Maximum Likelihood Estimation & Inference

Proposed by R. A. Fisher (1921/1925)

Controversial – contrary to Karl Pearson's "method of moments"

Consider : random data $\{x_1, ---, x_n\}$

parameter vector

$$\underset{\sim}{\theta} = (\theta_1, ---, \theta_K)'$$

Objective : Estimate $\underset{\sim}{\theta}$

Role of $\underset{\sim}{\theta}$: $\underset{\sim}{\theta}$ determines the x_i values through their joint p.d.f., $p(\underset{\sim}{x} \mid \underset{\sim}{\theta})$, or $p(x_1, ---, x_n \mid \theta_1, --, \theta_K)$.

We can view $p(\underset{\sim}{x} \mid \underset{\sim}{\theta})$ in two ways –

(i) As a fctn. of $\{x_1, ---, x_n\}$, given $\underset{\sim}{\theta}$.

(ii) As a fctn. of $(\theta_1, ----, \theta_K)$, given $\underset{\sim}{x}$.

Latter is called the Likelihood Function.

$$L(\underset{\sim}{\theta}) = L(\underset{\sim}{\theta} \mid x_1, ----, x_n) = p(x_1, ---, x_n \mid \underset{\sim}{\theta}).$$

Definition : The Maximum Likelihood Estimator of $\underset{\sim}{\theta}$ (say, $\tilde{\underset{\sim}{\theta}}$) is that value of $\underset{\sim}{\theta}$ such that $L(\tilde{\underset{\sim}{\theta}}) \geq L(\hat{\underset{\sim}{\theta}})$, for all other $\hat{\underset{\sim}{\theta}}$.

Motivation : Given a paricular sample of data, what value of $\underset{\sim}{\theta}$ is most likely to have generated it from the population? This is $\tilde{\underset{\sim}{\theta}}$.

Note : (i) $\tilde{\underset{\sim}{\theta}}$ need not be unique.

(ii) $\tilde{\underset{\sim}{\theta}}$ should locate the global max of L.

(iii) If sample data are independent then $L(\underset{\sim}{\theta} \mid \underset{\sim}{x}) = p(\underset{\sim}{x} \mid \underset{\sim}{\theta})$

$$= \prod_i p(x_i \mid \underset{\sim}{\theta})$$

(iv) Any monotonic transformation of $L(\underset{\sim}{\theta})$ leaves location of extrem unchanged. (e.g. : $\log L(\underset{\sim}{\theta})$)

Some Basic Concepts & Notation:

(i) "Gradient Vector" ("Score Vector"): $\underset{(k\times 1)}{\left[\dfrac{\partial \log L(\underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}}\right]}$

(ii) "Hessian Matrix": $\underset{(k\times k)}{\left[\dfrac{\partial^2 \log L(\underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}\, \partial \underset{\sim}{\theta}'}\right]}$

(iii) "Likelihood Equation(s)":

$$\frac{\partial \log L(\underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}} = \underset{\sim}{0} \qquad (k\times 1)$$

To obtain the MLE, $\tilde{\underset{\sim}{\theta}}$, we solve the Likelihood Equations, & then check the second-order condition(s) to make sure that we have maximized (not minimized) $L(\underset{\sim}{\theta})$. If the Hessian matrix is at least negative semi-definite, then $\log L$ is concave, & this is sufficient (not necessary) for a maximum.

Example: Geometric Distribution.

Independent Bernoulli trials Y_i = no. of failures before first success:

$$\Pr.(Y_i = y_i) = (1-p)^{y_i}\, p$$

where $p = \Pr.(\text{success})$

$$L = p(y_1, y_2, \ldots, y_n) = \prod_{i=1}^{n} p(y_i)$$

$$= (1-p)^{\Sigma y_i}\, p^n \qquad (\Sigma y_i \text{ sufficient})$$

$$\log L = \Sigma y_i \log(1-p) + n\log(p)$$

$$(\partial \log L/\partial p) = -\frac{n\bar{y}}{1-p} + n/p \qquad \text{(i)}$$

$$(\partial^2 \log L/\partial p^2) = -\frac{n\bar{y}}{(1-p)^2} - n/p^2 \qquad \text{(ii)}$$

From (i): $\dfrac{-n\bar{y}}{1-\tilde{p}} + n/\tilde{p} = 0$

$$\Rightarrow \tilde{p} = \left(\frac{1}{1+\bar{y}}\right) \qquad \#$$

From (ii): $(\partial^2 \log L/\partial p^2) < 0$, everywhere.

