MATH3802
Chapter 4: Model fitting

Peter Thwaites

with thanks to Stuart Barber for previous years' slides and notes

March 2017

4.1 Model selection

- How to choose between AR, MA, ARMA, and ARIMA?
- We know

Model	ρ_k
MA(q)	$\rho_k = 0 \quad \forall k > q$
AR(p)	$\rho_k = \alpha_1 \rho_{k-1} + \cdots + \alpha_p \rho_{k-p}$
ARMA(p, q)	As for AR(p), except for the first q values

- This makes it easy to identify MA(q).
- How to spot AR(p)?

Lemma 4.1 Let $y_1, \dots, y_p$ be the roots of $\alpha(y) = 1 - \alpha_1 y - \cdots - \alpha_p y^p$. Then for AR($p$),

$$\rho_k = \sum_{j=1}^{p} \frac{c_j}{y_j^k}, \tag{1}$$

for suitable constants $c_1, \dots, c_p$.

How to recognise solutions to (1)?

We have

$$\rho_k = \sum_{j=1}^{p} \frac{c_j}{y_j^k}. \tag{1}$$

AR(1) $\rho_k = \alpha^k$. Hence

exponential decay ($\alpha > 0$) or
exponential decay with alternating signs ($\alpha < 0$).

AR(p), $p > 1$ Damped oscillations with

$$\rho_k = c^k \cos(\phi k).$$

for some constants ϕ, c, $|c| < 1$.

ARIMA(p, d, q) Here, ρ_k decays very slowly; difference until ρ_k indicates AR, MA, or ARMA.

Partial autocorrelation

Definition 4.1 (Partial autocorrelation) For $k = 1, 2, \ldots$, use the Yule-Walker equations to fit an AR(k) model (using $\rho_1, \ldots, \rho_k$), obtaining coefficients $\alpha_{k1}, \ldots, \alpha_{kk}$. Then α_{kk} is called the *lag-k partial autocorrelation* (pacf) of $\{X_t\}$.

One can show that the pacf satisfies

Model	α_{kk}
AR(p)	cut off after lag p
MA(q), ARMA(p, q)	exponential decay / damped oscillations

Results

- if $X_1, \ldots, X_n$ is MA(q), then for $k > q$

$$\widehat{\rho}_k \sim N\left(0, \frac{1}{n}\left[1 + 2\sum_{i=1}^{q} \widehat{\rho}_i^2\right]\right).$$

- If $X_1, \ldots, X_n$ is AR(p), then for $k > p$

$$\widehat{\alpha}_{kk} \sim N\left(0, \frac{1}{n}\right).$$

Run the R script sim.arima.R, available in the VLE.

4.2 Parameter estimation

So far, we have assumed our stationary processes are zero mean.

We can cope with a non-zero mean μ for $\{X_t\}$ by modelling $Y_t = X_t - \mu$. For example, our AR(1) model becomes

$$X_t - \mu = \alpha(X_{t-1} - \mu) + \varepsilon_t \qquad t = 1, 2, \ldots, n.$$

We estimate μ by $\widehat{\mu} = n^{-1} \sum_{t=1}^{n} X_t$, subtract $\widehat{\mu}$ from the observations to estimate $\{Y_t\}$ and proceed with fitting a model to the $\{Y_t\}$.

Method of moments

Choose parameters such that the means and correlations for model and sample coincide. For AR(p) models, this gives the Yule-Walker equations.

Example 4.1 (AR(1)) From the YW equations, $\widehat{\alpha}_1 = \widehat{\rho}_1$.

We can estimate σ_ε^2 via the method of moments. Assuming $\{X_t\}$ is stationary,

$$\widehat{\sigma}_\varepsilon^2 = \widehat{\sigma}_X^2(1 - \widehat{\alpha}_1^2).$$

We estimate σ_X^2 by the sample variance, ie $\widehat{\sigma}_X^2 = s_x^2$.

Least squares estimation (LSE)

Choose parameters to minimise the residual sum of squares

Example 4.2 (AR(1)) We estimate α by minimising

$$S(\alpha) = \sum_{t=1}^{n} (X_t - \alpha X_{t-1})^2 .$$

Hence

$$\widehat{\alpha} \approx \widehat{\rho}_1 \qquad \widehat{\sigma}_\varepsilon^2 \approx s_X^2(1 - \widehat{\alpha}^2)$$

— the same as MOM.

Maximum likelihood estimation (MLE)

To use this method, we need to assume a distribution for the ε_t.
We then minimise the *likelihood* $L(\alpha|X) = f(X|\alpha)$.

Example 4.3 (AR(1)) Assume

$$X_0 = 0, X_t = \alpha X_{t-1} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2).$$

We get

$$\widehat{\alpha} \approx \widehat{\rho}_1 \qquad \widehat{\sigma}_\varepsilon^2 \approx s_X^2(1 - \widehat{\alpha}^2),$$

ie, for the AR(1) process with normally distributed innovations, the MLE and LSE are the same.

