Nonlinear Models

ARIMA/ARFIMA processes are *linear*.

- A linear combination of i.i.d. shocks.
- Depend linearly on past values.
- The autocorrelations contain all information about dependence.

Plenty of evidence that this type of model is not adequate to describe observed series.

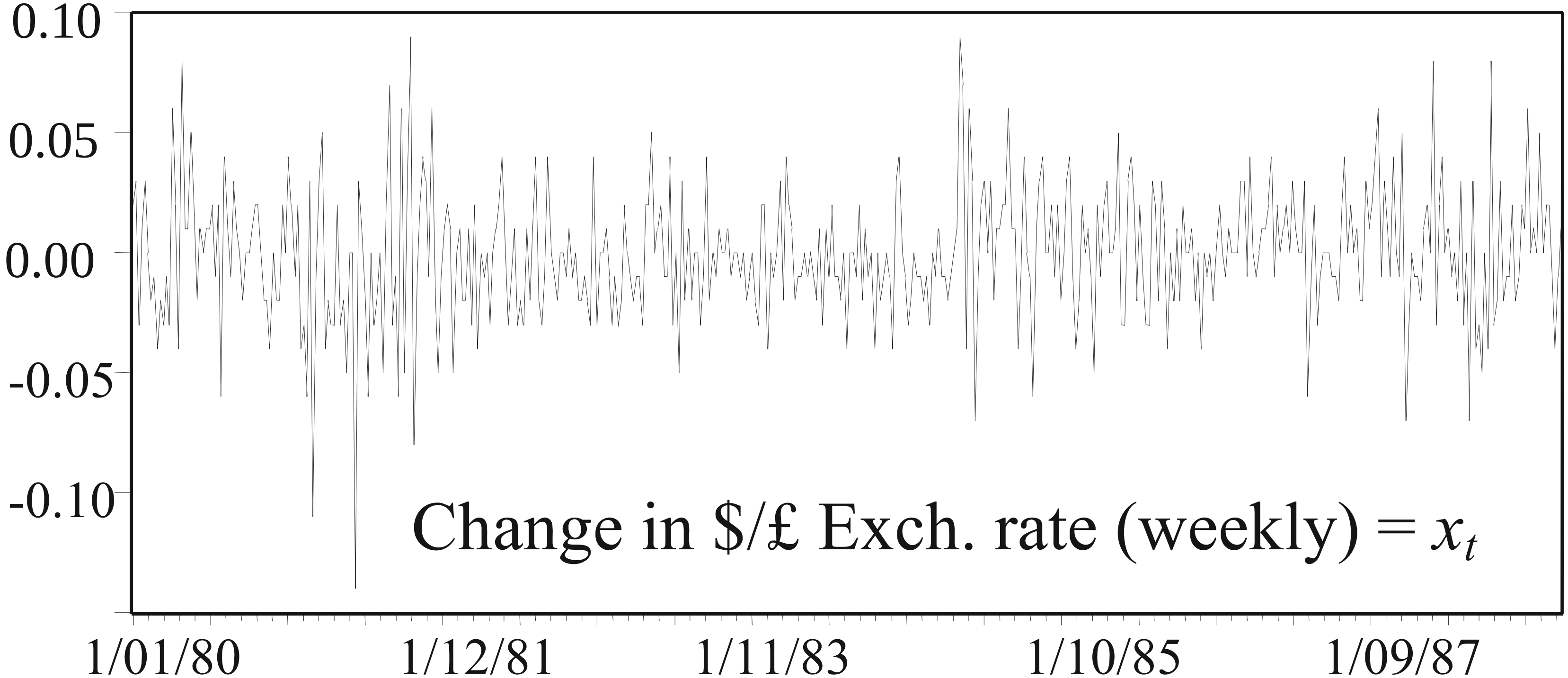
Recent time series research has focused on nonlinear models.

These classes of nonlinear model are currently the most important:

1. Conditional variance models (ARCH/GARCH)
2. Markov-switching models.
3. Nonlinear autoregressions, including
    - self-exciting threshold (SETAR) models
    - Bilinear AR models.

ARCH and GARCH Models

The problem:

