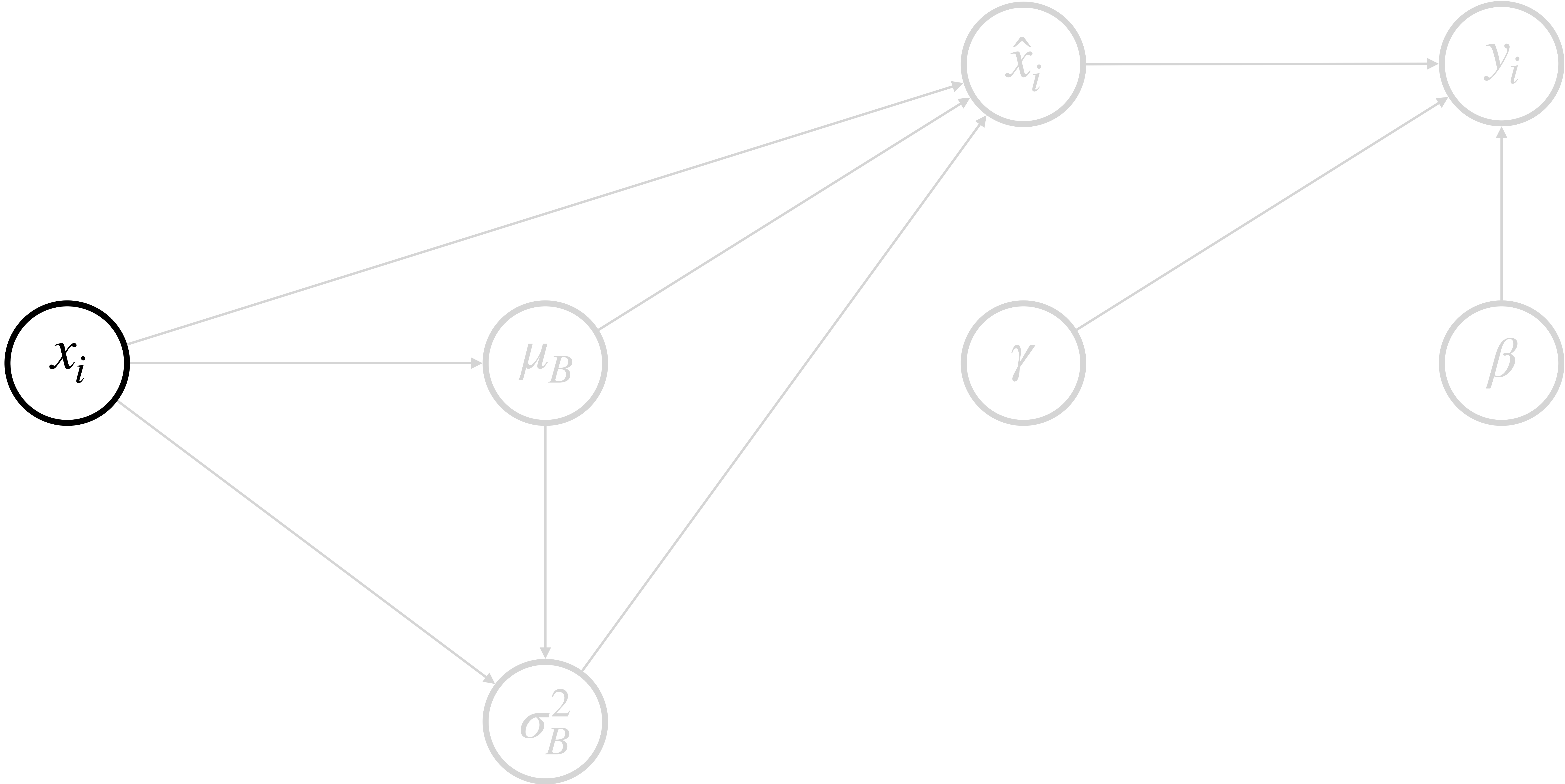
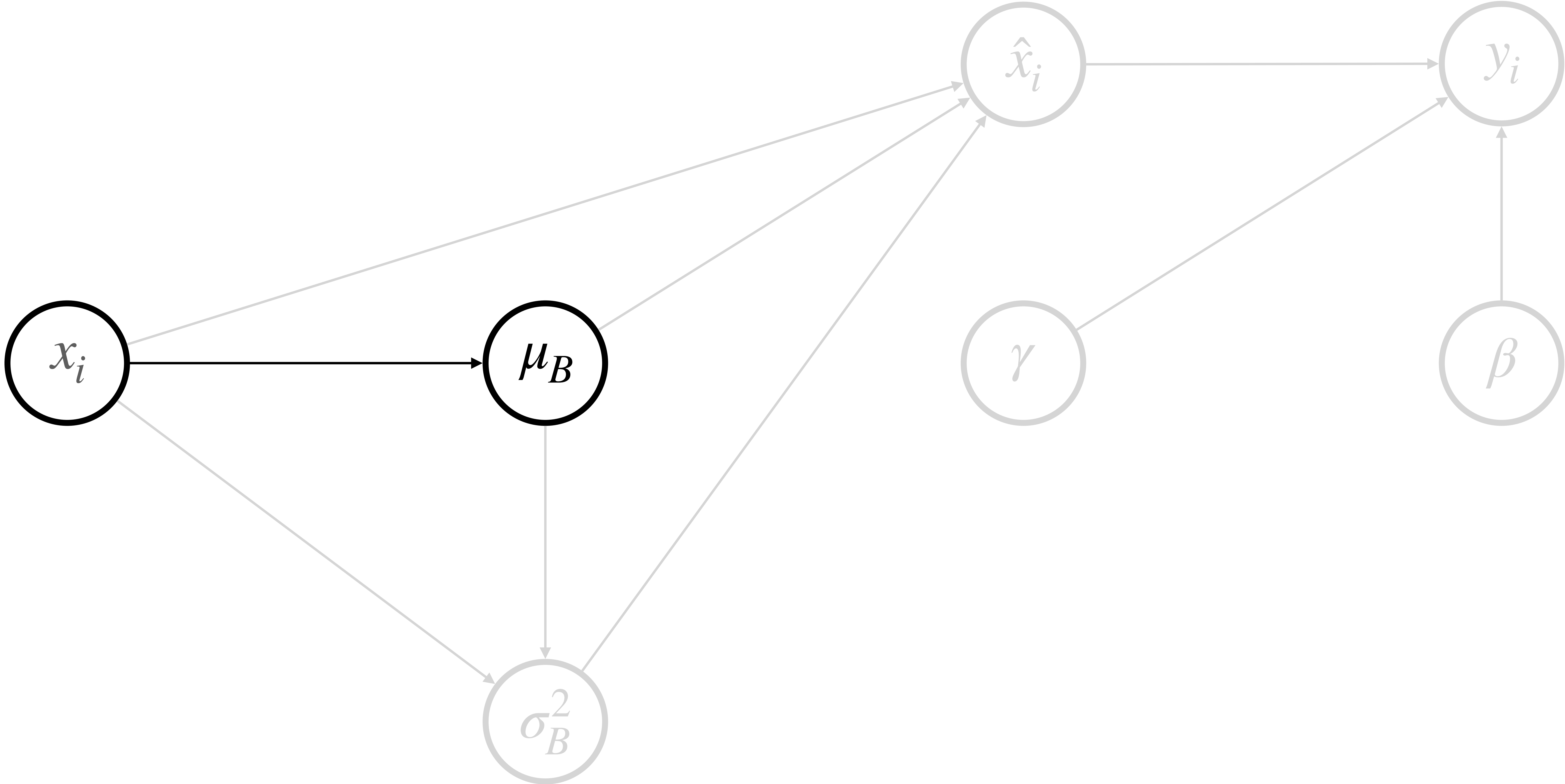
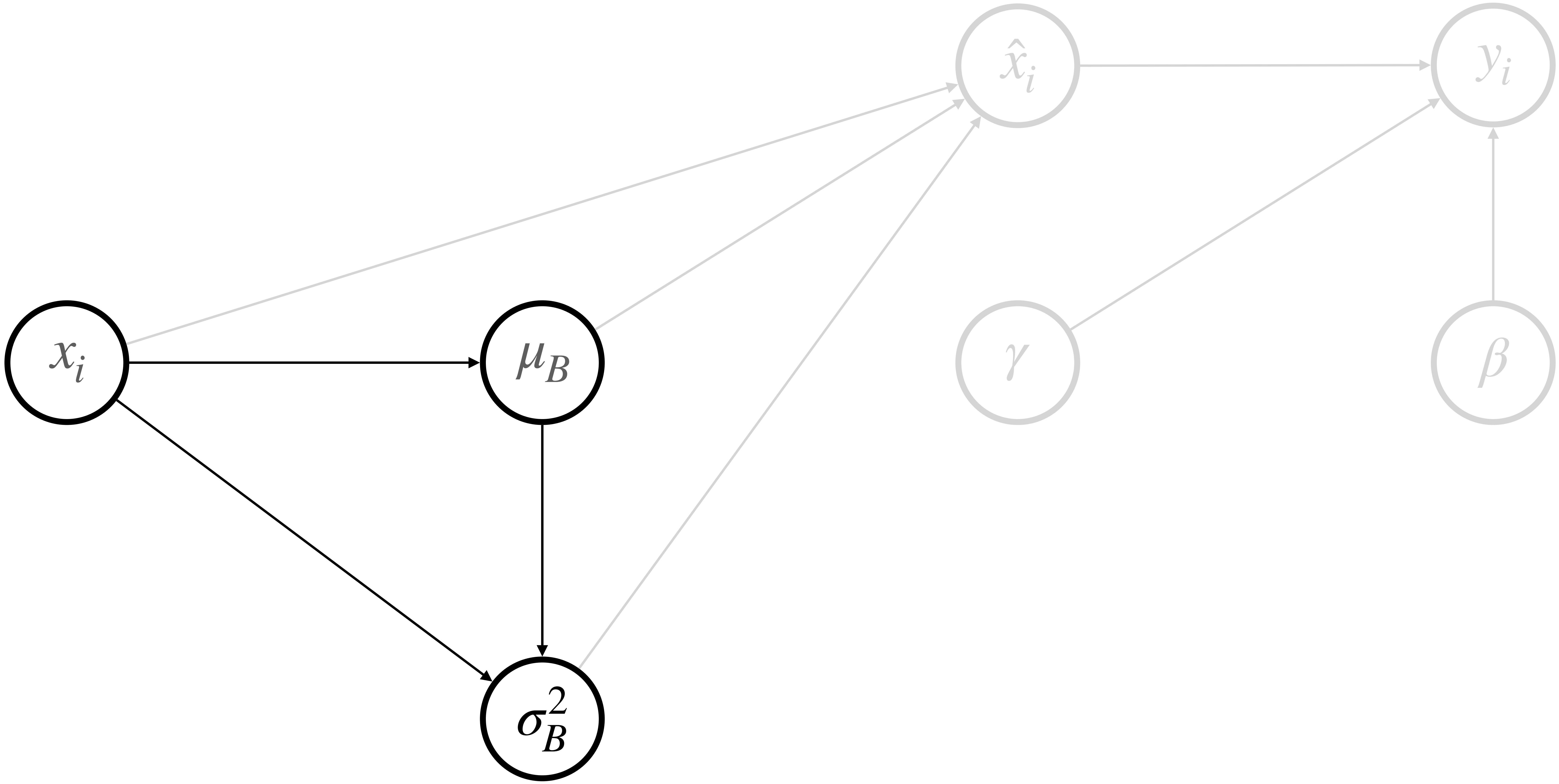

