Time Series Analysis

- A time series model is used to describe the relationship between a variable and past values of itself.
  - The time series model is a mathematical description of the time dependent behavior of y
  - Useful for:
    - prediction of time dependent data
    - Modelling of dynamic systems

- Time series models can be fitted using regression techniques. It is convenient to write the model in a form such that the parameters enter linearly. In such cases, standard multiple linear regression tools may be used to fit the model.

Time Series Models

- Models for a single variable in terms of its own past history
  - e.g. IBM daily closing stock price – model current price in terms of past closing prices
  - A simple lagged regression model
- $y_t = b_1 y_{t-1} + b_2 y_{t-2} + \ldots + b_p y_{t-p} + e_t$
  - Use for prediction of future y's

- Dynamic models including input variables
  - e.g. viscosity of polymer exiting a CSTR in terms of the reactor temperature T and catalyst flow F
- $Y_t = a_1 Y_{t-1} + b_1 T_{t-1} + b_2 F_{t-1} + e_t$
- Dynamic models used for controller design (process identification)

An Example

- **A series of measurements of y were taken at regular intervals. The data are:**

y = {2 , 1 , -1 , -2 , 0 , 1 , 3}

Use these data to fit the model

$$y_t = b_1 y_{t-1} + b_2 y_{t-2} + \varepsilon$$

$$\mathbf{y} = \begin{bmatrix} 2 \\ 1 \\ -1 \\ -2 \\ 0 \\ 1 \\ 3 \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} & \\ \mathbf{2} & \\ \mathbf{1} & 2 \\ \mathbf{-1} & \mathbf{1} \\ \mathbf{-2} & \mathbf{-1} \\ \mathbf{0} & \mathbf{-2} \\ 1 & \mathbf{0} \end{bmatrix}$$

Example (Cont.)

$$\mathbf{y} = \begin{bmatrix} \mathbf{-1} \\ \mathbf{-2} \\ \mathbf{0} \\ \mathbf{1} \\ \mathbf{3} \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} \mathbf{1} & \mathbf{2} \\ \mathbf{-1} & \mathbf{1} \\ \mathbf{-2} & \mathbf{-1} \\ \mathbf{0} & \mathbf{-2} \\ 1 & \mathbf{0} \end{bmatrix}$$

$$\mathbf{b} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$$

$$= \begin{bmatrix} \mathbf{0}.9508 \\ -0.885 \end{bmatrix}$$

Nonlinear Regression

The general form of a nonlinear model is:

$$y = \eta(\mathbf{x}, \boldsymbol{\beta}) + \varepsilon$$

In a linear model,

$$\eta(\mathbf{x}, \boldsymbol{\beta}) = \mathbf{x}'\boldsymbol{\beta}$$

In a nonlinear model, the expectation function can have any form.

Example:

$$y = \frac{e^{-\beta_1 x}}{1 + \beta_2 x} + \varepsilon$$

Fitting Nonlinear Models

The approach is exactly the same as for linear models.

We use the same objective function:

$$\begin{aligned} S &= \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} \\ &= \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \\ &= \sum_{i=1}^{n} (y_i - \eta(\mathbf{x}_i, \boldsymbol{\beta}))^2 \end{aligned}$$

We minimize S over β.

The Optimization Problem

•The big difference between linear and nonlinear regression is that in general, the optimization problem for **a nonlinear model does not have an exact analytical solution**.

•Therefore, we have to use a numerical optimization algorithm such as:

- •Gauss-Newton
- •Steepest Descent
- •Conjugated Gradients
- •Any other optimization algorithm

Problems with Numerical Optimization

- Failure to converge
- Finding only a local minimum and not the global minimum
- Requires good starting guesses for the parameters
- Can be sensitive to the choice of convergence criteria and other "tuning parameters" of the algorithm
- Sometimes requires specification of the derivatives of the model with respect the the parameters.

Inputs for Nonlinear Regression

When using an optimization algorithm to solve nonlinear regression problems, one needs to be able to specify:

1. an expectation function (i.e. the form of the model)
2. data
3. starting guesses for β
4. stopping criteria
5. possibly other "tuning" parameters associated with the optimization algorithm

Confidence Intervals in Nonlinear Regression

•The inference results in nonlinear regression are usually approximated by linearizing the model and using the expressions from the linear case.

•The approach is to approximate the Nonlinear model by a linear model using a Taylor Series approximation, and then to simply use the linear inference results.

