Financial Times Series

Lecture 8

High-Frequency Data

- By HF data we mean data taken at fine time intervals, typically intraday, like every five minutes or we may even record the prices of an asset at every transaction (of the asset)

- HF data allows us to study the so called market micro structure, i.e. how the trading mechanisms affect the pricing of assets

Nonsynchronous trading

- Two assets are typically not traded at the same point in time
- Also the trading intensities vary between stocks
- When we use stock prices to fit models we typically use the closing price and (wrongly) assume closing prices are 24 hours apart

Effects of nonsynchronous trading

- Lag-1 cross correlation between stock returns
- Lag-1 serial correlation in a portfolio return
- Negative correlations in the return series of a single stock

Lag-1 cross correlation between stock returns

- Assume that we have two independent stocks A and B
- Assume that the trading frequency of A is higher than that of B
- If market affecting news arrives late in the day it is more likely to affect A than B
- It may be the case that the news does not affect B until the next day…

Lag-1 serial correlation in a portfolio return

- If a portfolio holds both stocks A and B as above then the different trading frequencies may introduce serial correlations within the portfolio return series

- An example of how negative correlations may occur (due to nonsynchronous trading) within a portfolio is derived on p 233-235 in Tsay 3rd ed

BID-ASK Spread

- As in all markets a market maker buys his goods at one price and sells his goods at a higher price

- Same thing goes for stocks, at a fixed point in time me and you as investors buy one stock at a price higher than the price at which we sell the same stock

BID-ASK Spread

- The price at which we sell (i.e. the market maker buys) the stock is called the bid price and we denote this price by P_b

- The price at which we buy (i.e. the market maker sells) the stock is called the ask price and we denote this price by P_a

- The difference $P_a - P_b$ is called the bid-ask spread and is the market maker's primary source of compensation

BID-ASK Spread implications

- The bid-ask spread introduces negative 1-lag correlation in the return series of a single stock

- The observed price P_t of a stock at time t may be modelled (Roll 1984)

$$P_t = {P^*}_t + I_t \frac{S}{2}$$

where ${P^*}_t$ is the fundamental price in a frictionless market, $P(I_t = 1) = P(I_t = -1) = \frac{1}{2}$, $\{I_t\}$ is a sequence of independent r.v.'s and $S = P_a - P_b$

BID-ASK Spread implications

- If a trade does not induce a change in ${P^*}_t$ the observed price changes are

$$\Delta P_t = (I_t - I_{t-1})\frac{S}{2}$$

- We note that

$$E[\Delta P_t] = 0, Var(\Delta P_t) = \frac{S^2}{2},$$
$$Cov(\Delta P_t, \Delta P_{t-1}) = -\frac{S^2}{4} \text{ and}$$
$$Cov(\Delta P_t, \Delta P_{t-l}) = 0 \text{ for } l > 1$$

- So there is a negative 1-lag correlations in the return series which is referred to as the bid-ask bounce

Transactions data

- Let t_i (typically measured in seconds from midnight) be a point in time where a stock is traded
- At that point in time several measurements, such as transaction price, bid-ask spread, trading volume and so forth, are made
- The recorded data is referred to as transactions data

Characteristics of transactions data

- Time points are not equidistant and it is of interest to model duration, i.e. the time beween trades

