$$P(y^i_{hlt} = 1 | x_{hlt}, \beta_h) = \frac{\exp(\beta^T x_{hlti})}{\sum_{j'} \exp(\beta^T x_{hltj'})}$$

$$KL[q \| p] = \int \log \frac{q(\beta_h | \lambda)}{p(\beta_h | D, \vartheta, \Omega)} = E\left[\log \frac{q(\beta_h | \lambda)}{p(\beta_h | D, \vartheta, \Omega)}\right]$$

$$= -\mathcal{L}(\lambda, \vartheta, \Omega) + \log p(D | \vartheta, \Omega)$$

$$\beta_h | \vartheta, \Omega \overset{iid}{\sim} N_K(\vartheta, \Omega)$$

$$\vartheta | \beta_0, \Omega_0 \sim N_K(\beta_0, \Omega_0)$$

$$\Omega | S, r \sim W^{-1}(S^{-1}, r)$$

$$p(\beta_h, \vartheta, \Omega | D) = \frac{p(\vartheta) p(\Omega) \prod^H p(\beta_h | \vartheta, \Omega) \prod^{T_h} p(y_{ht} | x_{ht}, \beta_h)}{\int p(\vartheta) p(\Omega) \prod^H p(\beta_h | \vartheta, \Omega) \prod^{T_h} p(y_{ht} | x_{ht}, \beta_h) \, d\beta_h \, d\vartheta \, d\Omega}$$

$$p(y_{ht} | x_{ht}, \beta_{ht}) = \prod^J \left[\frac{\exp(x^T_{htj} \beta_h)}{\sum_{j'} \exp(x^T_{htj'} \beta_h)}\right]^{y^j_{ht}}$$

$$\mathcal{L}(\lambda; \vartheta, \Omega) := -E_{q_\lambda} \log\left[\frac{q(\beta_h | \lambda)}{p(\beta_h, D | \vartheta, \Omega)}\right]$$

$$q(\beta_h | \mu_h, \Sigma_h) \overset{iid}{\sim} N(\mu_h, \Sigma_h)$$

$$H(q) + \sum_{h=1}^H E_q \log p(\beta_h | \vartheta, \Omega) + \sum_{h=1}^H \sum_{t=1}^{T_h} E_q \log p(y_{ht} | x_{ht}, \beta_h)$$

$$H(q) + \sum_{h=1}^H E_q[\log p(\beta_h | \vartheta, \Omega)] + \sum_{h=1}^H \sum_{t=1}^{T_h} E_q\left[\sum^J y^j_{ht} (x^T_{htj} \beta_h) - \cancel{\log(\textstyle\sum^J x^T_{htj} \beta_h)} \log\left(\sum_{j=1}^J \exp(x^T_{htj} \beta_h)\right)\right]$$

$$H(q) + \sum_{h=1}^H E_q[\log p(\beta_h | \vartheta, \Omega)] + \sum_{h=1}^H \sum_{t=1}^{T_h} \left[\sum^J y^j_{ht} x^T_{htj} \mu_h - \underbrace{E_q\left[\log\left(\sum_{j=1}^J \exp(x^T_{htj} \beta_h)\right)\right]}_{\text{has no closed form}}\right]$$

Delta method : $E\, g(v) \approx g(E(v)) + \frac{1}{2} tr\left[\left(\frac{\partial g(E(v))}{\partial v \partial v^T}\right) cov(v)\right]$

$$E_q\left[\log\left(\sum_{j=2}^{J} \exp(x_{htj}^T \beta_h)\right)\right] = \log\left(\sum_{j=2}^{J} \exp(x_{htj}^T \mu_h)\right) + \frac{1}{2} tr\left[\frac{\partial\left[\log\left(\sum^{J} \exp(x_{htj}^T \mu_h)\right)\right]}{\partial \mu_h \partial \mu_h^T} \Sigma_h\right]$$

$$= \log\left(\sum_{j=2}^{J} \exp(x_{htj}^T \mu_h)\right) + \frac{1}{2} tr\left[\frac{\partial\left[\log\left(\sum \exp(x_{htj}^T \mu_h)\right)\right]}{\partial \mu_h \partial \mu_h^T}\right] \exp(\sigma)$$

$$= \log\left(\sum_{j=2}^{J} \exp(x_{htj}^T \mu_h)\right) + \frac{1}{2}\left[(x_{htj}^T \mu_h)^{-1} (x_{ht} \odot x_{htj})^T \exp(x_{htj} \mu_h) - (x_{htj}^T \mu_h)^{-2} (x_{htj}^T \exp(x_{htj} \mu_h)) \odot (x_{htj}^T \exp(x_{htj} \mu_h))\right.$$

Jensen inequality $\quad \varphi(E[X]) \leq E[\varphi(X)]$

$$H(q) + \sum^{H} E_q[\log p(\beta_h | \varphi, \Omega)] + \sum_{h=1}^{H} \sum_{t=1}^{T_h} \left[\sum^{J} y_{ht}^j x_{htj}^T \mu_h - E_q\left[\log\left(\sum_{j=2}^{J} \exp(x_{htj}^T \beta_h)\right)\right]\right]$$

$$= H(q) + \sum^{H} E_q[\log p(\beta_h | \varphi, \Omega)] + \sum^{H} \sum^{T} \left[\sum^{J} y_{ht}^j x_{htj}^T \mu_h - \log\left(\sum^{J} \exp(x_{htj}^T \mu_h + \frac{1}{2} x_{htj}^T \Sigma_h x_{htj})\right)\right]$$

$$= \frac{1}{2} \sum^{H} \log[(2\pi e)^K |\Sigma_h|] - \frac{H}{2} \log((2\pi)^K |\Omega|) - \frac{1}{2} tr\left[\Omega^{-1} \sum^{H} \{\Sigma_h + (\mu_h - \varphi)(\mu_h - \varphi)^T\}\right]$$

$$+ \sum^{H} \sum^{T} \left[\sum^{J} y_{ht}^j x_{htj}^T \mu_h - \log\left(\sum^{J} \exp(x_{htj}^T \mu_h + \frac{1}{2} x_{htj}^T \Sigma_h x_{htj})\right)\right]$$