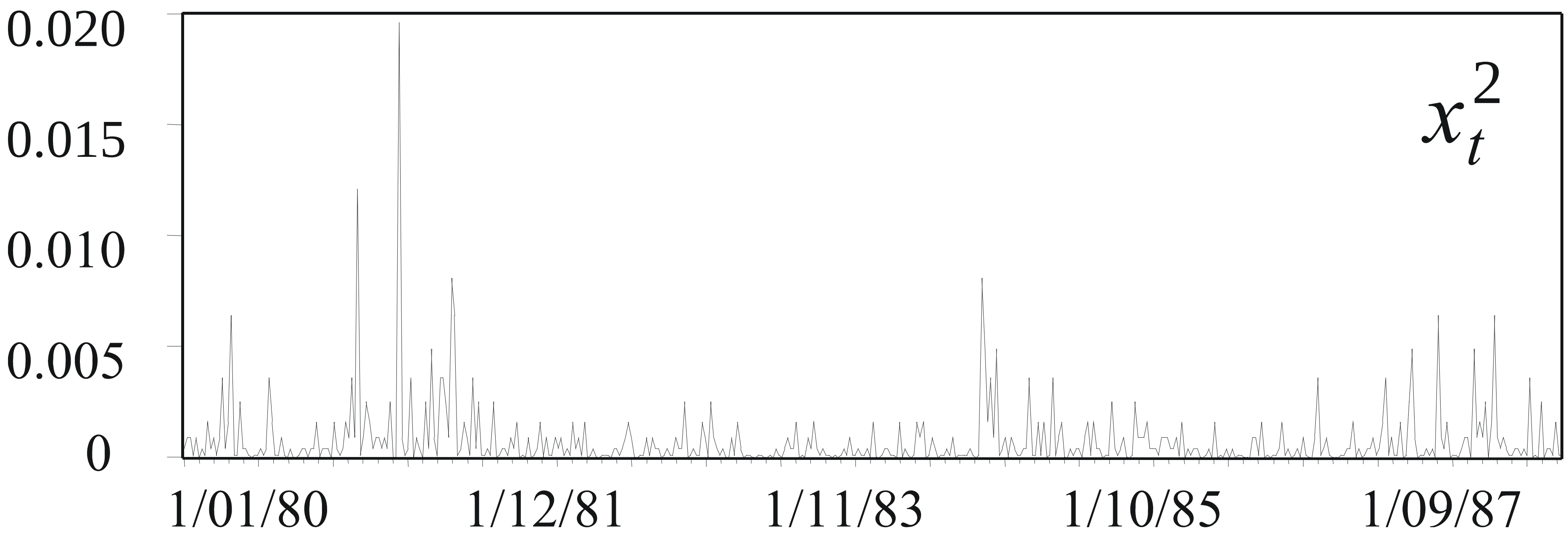


This series has mean = -0.000917(0.001273) and is uncorrelated.

Serial correlation test with 12 lags:

Box-Pierce test: $Q(12) = 13.17$.

But there is an obvious pattern in the *volatility*.



The serial correlation test of the squares is

$$Q(12) = 34.645.$$

The ARCH (Auto-Regressive Conditional Heteroscedasticity) model can capture such effects.

Suppose

$$x_t = h_t^{1/2} u_t$$

where u_t is an i.i.d. process and

$$h_t = \gamma + \alpha_1 x_{t-1}^2 + \cdots + \alpha_p x_{t-p}^2.$$

with $\alpha_0 > 0$, $\alpha_j \geq 0$ for each j and $\alpha_1 + \ldots + \alpha_p < 1$.

Note that

$$h_t = E(x_t^2 | \mathcal{F}_{t-1})$$

where $\mathcal{F}_t = \sigma(x_t, x_{t-1}, x_{t-2}, \ldots)$. This is the ARCH($p$) model.

Subject to the above restrictions, the process is stationary.

Taking expectations of both sides, and using LIE, note that

$$E(x_t^2) = E(h_t) = \gamma + \alpha_1 E(x_{t-1}^2) + \cdots + \alpha_p E(x_{t-p}^2) = \frac{\alpha_0}{1 - \alpha_1 - \ldots - \alpha_p} < \infty.$$

Defining

$$v_t = x_t^2 - h_t = h_t(u_t^2 - 1),$$

note that we can equivalently write the model as

$$x_t^2 = \gamma + \alpha_1 x_{t-1}^2 + \cdots + \alpha_p x_{t-p}^2 + v_t$$

and hence ARCH can be thought of as an AR(p) model in the squared process.

However, whereas

$$E(v_t|\mathcal{F}_{t-1}) = 0$$

by construction, note that

$$E(v_t^2|\mathcal{F}_{t-1}) = h_t^2 E((u_t^2 - 1)^2) = h_t^2(\mu_4 - 1)$$

where $\mu_4 = E(u_t^4)$.

Note that v_t is accordingly not a serially independent process, but is highly conditionally heterosedastic.

GARCH (Generalised ARCH)

Consider the GARCH(1,1) case.

$$h_t = \gamma + \alpha_1 x_{t-1}^2 + \beta_1 h_{t-1}$$

where $\alpha_0 > 0, \alpha_1, \beta_1 \geq 0$, If $\alpha_1 + \beta_1 < 1$, then this process is stationary with

$$E(x_t^2) = E(h_t) = \gamma + \alpha_1(x_{t-1}^2) + \beta_1 E(h_{t-1}) = \frac{\gamma}{1 - \alpha_1 - \beta_1} < \infty$$

Interpretation

1. By analogy with ARMA, this is a case of ARCH(∞):

$$\begin{aligned} h_t &= \frac{\gamma}{1-\beta_1} + \alpha_1 x_{t-1}^2 + \alpha_1 \beta_1 x_{t-2}^2 + \alpha_1 \beta_1^2 x_{t-3}^2 + \cdots \\ &= \omega + \sum_{j=1}^{\infty} \theta_j x_{t-j}^2 \end{aligned}$$

Note the implication: subject to the stationarity condition $\alpha_1 + \beta_1 < 1$,

$$\sum_{j=1}^{\infty} \theta_j = \alpha_1 \sum_{j=0}^{\infty} \beta_j = \frac{\alpha_1}{1-\beta_1} < 1$$

2. Write

$$x_t^2 = h_t + v_t = \gamma + \alpha_1 x_{t-1}^2 + \beta_1 h_{t-1} + v_t = \alpha_0 + \delta_1 x_{t-1}^2 + v_t - \beta_1 v_{t-1}$$

where $\delta_1 = \alpha_1 + \beta_1$. This is the ‘ARMA in squares’ representation.