4.3 Invertibility

Consider the MA(1) process $X_t = \varepsilon_t + \beta\varepsilon_{t-1}$ where ε is a white noise process with variance σ_ε^2.

Using the method of moments,

$$\widehat{\beta} = \frac{1 \pm \sqrt{1 - 4\widehat{\rho}_1^2}}{2\widehat{\rho}_1}, \qquad \widehat{\sigma}_\varepsilon^2 = s_X^2/(1+\widehat{\beta}^2).$$

How can we distinguish between these solutions?

AR representation

Write X_t as $X_t = (1+\beta B)\varepsilon_t$ where B is the backshift operator. Then

$$X_t = \varepsilon_t + \beta X_{t-1} - \beta^2 X_{t-2} + \beta^3 X_{t-3} - \cdots .$$

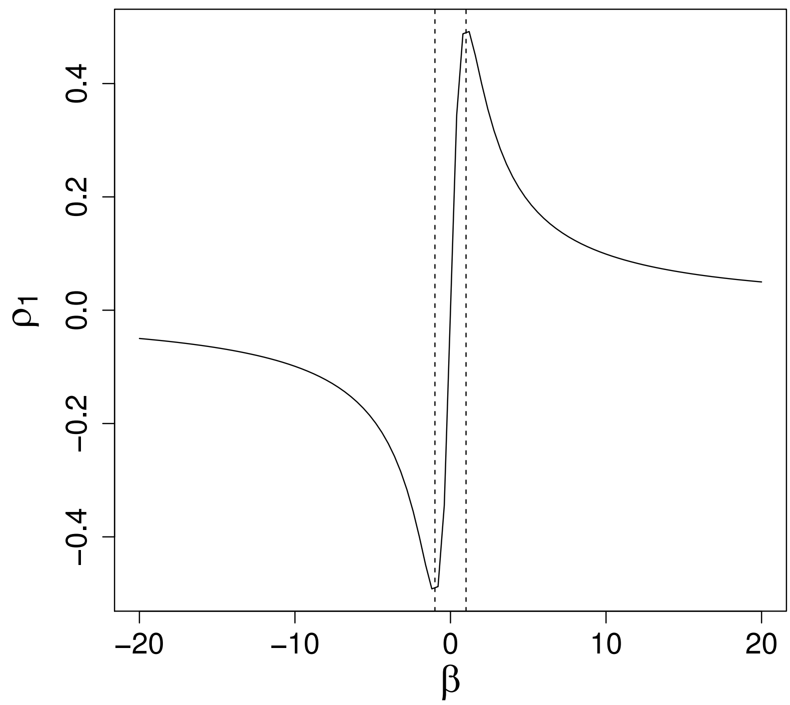
Consider two cases:

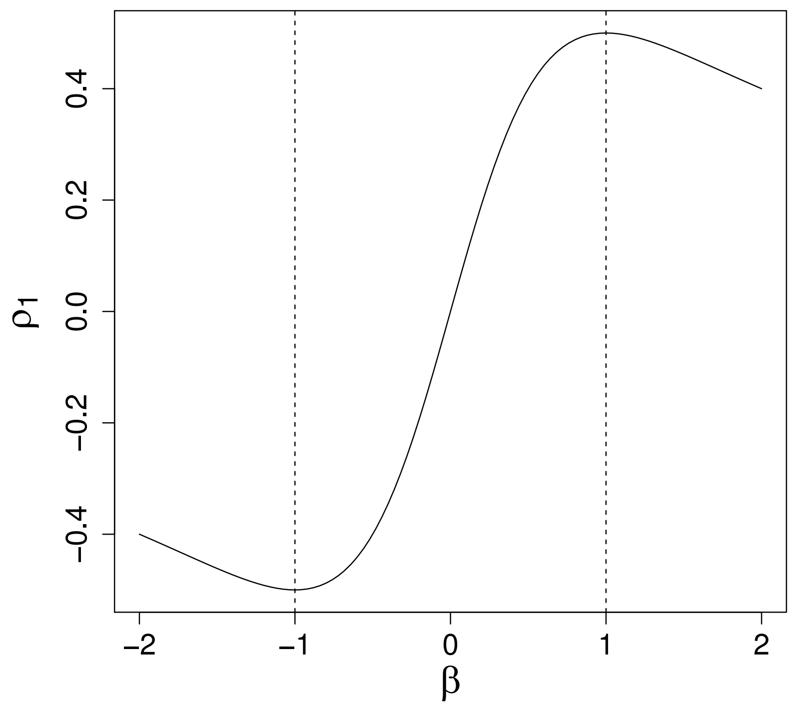
1. $|\beta| > 1$
2. $|\beta| < 1$.

