CH5115 QUIZ-2.

BY: S. VISHAL
CH18B020.

(4) $y[k] = f(x[k], \theta) + e[k]$ ie $e[k] \sim$ GWN
$\sigma_e^2$ known, $\mu = 0$

The given model is generic and could be non-linear

However to obtain a BLUE estimator, we need the DGP to be linear of the form.

$$\underline{y} = L\theta_0 + e$$ where e[k] should be zero mean R.V. so that,

$$E(y) = L\theta_0.$$

$$\Rightarrow \theta_0 = AE(\underline{y}) \quad (AL = I.)$$

Then we can obtain estimator as $\hat{\theta}(y) = Ay$.

the solution for A is known and is of the form

$$A^* = (L^T \Sigma_v^{-1} L)^{-1} (\Sigma_v^{-1} L).$$

So to proceed we need to linearise the model.

Obviously it will be an approximation of the original non-linear model. YES! It is possible to use BLUE in such a case.

Since we know the initial optimal estimate of $\theta$, we will also know $f(x[k], \theta)$. Assuming there won't be ~~much~~ large deviation in $\theta$ we can use a Taylor Series expansion upto the linear term.

$$f(x,\theta) \simeq f(x_0,\theta_i) + \left.\frac{\partial f}{\partial x}\right|_{x=x_0,\theta=\theta_i}(x-x_0) + \left.\frac{\partial f}{\partial \theta}\right|_{x=x_0,\theta=\theta_i}(\theta-\theta_i)$$

$\theta_i$ is the initial optimal estimate of $\theta$.

$x_0$ should be an $x[k]$ such that it is close to the new $x(k)$.

Vectorially,

$$f(\underline{x},\underline{\theta}) = f(\underline{x_0},\underline{\theta_i}) + \left(\frac{\partial f}{\partial \underline{x}}\right)^T[\underline{x}-\underline{x_0}] + \left(\frac{\partial f}{\partial \underline{\theta}}\right)^T[\underline{\theta}-\underline{\theta_i}]$$

Note that $f(\underline{x_0},\underline{\theta_i})$ is a known constant

$\left(\frac{\partial f}{\partial \underline{x}}\right)^T$ is also a constant. It is the derivative of $f$ wrt $\underline{x}$ evaluated at reqd. $\underline{x_0}$ and optimal $\underline{\theta_i}$

$\left(\frac{\partial f}{\partial \underline{\theta}}\right)^T$ is the derivative of $f$ wrt $\underline{\theta}$ evaluated at $\underline{\theta_i}$ and $\underline{x_0}$

$$f(\underline{x},\underline{\theta}) = \underline{y_0} + \left(\frac{\partial f}{\partial \underline{x}}\right)^T(x-x_0) + \left(\frac{\partial f}{\partial \underline{\theta}}\right)^T[\underline{\theta}-\underline{\theta_i}]$$

where $\underline{y_0} = f(x_0,\theta_i)$

Let us call $\underline{y} - \underline{y}_0 = \underline{\tilde{y}}$, $\underline{x} - \underline{x}_0 = \underline{\tilde{x}}$.

and $\underline{\theta} - \underline{\theta}_i = \tilde{\theta}$.

Substituting the value of $f(\underline{x}, \underline{\theta})$ in DESP

$$\underline{y} = \underline{y}_0 + \left(\frac{\partial f}{\partial \underline{x}}\right)^T \underline{\tilde{x}} + \left(\frac{\partial f}{\partial \underline{\theta}}\right)^T \tilde{\theta} + \underline{v}$$

$$\Rightarrow \underline{\tilde{y}} = \left(\frac{\partial f}{\partial \underline{x}}\right)^T \underline{\tilde{x}} + \left(\frac{\partial f}{\partial \underline{\theta}}\right)^T \underline{\tilde{\theta}} + \underline{V}$$

$$\Rightarrow \left[\underline{\tilde{y}} - \left(\frac{\partial f}{\partial \underline{x}}\right)^T \underline{\tilde{x}}\right] = \left(\frac{\partial f}{\partial \underline{\theta}}\right)^T \underline{\tilde{\theta}} + \underline{V}$$

Let $\tilde{y} - \left(\frac{\partial f}{\partial \underline{x}}\right)^T \tilde{x} = \delta$

$$\Rightarrow \underline{\delta} = \left(\frac{\partial f}{\partial \underline{\theta}}\right)^T \underline{\tilde{\theta}} + \underline{V}$$

$$\Rightarrow \underline{\delta} = L \tilde{\theta} + \underline{V}$$

$\therefore$ This is a linearised form where $L = \left(\frac{\partial f}{\partial \underline{\theta}}\right)^T$ evaluated at $\theta_i$

$$\hat{\tilde{\theta}} = (L^T \Sigma_v^{-1} L)^{-1} (\Sigma_v^{-1} L) \underline{\delta}$$

* So as and when new data points arrive, RegressRaja can update $\theta$ by computing $\tilde{\hat{\theta}}$ ~~wh~~ and adding it to $\theta_i$. $\boxed{\hat{\theta} = \theta_i + \tilde{\hat{\theta}}}$

* This approximation will be acceptable only in cases of small deviations in $\theta_i$.

* However values of $x[k]$ can be close to any of the previous values whose $\theta_i$ we confidently know. That is, to compute $\tilde{x} = x[k] - x_0$, we can choose any $x_0$, so we can choose the $x_0$ closest to $x[k]$.

* So $x_0$ not need not be a vector of equal values but rather different values adjusted according to the corresponding element in $x$.

* This also means that $\left(\frac{\partial f}{\partial x}\right)$ vector's elements should be evaluated at appropriate $x_0$s as present in the $x_0$ vector. This $\left(\frac{\partial f}{\partial \theta}\right)$ ~~than~~ vector should also be evaluated at $x$ given by $x_0$.

* However $\theta$ will remain fixed for all elements in $\frac{\partial f}{\partial \theta}$ ~~vector~~ and $\frac{\partial f}{\partial x}$.

* If ~~$x_i = x_0$~~ $\underline{x}$ is within the domain of $f(x, \theta_i)$ we can simply substitute ~~$x = x_0$~~ $x_0 = \underline{x}$

$$\Rightarrow \tilde{x} = 0$$

$$\Rightarrow \quad \delta = \tilde{y}.$$

$$\therefore \quad \underline{\tilde{y}} = L\,\underline{\tilde{\theta}} + \underline{v}$$

$$\Rightarrow \quad \boxed{\hat{\tilde{\theta}} = \left(L^T \Sigma_v^{-1} L\right)^{-1} \left(\Sigma_v^{-1} L\right) \underline{\tilde{y}}}$$

* Algorithms in such a case:
  → Set $\hat{\theta} = \theta_i$
  i) compute $L$ based on $\underline{x}$ & $\hat{\theta}$
  ii) Compute $\underline{\tilde{y}}$ ~~based on~~ $= \underline{y} - f(\underline{x}, \theta_i)$
  iii) $\underline{\hat{\tilde{\theta}}} = \left(L^T \Sigma_v^{-1} L\right)^{-1} \left(\Sigma_v^{-1} L\right) \underline{\tilde{y}}$
  iv) $\underline{\hat{\theta}} = \underline{\hat{\theta}} + \underline{\hat{\tilde{\theta}}}$
  v) Repeat i) → v)

* As ~~$\hat{\theta}$ becomes~~ $\|\underline{\hat{\theta}} - \underline{\theta_i}\|$ becomes larger the error in the approximation increases.

* NOTE: $\underline{x}$ and $\underline{y}$ are the obtained new values from DGP