x_i
μ_B
σ_B^2
$\hat{x}_i$
γ
y_i
β

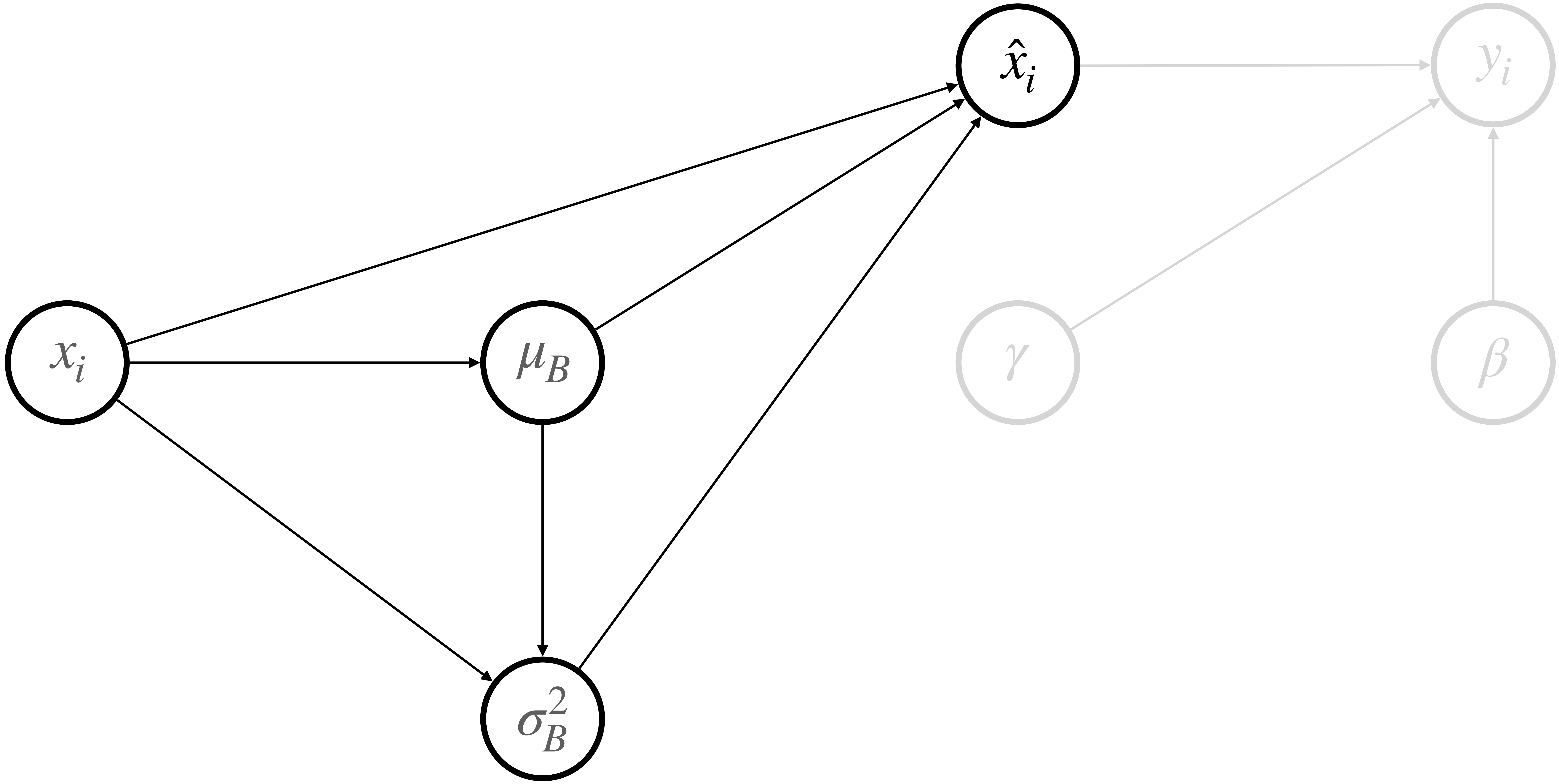


$$\mu_B = \frac{1}{m}\sum_{i=1}^{m} x_i$$



$$\mu_B = \frac{1}{m}\sum_{i=1}^{m} x_i$$

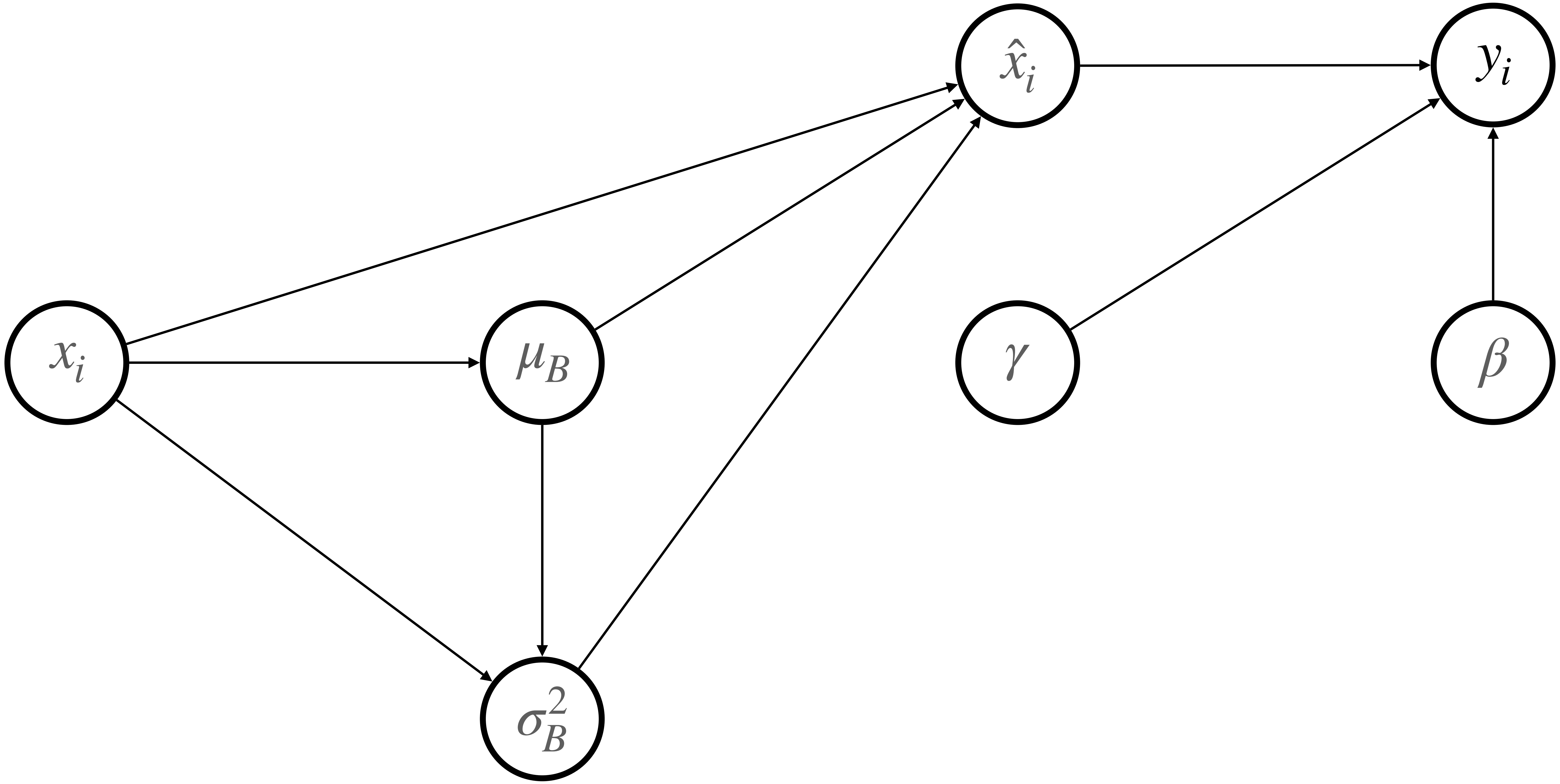
$$\sigma_B^2 = \frac{1}{m}\sum_{i=1}^{m} (x_i - \mu_B)^2$$