[Show that $E(Y_i) = (1-p)/p$. Note that $(1-\tilde{p})/\tilde{p} = \bar{y}$, so $\bar{y}$ is a natural estimator of $E(Y_i)$ here. Relate this to "invariance" later.

Example:

$$y_i = \beta x_i + \varepsilon_i \quad ; \quad \varepsilon_i \sim \text{iid } N(0, \sigma^2)$$

$$L(\beta, \sigma^2) = p(y_1, \dots, y_n \mid \beta, \sigma^2)$$
$$= \prod_{i=1}^{n} p(y_i \mid \beta, \sigma^2)$$

& $y_i \sim$ i.i.d. $N(\beta x_i, \sigma^2)$.

So, $$L = \prod_{i=1}^{n} \left(\frac{1}{\sigma\sqrt{2\pi}}\right) \exp\left[-\frac{1}{2\sigma^2}(y_i - \beta x_i)^2\right]$$

$$= (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta x_i)^2\right]$$

& $$\log L = -n/2 \log(2\pi) - n/2 \log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta x_i)^2 .$$

(i) $$\left(\frac{\partial \log L}{\partial \beta}\right) = \frac{1}{\sigma^2}\sum_i (x_i y_i - \beta x_i^2) = 0$$

(ii) $$\left(\frac{\partial \log L}{\partial \sigma^2}\right) = -(n/2\sigma^2) + \left(\frac{1}{2\sigma^4}\right)\sum_i (y_i - \beta x_i)^2 = 0.$$

From (i): $\sum_i x_i y_i = \tilde{\beta} \sum_i x_i^2$

or, $$\boxed{\tilde{\beta} = \left(\sum_i x_i y_i\right) / \left(\sum_i x_i^2\right).}$$

Substituting in (ii):

$$(n/2\tilde{\sigma}^2) = (1/2\tilde{\sigma}^4)\sum_i (y_i - \tilde{\beta} x_i)^2$$

or, $$\boxed{\tilde{\sigma}^2 = \frac{1}{n}\sum_i (y_i - \tilde{\beta} x_i)^2}$$

Now, consider 2nd-order condition:

(i) $$\left(\frac{\partial^2 \log L}{\partial \beta^2}\right) = -1/\sigma^2 \sum_i x_i^2$$

(ii) $$\left(\frac{\partial^2 \log L}{\partial \beta\, \partial \sigma^2}\right) = -\frac{1}{\sigma^4}\sum_i (x_i y_i - \beta x_i^2)$$
$$= \left(\frac{\partial^2 \log L}{\partial \sigma^2\, \partial \beta}\right)$$

(iii) $$\left(\frac{\partial^2 \log L}{\partial \sigma^2\, \partial \sigma^2}\right) = (n/2\sigma^4) - \left(\frac{1}{\sigma^6}\right)\sum_i (y_i - \beta x_i)^2$$

Substituting $\tilde{\beta}$ for β, & $\tilde{\sigma}^2$ for σ^2:

(a) $$\left(\frac{\partial^2 \log L}{\partial \beta^2}\right)\bigg|_{\tilde{\beta}, \tilde{\sigma}^2} = -\frac{1}{\tilde{\sigma}^2}\sum_i x_i^2$$

(b) $$\left(\frac{\partial^2 \log L}{\partial \beta\, \partial \sigma^2}\right)\bigg|_{\tilde{\beta}, \tilde{\sigma}^2} = -\frac{1}{\tilde{\sigma}^4}\sum_i (x_i y_i - \tilde{\beta} x_i^2)$$
$$= -\frac{1}{\tilde{\sigma}^4}\left[\sum_i x_i y_i - \sum_i x_i y_i\right] = 0.$$

(b) $$\left(\frac{\partial^2 \log L}{\partial\sigma^2\,\partial\sigma^2}\right)\Big|_{\tilde\beta,\tilde\sigma^2} = \left(n/2\tilde\sigma^4\right) - \frac{1}{\tilde\sigma^6}\sum_i (y_i - \tilde\beta x_i)^2$$

$$= (n/2\tilde\sigma^4) - (n\tilde\sigma^2/\tilde\sigma^6) = -(n/2\tilde\sigma^4).$$

So, the Hessian Matrix is:

$$H(\tilde\beta,\tilde\sigma^2) = \begin{bmatrix} -\sum_i x_i^2/\tilde\sigma^2 & 0 \\ 0 & -n/2\tilde\sigma^4 \end{bmatrix}$$

$$|a_{11}| = a_{11} = -\sum_i x_i^2/\tilde\sigma^2 < 0$$

$$|H| = \left(-\sum_i x_i^2/\tilde\sigma^2\right)\left(-n/2\tilde\sigma^4\right) > 0$$

So, H is negative-definite.