Note in particular that δ_1 is the autoregressive coefficient, hence the stationarity analogy with models of the mean is reinforced.

Again, care is necessary in the interpretation!

3. GARCH(p,q)

Letting $\delta(L) = 1 - \delta_1 L - \cdots - \delta_p L^p$ and $\beta(L) = 1 - \beta_1 L - \cdots - \beta_p L^p$, write the "ARMA-in-squares" representation as

$$\delta(L)x_t^2 = \gamma + \beta(L)v_t$$

which is equivalent to the ARCH(∞) representation

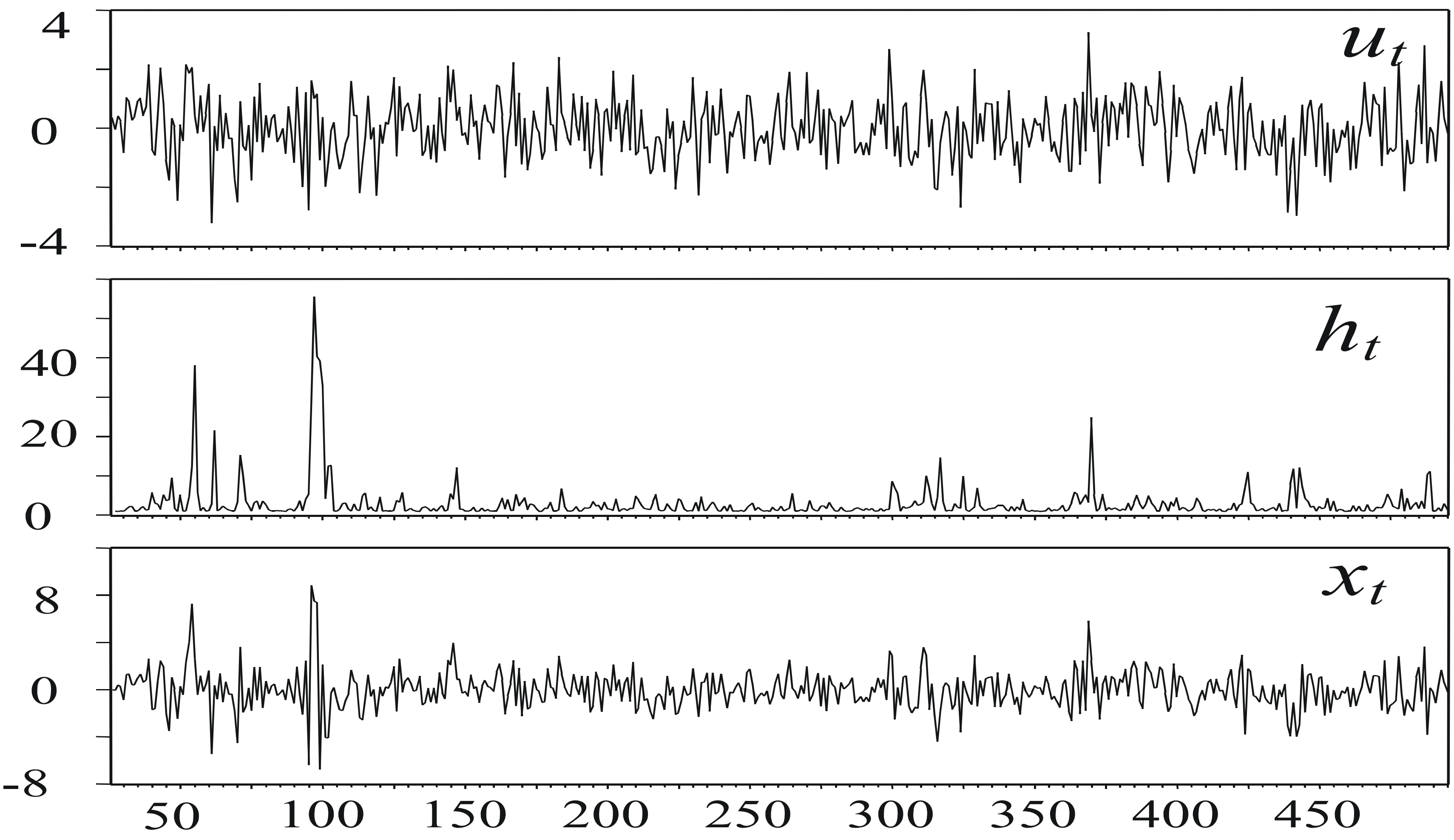
$$h_t = x_t^2 - v_t = \omega + \left(1 - \frac{\delta(L)}{\beta(L)}\right)x_t^2 = \omega + \theta(L)x_t^2$$

- Observe that $\theta_0 = 0$ by construction.
- Note also, $\omega = \gamma/\beta(1)$.