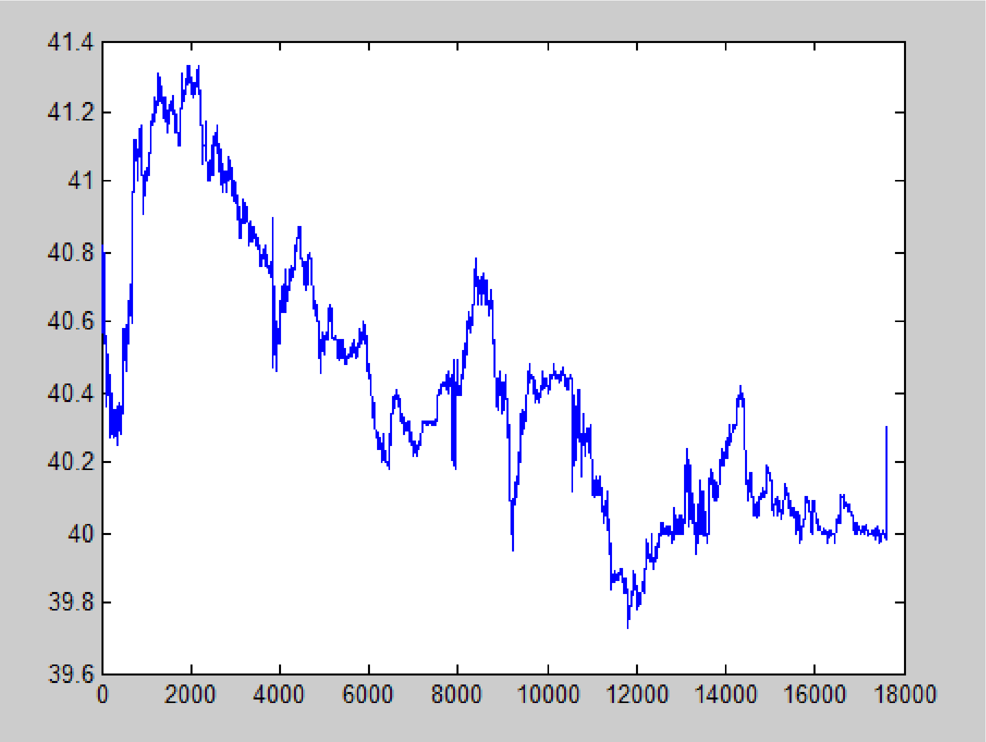
- Prices are discrete valued

- Transaction intensities show daily periodic (diurnal) patterns since trading is typically more intense just after opening and just before closing than it is mid-day

- There may be several trades at different prices taking place at the same point in time since time is measured in seconds