Obviously generalizes to multiple regression model:

$$\tilde\beta = (X'X)^{-1}X'y$$

$$\tilde\sigma^2 = (y - X\tilde\beta)'(y - X\tilde\beta)/n.$$

Sometimes the information we have is about some random variable, z, not the observed $\underset{\sim}{x}$. We need to derive $p(\underset{\sim}{x})$ from $p(\underset{\sim}{z})$ in order to form $L(\underset{\sim}{\theta}\,|\,\underset{\sim}{x})$.

Example – we derived density of y_i from assumed density of ε_i in regression example.

Definition: Let $\underset{\sim}{z}$ and $\underset{\sim}{x}$ be functionall[y] related random vectors, each $(n \times 1)$.

Then, to derive $p(\underset{\sim}{x})$ from $p(z)$, we use:

$$p(\underset{\sim}{x}) = p(\underset{\sim}{z})\,J_+$$

where $J_+ = \text{abs.}\left|\frac{\partial z_i}{\partial x_j}\right|$ $(n \times n)$

is the Jacobian of the mapping from $\underset{\sim}{z}$ to [x].

(In regression example $\partial\varepsilon_i/\partial y_i = 1$; & $\partial\varepsilon_i/\partial y_j = 0$ $(i \neq j)$, so $J_+ = 1$.)

<u>Example</u>: $p(z) = \frac{1}{2} \; ; \; 0 \le z \le 2$

$= 0 \; ; \;$ otherwise

Let $x = 4z \; ; \; z = x/4 \; ; \; \frac{dz}{dx} = 1/4$.

So, $p(x) = p(z) \left| \frac{1}{4} \right| = (1/2)(1/4) = 1/8$

& $0 \le x \le 8$.

<u>Example</u>: $z \sim N(0, 1) \; ; \; -\infty < z < \infty$

$x = \sigma z + \mu$

So, $z = (x - \mu)/\sigma \; ; \; \frac{\partial z}{\partial x} = 1/\sigma$

$p(x) = p(z) \, |1/\sigma|$

$= \left| \frac{1}{\sigma} \right| \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$

So, if $\sigma > 0$:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \; ; \; -\infty < x < \infty$$

$\sim N[\mu, \sigma^2]$.

Recall that $L(\underset{\sim}{\theta} \mid \underset{\sim}{x})$ is the <u>joint</u> data density, viewed as a fctn. of $\underset{\sim}{\theta}$. If the x_i's are <u>independent</u>, then we have $L(\underset{\sim}{\theta} \mid \underset{\sim}{x}) = \prod_i p(x_i \mid \underset{\sim}{\theta})$. If not, we need to know the joint data density. For example

<u>Definition</u>: Let $\underset{\sim}{x}$ be an $(n \times 1)$ random vector. Then $\underset{\sim}{x}$ is Multivariate Normally distributed, $N(\underset{\sim}{\mu}, V)$, if:

$$p(\underset{\sim}{x} \mid \underset{\sim}{\mu}, V) = (2\pi)^{-n/2} |V|^{-1/2} \exp\left[-\tfrac{1}{2}(\underset{\sim}{x} - \underset{\sim}{\mu})' V^{-1} (\underset{\sim}{x} - \underset{\sim}{\mu})\right]$$

(Clearly V must be p.d.s. Also, if $n = 1$ then $V = \sigma^2$ & we get usual normal density.

<u>Example</u>:

$$y = X\beta + \varepsilon$$

$\varepsilon \sim MVN(0, \sigma^2 \Omega) \; ; \; \Omega$ known

So, $p(\varepsilon) = (2\pi)^{-n/2} |\sigma^2 \Omega|^{-1/2} \exp\left[-\tfrac{1}{2}\varepsilon'(\sigma^2\Omega)^{-1}\varepsilon\right]$

Consider the Jacobian for $\varepsilon \to y$:

$\partial \varepsilon_i / \partial y_i = 1 \; ; \; \partial \varepsilon_i / \partial y_j = 0 \quad (i \ne j)$

So $J_+ = \text{abs.} |I_n| = 1$.

