less well explored. Engle and Russell (1998) account for diurnality by decomposing expected duration into deterministic and stochastic components. Currently there is no dominant method to adjust for diurnality. To eliminate the calendar time effect, we adopt the cubic spline method of Engle and Russell (1998).

Data for the US Treasury markets is drawn from the Cantor-Fitzgerald eSpeed database which captures approximately half the secondary trade in US bonds for these maturities; see Dungey et al (2013). The 2 year US Treasury market is sampled for all trading days in the

month of May 2006 and diurnally adjusted using a cubic spline with hourly nodes. The bulk of the trade in this market occurs between 7:30EST and 16:00EST, but we discard the first half hour of morning trade. From this diurnalised data we consider the subsample of the final five trading days in the month (May 24, 25, 26, 30 and 31) comprising 3042 observations. The diurnalised data have a sample Ljung-Box $Q(25)$ statistic of 180.95, compared with a critical value of 37.65 (the raw data have a $Q(25)$ of 293.22). The state variable used in semiparametric estimation is the lagged value of the ratio of the diurnally adjusted durations to the parametric estimator for conditional duration. The (p, q) order of the models indicated via the *AIC* and *SBC* criteria are the EACD(1, 3), WACD(1, 3), GACD(1, 3) and LogWACD(1, 1).

The sheer number of transactions make the results difficult to present. Instead we present a one example day of the diurnalised and fitted durations, given in Figure 1. Figure 1(a) shows the diurnalised Treasury note data for May 24, 2006. Panels (b) to (e) of Figure 1 present the parametric and semiparametric fitted conditional durations for that day using each of the alternative distributional forms. Again it is clear from the example that the semiparametric estimates are able to provide a greater degree of volatility than the parametric approaches. This is also represented in the results for the MSE results presented in Table 1. The Ljung-Box tests suggest that all but the LogWACD models pass the model specification test at the 5 percent level. The semiparametric estimates provide an improvement of between 6.62% (WACD) and 9.02% (GACD) in the MSE for the corresponding models.

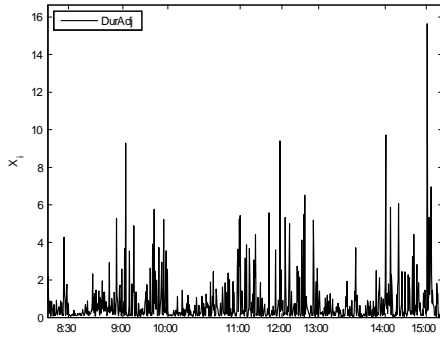
5 Conclusions

The developments in the literature on modelling price volatility have parallels in the literature on duration modelling. Building on this insight this paper has developed a semiparametric form of a number of parametric conditional duration models, developed from recent work by Mishra et al. (2010). The paper developed the asymptotic theory for the case of semiparametric estimation of durations in high frequency financial data. The application to a new sample of US Treasury market transactions, support the ability of the semiparametric specifications to obtain better coverage of the original durations than the parametric models. This contribution to the literature goes some way to the need to obtain models which are more able to capture the abrupt changes in duration observed in the data. We conjecture that the proposal of Bauwens et al. (2004) to improve duration modelling by focusing probability distributions with more mass on small, but not too small, durations, can be alternatively addressed in our framework by local bandwidth choice.

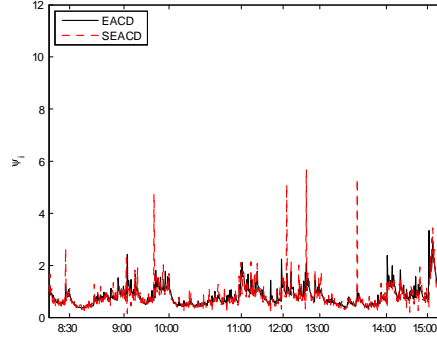
Figure 1: Duration data for 2 year US Treasury notes and fitted parametric (solid line) and semiparametric (dashed line) models

