Estimation with normal regression model

Model

Normal regression model has the pdf

$$f\left(y_t|\psi_t,\Theta\right)=\frac{1}{\sqrt{2\pi r}}\exp\left\{-\frac{1}{2r}\left(y_t-\psi_t^{'}\theta\right)^2\right\}$$

The square can be edited as follows

$$\left(y_t-\psi_t^{'}\theta\right)^2=\left(y_t-\theta^{'}\psi_t\right)\left(y_t-\psi_t^{'}\theta\right)=$$

$$=\left(-\left[-1,\,\theta^{'}\right]\left[\begin{array}{c}y_t\\ \psi_t\end{array}\right]\right)\left(-\left[y_t,\,\psi_t^{'}\right]\left[\begin{array}{c}-1\\ \theta\end{array}\right]\right)=$$

$$=\left[-1,\,\theta^{'}\right]\left(\left[\begin{array}{c}y_t\\ \psi_t\end{array}\right]\left[y_t,\,\psi_t^{'}\right]\right)\left[\begin{array}{c}-1\\ \theta\end{array}\right]=$$

$$=\left[-1,\,\theta^{'}\right]D_t\left[\begin{array}{c}-1\\ \theta\end{array}\right]$$

where $D_t=\left[\begin{array}{c}y_t\\ \psi_t\end{array}\right]\left[y_t,\,\psi_t^{'}\right]$ is so called data matrix.

The regression model then is

$$f\left(y_t|\psi_t,\Theta\right)=\frac{1}{\sqrt{2\pi r}}\exp\left\{-\frac{1}{2r}\left[-1,\,\theta'\right]D_t\left[\begin{array}{c}-1\\ \theta\end{array}\right]\right\}$$

The conjugate distribution to this model (i.e. the distribution of the parameters that is reproducible during estimation) is GaussWishart (GW) with the pdf

$$f\left(\Theta|d\left(t\right)\right)\propto r^{-0.5\kappa_t}\exp\left\{-\frac{1}{2r}\left[-1,\,\theta'\right]V_t\left[\begin{array}{c}-1\\ \theta\end{array}\right]\right\}$$

where V_t is a square, positive definite matrix called information matrix.

Statistics

Now, substituting into the Bayes rule

$$\underbrace{f\left(\Theta|d\left(t\right)\right)}_{\text{GW}}\propto\underbrace{f\left(y_t|\psi_t,\Theta\right)}_{\text{model}}\underbrace{f\left(\Theta|d\left(t-1\right)\right)}_{\text{GW}}$$

we get

$$f\left(\Theta|d\left(t\right)\right)\propto r^{-0.5\kappa_t}\exp\left\{-\frac{1}{2r}\left[-1,\,\theta'\right]V_t\left[\begin{array}{c}-1\\ \theta\end{array}\right]\right\}\propto$$

$$\propto r^{-0.5} \exp\left\{-\frac{1}{2r}\left[-1,\, \theta^{'}\right] D_t \left[\begin{array}{c} -1 \\ \theta \end{array}\right]\right\} \times r^{-0.5\kappa_{t-1}} \exp\left\{-\frac{1}{2r}\left[-1,\, \theta^{'}\right] V_{t-1} \left[\begin{array}{c} -1 \\ \theta \end{array}\right]\right\} =$$

$$r^{-0.5(\kappa_{t-1}+1)} \exp\left\{-\frac{1}{2r}\left[-1,\, \theta^{'}\right] (V_{t-1}+D_t) \left[\begin{array}{c} -1 \\ \theta \end{array}\right]\right\}$$

By comparison of of the first and last expression we get the well known update of statistics

$$V_t = V_{t-1} + D_t$$

$$\kappa_t = \kappa_{t-1} + 1$$

The posterior pdf at time N is

$$f\left(\Theta | d\left(N\right)\right) \propto r^{-0.5\kappa_N} \exp\left\{-\frac{1}{2r}\left[-1,\, \theta^{'}\right] V_N \left[\begin{array}{c} -1 \\ \theta \end{array}\right]\right\} \tag{1}$$

Point estimates

The point estimates of the parameters θ and r lay at its maximum.

– parameter θ

The maximum will be at the minimum of $\left[-1,\, \theta^{'}\right] V_N \left[\begin{array}{c} -1 \\ \theta \end{array}\right]$. It can be found by completion to square. We express

$V_N = \begin{bmatrix} V_y & V_{yp}^{'} \\ V_{yp} & V_p \end{bmatrix}$ and we have

$$\left[-1,\, \theta^{'}\right] \begin{bmatrix} V_y & V_{yp}^{'} \\ V_{yp} & V_p \end{bmatrix} \begin{bmatrix} -1 \\ \theta \end{bmatrix} = V_y - 2\theta^{'} V_{yp} + \theta^{'} V_p \theta =$$

$$= \underbrace{\theta^{'} V_p \theta - 2\theta^{'} V_p V_p^{-1} V_{yp} + V_{yp}^{'} V_p V_{yp}}_{\left(\theta - V_p^{-1} V_{yp}\right)^{'} \left(\theta - V_p^{-1} V_{yp}\right) = 0} + \underbrace{V_y - V_{yp}^{'} V_p V_{yp}}_{\Lambda} = \Lambda$$

for optimal

$$\hat{\theta} = V_p^{-1} V_{yp}$$

– parameter r

Now, putting the minimized quadratic form back to the posterior (1), we get

$$f\left(\Theta | d\left(N\right)\right) \propto r^{-0.5\kappa_N} \exp\left\{-\frac{1}{2r}\Lambda\right\}.$$