Simulations

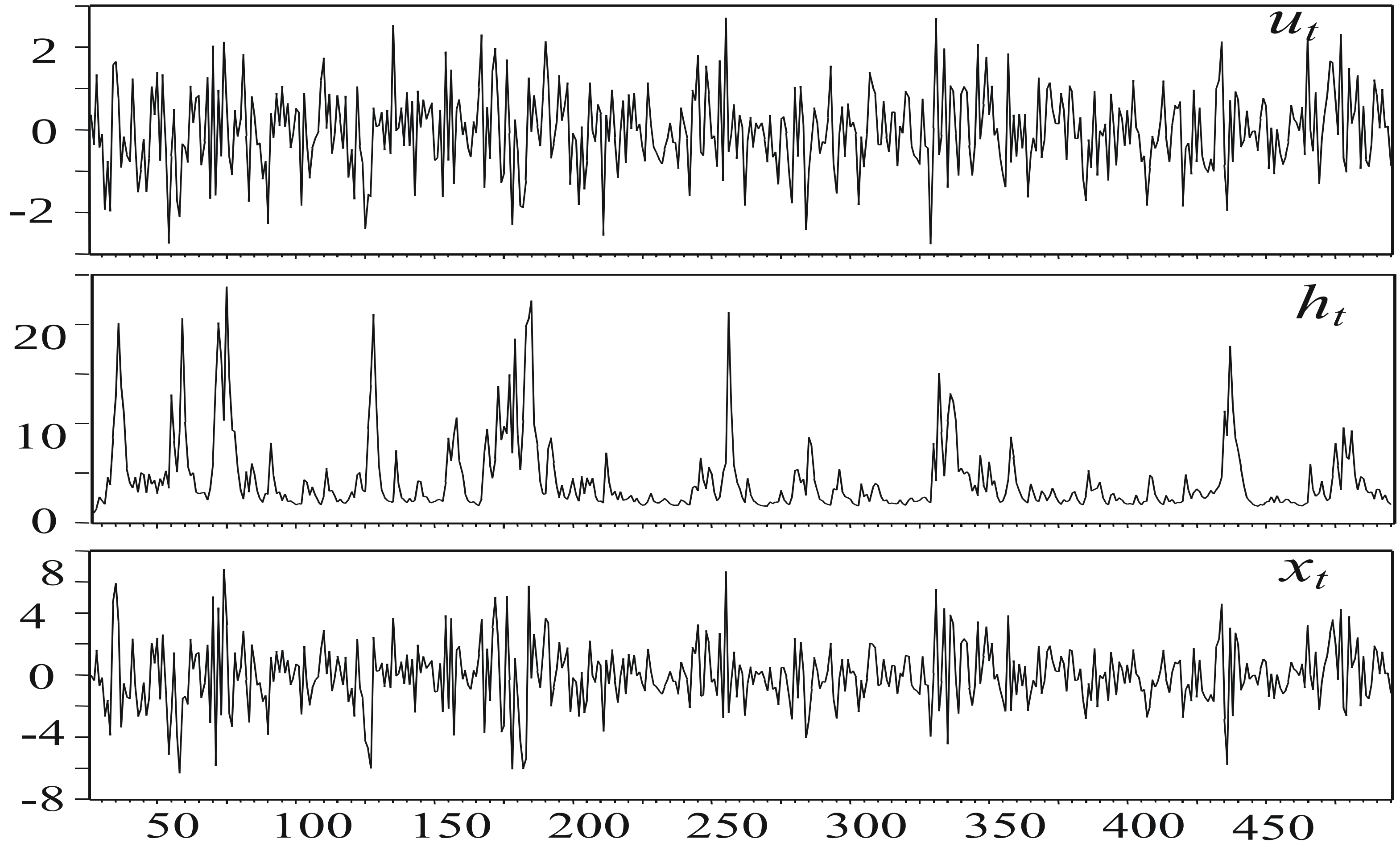
Case 1: ARCH,

$$h_t = 1 + 0.7x_{t-1}^2$$



Case 2: GARCH,

$$h_t = 1 + 0.4x_{t-1}^2 + 0.4h_{t-1}$$



Testing for ARCH

The LM test is the standard test for ARCH in regression residuals.

Implementation: To test for ARCH(m),

1. Regress x_t^2 onto $x_{t-1}^2, \dots, x_{t-m}^2$ and get R^2.
2. Calculate statistic LM = TR^2 where T = sample size.

LM is distributed as $\chi^2(m)$ in large samples when H0: x_t i.i.d$(0,\sigma^2)$ is true.

- This is also the optimal test for GARCH, with $m = \max(p,q)$.
- To test for ARCH effects in time series, first fit an ARMA or ARIMA process, *then* test the residuals for ARCH.

Example

Testing for ARCH in the $/£ exchange rate: (recall there is no autocorrelation in this series, so assume ARMA(0,0)).

In the test for ARCH(1), LM = 2.798 (p = 0.094).

but LM(6) = 19.99 (p = 0.0027) (biggest coefficient is at lag 3).

Estimating the ARCH Model

Example: Assume the AR(1)/ARCH(1) model,

$$\begin{aligned} x_t &= \lambda_0 + \lambda_1 x_{t-1} + \varepsilon_t \\ u_t &= h_t^{1/2} e_t, \quad e_t \sim N(0,1) \\ h_t &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 \\ &= \alpha_0 + \alpha_1 (x_{t-1} - \lambda_0 - \lambda_1 x_{t-2})^2 \end{aligned} \qquad (*)$$

Method of Maximum Likelihood.

As with ARMA models, construct the likelihood function by sequential conditioning.

Conditional density of an observation is

$$\phi(x_t | x_{t-1}, x_{t-2}, \ldots) = \phi(x_t | x_{t-1}, x_{t-2}) = \frac{1}{\sqrt{2\pi h_t}} e^{-\frac{1}{2}(\varepsilon_t^2 / h_t)}.$$

Note, by (*) this formula depends only on x_t, x_{t-1}, x_{t-2}, where x_{t-1}, x_{t-2} are held conditionally fixed.