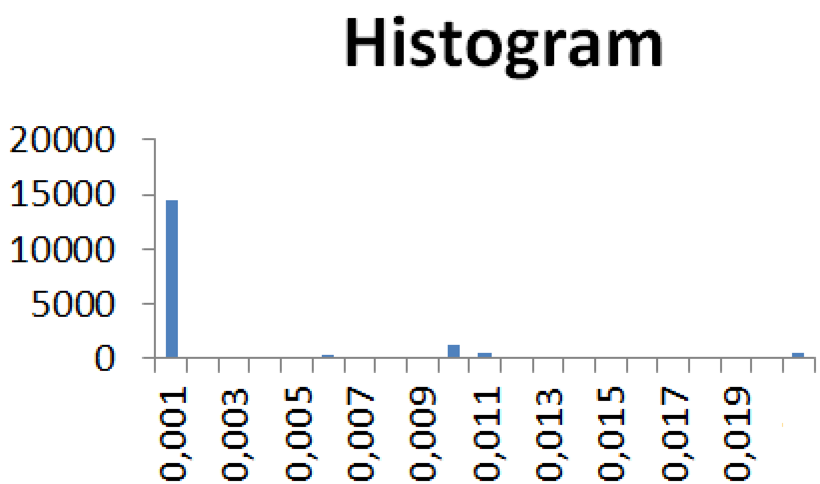
ITT data

- All prices during 131202 8:40:09am to 131205 4:11:35pm



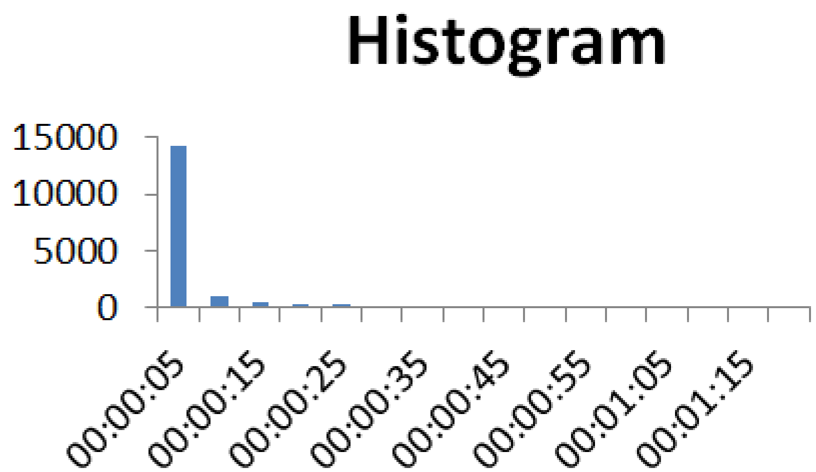
ITT data

- There are 17623 observed transactions and 11326 of them do not cause a price change