$$\mu_B = \frac{1}{m}\sum_{i=1}^{m} x_i$$

$$\sigma_B^2 = \frac{1}{m}\sum_{i=1}^{m} (x_i - \mu_B)^2$$

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}$$

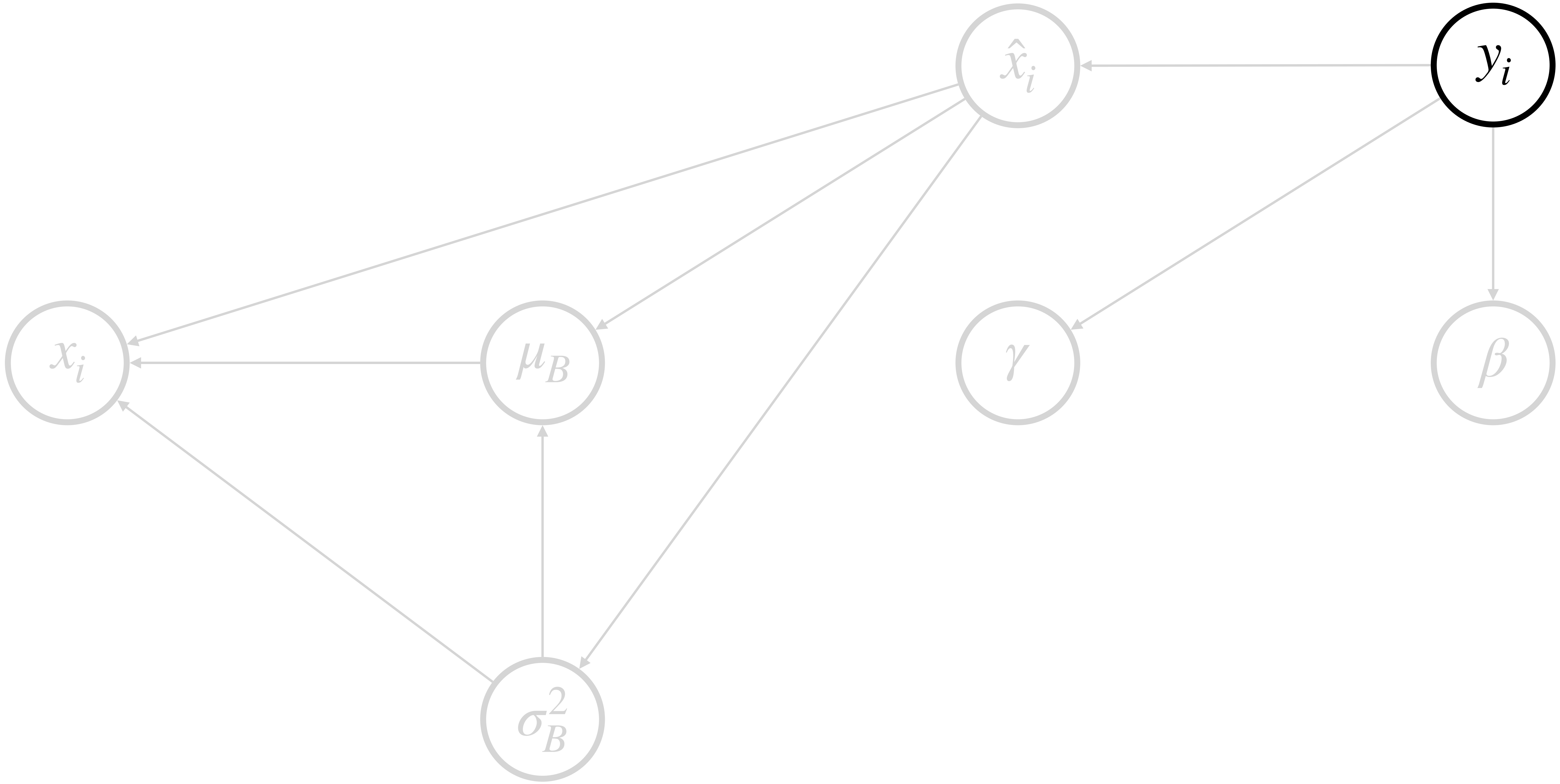


$$\mu_B = \frac{1}{m}\sum_{i=1}^{m} x_i$$

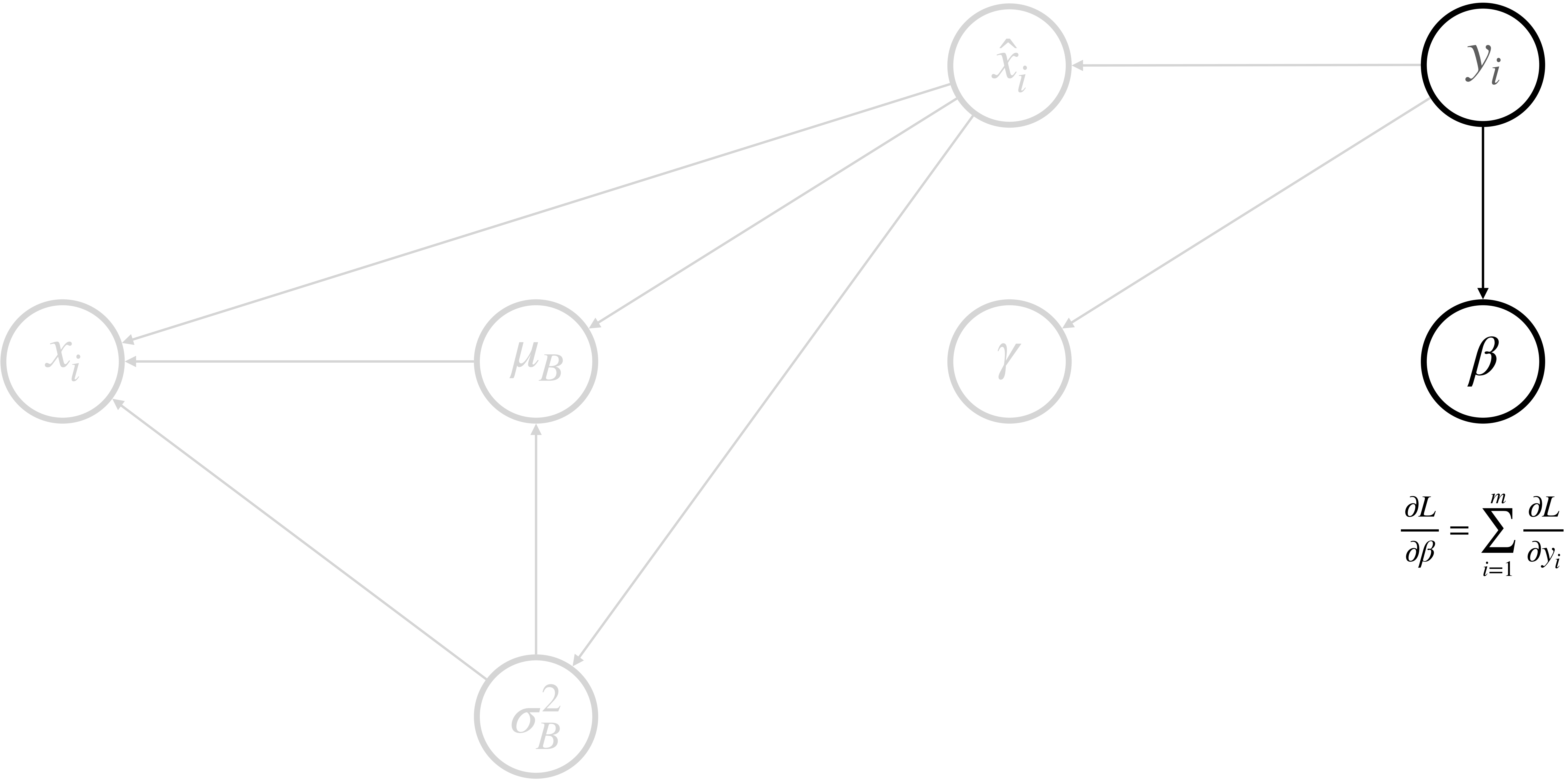
$$\sigma_B^2 = \frac{1}{m}\sum_{i=1}^{m} (x_i - \mu_B)^2$$