So, $p(y \mid \beta, \sigma^2, \Omega) = (2\pi)^{-n/2} \left| \sigma^2 \Omega \right|^{-1/2}$
$\cdot \exp\left[-\tfrac{1}{2}(y - X\beta)'(\sigma^2\Omega)^{-1}(y - X\beta)\right]$

$= (2\pi\sigma^2)^{-n/2} |\Omega|^{-1/2} \exp\left[-\frac{1}{2\sigma^2}(y - X\beta)\Omega^{-1}(y - X\beta)\right]$

$= L(\beta, \sigma^2 \mid \underset{\sim}{y})$;

& $\log L = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log(\sigma^2) - \frac{1}{2}\log|\Omega|$
$- \frac{1}{2\sigma^2}(y - X\beta)'\Omega^{-1}(y - X\beta)$

(i) $(\partial \log L / \partial \beta) = -\frac{1}{2\sigma^2}\frac{\partial}{\partial \beta}\left[y'\Omega^{-1}y + \beta'X'\Omega^{-1}X\beta - 2y'\Omega^{-1}X\beta\right]$

$= -\frac{1}{2\sigma^2}\left[2X'\Omega^{-1}X\beta - 2X'\Omega^{-1}y\right]$

$\overset{=0}{}$

$\Rightarrow \boxed{\tilde{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y}$

(ii) $(\partial \log L / \partial \sigma^2) = -\frac{n}{2}(1/\sigma^2) - (\tfrac{1}{2})(-1)(\sigma^2)^{-2} \cdot (y - X\beta)'\Omega^{-1}(y - X\beta)$

$\overset{=0}{}$

$\Rightarrow (n/2\tilde{\sigma}^2) = (\frac{1}{2\tilde{\sigma}^4})(y - X\tilde{\beta})'\Omega^{-1}(y - X\tilde{\beta})$

$\Rightarrow \boxed{\tilde{\sigma}^2 = \frac{1}{n}(y - X\tilde{\beta})'\Omega^{-1}(y - X\tilde{\beta})}$

(Confirm 2nd-order condition.)

<u>Note</u>: <u>(i)</u> $\tilde{\beta}$ = GLS estimator of β.

<u>(ii)</u> $\tilde{\sigma}^2$ is a biased estimator.

<u>(iii)</u> If $\Omega = I$, we just have OLS

This is for the case where Ω is known.

Usually, Ω is unknown, but $\Omega = \Omega(\theta)$ where θ has small dimension. Then we solve:

$$\left.\begin{aligned} (\partial \log L / \partial \beta) &= 0 \\ (\partial \log L / \partial \sigma^2) &= 0 \\ (\partial \log L / \partial \theta) &= 0 \end{aligned}\right\}$$

simultaneously, for the MLE's of all of the parameters.

e.g.: AR(1) errors –

$$\Omega = \Omega(\rho) = \frac{1}{1-\rho^2}\begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & & \\ \vdots & & & \ddots & \\ \rho^{n-1} & \rho^{n-2} & & & 1 \end{bmatrix}$$

Sometimes we can determine the MLE by inspection – avoid lots of calculus!

<u>Example</u>: $y = X\beta + \varepsilon$; $\varepsilon \sim N(0, \sigma^2 I_n)$

Estimate β, s.t. $R\beta = q$.

$$L(\beta, \sigma^2) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right]$$

$$\log L = -\tfrac{n}{2}\log(\sigma^2) - \tfrac{n}{2}\log(2\pi) - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)$$

& we want to <u>maximize</u> $\log L$, s.t. $R\beta = q$.

That is, <u>minimize</u> $(y - X\beta)'(y - X\beta)$, s.t. $R\beta = q$.

$\Rightarrow$ <u>Restricted Least Squares</u> estimator of β.

($\tilde{\sigma}^2 = (y - X\tilde{\beta})'(y - X\tilde{\beta})/n$; $\tilde{\beta} = MLE = RLS$)

<u>Example</u>: $y = X\beta + \varepsilon$

Errors follow <u>multivariate Student-t</u> dist'n.:

$$p(\underset{\sim}{\varepsilon} \mid \nu) = \frac{c}{\sigma}\left[\nu + \underset{\sim}{\varepsilon}'\underset{\sim}{\varepsilon}/\sigma^2\right]^{-(n+\nu)/2}$$

& $E(\underset{\sim}{\varepsilon}) = 0$; $V(\underset{\sim}{\varepsilon}) = \left(\frac{\sigma^2 \nu}{\nu - 2}\right) I_n$

(if $\nu > 2$)

