

Lecture Note of Bus 41202, Spring 2008: Nonlinear Models & Market Microstructure

Does nonlinearity exist in financial TS?

Yes, especially in volatility & high-freq data

We focus on simple nonlinear models & neural networks

What is a linear time series?

$$x_t = \mu + \sum_{i=0}^{\infty} \psi_i a_{t-i}$$

where μ is a constant, ψ_i are real numbers with $\psi_0 = 1$, and $\{a_t\}$ is an iid $(0, \sigma_a^2)$.

General concept: Let F_{t-1} be info. available at time $t - 1$.

Conditional mean:

$$\mu_t = E(x_t | F_{t-1}) \equiv g(F_{t-1}),$$

Conditional variance:

$$\sigma_t^2 = \text{Var}(x_t | F_{t-1}) \equiv h(F_{t-1})$$

where $g(\cdot)$ and $h(\cdot)$ are well-defined functions with $h(\cdot) > 0$.

For a linear series, $g(\cdot)$ is a linear function of F_{t-1} and $h(\cdot) = \sigma_a^2$.

Statistics literature: focuses on $g(\cdot)$

See the book by Tong (Oxford University Press, 1990)

Econometrics literature: focuses on $h(\cdot)$

Some specific models

TAR model: a piecewise linear model in the threshold space.

Example: 2-regime AR(1) model

$$x_t = \begin{cases} -1.5x_{t-1} + a_t & \text{if } x_{t-1} < 0, \\ 0.5x_{t-1} + a_t & \text{if } x_{t-1} \geq 0, \end{cases}$$

where a_t 's are iid $N(0, 1)$.

Here the delay is 1 time period, x_{t-1} is the **threshold** variable, and the threshold is 0. The threshold divides the x_{t-1} -space into two regimes with Regime 1 denoting $x_{t-1} < 0$.

What is so special about this model? See the time plot.

Special features of the model: (a) asymmetry in rising and declining patterns, (more obs are positive than negative) (b) the mean of x_t is not zero even though there is no constant term in the model, (c) the lag-1 coefficient may be greater than 1 in absolute value.

Financial application:

Modeling asymmetry in volatility (recall EGARCH model)

Example: Daily log returns of IBM stock from July 3, 1962 to December 31, 2003 for 10,446 observations. See Figure 4.3 of the text (p. 162).

AR(2)-GARCH(1,1) model:

$$\begin{aligned} r_t &= 0.062 - 0.024r_{t-2} + a_t, & a_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= 0.037 + 0.077a_{t-1}^2 + 0.913\sigma_{t-1}^2 \end{aligned}$$

Std residuals: $Q(10) = 5.19(0.88)$ and $Q(20) = 24.38(0.23)$

Sq. std. res.: $Q(10) = 11.67(0.31)$ and $Q(20) = 18.25(0.57)$.

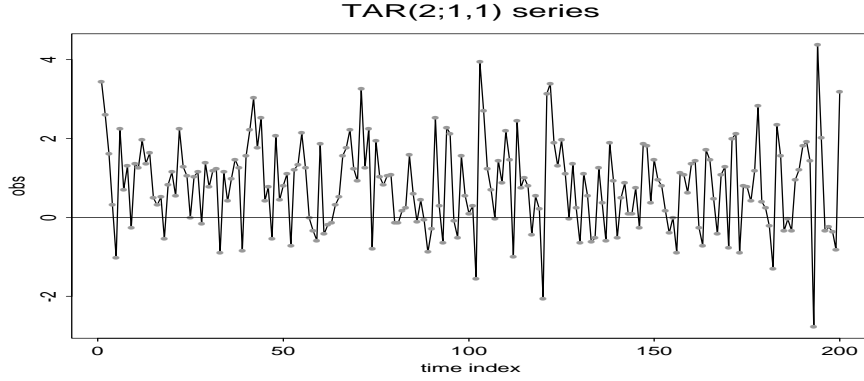


Figure 1: A simulated two-regime TAR process

AR(2)-TAR-GARCH(1,1) model

$$\begin{aligned}
 r_t &= 0.033 - .023r_{t-2} + a_t, & a_t &= \sigma_t \epsilon_t \\
 \sigma_t^2 &= 0.098a_{t-1}^2 + 0.954\sigma_{t-1}^2 \\
 &+ (0.060 - .052a_{t-1}^2 - 0.069\sigma_{t-1}^2)I(a_{t-1} > 0).
 \end{aligned}$$