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}$$

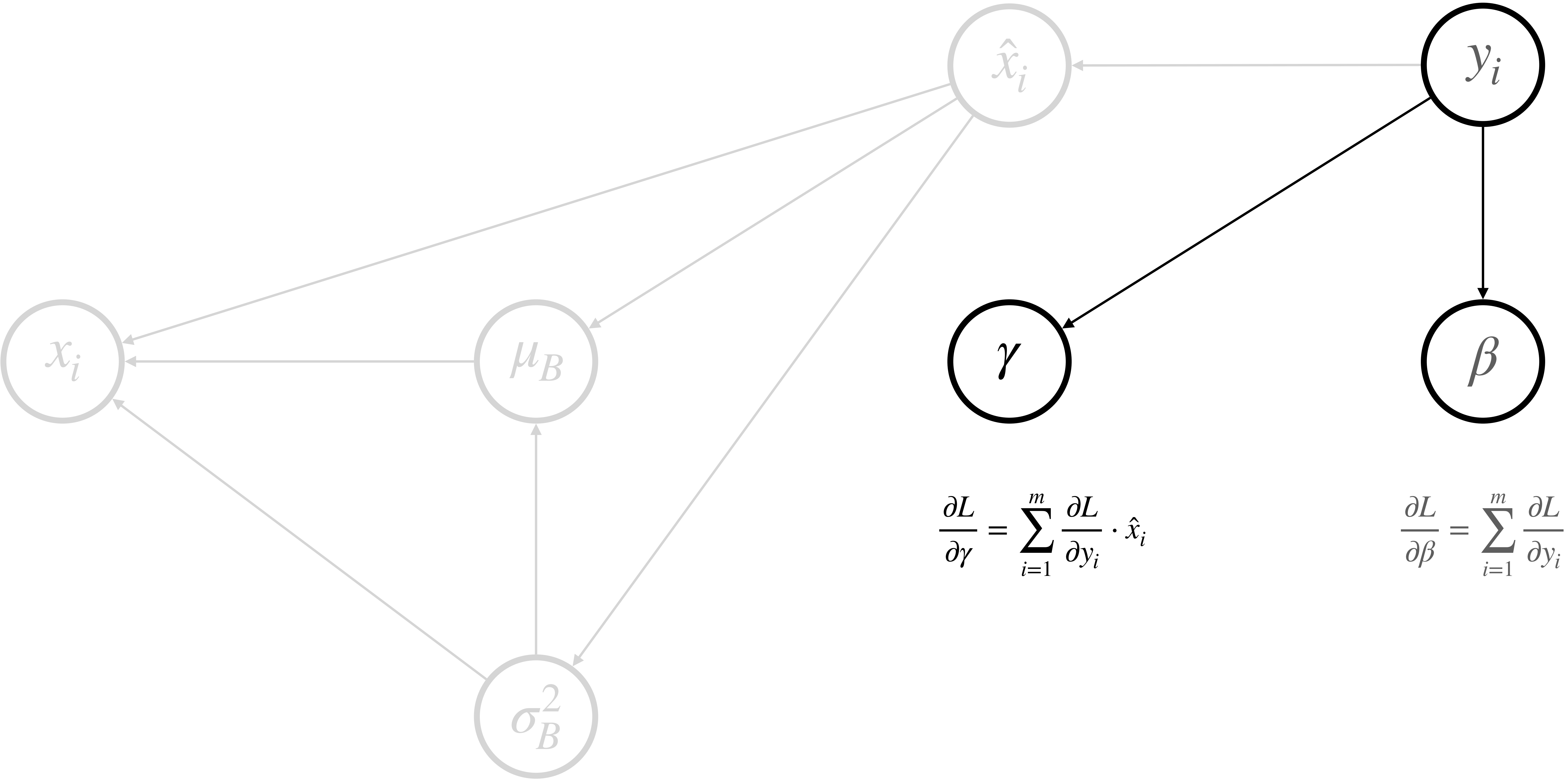
$$y_i = \gamma \hat{x}_i + \beta$$



$\hat{x}_i$
y_i
x_i
μ_B
γ
β
σ_B^2

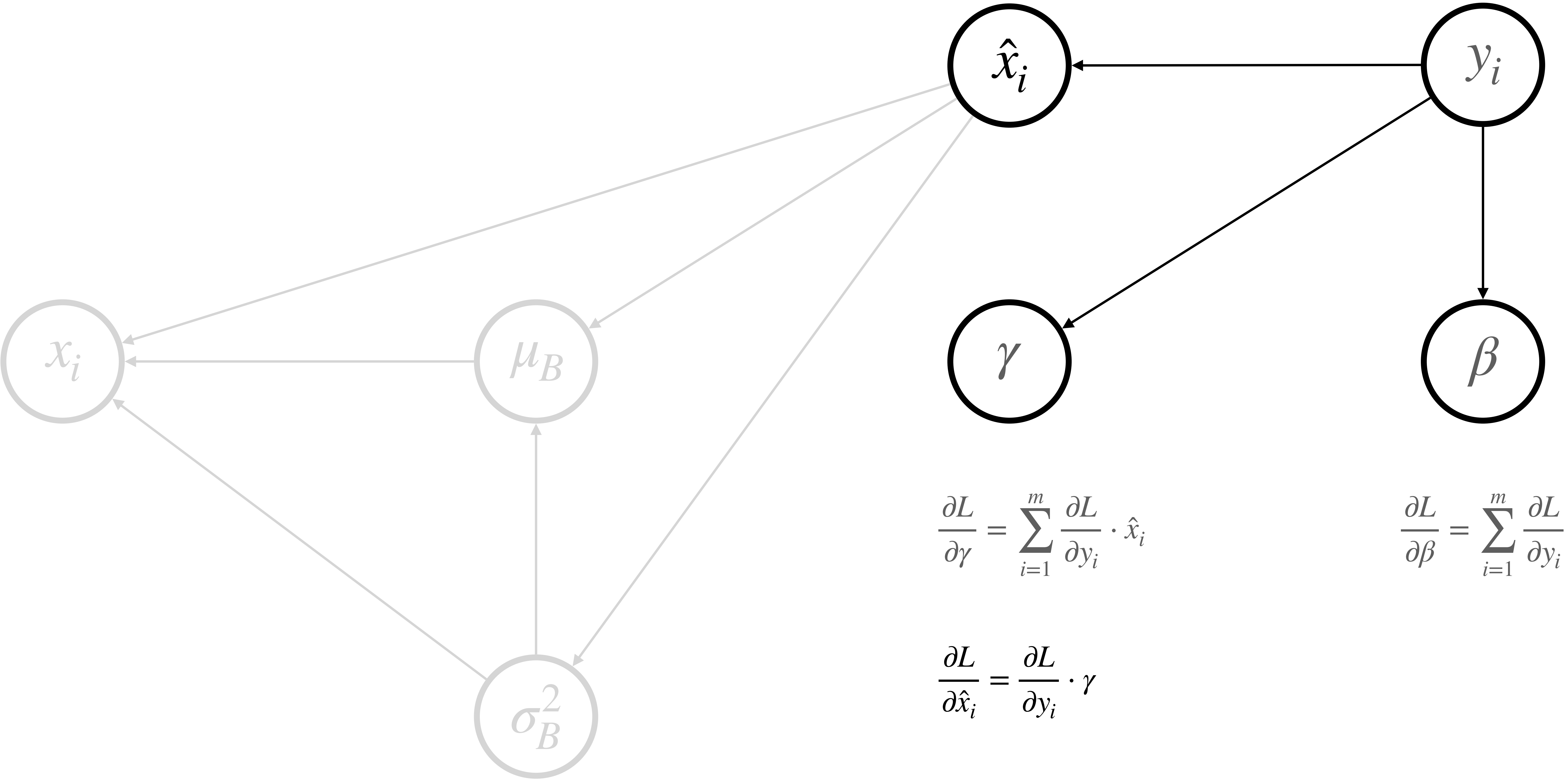


$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^{m} \frac{\partial L}{\partial y_i}$$



$$\frac{\partial L}{\partial \gamma} = \sum_{i=1}^{m} \frac{\partial L}{\partial y_i} \cdot \hat{x}_i$$

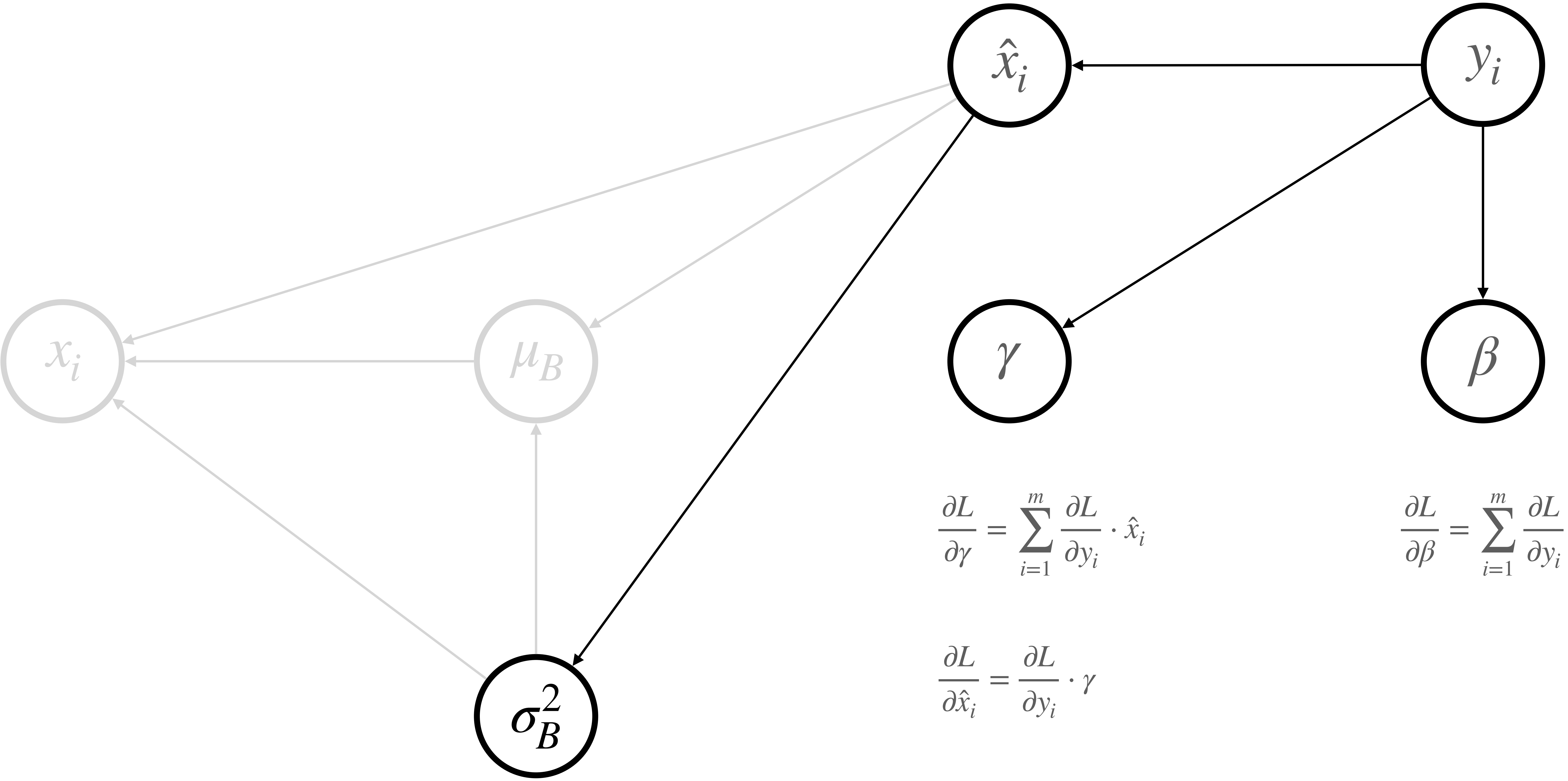
$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^{m} \frac{\partial L}{\partial y_i}$$



$$\frac{\partial L}{\partial \gamma} = \sum_{i=1}^{m} \frac{\partial L}{\partial y_i} \cdot \hat{x}_i$$

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^{m} \frac{\partial L}{\partial y_i}$$

$$\frac{\partial L}{\partial \hat{x}_i} = \frac{\partial L}{\partial y_i} \cdot \gamma$$