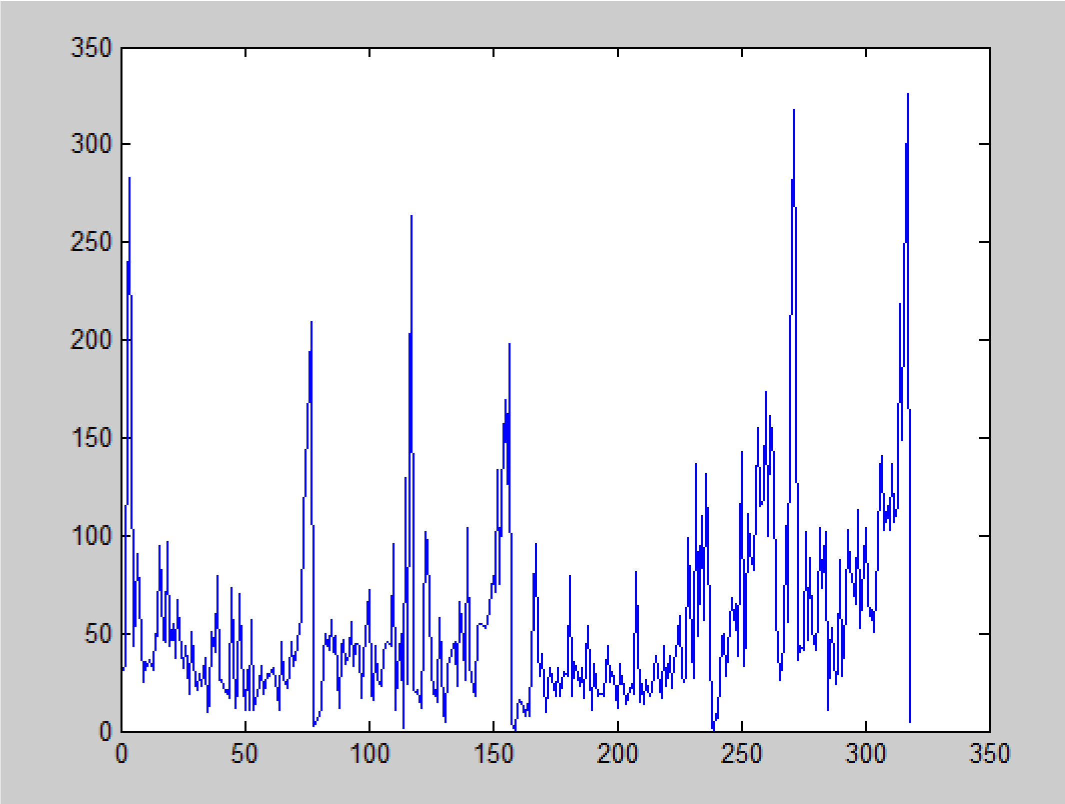
ITT data

- Most durations are less than five seconds



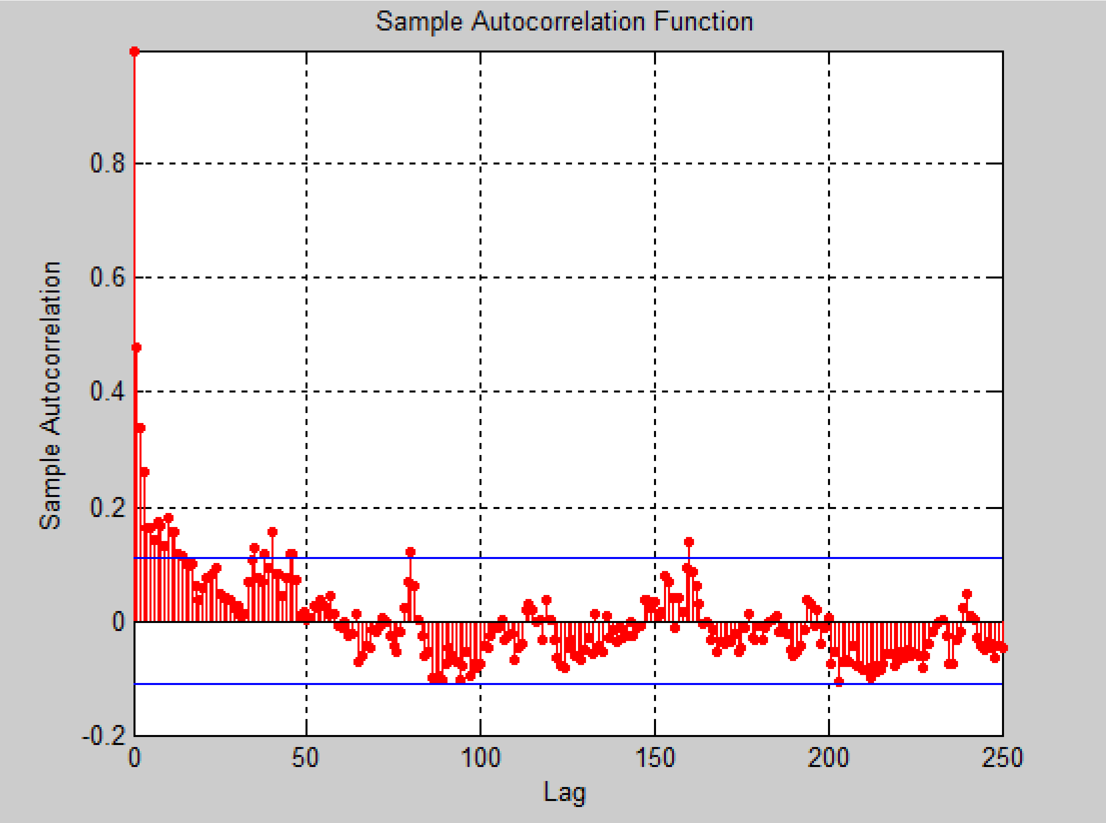
ITT data

- Number of trades in five minute intervals (only 9:30am-4pm data)