The approximate MLE is derived as follows:

1. Sacrifice two observations. Treat x_1, x_2 as fixed, and start sample at $t = 3$.
2. With x_1, x_2 fixed, $\phi(x_3|x_1,x_2) = \phi(x_3)$.
3. The joint density of the sample can be calculated recursively,

$$\Phi(x_3, x_4, \ldots, x_T) = \phi(x_3) \times \phi(x_3|x_2) \times \ldots \times \phi(x_T|x_{T-1}, x_{T-2})$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^{T-2} \left(\prod_{t=3}^{T} h_t^{-1}\right) \exp\left\{-\frac{1}{2}\sum_{t=2}^{T}(\varepsilon_t^2/h_t)\right\}$$

4. Hence, the approximate log-likelihood function is

$$L(\lambda_0, \lambda_1, \alpha_0, \alpha_1) = K - \frac{1}{2}\sum_{t=3}^{T}\left(\log h_t + \frac{\varepsilon_t^2}{h_t}\right)$$

... to be maximised w.r.t. $\lambda_0, \lambda_1, \alpha_0, \alpha_1$.

Points:

- While x_1 and x_2 are not really fixed, the approximation involved is of small order when T is large.
- the ML estimates are as usual CAN, and asymptotically efficient when the model is correct.
- This method can also be justified as QML. Gaussianity not essential!
- Handy formulae exist for the asymptotic standard errors, covariance matrix, etc.
- Tests (Wald, LM, LR) for restrictions on the parameters can be easily derived.

Estimating the GARCH Model

The ARMA and GARCH models depend on an infinite number of lags. The pre-sample data have to be approximated in some way.

Neatest approach: express the model in terms of residuals and set pre-sample values to 0.

Example: in the ARMA(1,1)/GARCH(1,1) model the log-likelihood function is (as above)

$$L(\lambda_0, \lambda_1, \theta_1, \alpha_0, \alpha_1, \beta_1) = K - \frac{1}{2}\sum_{t=2}^{T}\left(\log h_t + \frac{\varepsilon_t^2}{h_t}\right)$$

where now

$$\varepsilon_t - x_t - \lambda_0 - \lambda_1 x_{t-1} - \theta_1 \varepsilon_{t-1} \qquad (t \geq 2) \qquad (*)$$

and (recall) $h_t = \varepsilon_t^2 - v_t$ where

$$v_t = \varepsilon_t^2 - \alpha_0 - (\alpha_1 + \beta_1)\varepsilon_{t-1}^2 - \beta_1 v_{t-1} \qquad (t \geq 3) \qquad (**)$$

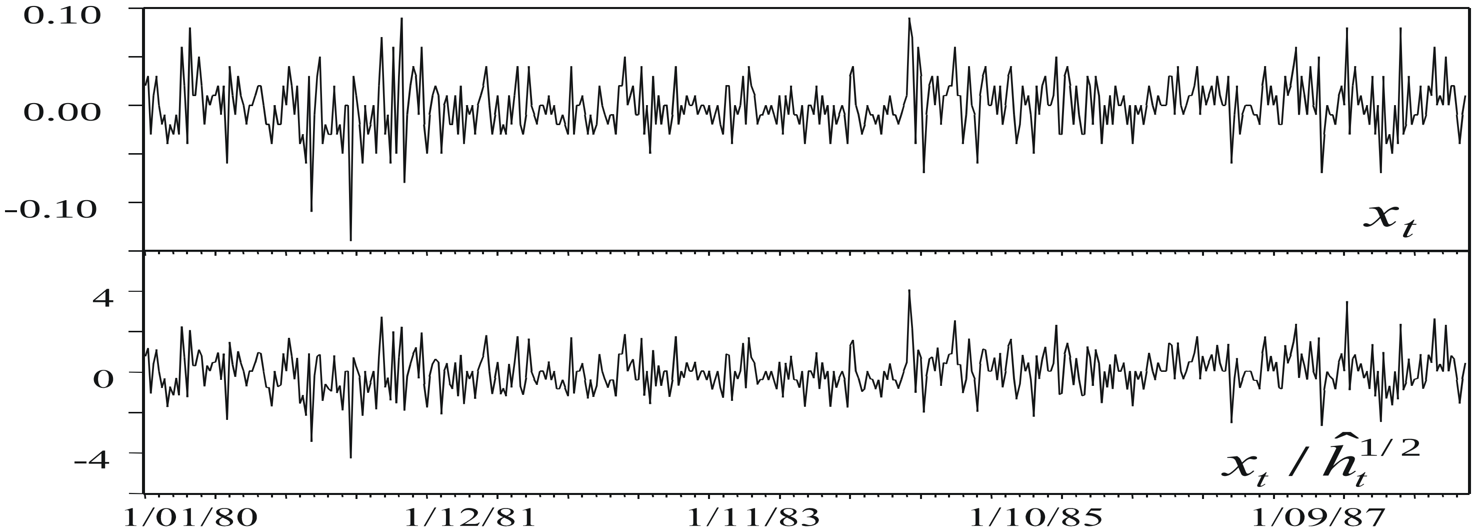
To generate ε_t and h_t for $t = 2, 3, \ldots, T,$ use equations (*) and (**) recursively, with $\varepsilon_1 = v_1 = 0.$

- The recurrsive formulae can be used to evaluate L at any desired values of $(\lambda_0, \lambda_1, \theta_1, \alpha_0, \alpha_1, \beta_1)$.
- The approximation is again of small order with large T.