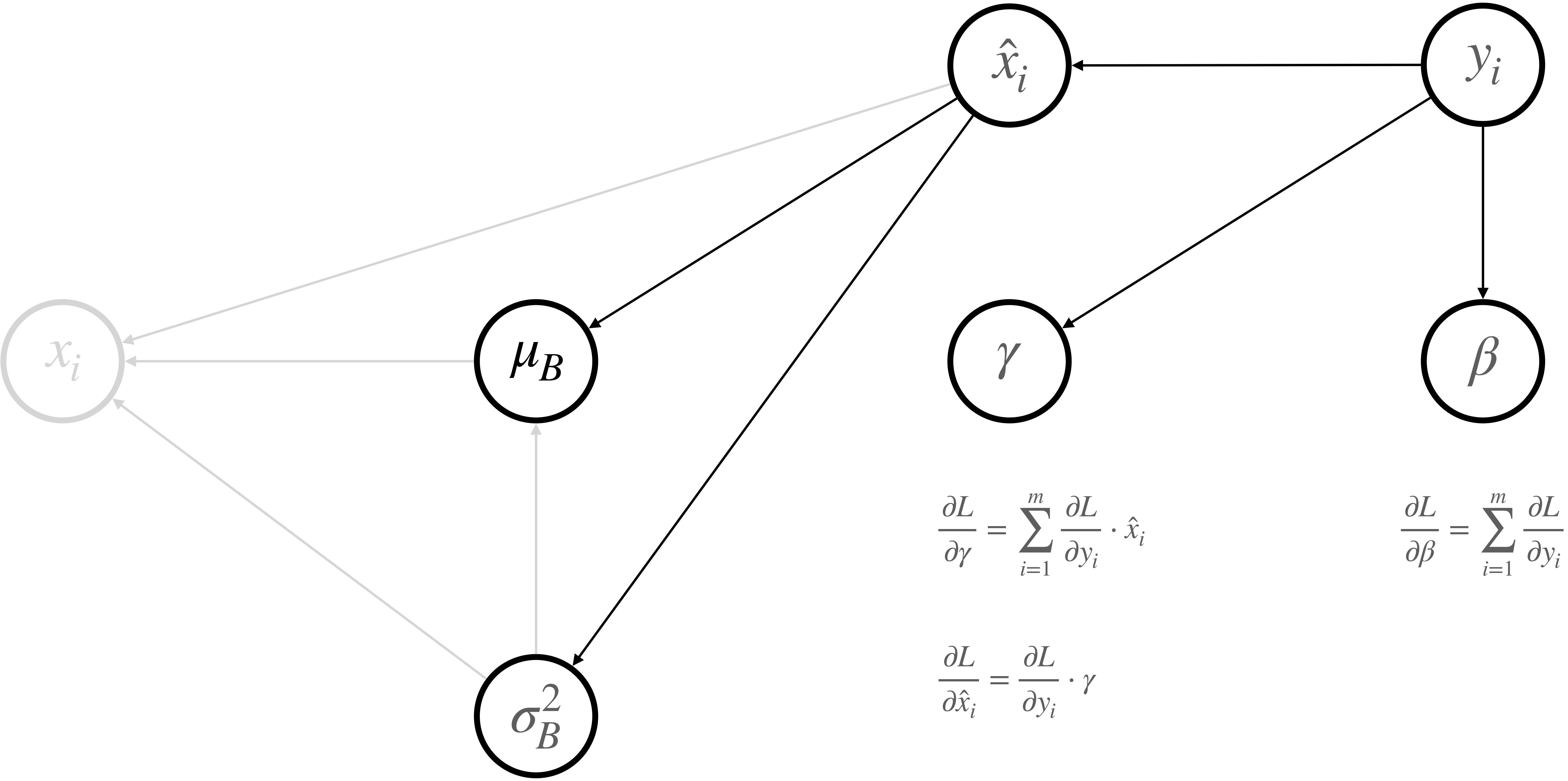


$$\frac{\partial L}{\partial \gamma} = \sum_{i=1}^{m} \frac{\partial L}{\partial y_i} \cdot \hat{x}_i$$

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^{m} \frac{\partial L}{\partial y_i}$$

$$\frac{\partial L}{\partial \hat{x}_i} = \frac{\partial L}{\partial y_i} \cdot \gamma$$

$$\frac{\partial \hat{x}_i}{\partial \sigma_B^2} = (x_i - \mu_B) \cdot \left(-\frac{1}{2}\right) \cdot \left(\sigma_B^2 + \epsilon\right)^{-\frac{3}{2}}$$



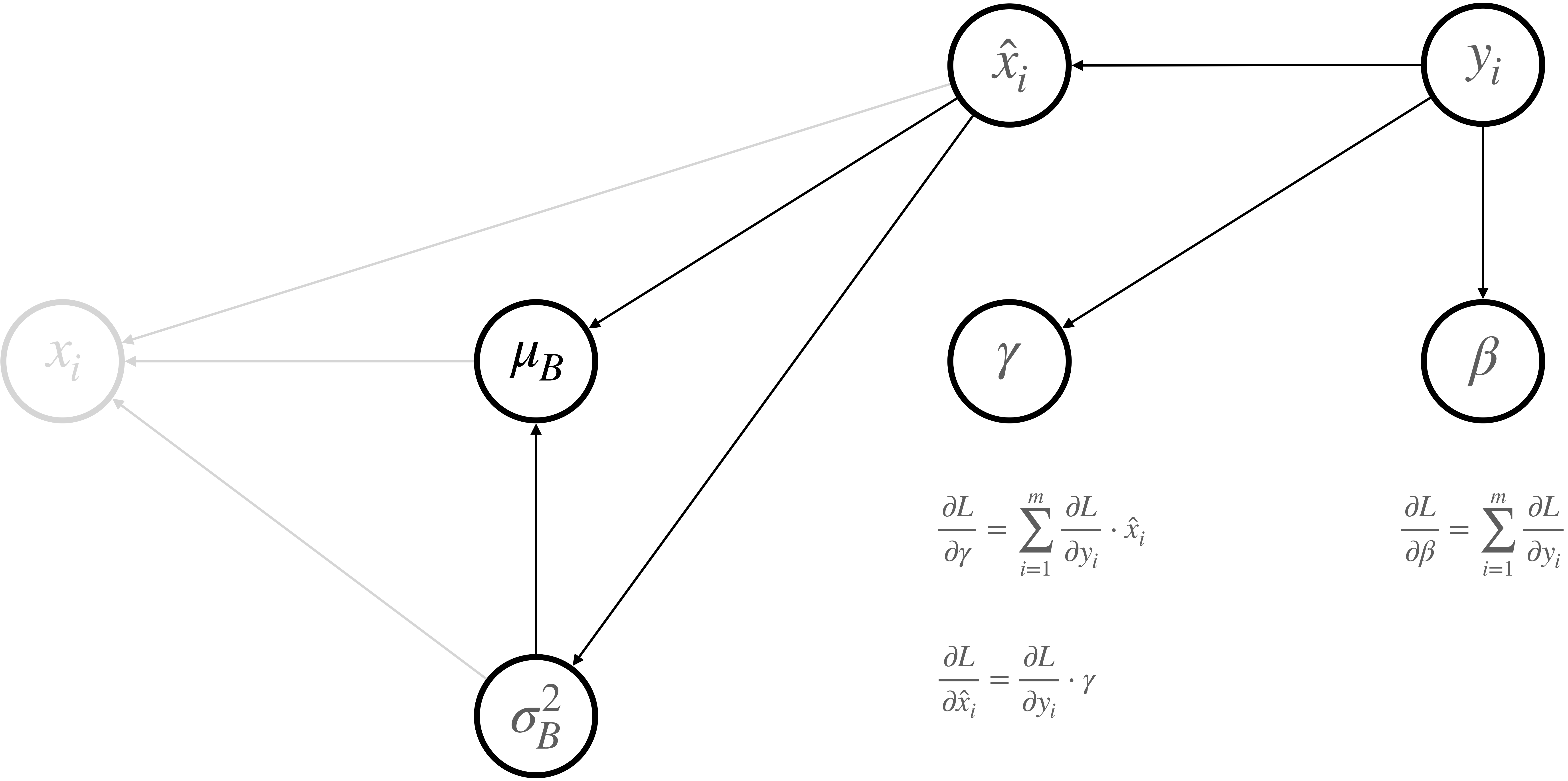
$$\frac{\partial L}{\partial \gamma} = \sum_{i=1}^{m} \frac{\partial L}{\partial y_i} \cdot \hat{x}_i$$

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^{m} \frac{\partial L}{\partial y_i}$$

$$\frac{\partial L}{\partial \hat{x}_i} = \frac{\partial L}{\partial y_i} \cdot \gamma$$

$$\frac{\partial \hat{x}_i}{\partial \sigma_B^2} = (x_i - \mu_B) \cdot \left(-\frac{1}{2}\right) \cdot \left(\sigma_B^2 + \epsilon\right)^{-\frac{3}{2}}$$

$$\frac{\partial \hat{x}_i}{\partial \mu_B} = \frac{-1}{\sqrt{\sigma_B^2 + \epsilon}}$$



$$\frac{\partial L}{\partial \gamma} = \sum_{i=1}^{m} \frac{\partial L}{\partial y_i} \cdot \hat{x}_i$$

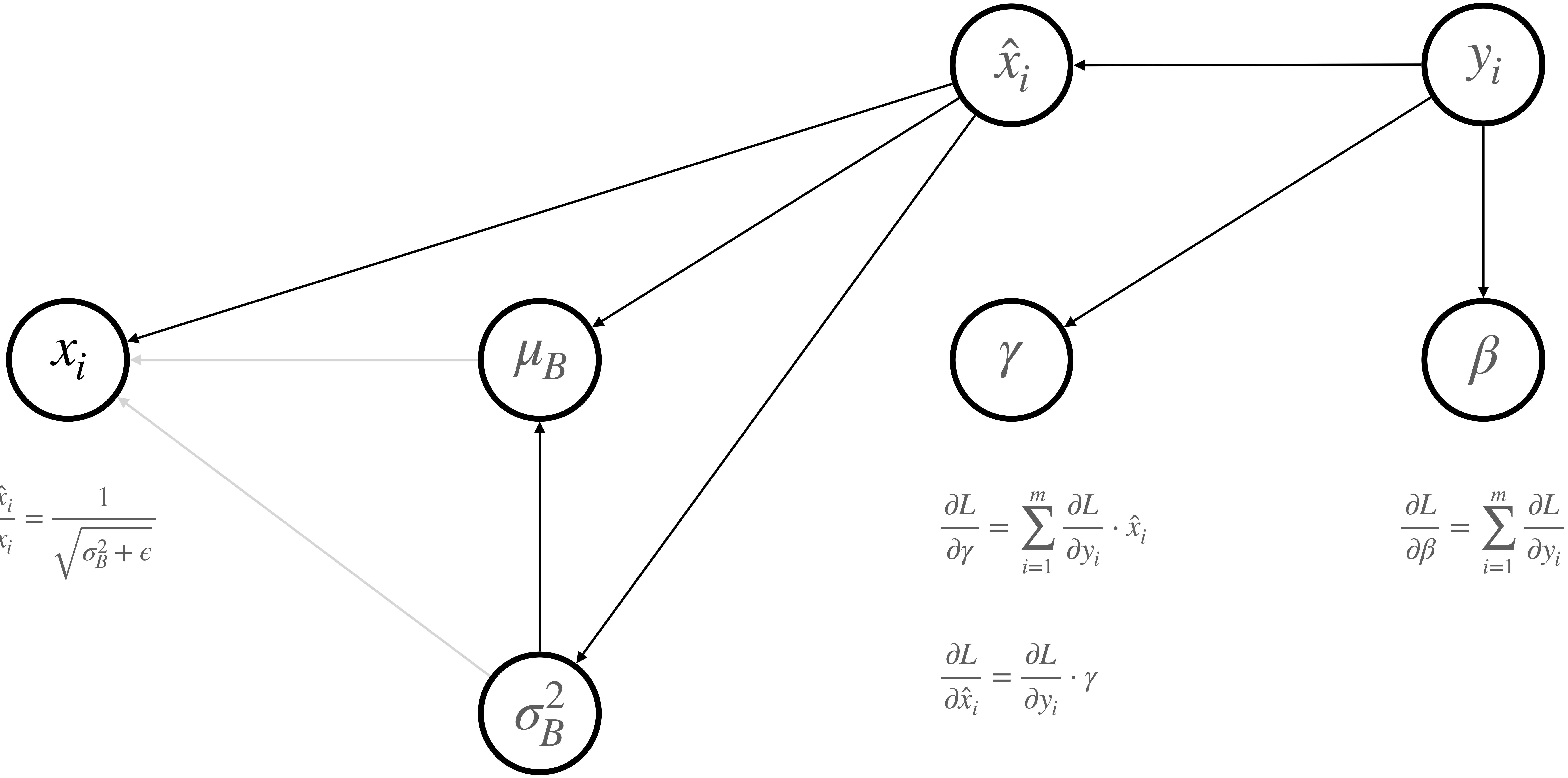
$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^{m} \frac{\partial L}{\partial y_i}$$