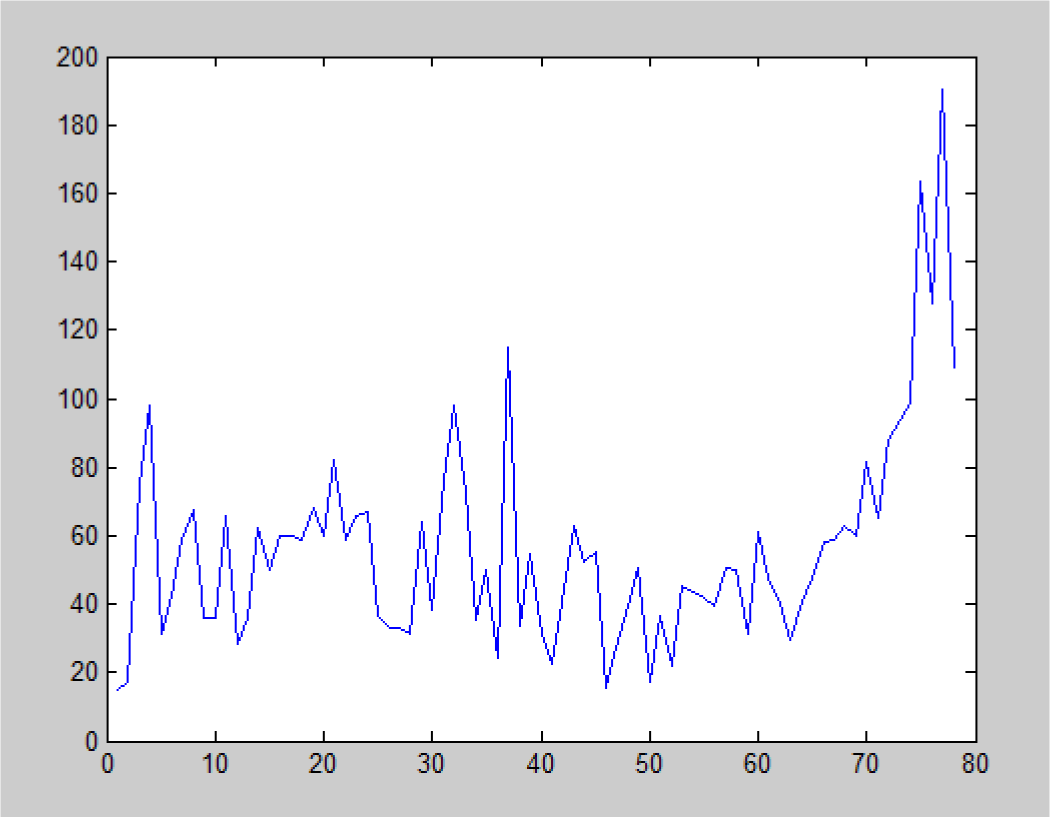
ITT data

- Correlation at lag 78? There are 78 five minute intervals in a trading day…



ITT data

- Mean number of trades for five minute periods indicate that closing time is busier than mid-day



Classification of movements

	ith trade		
$(i-1)$th trade	+	0	-
+	485	1552	1110
0	1508	8238	1580
-	1155	1535	458

- We note that consecutive increases and decreases are rare, as only 485/17621=2.8% and 458/17621=2.6% are the respective categories
- There are slightly more ”up to unchanged” than ”up to down” movements
- As we have seen before; most transactions leave the price unchanged

Increase followed by increase vs. bid-ask bounce

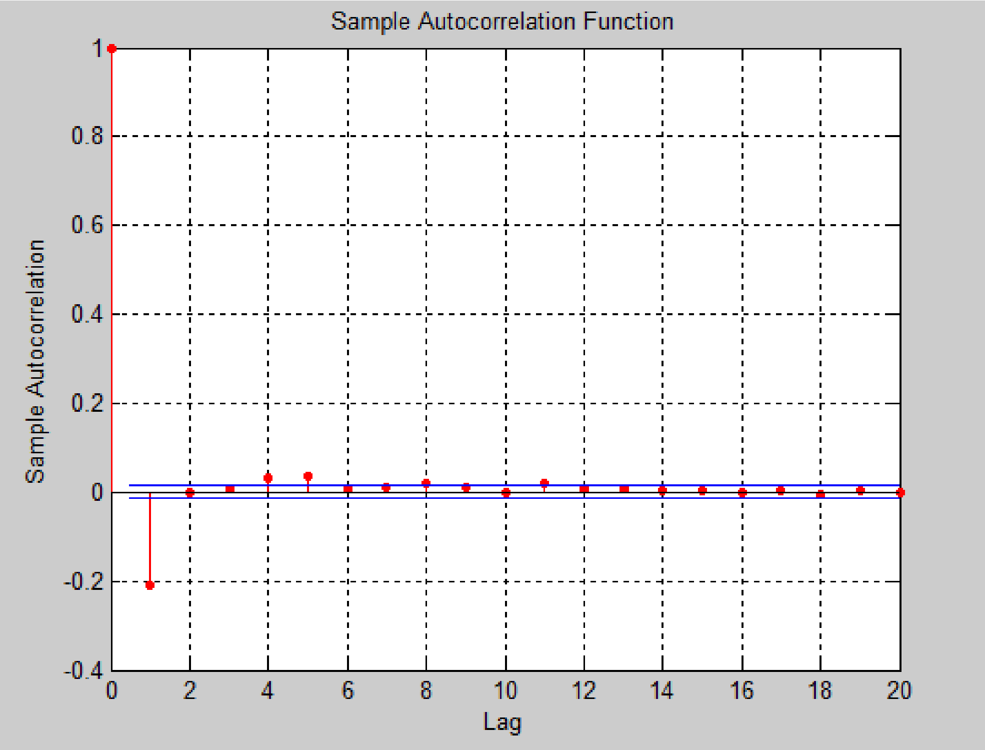
- That an increase is likely to be followed by a decrease and vice versa is an example of the bid-ask bounce in action