Once again, $J_T = 1$, &

$$L(\beta, \sigma^2) = p(\underset{\sim}{y} \mid \beta, \sigma^2) = p(\underset{\sim}{\varepsilon} \mid \nu)$$
$$= \frac{c}{\sigma}\left[\nu + (y - X\beta)'(y - X\beta)/\sigma^2\right]^{-(n+\nu)/2}$$

& $\log L = \log(c) - \frac{1}{2}\log\sigma^2 - \left(\frac{n+\nu}{2}\right)\log\left[\nu + (y - X\beta)'(y - X\beta)/\sigma^2\right]$

Immediately, to maximize $\log L$, we see we need to minimize $(y - X\beta)'(y - X\beta)$ w.r.t. β, & so $\tilde{\beta} = OLS$ in this case (just like Normal)

(Of course the MLE of σ^2 still needs to be considered.)

What can we observe about <u>properties</u> of MLE so far?

(i) May be biased / unbiased in finite samples

(ii) May be efficient / inefficient in finite samples.

(iii) If MLE is <u>unique</u>, then it must be a function of a <u>sufficient statistic</u>

(Geometric: $\tilde{p} = 1/(1 + \bar{y})$; $\sum y_i$ & hence $\bar{y}$ are sufficient.)

Also: MLE's are invariant to continuous transformations. So, if $g(\cdot)$ is any continuous fctn., & if $\tilde{\theta}$ is MLE for θ, then $g(\tilde{\theta})$ is MLE for $g(\theta)$.

e.g. If $\tilde{\sigma}^2$ is MLE for σ^2, then $\sqrt{\tilde{\sigma}^2}$ is MLE for σ.

If $\tilde{\beta}$ is MLE for β $(k\times 1)$, then $(\tilde{\beta}_1/\tilde{\beta}_2)$ is MLE for (β_1/β_2).

Clearly, this result can save a lot of work!

(Note: Some texts say that $g(\cdot)$ must be a 1-1 mapping, this is <u>not</u> needed for invariance.)

Geometric - $\tilde{p} = \frac{1}{1+\bar{y}}$ = MLE for p.
$E(Y_i) = (1-p)/p$, so MLE for $E(Y_i)$ is $(1-\tilde{p})/\tilde{p} = \bar{y}$. More than just a "natural" estimator for $E(Y_i)$.

Now, let's consider a more complex example that illustrates the need to take care with respe to the Jacobian, & illustrates the MVN error cas when Ω is <u>unknown</u>.

<u>Example</u>:

$$y_t = x_t'\beta + u_t$$
$$u_t = \rho u_{t-1} + \varepsilon_t \quad ; \; t=1,\ldots,n$$
$$\varepsilon_t \sim \text{i.i.d. } N(0,\sigma_\varepsilon^2) \quad ; \; |\rho|<1$$

So: $E(\underset{\sim}{u}) = \underset{\sim}{0}$; $V(\underset{\sim}{u}) = \sigma_\varepsilon^2\,\Omega(\rho)$

where
$$\Omega(\rho) = \frac{1}{1-\rho^2}\begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \vdots & & \ddots & & \vdots \\ \rho^{n-1} & & & \cdots & 1 \end{bmatrix}$$

What do we usually do to estimate this model

Note that
$$\Omega^{-1/2} = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & \cdots & & 0 \\ -\rho & 1 & \cdots & & 0 \\ 0 & -\rho & 1 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \end{bmatrix}$$

& transform both sides of model by $\Omega^{-1/2}$

$$y_t^* = \begin{bmatrix} \sqrt{1-\rho^2}\; y_1 \\ y_2 - \rho y_1 \\ \vdots \\ y_n - \rho y_{n-1} \end{bmatrix} \quad ; \underline{etc}$$

Then fit the model :

$$y_t^* = x_t^{*\prime}\beta + \varepsilon_t$$

or,

$$\left\{\begin{array}{l} y_t = \rho y_{t-1} + x_t'\beta - x_{t-1}'\rho\beta + \varepsilon_t \qquad (t = 2, ---, n) \\ \sqrt{1-\rho^2}\, y_1 = x_1'\sqrt{1-\rho^2}\,\beta + \varepsilon_1 \end{array}\right.$$

(This is effectively a NLLS problem.)

Now, is this the same as MLE? Actually, <u>not quite</u> !