$$\frac{\partial L}{\partial \hat{x}_i} = \frac{\partial L}{\partial y_i} \cdot \gamma$$

$$\frac{\partial \hat{x}_i}{\partial \sigma_B^2} = (x_i - \mu_B) \cdot \left(-\frac{1}{2}\right) \cdot \left(\sigma_B^2 + \epsilon\right)^{-\frac{3}{2}}$$

$$\frac{\partial \hat{x}_i}{\partial \mu_B} = \frac{-1}{\sqrt{\sigma_B^2 + \epsilon}}$$

$$\frac{\partial \sigma_B^2}{\partial \mu_B} = \frac{\sum_{i=1}^{m} -2(x_i - \mu_B)}{m}$$



$$\frac{\partial \hat{x}_i}{\partial x_i} = \frac{1}{\sqrt{\sigma_B^2 + \epsilon}}$$

$$\frac{\partial L}{\partial \gamma} = \sum_{i=1}^{m} \frac{\partial L}{\partial y_i} \cdot \hat{x}_i$$

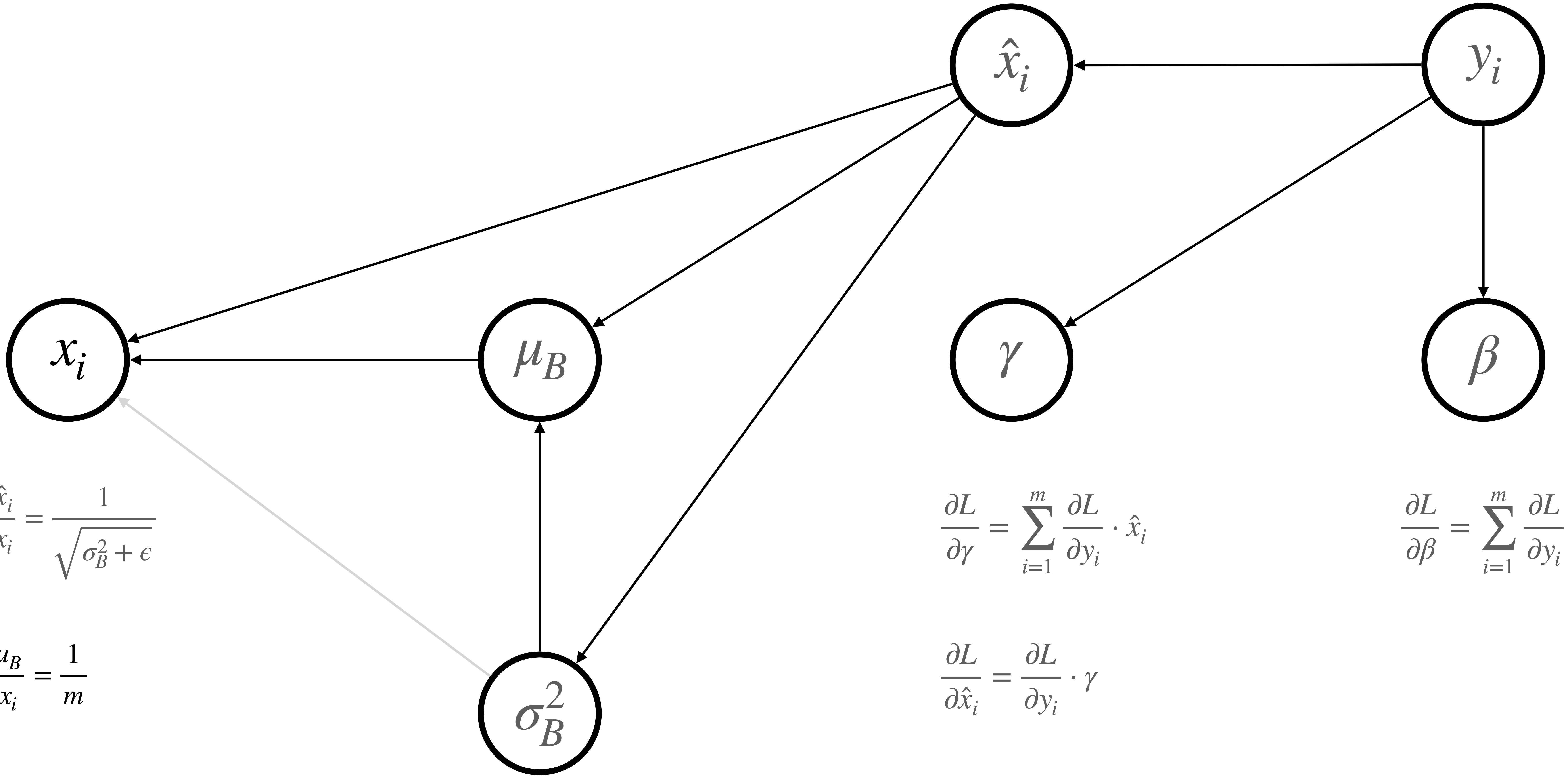
$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^{m} \frac{\partial L}{\partial y_i}$$

$$\frac{\partial L}{\partial \hat{x}_i} = \frac{\partial L}{\partial y_i} \cdot \gamma$$

$$\frac{\partial \hat{x}_i}{\partial \sigma_B^2} = (x_i - \mu_B) \cdot \left(-\frac{1}{2}\right) \cdot \left(\sigma_B^2 + \epsilon\right)^{-\frac{3}{2}}$$

$$\frac{\partial \hat{x}_i}{\partial \mu_B} = \frac{-1}{\sqrt{\sigma_B^2 + \epsilon}}$$

$$\frac{\partial \sigma_B^2}{\partial \mu_B} = \frac{\sum_{i=1}^{m} -2(x_i - \mu_B)}{m}$$



$$\frac{\partial \hat{x}_i}{\partial x_i} = \frac{1}{\sqrt{\sigma_B^2 + \epsilon}}$$

$$\frac{\partial \mu_B}{\partial x_i} = \frac{1}{m}$$

$$\frac{\partial L}{\partial \gamma} = \sum_{i=1}^{m} \frac{\partial L}{\partial y_i} \cdot \hat{x}_i$$

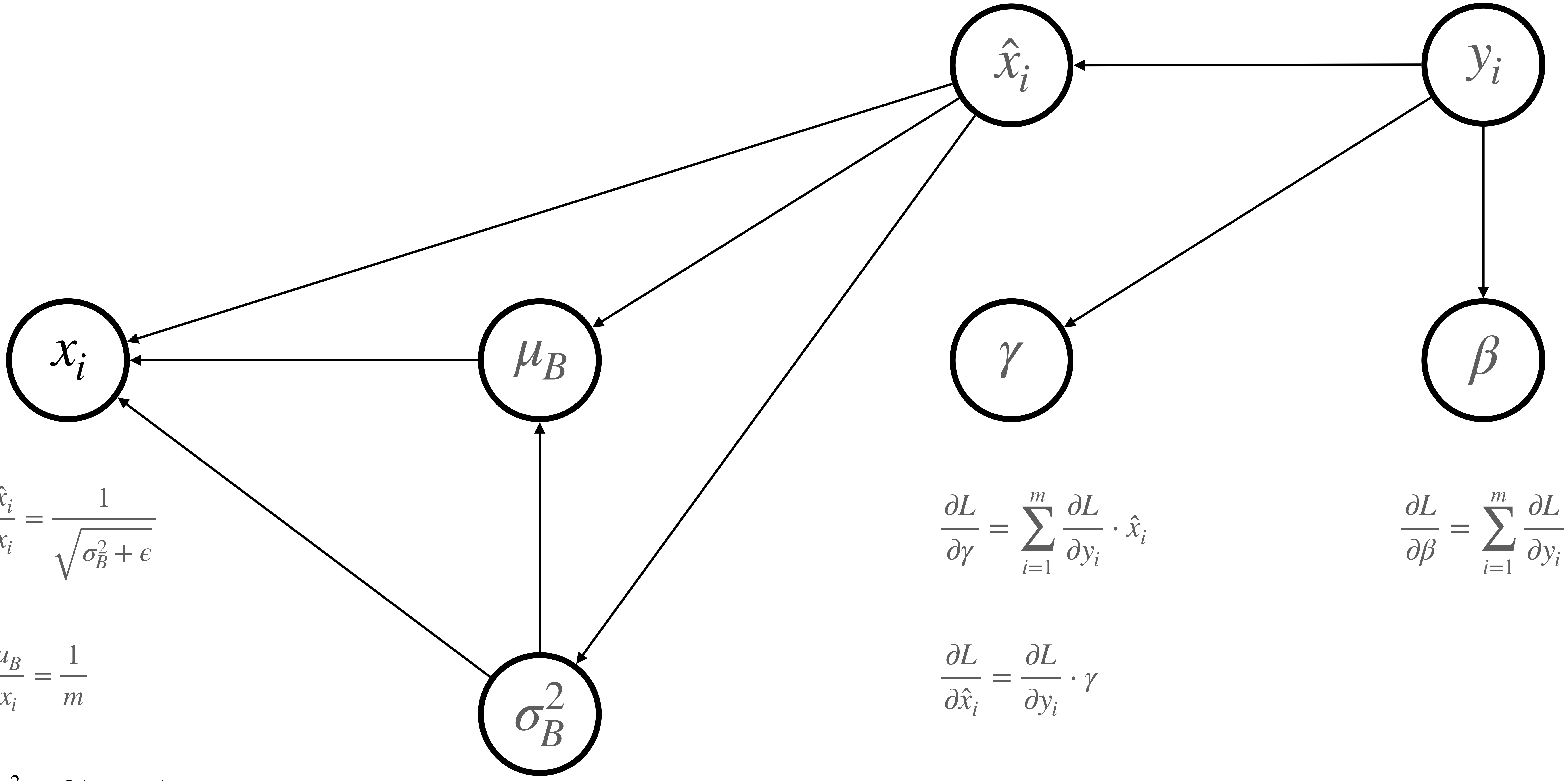
$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^{m} \frac{\partial L}{\partial y_i}$$

