STAT 804: Lecture 7

Model identification summary

Simplest model identification tactic – look for pure MA or pure AR model. To do so:

1. compute sample autocorrelation function (ACF). Plot $\widehat{ACF}(h)$ against h. If the process is an $MA(q)$ then the ACF will be 0 after lag q.

2. compute sample partial ACF (PACF). Plot estimated $PACF(h)$ against h. If the process is an $AR(p)$ then the PACF will be 0 after lag p.

Non stationary processes

If X is not stationary: transform X to find related stationary series.

In this course: two sorts of non-stationarity — non constant mean and integration.

Non constant mean: If $\mathrm{E}(X_t)$ is not constant we will hope to model $\mu_t = \mathrm{E}(X_t)$ using a small number of parameters and then model $Y_t = X_t - \mu_t$ as a stationary series.

Three common structures for μ_t: linear, polynomial and periodic.

Linear trend: Suppose

$$\mu_t = \alpha + \beta t$$

We seek to estimate α and β in order to analyze $Y_t = X_t - \alpha - \beta t$.

Method 1: regression (detrending). Regress X_t on t to get $\hat{\alpha}$ and $\hat{\beta}$ and analyze

$$\hat{Y}_t = X_t - \hat{\alpha} - \hat{\beta} t$$

We hope

$$\hat{Y} \approx Y$$

Method 2: differencing. Define

$$\begin{aligned} W_t &= X_t - X_{t-1} \\ &= [(I - B)X]_t \end{aligned}$$

Then

$$\begin{aligned} W_t &= [(\alpha + \beta t + Y_t) - (\alpha + \beta(t-1) + Y_{t-1})] \\ &= \beta + Y_t - Y_{t-1} \end{aligned}$$

Linear filter applied to series Y so stationary if Y is.

BUT, it might be stationary even if Y is not.

Suppose that ϵ_t is an iid mean 0 sequence. Then

$$Y_t = \sum_1^t \epsilon_j$$

is a **random walk**.

Notice that $Y_t - Y_{t-1} = \epsilon_t$ is stationary.

To see that Y_t, the random walk, is not stationary notice that Var$(Y_t) = t$Var(ϵ_1).

Random walk models common in Economics.

In physics: used in limit of very small time increments – leads to Brownian motion.

Definition: X satisfies ARIMA(p, d, q) model if

$$\phi(B)\nabla^d X = \psi(B)\epsilon$$

where

1. ϵ is white noise,

2. ϕ is a polynomial of degree p,

3. ψ is a polynomial of degree q,

4. $\nabla = I - B$ is the differencing operator.

Remark: If $X_t = \mu_t + Y_t$ where Y is stationary and μ is a polynomial of degree less than or equal to p then $\nabla^d X$ is stationary. (So a cubic shaped trend could be removed by differencing 3 times.)

WARNING: it is a common mistake in students' data analyses to over difference. When you difference a stationary $ARMA(p,q)$ you introduce a unit root in the defining polynomial – the result cannot be written as an infinite order moving average.

Detrending: Define a response vector

$$U = \begin{bmatrix} X_0 \\ \vdots \\ X_{T-1} \end{bmatrix}$$

and a design matrix by

$$V = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & T-1 \end{bmatrix}$$

Write

$$U = V\theta + Y$$

with

$$Y = \begin{bmatrix} Y_0 \\ \cdots \\ Y_{T-1} \end{bmatrix}$$

and

$$\theta = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}.$$

Then we estimate by

$$\hat{\theta} = (V^T V)^{-1} V^T U$$

Our detrended series is

$$\begin{aligned} \hat{Y} &= U - V\hat{\theta} \\ &= (I - V(V^T V)^{-1} V^T)U \end{aligned}$$

Since Cov(Y) $\neq \sigma^2 I$ ordinary least squares is not strictly appropriate.

In general in the model

$$U = V\theta + W$$

where W has mean 0 and variance covariance matrix Σ, the minimum variance linear estimator of θ is the generalized least squares estimate

$$\hat{\theta}_{GLS} = (V^T\Sigma^{-1}V)^T V^T \Sigma^{-1} U\,.$$

This estimate is unbiased and has variance

$$(V^T\Sigma^{-1}V)^{-1}$$

For this model the ordinary least squares estimate is also unbiased and has variance

$$(V^TV)^{-1}V^T\Sigma V(V^TV)^{-1}$$

Problem in our context (almost always a problem): can only use $\hat{\theta}_{GLS}$ if you know Σ.

In our context you won't know Σ until you have removed a trend, selected a suitable $ARMA$ model and estimated the parameters.

Natural proposal: follow an iterative process:

1. Fit linear model by ordinary least squares.

2. Compute the residual series.

3. Identify a suitable $ARMA(p, q)$ model.

4. Fit the parameters of this model.

5. Using the estimates compute an estimated variance covariance of the residual process.

6. Refit linear model in step 1 using Generalized Least Squares with the variance covariance as estimated in the previous step, then go back to 2).