Further simplification:

$$\begin{aligned}
 r_t &= 0.068 - .027r_{t-2} + a_t, & a_t &= \sigma_t \epsilon_t \\
 \sigma_t^2 &= (1.0 - 0.926)a_{t-1}^2 + 0.926\sigma_{t-1}^2 \\
 &+ (0.045 - .038a_{t-1}^2 + 0.021\sigma_{t-1}^2)I(a_{t-1} > 0).
 \end{aligned}$$

Rewrite the TAR-GARCH(1,1) as

$$\sigma_t^2 = \begin{cases} 0.074a_{t-1}^2 + 0.926\sigma_{t-1}^2 & \text{if } a_{t-1} \leq 0 \\ 0.045 + 0.036a_{t-1}^2 + 0.947\sigma_{t-1}^2 & \text{if } a_{t-1} > 0, \end{cases}$$

Discussion: The asymmetry in volatility is clearly seen. When $a_{t-1} < 0$, the volatility follows an IGARCH model without any drift. However, the volatility follows a GARCH(1,1) model when a_{t-1} is positive. But the persistent coefficient is close to unity for $a_{t-1} > 0$.

A different formulation: Recall GJR (or TGARCH) model. A TGARCH(m, s) model assumes the form

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^s (\alpha_i + \gamma_i N_{t-i}) a_{t-i}^2 + \sum_{j=1}^m \beta_j \sigma_{t-j}^2,$$

where N_{t-i} is an indicator for *negative* a_{t-i} , that is,

$$N_{t-i} = \begin{cases} 1 & \text{if } a_{t-i} < 0, \\ 0 & \text{if } a_{t-i} \geq 0, \end{cases}$$

and α_i, γ_i and β_j are non-negative parameters satisfying conditions similar to those of GARCH models.

Note that the impact of a_{t-i} on σ_t^2 is

$$\begin{cases} \alpha_i a_{t-i}^2 & \text{if } a_{t-1} \geq 0 \\ (\alpha_i + \gamma_i) a_{t-i}^2 & \text{if } a_{t-1} < 0. \end{cases}$$

Example. Consider the monthly log returns of IBM stock from 1926 to 2003.

The fitted TGARCH(1,1) model with “ged” innovations is

$$\begin{aligned} r_t &= 0.0121 + a_t, & a_t &= \sigma_t \epsilon_t, & \epsilon_t &\sim \text{ged}(1.51), \\ \sigma_t^2 &= 3.45 \times 10^{-4} + (0.0658 + 0.0843 N_{t-1}) a_{t-1}^2 + 0.8182 \sigma_{t-1}^2. \end{aligned}$$

The model seems adequate.

Markov switching model

Two-state MS model:

$$x_t = \begin{cases} c_1 + \sum_{i=1}^p \phi_{1,i} x_{t-i} + a_{1t} & \text{if } s_t = 1, \\ c_2 + \sum_{i=1}^p \phi_{2,i} x_{t-i} + a_{2t} & \text{if } s_t = 2, \end{cases}$$

where s_t assumes values in $\{1,2\}$ and is a first-order Markov chain with trans. prob.

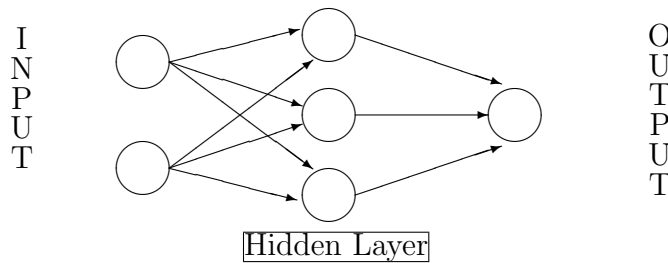
$$P(s_t = 2 | s_{t-1} = 1) = w_1, \quad P(s_t = 1 | s_{t-1} = 2) = w_2,$$

where $0 \leq w_1 \leq 1$ is the probability of switching out State 1 from time $t - 1$ to time t . A large w_1 means that it is easy to switch out State 1, i.e. cannot stay in State 1 for long. The inverse, $1/w_1$, is the expected duration (number of time periods) to stay in State 1. Similar idea applies to w_2 .

Example: Growth rate of US quarterly real GNP 47-91.