$$\frac{\partial L}{\partial \hat{x}_i} = \frac{\partial L}{\partial y_i} \cdot \gamma$$

$$\frac{\partial \hat{x}_i}{\partial \sigma_B^2} = (x_i - \mu_B) \cdot \left(-\frac{1}{2}\right) \cdot \left(\sigma_B^2 + \epsilon\right)^{-\frac{3}{2}}$$

$$\frac{\partial \hat{x}_i}{\partial \mu_B} = \frac{-1}{\sqrt{\sigma_B^2 + \epsilon}}$$

$$\frac{\partial \sigma_B^2}{\partial \mu_B} = \frac{\sum_{i=1}^{m} - 2(x_i - \mu_B)}{m}$$



$$\frac{\partial \hat{x}_i}{\partial x_i} = \frac{1}{\sqrt{\sigma_B^2 + \epsilon}}$$

$$\frac{\partial \mu_B}{\partial x_i} = \frac{1}{m}$$

$$\frac{\partial \sigma_B^2}{\partial x_i} = \frac{2(x_i - \mu_B)}{m}$$

$$\frac{\partial \hat{x}_i}{\partial \sigma_B^2} = (x_i - \mu_B) \cdot \left(-\frac{1}{2}\right) \cdot \left(\sigma_B^2 + \epsilon\right)^{-\frac{3}{2}}$$

$$\frac{\partial \hat{x}_i}{\partial \mu_B} = \frac{-1}{\sqrt{\sigma_B^2 + \epsilon}}$$

$$\frac{\partial \sigma_B^2}{\partial \mu_B} = \frac{\sum_{i=1}^{m} - 2(x_i - \mu_B)}{m}$$

$$\frac{\partial L}{\partial \gamma} = \sum_{i=1}^{m} \frac{\partial L}{\partial y_i} \cdot \hat{x}_i$$

$$\frac{\partial L}{\partial \hat{x}_i} = \frac{\partial L}{\partial y_i} \cdot \gamma$$

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^{m} \frac{\partial L}{\partial y_i}$$

$$
\begin{tikzcd}
 & & \hat{x}_i \arrow[dll] \arrow[dl] \arrow[ddl] & y_i \arrow[l] \arrow[dl] \arrow[d] \\
x_i & \mu_B \arrow[l] & \gamma & \beta \\
 & \sigma_B^2 \arrow[ul] \arrow[u] & &
\end{tikzcd}
$$

$$\frac{\partial \hat{x}_i}{\partial x_i} = \frac{1}{\sqrt{\sigma_B^2 + \epsilon}}$$

$$\frac{\partial \mu_B}{\partial x_i} = \frac{1}{m}$$

$$\frac{\partial \sigma_B^2}{\partial x_i} = \frac{2(x_i - \mu_B)}{m}$$

$$\frac{\partial \hat{x}_i}{\partial \sigma_B^2} = (x_i - \mu_B) \cdot \left(-\frac{1}{2}\right) \cdot \left(\sigma_B^2 + \epsilon\right)^{-\frac{3}{2}}$$

$$\frac{\partial \hat{x}_i}{\partial \mu_B} = \frac{-1}{\sqrt{\sigma_B^2 + \epsilon}}$$

$$\frac{\partial \sigma_B^2}{\partial \mu_B} = \frac{\sum_{i=1}^{m} -2(x_i - \mu_B)}{m}$$

$$\frac{\partial L}{\partial \gamma} = \sum_{i=1}^{m} \frac{\partial L}{\partial y_i} \cdot \hat{x}_i$$

$$\frac{\partial L}{\partial \hat{x}_i} = \frac{\partial L}{\partial y_i} \cdot \gamma$$

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^{m} \frac{\partial L}{\partial y_i}$$

$$\mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \boldsymbol{x_i} \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{in} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mj} & \cdots & x_{mn} \end{bmatrix}_{m\times n}$$

$$\mu_B = \begin{bmatrix} \mu_{B1} & \mu_{B2} & \cdots & \mu_{Bj} & \cdots & \mu_{Bn} \end{bmatrix}_{1\times n}$$

$$\sigma_B^2 = \begin{bmatrix} \sigma_{B1}^2 & \sigma_{B2}^2 & \cdots & \sigma_{Bj}^2 & \cdots & \sigma_{Bn}^2 \end{bmatrix}_{1\times n}$$

$$\hat{\mathbf{X}} = \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \vdots \\ \hat{\boldsymbol{x}}_{\boldsymbol{i}} \\ \vdots \\ \hat{x}_m \end{bmatrix} = \begin{bmatrix} \hat{x}_{11} & \hat{x}_{12} & \cdots & \hat{x}_{1j} & \cdots & \hat{x}_{1n} \\ \hat{x}_{21} & \hat{x}_{22} & \cdots & \hat{x}_{2j} & \cdots & \hat{x}_{2n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \hat{x}_{i1} & \hat{x}_{i2} & \cdots & \hat{x}_{ij} & \cdots & \hat{x}_{in} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \hat{x}_{m1} & \hat{x}_{m2} & \cdots & \hat{x}_{mj} & \cdots & \hat{x}_{mn} \end{bmatrix}_{m\times n}$$

$$\gamma = \begin{bmatrix} \gamma_1 & \gamma_2 & \cdots & \gamma_j & \cdots & \gamma_n \end{bmatrix}_{1\times n}$$

$$\beta = \begin{bmatrix} \beta_1 & \beta_2 & \cdots & \beta_j & \cdots & \beta_n \end{bmatrix}_{1\times n}$$

$$\begin{cases} \mu_B = \frac{1}{m}\sum_{i=1}^m x_i \\ \sigma_B^2 = \frac{1}{m}\sum_{i=1}^m (x_i - \mu_B)^2 \\ \hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \\ y_i = \gamma \hat{x}_i + \beta \end{cases}$$

$$\begin{cases} \frac{\partial L}{\partial \gamma} = \sum_{i=1}^m \frac{\partial L}{\partial y_i} \cdot \hat{x}_i \\ \frac{\partial L}{\partial \beta} = \sum_{i=1}^m \frac{\partial L}{\partial y_i} \\ \frac{\partial L}{\partial \hat{x}_i} = \frac{\partial L}{\partial y_i} \cdot \gamma \\ \frac{\partial L}{\partial \sigma_B^2} = \sum_{i=1}^m \frac{\partial L}{\partial \hat{x}_i} \cdot (x_i - \mu_B) \cdot \left(-\frac{1}{2}\right) \cdot \left(\sigma_B^2 + \epsilon\right)^{-\frac{3}{2}} \\ \frac{\partial L}{\partial \mu_B} = \sum_{i=1}^m \frac{\partial L}{\partial \hat{x}_i} \cdot \frac{-1}{\sqrt{\sigma_B^2 + \epsilon}} + \frac{\partial L}{\partial \sigma_B^2} \cdot \frac{\sum_{i=1}^m -2(x_i - \mu_B)}{m} \\ \frac{\partial L}{\partial x_i} = \frac{\partial L}{\partial \hat{x}_i} \cdot \frac{1}{\sqrt{\sigma_B^2 + \epsilon}} + \frac{\partial L}{\partial \sigma_B^2} \cdot \frac{2(x_i - \mu_B)}{m} + \frac{\partial L}{\partial \mu_B} \cdot \frac{1}{m} \end{cases}$$

$$\begin{aligned} \frac{\partial L}{\partial \sigma_B^2} &= \sum_{i=1}^m \frac{\partial L}{\partial \hat{x}_i} \cdot (x_i - \mu_B) \cdot \left(-\frac{1}{2}\right) \cdot \left(\sigma_B^2 + \epsilon\right)^{-\frac{3}{2}} \\ &= -\frac{1}{2} \cdot \sum_{i=1}^m \frac{\partial L}{\partial \hat{x}_i} \cdot \frac{\hat{x}_i}{\sigma_B^2 + \epsilon} \end{aligned}$$

$$\begin{aligned} \frac{\partial L}{\partial \mu_B} &= \sum_{i=1}^m \frac{\partial L}{\partial \hat{x}_i} \cdot \frac{-1}{\sqrt{\sigma_B^2 + \epsilon}} + \frac{\partial L}{\partial \sigma_B^2} \cdot \frac{\sum_{i=1}^m -2(x_i - \mu_B)}{m} \\ &= \sum_{i=1}^m \frac{\partial L}{\partial \hat{x}_i} \cdot \frac{-1}{\sqrt{\sigma_B^2 + \epsilon}} + \sum_{i=1}^m \frac{\partial L}{\partial \hat{x}_i} \cdot \frac{\hat{x}_i}{\sigma_B^2 + \epsilon} \cdot \frac{\sum_{i=1}^m (x_i - \mu_B)}{m} \\ &= \sum_{i=1}^m \frac{\partial L}{\partial \hat{x}_i} \cdot \frac{1}{\sqrt{\sigma_B^2 + \epsilon}} \cdot \left(\frac{\hat{x}_i}{\sqrt{\sigma_B^2 + \epsilon}} \cdot \frac{\sum_{i=1}^m (x_i - \mu_B)}{m} - 1 \right) \\ &= \sum_{i=1}^m \frac{\partial L}{\partial \hat{x}_i} \cdot \frac{1}{\sqrt{\sigma_B^2 + \epsilon}} \cdot \left(\frac{\hat{x}_i}{\sqrt{\sigma_B^2 + \epsilon}} \cdot \left(\frac{\sum_{i=1}^m x_i}{m} - \frac{\sum_{i=1}^m \mu_B}{m} \right) - 1 \right) \\ &= -\sum_{i=1}^m \frac{\partial L}{\partial \hat{x}_i} \cdot \frac{1}{\sqrt{\sigma_B^2 + \epsilon}} \end{aligned}$$