The process is repeated until the estimates stop changing in any important way.

Folklore: There is evidence that the OLS estimator has a variance which is not too much different from GLS in common ARMA models.

Seasonal Non-stationarity

Every winter the measured (not reported) unemployment rate in Canada rises.

Simple model with this feature: non-stationary mean of form

$$\mu_{t+S} = \mu_t$$

Typically $S = 12$ for monthly data or 4 for quarterly data (common in economic data).

Normally, $\mu_1, \ldots, \mu_S$ are not the same.

Definition: **Deseasonalization** is the process of transforming X to eliminate this sort of seasonal variation in the mean.

Method A: Regression. Estimate

$$\hat{\mu}_t = \frac{X_t + X_{t+S} + \cdots}{\text{\# of terms}}$$

and then

$$\hat{Y}_t = X_t - \hat{\mu}_t$$

This is ordinary least squares regression with

$$\theta = \begin{bmatrix} \mu_0 \\ \vdots \\ \mu_{S-1} \end{bmatrix}$$

and

$$V = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}_{T \times S}$$

Method B: Seasonal differencing:

$$W_t = X_{t+S} - X_t$$

is stationary. BUT: if Y was an ARMA process then now W has a unit root.

Definition: A multiplicative ARIMA model written

$$ARIMA(p,d,q) \times (P,D,Q)_S$$

has the form:

$$\Phi(B^S)\phi(B)(I-B^S)^D(I-B)^d X = \Psi(B^S)\psi(B)\epsilon$$

where Φ, ϕ, Ψ, ψ polynomials of degree P, p, Q, q respectively with all roots outside the unit circle (and other technical conditions – see assignment 2, question 1) and ϵ is white noise.

As an example consider the model

$$(1-\Phi_1 B^{12})(1-\phi_1 B)X = (1-\Psi_1 B^{12})(1-\psi_1 B)\epsilon$$

which is an $ARIMA(1,0,1) \times (1,0,1)_{12}$ model.

Multiplying this out we get

$$\begin{aligned}(I-\phi_1 B-\Phi_1 B^{12}+\phi_1\Phi_1 B^{13})X \\ =(I-\psi_1 B-\Psi_1 B^{12}+\psi_1\Psi_1 B^{13})\epsilon\end{aligned}$$

which looks like an $ARMA(13,13)$.

Key point: new model has only 4 parameters (plus the variance of ϵ) instead of the 26 for a general $ARMA(13,13)$.

Fitting $ARIMA(p,d,q)$ models to data

Fitting I part easy: difference d times.

Same for seasonal multiplicative model.

Thus to fit an ARIMA(p,d,q) model to X compute $Y = (I-B)^d X$.

Note: shortens data set by d observations.

Then fit an $ARMA(p,q)$ model to Y.

So we assume that $d=0$.

Simplest case: fitting the AR(1) model

$$X_t = \mu + \rho(X_{t-1} - \mu) + \epsilon_t$$

Estimate 3 parameters: μ, ρ and $\sigma^2 = \text{Var}(\epsilon_t)$.

Our basic strategy will be:

- Estimate the parameters by maximum likelihood as if the series were Gaussian.

- Investigate the properties of the estimates for non-Gaussian data.

Generally the full likelihood is rather complicated.

So use conditional likelihoods and ad hoc estimates of some parameters to simplify the situation.

The likelihood: Gaussian data

If the errors ϵ are normal then so is the series X.

In general the vector $X = (X_0, \ldots, X_{T-1})^t$ has a $MVN(\mu, \Sigma)$ where $\Sigma_{ij} = C(i-j)$ and μ is a vector all of whose entries are μ.

The joint density of X is

$$f_X(x) = \frac{1}{(2\pi)^{T/2}\det(\Sigma)^{1/2}} \times \exp\left\{-\frac{1}{2}(x-\mu)^t\Sigma^{-1}(x-\mu)\right\}$$

so that the log likelihood is

$$\ell(\mu, a_1, \ldots, a_p, b_1, \ldots, b_q, \sigma) = -\frac{1}{2}\left[(x-\mu)^t\Sigma^{-1}(x-\mu) + \log(\det(\Sigma))\right]$$

Notice parameters on which quantity depends for an $ARMA(p,q)$.

It is possible to carry out full maximum likelihood by maximizing the quantity in question numerically. In general this is hard, however.

Here I indicate some standard tactics. In your homework I will be asking you to carry through this analysis for one particular model.

The $AR(1)$ model

Consider the model

$$X_t - \mu = \rho(X_{t-1} - \mu) + \epsilon_t$$

This model formula permits us to write down the joint density of X in a simpler way:

$$f_X = f_{X_{T-1}|X_{T-2},\ldots,X_0} f_{X_{T-2}|X_{T-3},\ldots,X_0} \cdots f_{X_1|X_0} f_{X_0}$$

Each of the conditional densities is simply

$$f_{X_k|X_{k-1},\ldots,X_0}(x_k|x_{k-1},\ldots,x_0) = g\left[x_k - \mu - \rho(x_{k-1} - \mu)\right]$$

where g is the density of an individual ϵ.