See Figure 4.4 of the textbook (p.166).

| State 1 | | | | | | | |
|---------|--------|----------|----------|----------|----------|------------|-------|
| Par | c_i | ϕ_1 | ϕ_2 | ϕ_3 | ϕ_4 | σ_i | w_i |
| Est | 0.909 | 0.265 | 0.029 | -0.126 | -0.110 | 0.816 | 0.118 |
| S.E | 0.202 | 0.113 | 0.126 | 0.103 | 0.109 | 0.125 | 0.053 |
| State 2 | | | | | | | |
| Est | -0.420 | 0.216 | 0.628 | -0.073 | -0.097 | 1.017 | 0.286 |
| S.E | 0.324 | 0.347 | 0.377 | 0.364 | 0.404 | 0.293 | 0.064 |



Discussion

- Regime 2, which has a negative expectation, denotes “recession” periods. The S.E. of the estimates are large due to the small number of data in the regime.
- The expected durations for Regime 1 and 2 are 8.5 and 3.5 quarters, respectively. $(1/w_i)$

Neural networks

- a semi-parametric approach to data analysis
- Structure of a network
 - Output layer
 - Input layer
 - Hidden layer
 - Nodes
- Activation function:
 - Logistic function:

$$\ell(z) = \frac{\exp(z)}{1 + \exp(z)}$$

– Heaviside (or threshold) function:

$$H(z) = \begin{cases} 1 & \text{if } z > 0 \\ 0 & \text{if } z \leq 0 \end{cases}$$

- Use $\ell(z)$ for the hidden layer

Feed-forward neural network:

Hidden node:

$$x_j = f_j(\alpha_j + \sum_{i \rightarrow j} w_{ij} x_i)$$

where $f_j(\cdot)$ is an activation function which is typically taken to be the logistic function

$$f_j(z) = \frac{\exp(z)}{1 + \exp(z)},$$

α_j is called the bias, the summation $i \rightarrow j$ means summing over all input nodes feeding to j , and w_{ij} are the weights.

Output node:

$$y = f_o(\alpha_o + \sum_{j \rightarrow o} w_{jo} x_j),$$

where the activation function $f_o(\cdot)$ is either linear or a Heaviside function. By a Heaviside function, we mean $f_o(z) = 1$ if $z > 0$ and $f_o(z) = 0$, otherwise.

General form:

$$y = f_o \left[\alpha_o + \sum_{j \rightarrow o} w_{jo} f_j \left(\alpha_j + \sum_{i \rightarrow j} w_{ij} x_i \right) \right].$$

With direct connections from the input layer to the output layer:

$$y = f_o \left[\alpha_o + \sum_{i \rightarrow o} w_{io} x_i + \sum_{j \rightarrow o} w_{jo} f_j \left(\alpha_j + \sum_{i \rightarrow j} w_{ij} x_i \right) \right],$$

Training and forecasting

Divide the data into training and forecasting subsamples.

Training: build a few network systems

Forecasting: based on the accuracy of out-of-sample forecasts to select the “best” network.

Example: Monthly log returns of IBM stock 26-99.

See text for details.

Some R and S-Plus commands:

You must load the library “nnet” into S-Plus before running the command **nnet**.

```
x=read.table('m-ibmln.txt')
y=x[4:864] % select the output: r(t)
# obtain the input variables: r(t-1), r(t-2), and r(t-3)
ibm.x_cbind(x[3:863],x[2:862],x[1:861])
# build a 3-2-1 network with skip layer connections
# and linear output.
ibm.nn_nnet(ibm.x,y,size=2,linout=T,skip=T,maxit=10000,
decay=1e-2,reltol=1e-7,abstol=1e-7,range=1.0)
# print the summary results of the network
summary(ibm.nn)
# compute \& print the residual sum of squares.
sse=sum((y-predict(ibm.nn,ibm.x))^2)
```

```

print(sse)
#eigen(nnet.Hess(ibm.nn,ibm.x,y),T)$values
# setup the input variables in the forecasting subsample
ibm.p=cbind(x[864:887],x[863:886],x[862:885])
# compute the forecasts
yh=predict(ibm.nn,ibm.p)
# The observed returns in the forecasting subsample
yo=x[865:888]
# compute \& print the sum of squares of forecast errors
ssfe=sum((yo-yh)^2)
print(ssfe)

```

Remark: One-step ahead Out-of-sample-forecasts using **nnet** command. A R script, **r-backnnet.txt**, is developed to carry out the evaluation of 1-step ahead out-of-sample forecasts. For illustration,

```

> source('r-backnnet.txt')
> m3=backnnet(x,y,nsample,orig,nl,nsk,miter)

```

Some references

A recent book: *Neural Networks in Finance: Gaining Predictive Edge in the Market* by Paul D. McNelis (2005, Elsevier). It uses Matlab.