Modelling the \$/£ rate

In view of LM test result, fit ARCH(3). We obtain (s.e.s in parentheses).

$$\hat{h}_t = \underset{(.000004)}{0.0005} + \underset{(.042)}{0.714}\, x_{t-1}^2 + \underset{(.041)}{0.720}\, x_{t-2}^2 + \underset{(.0479)}{0.247}\, x_{t-3}^2$$



Test for ARCH(10) in $x_t/\hat{h}_t^{1/2}$:

$$\text{LM} = 7.61 \text{ (p = 0.66)},$$

Normality (Jarque-Bera) tests: ($\chi^2(2)$ on H_0)

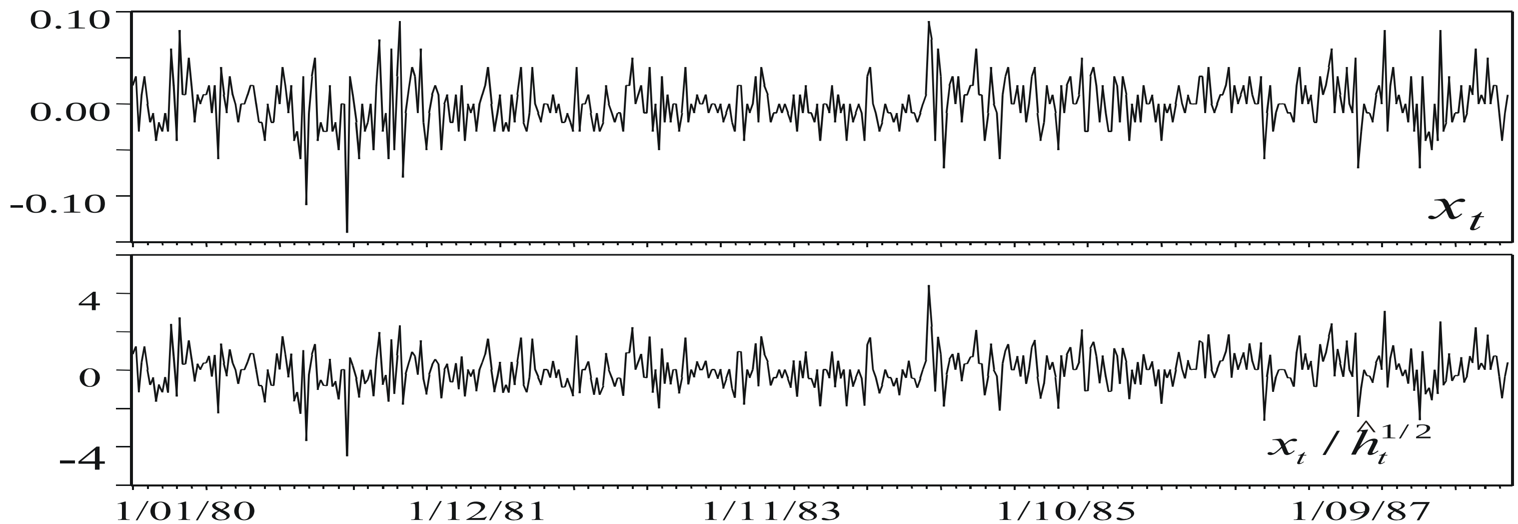
$$: x_t\text{: JB} = 83.45. \qquad x_t/\hat{h}_t^{1/2}\text{: JB} = 36.92.$$

$$\text{AIC} = -7.163, \text{ SIC} = -7.118$$

GARCH(1,1) results:

$$\hat{h}_t = \underset{(.000014)}{0.000032} + \underset{(.021)}{0.068}\, x_{t-1}^2 + \underset{(.033)}{0.89}\, \hat{h}_{t-1}$$

AIC = −7.167, SIC = −7.131



This equation is very slightly preferred on parsimony grounds, but note, JB = 50.07

Points:

- Neither model reduces the series to a *normal* i.i.d. - excess leptokurtosis evident.
- This is a potential problem if we assume normality to construct and interpret the MLE (but estimator should still be CAN, at least).
- Could construct MLE based on Student's *t* distribution (implemented in TSM).

Variations on the Theme

1. Leverage effects - TARCH. (Threshold ARCH). Different response to squared shocks depending on sign: for example,

$$h_t = \alpha_0 + \alpha_1 x_{t-1}^2 + \gamma_1 x_{t-1}^2 I(x_t < 0) + \beta_1 h_{t-1}$$

2. EGARCH - different functional form:

$$\log h_t = \alpha_0 + \alpha_1 g_t + \beta_1 \log h_{t-1}$$

where $g_t = g(e_t)$, such that $e_t = h_t^{-1/2} x_t$ is assumed iid$(0,1)$, and $E(g_t) = 0$.

- e.g. $g_t = |e_t| - E|e_t| + \mu e_t$, to allow for leverage effects.
- Proposed by Nelson (1991) in conjunction with a fatter tailed-than-normal error distribution (GED).