One day sample: May 24, 2006

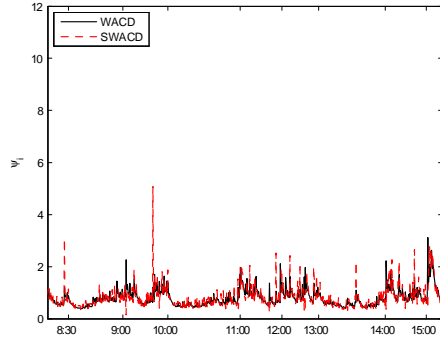
(a) Diurnalised data



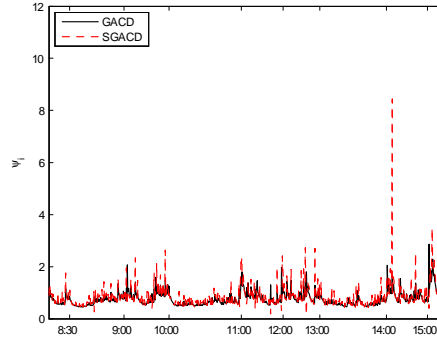
(b) EACD models



(c) WACD models



(d) GACD models



(e) LogWACD models

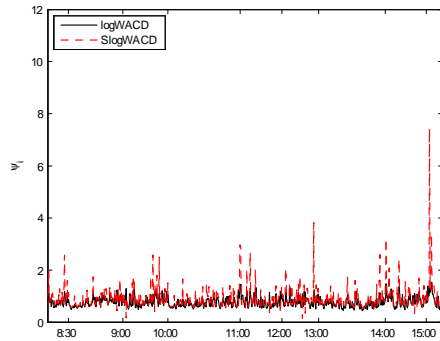


Table 1: Mean Squared Error results for different model specifications

	MSE	Q(25)
Parametric specifications		
EACD	12190	16.84
WACD	12089	14.96
GACD	12047	15.63
LogWACD	12074	88.94
Semiparametric specifications		
EACD	11165	21.47
WACD	11289	18.66
GACD	10960	18.74
LogWACD	11116	80.83

References

- [1] Bauwens, L. and P. Giot, 2000, “The Logarithmic ACD Model: an Application to the Bid-ask Quote Process of Three NYSE stocks,” *Annales d’Economie et de Statistique*, 60, 117-150.
- [2] Bauwens, L., Giot, P., Grammig, J., and D. Veredas, 2004, “A Comparison of Financial Duration Models via Density Forecasts,” *International Journal of Forecasting*, 20, 589–609.
- [3] Bauwens, L., and N. Hautsch, 2006, “Modelling Financial High Frequency Data Using Point Processes,” Discussion Paper, 2006-80, CORE, Université Catholique, Louvain-la-Neuve.
- [4] Dungey, M., Henry, O. and M. McKenzie, 2013, "Modelling Trade Duration in US Treasury Markets", *Quantitative Finance*, 13,1431-1442.
- [5] Engle R. and J. Russell, 1998, “Autoregressive Conditional Duration: a New Approach for Irregularly Spaced Transaction Data,” *Econometrica*, 66, 1127–1162.
- [6] Glad, I.K., 1998, “Parametrically Guided Non-parametric Regression,” *Scandinavian Journal of Statistics*, 25, 649-668.
- [7] Goodhart, C. A. E. and M. O’Hara, 1997, “High Frequency Data in Financial Markets: Issues and Applications,” *Journal of Empirical Finance*, 4, 73-114.
- [8] Grammig, J. and K.-O. Maurer, 2000, “Non-monotonic Hazard Functions and the Autoregressive Conditional Duration Model,” *Econometrics Journal*, 3, 16–38.

- [9] Hujer, R., S. Vuletic, 2007, “Econometrica Analysis of Financial Trade Processes by Discrete Mixture Duration Models,” *Journal of Economic Dynamics and Control*, 31, 635-667.
- [10] Lee, S.W. and B. E. Hansen, 1994, “Asymptotic Theory for the GARCH(1,1) Quasi-maximum Likelihood Estimator,” *Econometric Theory*, 10, 29-52.
- [11] Long, X., L. Su and A. Ullah, 2011, “Estimation and Forecasting of Dynamic Conditional Covariance: A Semiparametric Multivariate Model,” *Journal of Business and Economic Statistics*, 1, 109–125.
- [12] Lumsdaine, R. L., 1996, “Consistency and Asymptotic Normality of the Quasi-maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models,” *Econometrica*, 64, 575-596.
- [13] Lunde, A., 2000, “A Generalized Gamma Autoregressive Conditional Duration Model,” Discussion paper, Aarlborg University.
- [14] Madhavan, A., 2000, “Market Microstructure: a Survey,” *Journal of Financial Markets*, 3, 205–258.
- [15] Martins-Filho, C., S. Mishra, and A. Ullah, 2008, “A Class of Improved Parametrically Guided Nonparametric Regression Estimators,” *Econometric Reviews*, 27, 542-573.
- [16] Mishra, S., L. Su, and A. Ullah, 2010, “Semiparametric Estimator of Time Series Conditional Variance,” *Journal of Business and Economic Statistics*, 28, 275-290.
- [17] Pacurar, M., 2008, “Autoregressive Conditional Duration Models in Finance: A Survey of the Theoretical and Empirical Literature,” *Journal of Economic Surveys*, 22, 711-751.
- [18] Silverman, B.W. 1986, *Density Estimation for Statistics and Data Analysis*, New York, Chapman and Hall.
- [19] Zhang, M. Y., J. Russell, and R. S. Tsay, 2001, “A Nonlinear Autoregressive Conditional Duration Model with Applications to Financial Transaction Data,” *Journal of Econometrics*, 104, 179–207.