- Let $\{D_i\}$ be defined by; $D_i = -1$ if a transaction induces a negative price movement, $D_i = 0$ if a transaction does not induce a price movement and $D_i = 1$ if a transaction induces a positive price movement

- Below we see the autocorrelation of the series $\{D_i\}$ for the ITT data

Increase followed by increase vs. bid-ask bounce

- As the previously discussed model (Roll) indicated, the bid-ask spread introduces lag-1 negative correlation, i.e. the bid-ask bounce



Modeling price changes

- Using transactions data to model price changes could be done with e.g.;
- Ordered Probit Model, see Tsay ch.5
- Decomposition model, see below (and Tsay ch.5)

Decomposition model

- Under this model we assume that

$$y_i = P_{t_i} - P_{t_{i-1}} = A_i D_i S_i$$

where $A_i = 1$ if there is a price change and $A_i = 0$ otherwise, $D_i = 1$ if $A_i = 1$ and the price change is positive and $D_i = -1$ if $A_i = 1$ and the price change is negative and S_i is the (absolute) size of the price change (in no. of ticks)

Decomposition model

- We let $p_i = P(A_i = 1)$ and assume that

$$ln\left(\frac{p_i}{1-p_i}\right) = \boldsymbol{x_i}'\boldsymbol{\beta}$$

- We let $\delta_i = P(D_i = 1|A_i = 1)$ and assume that

$$ln\left(\frac{\delta_i}{1-\delta_i}\right) = \boldsymbol{y_i}'\boldsymbol{\gamma}$$

Decomposition model

- We also assume that $S_i|D_i = 1, A_i = 1$ and $S_i|D_i = -1, A_i = 1$ are distributed as $1 + g(\lambda_{u,i})$ and $1 + g(\lambda_{d,i})$, respectively, where g is a geometric distribution

- For the parameters $\lambda_{u,i}$ and $\lambda_{d,i}$ we assume that

$$ln\left(\frac{\lambda_{j,i}}{1-\lambda_{j,i}}\right) = \boldsymbol{z_i}'\boldsymbol{\theta_j}$$

Decomposition model

- The model may be fitted using ML (likelihood function is given in Tsay)

- We fit it for the ITT data with the same explanatory variables as Tsay uses for IBM (intraday) data, namely;

- Lagged values of A_i, D_i and S_i

Decomposition model

- We let

$$ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 A_{i-1},$$

$$ln\left(\frac{\delta_i}{1-\delta_i}\right) = \gamma_0 + \gamma_1 D_{i-1},$$

$$ln\left(\frac{\lambda_{j,i}}{1-\lambda_{j,i}}\right) = \theta_{j,0} + \theta_{j,1} S_{i-1}$$