3. APARCH (asymmetric power garch)

$$h_t^{\eta/2} = \alpha_0 + (\alpha_1 + \gamma_1 I(x_t < 0))|x_{t-1}|^\eta + \beta_1 h_{t-1}$$

Leverage effects, and η estimated as an additional parameter.

4. IGARCH This is the nonstationary ("integrated") version of GARCH:

The GARCH(1,1) case is

$$h_t = \gamma + \alpha_1 x_{t-1}^2 + \beta_1 h_{t-1}, \qquad \alpha > 0, \beta > 0, \alpha + \beta = 1$$

The GARCH(p,q) case is $(1-L)\delta(L)x_t^2 = \gamma + \beta(L)v_t$, or in ARCH($\infty$) representation,

$$h_t = \omega + \left(1 - \frac{\delta(L)(1-L)}{\beta(L)}\right)x_t^2$$

where $\delta(L).is$ of order $p-1$.

- The operator $(1-L)$ gives the appearance that x_t^2 is an ‘integrated’ process, hence ‘Integrated GARCH’.

- However, this is a misnomer!
  - IGARCH is wide-sense nonstationary since $E(x_t^2)$ is undefined.
  - e.g., IGARCH(1,1) has solution

$$h_t = \omega + (1-\beta_1)x_{t-1}^2 + (1-\beta_1)\beta_1 x_{t-2}^2 + (1-\beta_1)\beta_1^2 x_{t-3}^2 + \cdots$$

The *memory* of the process decays exponentially - it’s zero if $\beta_1 = 0$.

- We must be careful with adapting established time series concepts to conditional variance models!

FIGARCH

The "FI" generalization is FIGARCH:

$$h_t = \omega + \left(1 - \frac{\delta(L)(1-L)^d}{\beta(L)}\right)x_t^2$$

for $0 < d \leq 1$. This model solves as

$$h_t = \omega + \theta_1 u_{t-1}^2 + \theta_2 u_{t-2}^2 + \theta_3 u_{t-3}^2 + \cdots$$

where $\theta_j = O(j^{-d-1})$ but also $\sum_{j-1}^{\infty} \theta_j = 1$.

- Covariance nonstationary, and hyperbolic memory decay.
- *Not* long memory in the “usual” sense, i.e. $\sum_{j-1}^{\infty} \theta_j = \infty$.
- If $E(v_t) = \mu > 0$ and $\theta_j = O(j^{\delta-1})$ for $\delta > 0$ then

$$E(h_t) = E(\textstyle\sum_{j-1}^{t} \theta_j v_{t-j}) = O(t^{\delta})$$

  – diverges!

- Analogies between long memory processes in mean and variance are *misleading*.

HYGARCH model (Davidson 2004)

$$h_t = \omega + \left(1 - \frac{1 + a((1-L)^d - 1)}{1 - \beta L}\right)u_t^2$$

- Equivalent to FIGARCH when $a = 1$.
- Covariance stationary when $a < 1$.
- Stationary GARCH with root $a < 1$ when $d = 1$.

Regime-Switching

Consider a model written in generic form as

$$g_t(\theta(S_t)) \;\; = h_t(\theta(S_t))^{1/2} e_t$$

where $e_t \sim iid(0,1)$ and $S_t = 1,\ldots,M$ is a dummy variable denoting the regime prevailing at time t, and $\theta(S_t)$ represents the parameter values applying in regime S_t.

The vectors $\theta(1),\ldots,\theta(M)$ are to be estimated

In general, S_t is a random sequence that may either depend upon the observed variables, or be generated exogenously.

Example:

$$y_t = \alpha(S_t) + \lambda(S_t) y_{t-1} + u_t$$

where $u_t = \sigma e_t$ and in the above notation,

$$g_t(\alpha(S_t), \lambda(S_t)) \;\; = u_t.$$

Imagine a case $|\lambda(1)| < 1$ and $\lambda(2) = 1$, so that the process is either a random walk or a stationary process, depending on the regime.

Could also specify $u_t = h_t^{1/2} e_t$ o that the volatility process also depends on the regime.

$$h_t = \omega(S_t) + \beta(S_t) u_t^2.$$

Switching Mechanisms

(See Kim and Nelson (1999) for details.)

Simple Markov switching:

The switching is under the control of a Markov-chain updating mechanism with fixed transition probabilities.

$$p_{ji} = Pr(S_t = j|S_{t-1} = i). \tag{1}$$

subject to

$$\sum_{j=1}^{M} p_{ji} = 1.$$

Let

$$f(y_t|S_t = j, \Psi_{t-1}) \tag{2}$$

denote the conditional probability density of the dependent variable at time t when regime j is operating.

For example,

$$f(y_t|S_t = j, \Psi_{t-1}) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}(y_t - \alpha(j) - \lambda(j) y_{t-1})^2} \tag{3}$$

Points:

- Ψ_{t-1} represents the history of the whole process $\{Y_t, S_t\}$ up to date $t-1$.

- The distribution of Y_t can be thought of as described by a two-stage sequential procedure:

  - Draw a regime S_t according to (1)
  - Draw Y_t from (2).

It is also possible to have p_{ji} a function of predetermined variables - *explained switching*.