Related to Credit Risk

- Elmer & Borowski (1988, *Financial Management*): bankruptcy prediction
- Messier & Hansen (1988, *Management Science*): forecasting business failures

Related to Bond Rating: two articles in the book *Neural Networks in Capital Markets*, ed. A.P. Refens, Wiley, 1994. See the articles by Moody and Utans, and Singleton and Surkan.

Analysis of High-Frequency Financial Data & Market Microstructure

Market microstructure: Why is it important?

1. Important in market design & operation, e.g. to compare different markets (NYSE vs NASDAQ)
2. To study price discovery, liquidity, volatility, etc.
3. To understand costs of trading
4. Important in learning the consequences of institutional arrangements on observed processes, e.g.
 - Nonsynchronous trading
 - Bid-ask bounce
 - Impact of changes in tick size, after-hour trading, etc.
 - Impact of daily price limits (many foreign markets)

Nonsynchronous trading:

Key implication: may induce serial correlations even when the underlying returns are iid.

Setup: log returns $\{r_t\}$ are iid (μ, σ^2)

For each time index t , $P(\text{no trade}) = \pi$.

Cannot observe r_t if there is no trade.

What is the observed log return series r_t^o ?

It turns out r_t^o is given in Eq. (5.1),

$$r_t^o = \begin{cases} 0 & \text{with prob. } \pi \\ r_t & \text{with prob. } (1 - \pi)^2 \\ r_t + r_{t-1} & \text{with prob. } (1 - \pi)^2 \pi \\ \vdots & \vdots \\ \sum_{i=0}^k r_{t-i} & \text{with prob. } (1 - \pi)^2 \pi^k \\ \vdots & \vdots \end{cases}$$

One can use this relation to show that

$$\begin{aligned} \text{Var}(r_t^o) &= \sigma^2 + \frac{2\pi\mu^2}{1 - \pi} \\ \text{Cov}(r_t^o, r_{t-j}^o) &= -\mu^2\pi^j, \quad j \geq 1. \end{aligned}$$

Bid-ask bounce

Bid and ask quotes introduce **negative** lag-1 serial correlation.

Setup: simplest case of Roll(1984)

True price P_t^* is unchnaged, i.e. $P_t^* = P_{t-1}^*$

$S = P_a - P_b$ is the bid-ask spread

$$P_t = P_t^* + \begin{cases} S/2 & \text{with prob. } 0.5 \\ -S/2 & \text{with prob. } 0.5 \end{cases}$$

Then,

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

where I_t and I_{t-1} are independent binary variables with $P(I_t = 1) = 0.5$ and $P(I_t = -1) = 0.5$.

Note: $E(I_t) = 0$ and $\text{Var}(I_t) = 1$ for all t .

One can show that

$$\begin{aligned}\text{Var}(\Delta P_t) &= S^2/2 \\ \text{Cov}(\Delta P_t, \Delta P_{t-1}) &= -S^2/4 \\ \text{Cov}(\Delta P_t, \Delta P_{t-j}) &= 0, \quad j > 1.\end{aligned}$$

The result continues to hold if P_t^* follows a random walk model.

High-Frequency Financial Data

Observations taken with time intervals 24 hours or less

Some example:

1. Transaction (or tick-by-tick) data
2. 5-minute returns in FX
3. 1-minute returns on index futures and cash market

Some Basic Features of the Data:

1. Irregular time intervals
2. Leptokurtic or Heavy tails
3. Discrete values, e.g. price in multiples of tick size

4. Large sample size
5. Multi-dimensional variables, e.g. price, volume, quotes, etc.
6. Diurnal Pattern

An illustration

IBM stock transaction data from 11/01/1990 to 1/31/1991

- Source: Trades, Orders Reports and Quotes (TORQ)
- Trading days: 63
- Sample size: 60,328
- Intraday trades: 60,265.
- Data available: bid, ask, transaction prices, volume, time, etc.
- Zero durations: 6531 (about 11%).
- Kurtosis of adj-duration: 44.23(.02)

Frequencies of price change

| | | | | | | | |
|--------------|-----------|------|-------|-------|-------|------|----------|
| Number(tick) | ≤ -3 | -2 | -1 | 0 | 1 | 2 | ≥ 3 |
| Percentage | 0.66 | 1.33 | 14.53 | 67.06 | 14.53 | 1.27 | 0.63 |

Number of trades in 5-minute intervals

See Figures 5.1 and 5.2 on page 214 of the text.