$$\begin{aligned} \frac{\partial L}{\partial x_i} &= \frac{\partial L}{\partial \hat{x}_i} \cdot \frac{\partial \hat{x}_i}{\partial x_i} + \frac{\partial L}{\partial \sigma_B^2} \cdot \frac{\partial \sigma_B^2}{\partial x_i} + \frac{\partial L}{\partial \mu_B} \cdot \frac{\partial \mu_B}{\partial x_i} \\ &= \frac{\partial L}{\partial \hat{x}_i} \cdot \frac{1}{\sqrt{\sigma_B^2 + \epsilon}} + \frac{\partial L}{\partial \sigma_B^2} \cdot \frac{2(x_i - \mu_B)}{m} + \frac{\partial L}{\partial \mu_B} \cdot \frac{1}{m} \\ &= \frac{\partial L}{\partial \hat{x}_i} \cdot \frac{1}{\sqrt{\sigma_B^2 + \epsilon}} - \sum_{i=1}^m \frac{\partial L}{\partial \hat{x}_i} \cdot \frac{\hat{x}_i}{\sigma_B^2 + \epsilon} \cdot \frac{x_i - \mu_B}{m} - \sum_{i=1}^m \frac{\partial L}{\partial \hat{x}_i} \cdot \frac{1}{\sqrt{\sigma_B^2 + \epsilon}} \cdot \frac{1}{m} \\ &= \frac{1}{m \cdot \sqrt{\sigma_B^2 + \epsilon}} \cdot \left(m \frac{\partial L}{\partial \hat{x}_i} - \hat{x}_i \cdot \sum_{i=1}^m \frac{\partial L}{\partial \hat{x}_i} \cdot \hat{x}_i - \sum_{i=1}^m \frac{\partial L}{\partial \hat{x}_i} \right) \end{aligned}$$

$$\mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \boldsymbol{x_i} \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{in} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mj} & \cdots & x_{mn} \end{bmatrix}_{m\times n} \qquad \mu_L = \begin{bmatrix} \mu_{L1} \\ \mu_{L2} \\ \vdots \\ \mu_{Li} \\ \vdots \\ \mu_{Lm} \end{bmatrix}_{m\times 1} \qquad \sigma_L^2 = \begin{bmatrix} \sigma_{L1}^2 \\ \sigma_{L2}^2 \\ \vdots \\ \sigma_{Li}^2 \\ \vdots \\ \sigma_{Lm}^2 \end{bmatrix}_{m\times 1}$$

$$\hat{\mathbf{X}} = \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \vdots \\ \hat{\boldsymbol{x}}_{\boldsymbol{i}} \\ \vdots \\ \hat{x}_m \end{bmatrix} = \begin{bmatrix} \hat{x}_{11} & \hat{x}_{12} & \cdots & \hat{x}_{1j} & \cdots & \hat{x}_{1n} \\ \hat{x}_{21} & \hat{x}_{22} & \cdots & \hat{x}_{2j} & \cdots & \hat{x}_{2n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \hat{x}_{i1} & \hat{x}_{i2} & \cdots & \hat{x}_{ij} & \cdots & \hat{x}_{in} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \hat{x}_{m1} & \hat{x}_{m2} & \cdots & \hat{x}_{mj} & \cdots & \hat{x}_{mn} \end{bmatrix}_{m\times n} \qquad \gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_i \\ \vdots \\ \gamma_m \end{bmatrix}_{m\times 1} \qquad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_i \\ \vdots \\ \beta_m \end{bmatrix}_{m\times 1}$$

$$\begin{cases} \mu_L = \frac{1}{n}\sum_{j=1}^{n} x_j \\ \sigma_L^2 = \frac{1}{n}\sum_{i=1}^{n} (x_j - \mu_L)^2 \\ \hat{x}_j = \frac{x_j - \mu_L}{\sqrt{\sigma_L^2 + \epsilon}} \\ y_j = \gamma \hat{x}_j + \beta \end{cases}$$

$$\begin{cases} \frac{\partial L}{\partial \gamma} = \sum_{i=1}^{m} \frac{\partial L}{\partial y_i} \cdot \hat{x}_i \\ \frac{\partial L}{\partial \beta} = \sum_{i=1}^{m} \frac{\partial L}{\partial y_i} \\ \frac{\partial L}{\partial \hat{x}_i} = \frac{\partial L}{\partial y_i} \cdot \gamma \\ \frac{\partial L}{\partial \sigma_B^2} = \sum_{i=1}^{m} \frac{\partial L}{\partial \hat{x}_i} \cdot (x_i - \mu_B) \cdot \left(-\frac{1}{2}\right) \cdot \left(\sigma_B^2 + \epsilon\right)^{-\frac{3}{2}} \\ \frac{\partial L}{\partial \mu_B} = \sum_{i=1}^{m} \frac{\partial L}{\partial \hat{x}_i} \cdot \frac{-1}{\sqrt{\sigma_B^2 + \epsilon}} + \frac{\partial L}{\partial \sigma_B^2} \cdot \frac{\sum_{i=1}^{m} - 2(x_i - \mu_B)}{m} \\ \frac{\partial L}{\partial x_i} = \frac{\partial L}{\partial \hat{x}_i} \cdot \frac{1}{\sqrt{\sigma_B^2 + \epsilon}} + \frac{\partial L}{\partial \sigma_B^2} \cdot \frac{2(x_i - \mu_B)}{m} + \frac{\partial L}{\partial \mu_B} \cdot \frac{1}{m} \end{cases}$$

$$\begin{aligned} \frac{\partial L}{\partial \sigma_B^2} &= \sum_{i=1}^{m} \frac{\partial L}{\partial \hat{x}_i} \cdot (x_i - \mu_B) \cdot \left(-\frac{1}{2}\right) \cdot \left(\sigma_B^2 + \epsilon\right)^{-\frac{3}{2}} \\ &= -\frac{1}{2} \cdot \sum_{i=1}^{m} \frac{\partial L}{\partial \hat{x}_i} \cdot \frac{\hat{x}_i}{\sigma_B^2 + \epsilon} \end{aligned}$$

$$\begin{aligned} \frac{\partial L}{\partial \mu_B} &= \sum_{i=1}^{m} \frac{\partial L}{\partial \hat{x}_i} \cdot \frac{-1}{\sqrt{\sigma_B^2 + \epsilon}} + \frac{\partial L}{\partial \sigma_B^2} \cdot \frac{\sum_{i=1}^{m} - 2(x_i - \mu_B)}{m} \\ &= \sum_{i=1}^{m} \frac{\partial L}{\partial \hat{x}_i} \cdot \frac{-1}{\sqrt{\sigma_B^2 + \epsilon}} + \sum_{i=1}^{m} \frac{\partial L}{\partial \hat{x}_i} \cdot \frac{\hat{x}_i}{\sigma_B^2 + \epsilon} \cdot \frac{\sum_{i=1}^{m} (x_i - \mu_B)}{m} \\ &= \sum_{i=1}^{m} \frac{\partial L}{\partial \hat{x}_i} \cdot \frac{1}{\sqrt{\sigma_B^2 + \epsilon}} \cdot \left(\frac{\hat{x}_i}{\sqrt{\sigma_B^2 + \epsilon}} \cdot \frac{\sum_{i=1}^{m} (x_i - \mu_B)}{m} - 1 \right) \\ &= \sum_{i=1}^{m} \frac{\partial L}{\partial \hat{x}_i} \cdot \frac{1}{\sqrt{\sigma_B^2 + \epsilon}} \cdot \left(\frac{\hat{x}_i}{\sqrt{\sigma_B^2 + \epsilon}} \cdot \left(\frac{\sum_{i=1}^{m} x_i}{m} - \frac{\sum_{i=1}^{m} \mu_B}{m} \right) - 1 \right) \\ &= -\sum_{i=1}^{m} \frac{\partial L}{\partial \hat{x}_i} \cdot \frac{1}{\sqrt{\sigma_B^2 + \epsilon}} \end{aligned}$$

$$\begin{aligned} \frac{\partial L}{\partial x_i} &= \frac{\partial L}{\partial \hat{x}_i} \cdot \frac{\partial \hat{x}_i}{\partial x_i} + \frac{\partial L}{\partial \sigma_B^2} \cdot \frac{\partial \sigma_B^2}{\partial x_i} + \frac{\partial L}{\partial \mu_B} \cdot \frac{\partial \mu_B}{\partial x_i} \\ &= \frac{\partial L}{\partial \hat{x}_i} \cdot \frac{1}{\sqrt{\sigma_B^2 + \epsilon}} + \frac{\partial L}{\partial \sigma_B^2} \cdot \frac{2(x_i - \mu_B)}{m} + \frac{\partial L}{\partial \mu_B} \cdot \frac{1}{m} \\ &= \frac{\partial L}{\partial \hat{x}_i} \cdot \frac{1}{\sqrt{\sigma_B^2 + \epsilon}} - \sum_{i=1}^{m} \frac{\partial L}{\partial \hat{x}_i} \cdot \frac{\hat{x}_i}{\sigma_B^2 + \epsilon} \cdot \frac{x_i - \mu_B}{m} - \sum_{i=1}^{m} \frac{\partial L}{\partial \hat{x}_i} \cdot \frac{1}{\sqrt{\sigma_B^2 + \epsilon}} \cdot \frac{1}{m} \\ &= \frac{1}{m \cdot \sqrt{\sigma_B^2 + \epsilon}} \cdot \left(m \frac{\partial L}{\partial \hat{x}_i} - \hat{x}_i \cdot \sum_{i=1}^{m} \frac{\partial L}{\partial \hat{x}_i} \cdot \hat{x}_i - \sum_{i=1}^{m} \frac{\partial L}{\partial \hat{x}_i} \right) \end{aligned}$$