Case 1 is unrealistic so requiring $|\beta| < 1$ seems sensible. This condition is called *invertibility*; requiring it means we use

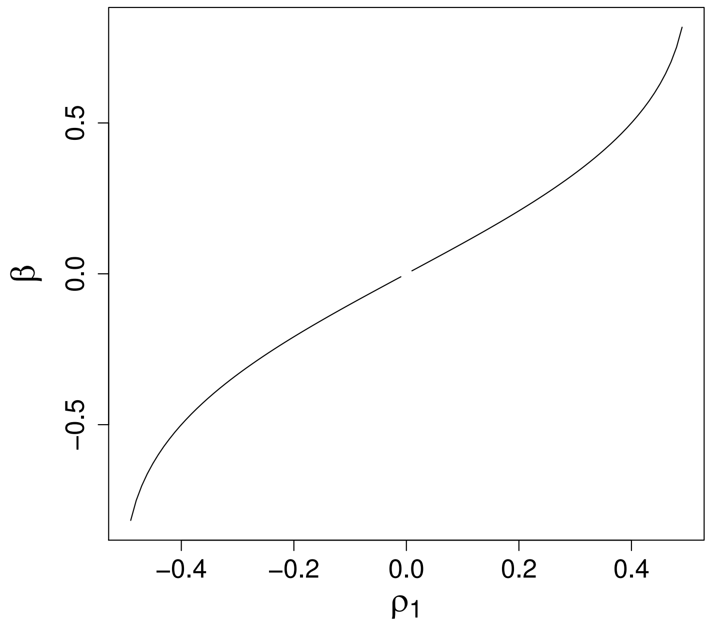
$$\widehat{\beta} = \frac{1 - \sqrt{1 - 4\widehat{\rho}_1^2}}{2\widehat{\rho}_1}$$

since then $|\widehat{\beta}| < 1$.

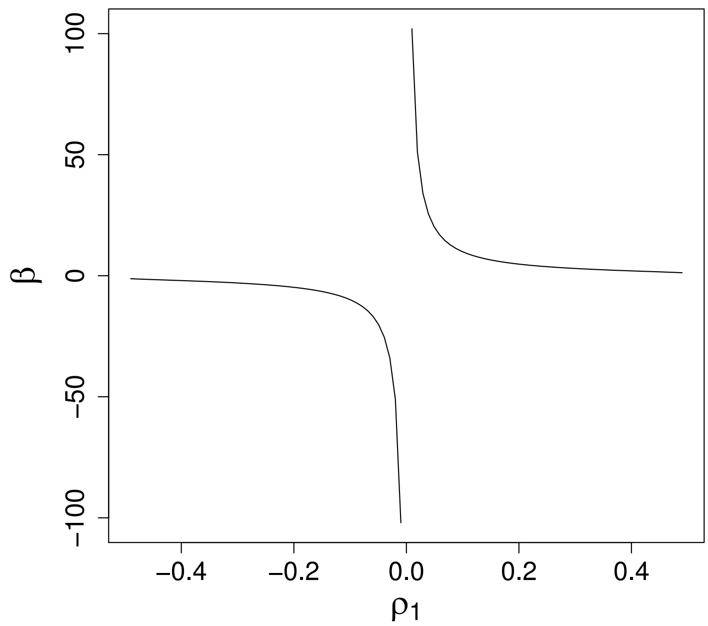

0.4
0.2
0.0
−0.2
−0.4
ρ1
−20
−10
0
10
20
β


0.4
0.2
0.0
−0.2
−0.4
ρ1
−2
−1
0
1
2
β

Invertible



Non invertible



MA(q)

For a general MA(q) process $X_t = \beta(B)\varepsilon$, where

$$\beta(y) = 1 + \beta_1 y + \beta_2 y^2 + \cdots + \beta_q y^q.$$

Suppose $\beta(y) = 0$ has roots $y_1^{-1}, \ldots, y_q^{-1}$. Then

$$\varepsilon_t = \sum_{i=0}^{\infty} h_i X_{t-i}$$

for $h_i = \sum_{k=1}^{q} c_k y_k^i$, for some $\{c_k\}$ (cf prop. 3.1).

Hence we require all the roots of $\beta(y)$ lie outside the unit circle.

Equivalent processes

The 2^q processes found using

$$\begin{aligned}
\beta_{(1)}(B) &= (1 - By_1)(1 - By_2)\cdots(1 - By_q) \\
\beta_{(2)}(B) &= (1 - By_1^{-1})(1 - By_2)\cdots(1 - By_q) \\
\beta_{(3)}(B) &= (1 - By_1)(1 - By_2^{-1})\cdots(1 - By_q) \\
&\vdots \\
\beta_{(2^q)}(B) &= (1 - By_1^{-1})(1 - By_2^{-1})\cdots(1 - By_q^{-1})
\end{aligned}$$

all have the same acf, but only one is invertible.