

## Estimation with normal regression model

### Model

Normal regression model has the pdf

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

The square can be edited as follows

$$\begin{aligned} (y_t - \psi_t' \theta)^2 &= (y_t - \theta' \psi_t) (y_t - \psi_t' \theta) = \\ &= \left( -[-1, \theta'] \begin{bmatrix} y_t \\ \psi_t \end{bmatrix} \right) \left( -[y_t, \psi_t'] \begin{bmatrix} -1 \\ \theta \end{bmatrix} \right) = \\ &= [-1, \theta'] \left( \begin{bmatrix} y_t \\ \psi_t \end{bmatrix} [y_t, \psi_t'] \right) \begin{bmatrix} -1 \\ \theta \end{bmatrix} = \\ &= [-1, \theta'] D_t \begin{bmatrix} -1 \\ \theta \end{bmatrix} \end{aligned}$$

where  $D_t = \begin{bmatrix} y_t \\ \psi_t \end{bmatrix} [y_t, \psi_t']$  is so called data matrix.

The regression model then is

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

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

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

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

## Statistics

Now, substituting into the Bayes rule

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

we get

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

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

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

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

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

## Point estimates

The point estimates of the parameters  $\theta$  and  $r$  lay at its maximum.

– parameter  $\theta$

The maximum will be at the minimum of  $\begin{bmatrix} -1, \theta' \end{bmatrix} V_N \begin{bmatrix} -1 \\ \theta \end{bmatrix}$ . It can be found by completion to square. We express

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

$$\begin{aligned} [-1, \theta'] & \begin{bmatrix} V_y & V'_{yp} \\ V_{yp} & V_p \end{bmatrix} \begin{bmatrix} -1 \\ \theta \end{bmatrix} = V_y - 2\theta' V_{yp} + \theta' V_p \theta = \\ & = \underbrace{\theta' V_p \theta - 2\theta' V_p V_p^{-1} V_{yp} + V'_{yp} V_p V_{yp}}_{(\theta - V_p^{-1} V_{yp})' (\theta - V_p^{-1} V_{yp}) = 0} + \underbrace{V_y - V'_{yp} V_p V_{yp}}_{\Lambda} = \Lambda \end{aligned}$$

for optimal

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

– parameter  $r$

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

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

The derivative is

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

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

## Estimation with categorical model

### Model

The model pf is

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

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

The conjugate distribution is the Dirichlet one

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

### Statistics

The Bayes rule gives

$$\begin{aligned} \prod_{y|\psi} \Theta_{y|\psi}^{S_{y|\psi; t}} &\propto \prod_{y|\psi} \Theta_{y|\psi}^{\delta(y|\psi; y_t|\psi_t)} \times \prod_{y|\psi} \Theta_{y|\psi}^{S_{y|\psi; t-1}} = \\ &= \prod_{y|\psi} \Theta_{y|\psi}^{S_{y|\psi; t-1} + \delta(y|\psi; y_t|\psi_t)}. \end{aligned}$$

From it, the statistics update is

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

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

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

### Point estimates

Point estimates of  $\Theta$  are arguments of its maxima.

Example

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

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

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

subtracting

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

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

from the first derivative

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

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

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

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

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

and from (2)

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