<u>Note</u>:

$$\left\{\begin{array}{l} y_t \mid y_{t-1} = \rho y_{t-1} + x_t'\beta - x_{t-1}'\rho\beta + \varepsilon_t \qquad (t=2, ---, n) \\ y_1 = x_1'\beta + \varepsilon_1/\sqrt{1-\rho^2} \end{array}\right.$$

We need to construct the Likelihood Function. We don't have the <u>joint</u> data density, & the observations are <u>not</u> independent (they are serially correlated). What can we do?

Recall : $p(x_2 \mid x_1) = p(x_2, x_1)/p(x_1)$

<u>or</u>: $p(x_2, x_1) = p(x_2 \mid x_1)\, p(x_1)$

Generalizing —

$$p(x_n, x_{n-1}, ---, x_1) = p(x_n \mid x_{n-1}) \cdot p(x_{n-1} \mid x_{n-2}) \cdot --- \cdot p(x_2 \mid x_1)\, p(x_1)$$

Now,

$$p(y_t \mid y_{t-1}) = \frac{1}{\sigma_\varepsilon \sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_\varepsilon^2}\left\{(y_t - \rho y_{t-1}) - (x_t' - \rho x_{t-1}')\beta\right\}^2\right] \qquad (t = 2, --, n)$$

& $\varepsilon_1 = (y_1 - x_1'\beta)\sqrt{1-\rho^2}$

So $(\partial \varepsilon_1 / \partial y_1) = \sqrt{1-\rho^2}$

& $p(y_1) = p(\varepsilon_1) + \sqrt{1-\rho^2}$

or,

$$p(y_1) = (1-\rho^2)^{1/2} \frac{1}{\sigma_\varepsilon\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_\varepsilon^2}\varepsilon_1^2\right]$$

$$= (1-\rho^2)^{1/2} \frac{1}{\sigma_\varepsilon\sqrt{2\pi}} \exp\left[\frac{-(1-\rho^2)}{2\sigma_\varepsilon^2}(y_1 - x_1'\beta)^2\right].$$

Now,

$$L(\beta, \rho, \sigma_\varepsilon^2) = p(y_n, ----, y_1 \mid \beta, \rho, \sigma_\varepsilon^2)$$

$$= p(y_n \mid y_{n-1})\, p(y_{n-1} \mid y_{n-2}) ---- p(y_2 \mid y_1)\, p(y_1)$$

$$= (2\pi)^{-(n-1)/2} (\sigma_\varepsilon^2)^{-(n-1)/2}$$

$$\cdot \exp\left\{-\frac{1}{2\sigma_\varepsilon^2}\sum_{t=2}^{n}\left[(y_t - \rho y_{t-1}) - (x_t' - \rho x_{t-1}')\beta\right]^2\right\}$$

$$\cdot (2\pi)^{-1/2} (\sigma_\varepsilon^2)^{-1/2} (1-\rho^2)^{1/2}$$

$$\cdot \exp\left[-\frac{(1-\rho^2)}{2\sigma_\varepsilon^2}(y_1 - x_1'\beta)^2\right]$$

We need to take logs, & maximize w.r.t. β, σ_ε^2 & ρ. (Beach & Mackinnon) (1978)

<u>Note</u>:

(i) If we use NLLS (Cochrane-Orcutt) we are omitting Jacobian term - $\log L$ is mis-specified & <u>all</u> parameter estimates are affected.

(ii) The first-order conditions (the likelihood equations) are highly non-linear in the parametes & cannot be solved explicitly. We need to solve them numerically. (This situation arises often)

<u>Example</u>: Non-Linear Model

$$y_t = \beta_1 + \beta_2 x_t^{\beta_1} + \varepsilon_t \quad ; \; \varepsilon_t \sim iid\ N(0, \sigma^2)$$

$J_t = 1$ & so

$$L(\beta_1, \beta_2, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_t (y_t - \beta_1 - \beta_2 x_t^{\beta_1})\right)$$

Using $\log L$, the likelihood equations are:

(i) $\frac{1}{\sigma^2}\sum_t\left[(y_t - \beta_1 - \beta_2 x_t^{\beta_1})(1 + \beta_2 x_t^{\beta_1}\log x_t)\right] = 0$

(ii) $\frac{1}{\sigma^2}\sum_t\left[(y_t - \beta_1 - \beta_2 x_t^{\beta_1})\, x_t^{\beta_1}\right] = 0$

(iii) $-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_t (y_t - \beta_1 - \beta_2 x_t^{\beta_1}) = 0$

Again, no 'closed-form' solution for $\tilde{\beta}_1$, $\tilde{\beta}_2$ & $\tilde{\sigma}^2$ - use numerical methods.