Another example: Transaction data of IBM stock in December 1999, from TAQ.

Two important changes

- Number of trades increased to sixfold (134,120 trades)
 - trades with zero time-duration became 23%
 - 42 trades in a single second in December 3, 1999.
- Tick size reduces to \$1/16; the percentage of trades without price change decreased from 67% to 46%.

Econometric models

1. Time Duration and Duration Models
2. Nonlinearity in Time Durations
3. A Model for Price Change and Duration
4. Hierarchical Models
5. Models for bid and ask quotes

Include statistical tools and methods useful in analyzing HF financial data

Data quality (need some cleaning)

- Trading hours (FX-24, US-market: 6.5 hours, but ...)
- Time stamp vs transaction time
- Missing values
- Order types (market or limit orders)

Important statistical issues:

1. Stationarity
2. Nonlinearity
3. Structural breaks

Price Change: Discrete values

- Ordered probit model: Hausman, Lo, & MacKinlay (1992)
- ADS model: Rydberg & Shephard (1998), McCulloch & Tsay (2000)

Look at a simple ADS decomposition:

- Price $P_t = P_0 + \sum_i^{N(t)} C_i$
- Number of transactions in $[0,t]$: $N(t)$
- $C_i = A_i D_i S_i$

– Action:

$$A_i = \begin{cases} 1 & \text{if } C_i \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

– Direction, given $A_i = 1$:

$$D_i = \begin{cases} 1 & \text{if } C_i > 0 \\ -1 & \text{if } C_i < 0 \end{cases}$$

– Size, given $A_i = 1$ and D_i : multiple of tick size

- Can be estimated by logistic regression

Model specification:

- Action A_i : Governed by a logistic regression

$$P(A_i = 1|F_{i-1}) = \text{logit}(F_{i-1})$$

- Direction given $A_i = 1$:

$$P(D_i = 1|F_{i-1}, A_i = 1) = \text{logit}(A_i, F_{i-1})$$

- Size given $A_i = 1$ and D_i :

$$P(S_i = s|A_i = 1, D_i = 1, F_{i-1}) \sim 1 + g(\lambda_{u,i})$$

$$P(S_i = s|A_i = 1, D_i = -1, F_{i-1}) \sim 1 + g(\lambda_{d,i})$$

where $g(\cdot)$ denotes a Geometric distribution and $\lambda_{j,i}$ is governed by a logistic equation:

$$\ln\left(\frac{\lambda_{j,i}}{1 - \lambda_{j,i}}\right) = \text{linear function of } F_{i-1}, A_i = 1, D_i.$$

Likelihood function:

$$P(C_i = s|F_{i-1}) =$$

$$P(S_i = s|A_i = 1, D_i, F_{i-1})P(D_i|A_i = 1, F_{i-1})P(A_i = 1|F_{i-1}).$$

A simple ADS model: IBM data 59,775 observations.

- Predictors: $\{A_{i-1}, D_{i-1}, S_{i-1}, V_{i-1}, x_{i-1}, BA_i\}$

1. V_{i-1} : volume of the previous trade (divided by 1000)

2. x_{i-1} : previous duration
3. BA_i : the prevailing bid-ask spread

- Model:

1. Action: $P(A_i|F_{i-1}) = p_i$, $\text{logit}(p_i) = \beta_0 + \beta_1 A_{i-1}$
2. Direction: $P(D_i = 1|A_i = 1, F_{i-1}) = \gamma_i$,
 $\text{logit}(\gamma_i) = \delta_0 + \delta_1 D_{i-1}$
3. Size: $\text{logit}(\lambda_{j,i}) = \theta_{j,0} + \theta_{j,1} S_{i-1}$ with $j = d$ or u .