The Recursive Equations

The probability of regime j prevailing at date t evolves according to

$$\Pr(S_t = j|\Psi_{t-1}) = \sum_{i=1}^{M} p_{ji} \Pr(S_{t-1} = i|\Psi_{t-1}) \tag{4}$$

and the conditional density of the switching process Y_t at the point y_t is then the weighted average

$$f(y_t|\Psi_{t-1}) = \sum_{j=1}^{M} f(y_t|S_t = j, \Psi_{t-1}) \Pr(S_t = j|\Psi_{t-1}).$$

To generate a sequence of these densities for $t = 1, 2, 3, \ldots$, applying these two relations, calls for a formula for

$$\Pr(S_t = j|\Psi_t).$$

- This is the probability that can be assigned to regime j *after* the out-turn $Y_t = y_t$ is known.
- Applying a version of Bayes' theorem for densities, this can be calculated as

$$\Pr(S_t = j|\Psi_t) = \frac{f(y_t|S_t = j, \Psi_{t-1}) \Pr(S_t = j|\Psi_{t-1})}{f(y_t|\Psi_{t-1})}$$

$$= \frac{f(y_t|S_t = j, \Psi_{t-1}) \Pr(S_t = j|\Psi_{t-1})}{\sum_{i=1}^{M} f(y_t|S_t = i, \Psi_{t-1}) \Pr(S_t = i|\Psi_{t-1})} \tag{5}$$

Estimation

The log-likelihood function to be maximized with respect to the parameters

$$\theta(1),\dots\theta(M),\, p_{ji},\, i = 1,\dots,M,\ j = 1,\dots,M-1,$$

is

$$L = \sum_{t=1}^{T} \log f(y_t|\Psi_{t-1}) = \sum_{t=1}^{T} \log\left(\sum_{j=1}^{M} f(y_t|S_t = j, \Psi_{t-1}) \Pr(S_t = j|\Psi_{t-1}) \right)$$

- This is calculated recursively, using the density formula (3) and equations (4) and (5).
- The series $Pr(S_t = j|\Psi_{t-1})$ for $j = 1,\dots,M-1$, the ‘filter probabilities’, are a by-product of the estimation.
- The ‘smoothed probabilities’ $Pr(S_t = j|\Psi_T)$ for $t = 1,\dots T$ can also be computed by a backwards recursion - see Kim and Nelson (1999) for details.

Hamilton's Model

Hamilton (Econometrica 1989) proposed a model of the business cycle in which the drift of a unit-root process switched randomly:

Suppose Y_t denotes quarterly GDP growth, with mean $\mu(1)$

$$Y_t = \mu(S_t) + \phi_1(Y_{t-1} - \mu(S_{t-1})) + \cdots + \phi_p(Y_{t-p} - \mu(S_{t-p})) + u_t$$

($p = 4$) where $\mu(1) > 0$ and $\mu(2) < 0$ (say).

- In effect, the drift at date t is

$$m_t = \mu(S_t) - \phi_1\mu(S_{t-1}) - \cdots - \phi_p\mu(S_{t-p}).$$

- This has 2^{p+1} possible discrete values, depending on the sequence of Markov states the system has passed through in the last p periods.
- However, except for replacing M by M^{p+1} and needing to compute $\Pr(S_t = j_0, \ldots S_{t-p} = j_p|\Psi_t)$ the basic idea is the same.
- Hamilton provides an updating algorithm to compute the likelihood in this case.

Nonlinear AR Models

Self-Exciting Threshold Autoregression (SETAR)

$$y_t = \alpha(1) + \phi(1)y_{t-1} + u_t, \quad y_{t-1} < y^*$$

$$y_t = \alpha(2) + \phi(2)y_{t-1} + u_t, \quad y_{t-1} \geq y^*$$

Here, the process itself determines the current regime, rather than an autonomous switching mechanism.

Smooth-Transition Model

$$y_t = G_t\alpha(1) + (1 - G_t)\alpha(2) + (G_t\phi(1) + (1 - G_t)\phi(2))y_{t-1} + u_t$$

where $G_t = \dfrac{1}{1 + e^{-\gamma(y_{t-1} - y^*)}}$.

- If γ is large enough, this is similar to the simple TAR model.
- A double-threshold model is also a possibility:

$$G_t = \frac{1}{1 + e^{-\gamma(y_{t-1} - y_1^*)(y_{t-1} - y_2^*)}}$$

so that G_t can be close to 0 for $y_1^* < y_{t-1} < y_2^*$ and close to 1 otherwise.

Bilinear Models

Extends the ARMA class with interaction terms.

Consider a simple example:

$$y_t = \phi_1 y_{t-1} + u_t + \theta_1 u_{t-1} + \psi_1 y_{t-1} u_{t-1}.$$

- Bilinear models have been advocated by authors such as Subba Rao (1981), Priestley (1988) as a "flexible functional form" which can approximate a range of different nonlinear structures.
- Solving out yields a "nonlinear moving average" model involving higher powers of lagged shocks.
- Hence, u_t needs to possess all its moments in general (e.g. Gaussian).

The fully general BL(p,q,m,r) class proposed by Subba Rao allows the "coefficients" of y_{t-j} for $j = 1,\ldots,r$ to be different lag distributions of order m.

TSM estimates a restricted class having the general form

$$y_t = \sum_{j=1}^{p} \phi y_{t-j} + \sum_{j=1}^{q} \theta_j u_{t-j} + \left(\sum_{k=1}^{r} \lambda_k u_{t-k} \right) \left(\sum_{j=1}^{p} \psi_j y_{t-j} \right) + u_t$$

with $\lambda_1 = 1$.