The derivative is

$$\frac{\partial f}{\partial r} = 0.5\kappa_N r^{-0.5\kappa_N - 1} \exp\left(-\frac{\Lambda}{2r}\right) = r^{-0.5\kappa_N} \exp\left(-\frac{\Lambda}{2r}\right) \frac{1}{2r^2} = 0$$

$$\kappa_N = r\frac{\Lambda}{r^2} \quad \rightarrow \quad \hat{r} = \frac{\Lambda}{\kappa_N} = \frac{V_y - V_{yp}^{'} V_p^{-1} V_{yp}}{\kappa_N}$$

Estimation with categorical model

Model

The model pf is

$$f\left(y_t|\psi_t,\Theta\right)=\Theta_{y_t|\psi_t}=\prod_{y|\psi}\Theta_{y|\psi}^{\delta(y|\psi;y_t|\psi_t)}$$

where $\delta\left(y|\psi;y_t|\psi_t\right)=1$ for $y|\psi=y_t|\psi_t$ and zero otherwise. The rightmost expression i formal and helps to derive the recursion for the statistics.

The conjugate distribution is the Dirichlet one

$$f\left(\Theta|d\left(t\right)\right)\propto\prod_{y|\psi}\Theta_{y|\psi}^{S_{y|\psi;t}}$$

Statistics

The Bayes rule gives

$$\prod_{y|\psi}\Theta_{y|\psi}^{S_{y|\psi;t}}\propto\prod_{y|\psi}\Theta_{y|\psi}^{\delta(y|\psi;y_t|\psi_t)}\times\prod_{y|\psi}\Theta_{y|\psi}^{S_{y|\psi};t-1}=$$

$$=\prod_{y|\psi}\Theta_{y|\psi}^{S_{y|\psi;t-1}+\delta(y|\psi;y_t|\psi_t)}.$$

From it, the statistics update is

$$S_{y|\psi;t}=S_{y|\psi;t-1}+\delta\left(y|\psi;y_t|\psi_t\right)$$

Remark

The statistic is a table, similarly as the model. This formula says: Take the entry with the index $y_t|\psi_t$ and increment it by one. The rest stays unchanged.

The final statistics determine the posterior pdf

$$f\left(\Theta|d\left(N\right)\right)\propto\prod_{y|\psi}\Theta_{y|\psi}^{S_{y|\psi;N}}$$

Point estimates

Point estimates of Θ are arguments of its maxima.

Example

For $f\left(\Theta|d\left(N\right)\right)=\Theta^{S_1}\Theta^{S_2}\left(1-\Theta_1-\Theta_2\right)^{S_3}$ we have

$$\frac{\partial f}{\partial\Theta_1}=S_1\Theta_1^{S_1-1}\Theta_2^{S2}\left(1-\Theta_1-\Theta_2\right)^{S_3}-\Theta_1^{S_1}\Theta_2^{S2}S_3\left(1-\Theta_1-\Theta_2\right)^{S_3-1}=0$$

$$\frac{\partial f}{\partial\Theta_2}=\Theta_1^{S_1}S_2\Theta_2^{S2-1}\left(1-\Theta_1-\Theta_2\right)^{S_3}-\Theta_1^{S_1}\Theta_2^{S2}S_3\left(1-\Theta_1-\Theta_2\right)^{S_3-1}=0$$

subtracting

$$S_1\Theta_1^{S_1-1}\Theta_2^{S2}\left(1-\Theta_1-\Theta_2\right)^{S_3}-\Theta_1^{S_1}S_2\Theta_2^{S2-1}\left(1-\Theta_1-\Theta_2\right)^{S_3}=0$$

$$S_1\Theta_2=S_2\Theta_1 \tag{2}$$

from the first derivative

$$S_1\left(1-\Theta_1-\Theta_2\right)=\Theta_1 S_3 \;\rightarrow\; S_1-S_1\Theta_1-S_1\Theta_2=\Theta_1 S_3$$

$$\Theta_1\left(S_3+S_1\right)+\Theta_2 S_1=S_1$$

substitute $S_1\Theta_2=S_2\Theta_1$ and we have

$$\Theta_1\left(S_1+S_3\right)+S_2\Theta_1=S_1$$

$$\Theta_1=\frac{S_1}{S_1+S_2+S_3}$$

and from (2)

$$\Theta_2=\frac{S_2}{S_1}\Theta_1=\frac{S_2}{S_1}\frac{S_1}{S_1+S_2+S_3}=\frac{S_2}{S_1+S_2+S_3}$$