- Results:

| | | | | |
|-----------|----------------|----------------|----------------|----------------|
| Parameter | β_0 | β_1 | δ_0 | δ_1 |
| Estimate | -1.057 | 0.962 | -0.067 | -2.307 |
| Std.Err. | 0.104 | 0.044 | 0.023 | 0.056 |
| Parameter | $\theta_{u,0}$ | $\theta_{u,1}$ | $\theta_{d,0}$ | $\theta_{d,1}$ |
| Estimate | 2.235 | -0.670 | 2.085 | -0.509 |
| Std.Err. | 0.029 | 0.050 | 0.187 | 0.139 |

Implication

1. Prob of price change:

$$P(A_i = 1|A_{i-1} = 0) = 0.258$$

$$P(A_i = 1|A_{i-1} = 1) = 0.476.$$

2. Interpretation: **Odds ratio**

Because A_{i-1} is also a binary variable, we have a 2×2 table:

| Outcome A_i | Independent variable A_{i-1} | |
|---------------|----------------------------------------------------------------------------|--------------------------------------------------------|
| | $A_{i-1} = 1$ | $A_{i-1} = 0$ |
| $A_i = 1$ | $P(A_i = 1) = \frac{\exp[\beta_0 + \beta_1]}{1 + \exp[\beta_0 + \beta_1]}$ | $P(A_i = 1) = \frac{\exp(\beta_0)}{1 + \exp(\beta_0)}$ |
| $A_i = 0$ | $P(A_i = 0) = \frac{1}{1 + \exp[\beta_0 + \beta_1]}$ | $P(A_i = 0) = \frac{1}{1 + \exp(\beta_0)}$ |

Odds Ratio: Row one divided by row 2:

$$OR = e^{\beta_1}.$$

3. Direction of price change:

$$P(D_i = 1 | F_{i-1}, A_i) = \begin{cases} 0.483 & \text{if } D_{i-1} = 0, \text{ i.e. } A_{i-1} = 0 \\ 0.085 & \text{if } D_{i-1} = 1, A_i = 1 \\ 0.904 & \text{if } D_{i-1} = -1, A_i = 1 \end{cases}$$

Bid-ask bounce

4. Weak evidence of price change cluster: price increases

$$S_i | (D_i = 1) \sim 1 + g(\lambda_{u,i}), \quad \lambda_{u,i} = 2.235 - 0.670S_{i-1}.$$

R command: `glm` stands for generalized linear model.

Duration and Duration Models

Focus on intraday time duration between transactions

Autoregressive conditional duration (ACD) model:

- Engle and Russell (1998)
- Intraday durations between trades, in seconds.

- Use ideas of GARCH models

Define

1. t_i : time of the i -th trade, starting at midnight, measured in seconds.
2. $X_i = t_i - t_{i-1}$
3. $f(t)$: Diurnal pattern of daily trading.
4. $x_i = X_i/f(t_i)$: adjusted time duration of i -th trade
5. F_i : information set available at t_i (inclusive)
6. ψ_i : Expected duration, $E(x_i|F_{i-1})$.

ACD(r, s) model:

$$\frac{x_i}{\psi_i} \sim \epsilon_i, \quad \epsilon_i \sim i.i.d. \quad g(\theta), \quad E(\epsilon_i) = 1$$

$$\psi_i = \omega_0 + \sum_{j=1}^r \gamma_j x_{i-j} + \sum_{j=1}^s \omega_j \psi_{i-j}.$$

The distribution of ϵ_i is either Standard Exponential or Standardized Weibull.

Refer to as an EACD or WACD model, respectively.

Let $\eta_i = x_i - \psi_i$.

- $\{\eta_i\}$ is a martingale difference sequence.

- ACD(r, s) model becomes

$$x_i = \omega_0 + \sum_{j=1}^{\max(r,s)} (\gamma_j + \omega_j)x_{i-j} - \sum_{j=1}^s \omega_j \eta_{i-j} + \eta_j.$$

Some properties of ACD model are easily available.

Remark: Duration models can be estimated via programs similar to GARCH models.

Focus on the first 5 days of November, 1990.

- Dates: November 1, 2, 5, 6, 7 of 1990
- Sample size: 3534 data points

A simple model: WACD(1,1)

$$x_i = \psi_i \epsilon_i, \quad \psi_i = 0.291 + 0.077x_{i-1} + 0.836\psi_{i-1}$$

where $\{\epsilon_i\}$ iid Weibull with $\hat{\alpha} = 0.878(0.011)$.

Implication:

- $E(x_i) = 3.34$ (vs 3.29, sample)
- Not strongly persistent: $\hat{\gamma}_1 + \hat{\omega}_1 \approx 0.91$
- Decaying intensity function.

Model diagnostics:

- ACF of $\hat{\epsilon}_i = \frac{x_i}{\hat{\psi}_i}$: $Q(12) = 5.7$ and $Q(24) = 19.9$
- ACF of $\hat{\epsilon}_i^2$: $Q(12) = 6.5$ and $Q(24) = 15.1$