Decomposition model

- ML gives

parameter	ML-estimate
β_0	-0.9813
β_1	1.0198
γ_0	-0.0026
γ_1	-0.8763
$\theta_{u,0}$	1.6226
$\theta_{u,1}$	-0.4229
$\theta_{d,0}$	1.5349
$\theta_{d,1}$	-0.4431

Interpreting the estimates

- For example we get that

$$P(A_i = 1 | A_{i-1} = 0) = \frac{e^{-0.9813}}{1 + e^{-0.9813}} = 0.2726$$

and

$$P(A_i = 1 | A_{i-1} = 1) = \frac{e^{-0.9813+1.0198}}{1 + e^{-0.9813+1.0198}} = 0.5096$$

- This indicates that most transactions are without price changes and that price changes may come in clusters

Interpreting the estimates

- We also get that

$$P(D_i = 1 | D_{i-1} = 0, A_{i-1} = 0) = \frac{e^{-0.0026}}{1+e^{-0.0026}} = 0.4993,$$

$$P(D_i = 1 | D_{i-1} = 1, A_{i-1} = 1) = \frac{e^{-0.0026-0.8763}}{1 + e^{-0.0026-0.8763}} = 0.2934$$

and

$$P(D_i = 1 | D_{i-1} = -1, A_{i-1} = 1) = \frac{e^{-0.0026+0.8763}}{1 + e^{-0.0026+0.8763}} = 0.7055$$

- This indicates that it is a 50/50 chance that a non-price change is followed by a price change, that consecutive increases are not so likely and the third probability once again establishes the bid-ask bounce

Duration models

- We may be interested in modeling how durations, i.e. time intervals between trades, behave

- We have seen that there may be deterministic patterns in intraday data so we work with adjusted duration series given by

$$\Delta t^*{}_i = \frac{\Delta t_i}{f(t_i)}$$

- The deterministic function f is typically some smoothing spline

- Below we have adjusted durations using the mean durations in each of the 78 trading day five minute windows

The ACD model

- Reminiscent of a GARCH

- Let $x_i = \Delta t^*{}_i$ and $\psi_i = E(x_i | F_{t-1})$ and assume that $x_i = \psi_i \epsilon_i$ for some i.i.d non-negative r.v.'s ϵ_i such that $E(\epsilon_i) = 1$

- In the model by Engle and Russel ϵ_i has exponential or Weibull distribution and;

$$\psi_i = \omega + \sum_{i=1}^{r} \gamma_j x_{i-j} + \sum_{j=1}^{s} \omega_j \psi_{i-j}$$

The ACD model

- The above model is called and ACD(r, s) and if the exponential or the Weibull distributions are used, models may be referred to as EACD or WACD, respectively

- Below we will focus on the EACD$(1,1)$ and try to fit it to our adjusted ITT durations

EACD(1,1)

- So we will assume that

$$x_i = \psi_i \epsilon_i \text{ and } \psi_i = \omega + \gamma_1 x_{i-1} + \omega_1 \psi_{i-1}$$