•It turns out that in nonlinear regression, we simply use the linear results but everywhere there is an X we use V instead, where V is:

V

$$\mathbf{V} = \begin{bmatrix} \dfrac{\partial \eta(\mathbf{x_1}, \hat{\boldsymbol{\beta}})}{\partial \beta_1} & \cdots & \dfrac{\partial \eta(\mathbf{x_1}, \hat{\boldsymbol{\beta}})}{\partial \beta_\mathrm{p}} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial \eta(\mathbf{x}_\mathrm{n}, \hat{\boldsymbol{\beta}})}{\partial \beta_1} & \cdots & \dfrac{\partial \eta(\mathbf{x}_\mathrm{n}, \hat{\boldsymbol{\beta}})}{\partial \beta_\mathrm{p}} \end{bmatrix}$$

Confidence intervals for parameters

The confidence interval for an individual parameter is:

$$\begin{aligned} &\hat{\beta}_\mathrm{i} \pm \mathrm{t}_{\nu,\alpha/2}\ \mathrm{se}\left(\hat{\beta}_\mathrm{i}\right) \\ &= \hat{\beta}_\mathrm{i} \pm \mathrm{t}_{\mathrm{n-p},\alpha/2} \sqrt{\hat{\sigma}^2 \left[(\mathbf{V'V})^{-1}\right]_\mathrm{ii}} \end{aligned}$$

Confidence Interval for the mean response

The confidence interval for the mean response is:

$$\hat{y}\big|_{\mathbf{x_0}} \pm t_{n-p,\alpha/2} \sqrt{\hat{\sigma}^2 \mathbf{v_0}' (\mathbf{V}'\mathbf{V})^{-1} \mathbf{v_0}}$$

where $\mathbf{v}_0 = \begin{bmatrix} \frac{\partial \eta(\mathbf{x_0},\boldsymbol{\beta})}{\partial \beta_1} \\ \vdots \\ \frac{\partial \eta(\mathbf{x_0},\boldsymbol{\beta})}{\partial \beta_p} \end{bmatrix}$ is the linear approximation to the nonlinear function at $\mathbf{x_0}$.

New Observation

The confidence interval for a new observation is:

$$\hat{y}\big|_{\mathbf{x_0}} \pm t_{n-p,\alpha/2} \sqrt{\hat{\sigma}^2 \left(1 + \mathbf{v_0}' (\mathbf{V}'\mathbf{V})^{-1} \mathbf{v_0}\right)}$$

Confidence REGIONS

The joint confidence region (based on a linear approx)is:

$$\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)'(\mathbf{V}'\mathbf{V})\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)=S(\hat{\boldsymbol{\beta}})\frac{p}{n-p}F_{p,n-p,\alpha}$$

or, not using the linear approximation for the expectation function:

$$S(\boldsymbol{\beta})=S(\hat{\boldsymbol{\beta}})\left[1+\frac{p}{n-p}F_{p,n-p,\alpha}\right]$$

Generalized Least Squares

- It is used when the variances of the errors are not constant and/or not independent.

Generalized Least Squares

The model is still expressed as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

However, the objective function changes and becomes

$$\mathrm{S} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

And the solution is:

$$\boldsymbol{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$$

and

$$\mathrm{V}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$$

This reduces to OLS if $\mathrm{V}(\varepsilon) = \mathbf{I}\sigma^2$.

Multi-Response Estimation

- Multi-response estimation is used when there is more than one response measured and there are common parameters in the models for each response

- Each response is modelled as

$$y_{iu} = f_i(x_u, \beta) + \varepsilon_{iu}$$

i = 1,2,…,r responses

u = 1,2, …, n experiments

  - and β is a vector of parameters some of which are common to more than one model (eg. Kinetic rate constants)

- What problems would arise if we fit each model separately?

The Optimization Problem

The optimization problem we solve in the case of a multiresponse estimation problem is:

minimize $S = |\mathbf{E}'\mathbf{E}|$

where $\mathbf{E}$ is an n x r matrix of residuals.

$$[E'E] = \begin{vmatrix} e_1'e_1 & e_1'e_2 & e_1'e_3 \\ e_2'e_1 & e_2'e_2 & e_2'e_3 \\ e_3'e_1 & e_3'e_2 & e_3'e_3 \end{vmatrix}$$

Checking Multi-response Models

Things to check:

1. Fit each response separately and check the adequacy of the model.
2. Check for mutual consistency of information on the parameters from each response (do the confidence intervals/regions overlap?).
3. Check for dependencies among y's. Use only independently measured y's in estimation.