For iid $N(0,\sigma^2)$ errors this gives log like

$$\ell(\mu,\rho,\sigma) = -\frac{1}{2\sigma^2}\sum_1^{T-1}\left[x_k-\mu-\rho(x_{k-1}-\mu)\right]^2 - (T-1)\log(\sigma) + \log(f_{X_0})$$

Now for a stationary series I showed that $X_t \sim N(\mu,\sigma^2/(1-\rho^2))$ so that

$$\log(f_{X_0}(x_0)) = -\frac{1-\rho^2}{2\sigma^2}(x_0-\mu)^2 - \log(\sigma) + \log(1-\rho^2)$$

This makes

$$\ell(\mu,\rho,\sigma) = -\frac{1}{2\sigma^2}\left\{\sum_1^{T-1}\left[x_k-\mu-\rho(x_{k-1}-\mu)\right]^2 + (1-\rho^2)(x_0-\mu)^2\right\} - T\log(\sigma) + \log(1-\rho^2)$$

Can maximize over μ and σ explicitly. First

$$\frac{\partial}{\partial\mu}\ell = \frac{1}{\sigma^2}\left\{\sum_1^{T-1}\left[x_k - \mu - \rho(x_{k-1} - \mu)\right](1-\rho) \right.$$
$$\left. +(1-\rho^2)(x_0 - \mu)\right\}$$

Set this equal to 0 to find

$$\begin{aligned}\hat{\mu}(\rho) &= \frac{(1-\rho)\sum_1^{T-1}(x_k - \rho x_{k-1}) + (1-\rho^2)x_0}{1-\rho^2 + (1-\rho)^2(T-1)} \\ &= \frac{\sum_1^{T-1}(x_k - \rho x_{k-1}) + (1+\rho)x_0}{1+\rho+(1-\rho)(T-1)}\end{aligned}$$

Notice that this estimate is free of σ and that if T is large we may drop the 1 in the denominator and the term involving x_0 in the denominator and get

$$\hat{\mu}(\rho) \approx \frac{\sum_1^{T-1}(x_k - \rho x_{k-1})}{(T-1)(1-\rho)}$$

Finally, the numerator is actually

$$\begin{aligned}\sum_0^{T-1} x_k - x_0 - \rho(\sum_0^{T-1} x_k - x_{T-1}) \\ = (1-\rho)\sum_0^{T-1} x_k - x_0 + \rho x_{T-1}\end{aligned}$$

The last two terms here are smaller than the sum; if we neglect them we get

$$\hat{\mu}(\rho) \approx \bar{X}\,.$$

Now compute

$$\frac{\partial}{\partial\sigma}\ell = \frac{1}{\sigma^3}\left\{\sum_1^{T-1}\left[x_k - \mu - \rho(x_{k-1}-\mu)\right]^2 + (1-\rho^2)(x_0-\mu)^2\right\} - \frac{T}{\sigma}$$

and set this to 0 to find

$$\hat{\sigma}^2(\rho) = \frac{\sum_1^{T-1}\left[x_k - \hat{\mu}(\rho) - \rho(x_{k-1}-\hat{\mu}(\rho))\right]^2}{T} + \frac{(1-\rho^2)(x_0-\hat{\mu}(\rho))^2}{T}$$

When ρ is known: can check that $(\hat{\mu}(\rho), \hat{\sigma}(\rho))$ maximizes $\ell(\mu, \rho, \sigma)$.

To find $\hat{\rho}$ plug $\hat{\mu}(\rho)$ and $\hat{\sigma}(\rho)$ into ℓ.

Get *profile likelihood*

$$\ell(\hat{\mu}(\rho), \rho, \hat{\sigma}(\rho))$$

and maximize over ρ.

Having thus found $\hat{\rho}$ the mles of μ and $\hat{\sigma}$ are simply $\hat{\mu}(\hat{\rho})$ and $\hat{\sigma}(\hat{\rho})$.

It is worth observing that fitted residuals can then be calculated:

$$\hat{\epsilon}_t = (X_t - \hat{\mu}) - \hat{\rho}(X_{t-1} - \hat{\mu})$$

(There are only $T-1$ of them since you cannot easily estimate ϵ_0.)

Note, too, formula for $\hat{\sigma}^2$ simplifies to

$$\begin{aligned}\hat{\sigma}^2 &= \frac{\sum_1^{T-1} \hat{\epsilon}_t^2 + (1-\rho^2)(x_0 - \mu(\rho))^2}{T} \\ &\approx \frac{\sum_1^{T-1} \hat{\epsilon}_t^2}{T}.\end{aligned}$$