where $\epsilon_i \sim Exp(1)$

- For moment properties see p. 256-257 of Tsay 3rd ed

Estimation for EACD$(1,1)$

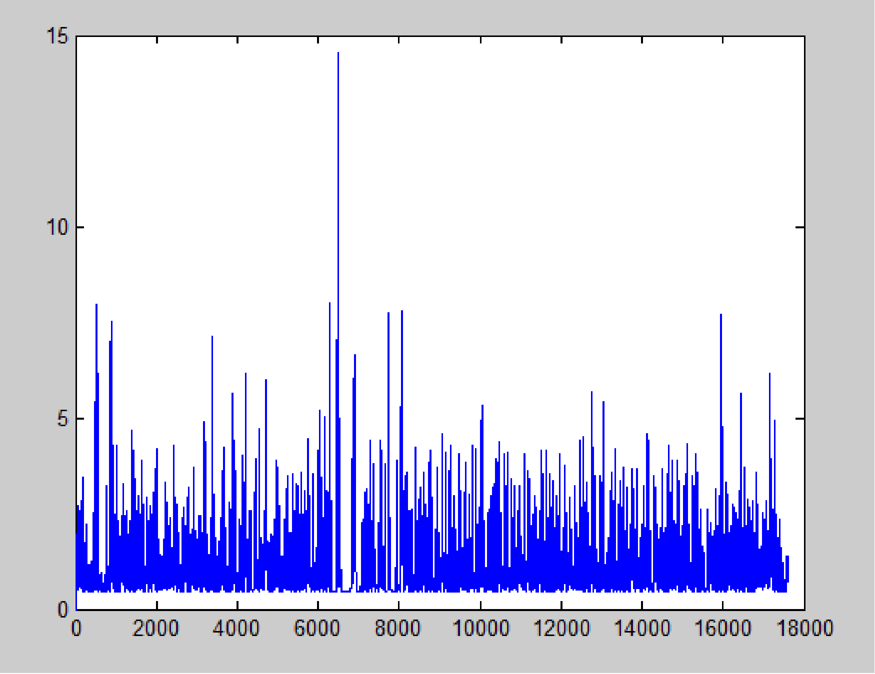
- Estimation is typically made using conditional ML

- For an EACD$(1,1)$ the conditional log-likelihood is given by (verify this)

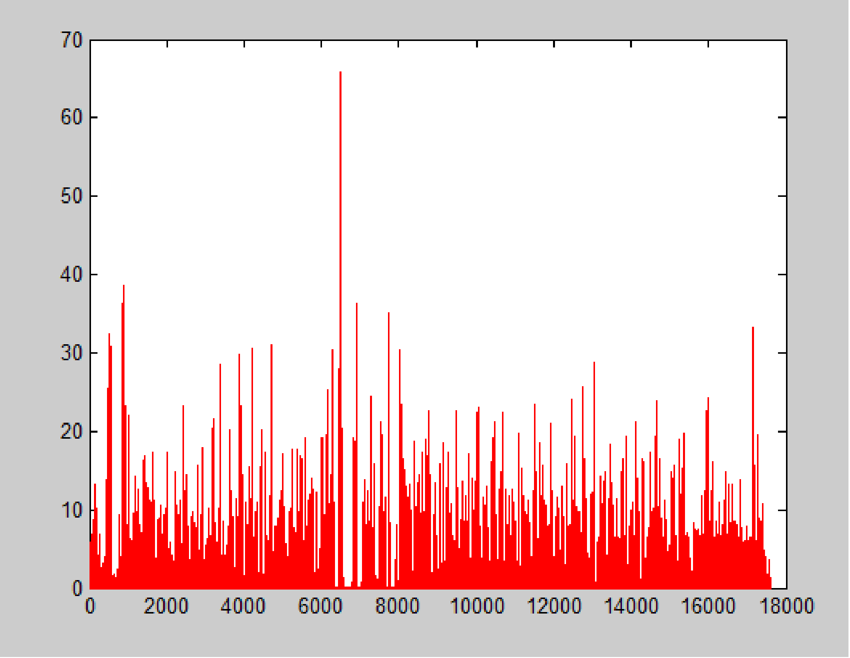
$$l\left(\omega, \gamma_1, \omega_1 | x_{i_0+1}, \dots, x_T\right) = -\sum_{i=i_0+1}^{T} ln(\psi_i) + \frac{x_i}{\psi_i}$$

Estimation for EACD(1,1)

- The parameters for the adjusted ITT durations are $\omega = 0.1145, \gamma_1 = 0.1582, \omega_1 = 0.7465$



Fitted model



Raw data

Residuals

- The residuals are given by $\hat{e}_i = x_i / \hat{\psi}_i$ where $\hat{\psi}_i$ are the values predicted by the model