Summary

This document contains a brief summary of the paper titled, "Autoregressive Conditional Duration - A New Model for Irregularly Spaced Transaction Data", by *Engle, R. F., and Russell, J*

# Introduction

With the arrival of HF data, we have reached the ultimate limit of data collection. Every transaction and every quote is being stored. In this new environment, what aspects should be considered in modeling ?

- Transactions happen at irregular intervals and all the HF data is being made available. Hence standard econometric modeling with uniform time spacing will lead to information loss
- Rate of arrival of transaction type may vary over the course of the day, week or year. Hence the choice of "optimal" interval is difficult
- There are transactions that suddenly shoot up whenever there is some news release or some unobservable event. One needs a way to model the after-effects of sudden spike in transactions

Given the above context, what's this paper about ?

- This paper proposes an alternate model to the fixed interval analysis. The arrival times are treated as random variables and with each arrival time, there are random variales called marks such as volume, bid-ask spread or price variable
- The conditional intensity is parametrized in terms of past events in way that seems particularly well suited to the transaction process
- The dependence of conditional intensity on the past durations motivates the authors to call the proposed model, "Autoregressive Conditional Duration" model.

# The Conditional Intensity Process

Consider a stochastic process that is simply a sequence of times $\{t_0, t_1, \ldots t_n, \ldots\}$. Associated with this process is a counting function $N(t)$ that is the number of events that would have occurred by time $t$. If there are characteristics associated with arrival times, such as price or volume, the process is called *marked point process*.

There are two characteristics of the process that have been mentioned in the paper :

- A point process on $[t_0, \infty]$ is said to *evolve without after-effects* if for any $t > t_0$ , the realization of points in $[t, \infty]$ is not dependent on the sequence of points in $[t_0, t]$
- A counting process is said to be conditionally orderly at time $t > t_0$ if for sufficient short interval of time and conditional on any event $P$ defined by the realization of the process on $[t_0, t]$, the probability of two or more events occurring is infinitesimal relative to the probability of one event

The model in the paper assumes a point process that *evolve with after-effects* and *conditionally orderly.* Hence the conditional intensity process is defined as

$$\lambda(t|N(t), t_1, \ldots, t_{N(t)}) = \lim_{\Delta t \to 0} \frac{P(N(t+\Delta t) > N(t)|N(t), t_1, \ldots, t_{N(t)})}{\Delta t}$$

The conditional intensity function is also called *hazard function.* The intensity function of self-exciting point process is one where past events impact the probability of future events. The loglikelihood can be expressed in terms of the conditional densities as follows

$$L = \sum_{i=1}^{N(t)} \log \lambda(t_i|N(i-1), t_0, \ldots, t_{i-1}) - \int_{t_0}^{T} \lambda(u|N(u), t_0, \ldots, t_{N(u)})du$$

There are many ways to parametrize the conditional intensity function

- Keep $\lambda$ fixed $\rightarrow$ you get *poisson process*
- Keep $\lambda_t$ as a function of only time and not the history of events $\rightarrow$ you get *non homogeneous poisson process*
- Make $\lambda$ dependent only on the number of events and not on the timing of the events $\rightarrow$ you get *pure birth process*
- Only the $m$ most recent arrival times are present in the conditional intensity function $\rightarrow$ you get a *m-memory self exciting counting process*
- Linear function of past arrivals

  $$\lambda(t|N(t), t_1, \ldots, t_{N(t)}) = \omega + \sum_{i=1}^{N(t)} \pi(t - t_i)$$

  These models are ineffective for transaction modeling as they imply that the marginal effect of let's say an event that occurred 10 minutes in the past is independent of the intervening history
- Model based on the past arrivals

  $$\lambda(t|N(t), t_1, \ldots, t_{N(t)}) = \omega + \sum_{i=1}^{N(t)} \pi(t_{N(t)+1-i} - t_{N(t)-i})$$

- Cox model with covariates $z_i$ as a vector of explanatory variables associated with arrival $i$

  $$\lambda(t|N(t), t_1, \ldots, t_{N(t)}) = \lambda(t)e^{\beta' z_{N(t)}}$$

  Lagged durations can also be used as covariates

# The ACD Model

The model mentioned in the paper is different from the one suggested in the literature. It is specified in terms of conditional density of the durations. Let $\psi_i$ be the expectation of the $i^{th}$ duration and denote the $i^{th}$ inter arrival by $x_i = t_i = t_{i-1}$.

$$E(x_i|x_{i-1}, \ldots x_1) = \psi_i(x_i|x_{i-1}, \ldots x_1) \equiv \psi_i \tag{1}$$

ACD class of models consist of parametrization of equation (1) with the assumption that

$$x_i = \psi_i \epsilon_i$$

where $\epsilon_i$ is iid with density $p(\epsilon, \phi)$. You can choose any form for expected duration and distribution of $\epsilon$. Let $p_0$ be the density of $\epsilon$ and let $S_0$ be the associated survival function. Define the baseline hazard function as

$$\lambda_0(t) = \frac{p_0(t)}{S_0(t)}$$

The conditional intensity model of ACD model can be expressed as

$$\lambda(t|N(t), t_1, \ldots, t_{N(t)}) = \lambda_0 \left(\frac{t - t_{N(t)}}{\psi_{N(t)+1}}\right) \frac{1}{\psi_{N(t)+1}}$$

ACD avatars

- Durations are conditionally exponential

$$\psi_i = \frac{1}{\psi_{N(t)+1}}$$

- ACD(m): An $m$ memory conditional intensity function

$$\psi_i = \omega + \sum_{j=0}^{m} \alpha_j x_{i-j}$$

- ACD(m,q): A more general model without the limited memory characteristic

$$\psi_i = \omega + \sum_{j=0}^{m} \alpha_j x_{i-j} + \sum_{j=0}^{q} \alpha_j x \psi_{i-j}$$

- EACD(m,q) : ACD(m,q) with exponentially distributed error terms
- WACD(m,q) : ACD(m,q) with Weibull distributed error terms

# Modeling the time between transactions of IBM stock using the ACD model

## Basic summary stats of the dataset used in the analysis

- IBM stock analyzed between Nov 1 1990 and Jan 31 1991
- 62 trading days considered
- 58942 limit and market orders considered
- 52405 unique time stamps
- min interarrival time is 1 sec and max interarrival time is 28.38 seconds
- Histogram of interarrivals does not concur with an exponential distribution. This does not mean that the conditional distribution is not exponential
- acf and pacf of interarrivals - The null that first 15 lags is 0 is rejected
- Significant value for Ljung-Box statistic on interarrival data
- The transaction volume varies across a day. Based on the spline fitting, diurnally adjusted data is obtained

### Modeling and inference

- EACD(1,1) and EACD(2,2) models are fit. Parameters of diurnal factors are individually insignificant but the likelihood ratio reject the null of no daily factor. The Ljung-Box statistic for diurnally adjusted data shows that diurnal factor is not the only reason for dependency
- Standardized durations of EACD(1,1) and EACD(2,2) are obtained. Based on the dispersion tests, the exponential distribution seems to be weak fit
- WACD(1,1) and WACD(2,2) models are fit. Parameters of diurnal factors are individually insignificant but the likelihood ratio reject the null of no daily factor. The Ljung-Box statistic is significantly reduced but based on dispersion tests, Weibull apppears to be a weak candidate for error modeling
- Good thing about using WACD models as compared to EACD model is that atleast the hazard function is taken care of. The hazard function for EACD is flat which blatantly violates what we see in the transaction data
- The paper also investigates whether episodes of high transactional intensity are due to liquidity traders or informed traders. A thinned process is obtained for analysis. EACD(2,2) and WACD(2,2) models are fit. The results find the occurrence of both liquidity and information based clustering.

## Takeaway

The paper models the duration between transactions.With the arrival of HF data, there needs to be a model that captures irregularly spaced timestamps. It is obvious that neither a standard Poisson process nor a non homogeneous Poisson process is going to be a good fit. The conditional intensity function has to depend on the past history of transaction times. The authors formulate a particular form for the conditional intensity function and explain the clustering phenomenon as well as some well known market microstructure patterns.