4. (1) $E(Y|X) = 1 \cdot E(Y=1|X) + 0 \cdot E(Y=0|X) = \sigma(\beta^T X)$. $E(\sigma(X\beta)|X) = \sigma(X\beta)$

$E(Y-\sigma(X\beta)|X) = E(Y|X) - E(\sigma(X\beta)|X) = \sigma(X\beta) - \sigma(X\beta) = 0$.

In GMM, we have $E(Y-\sigma(X\beta)) = 0$.

One way to construct a GMM estimator is to choose a suitable weighting matrix and then minimize the quadratic form of the moment condition. A common choice for the weighting matrix is the identity matrix

$\frac{1}{N} \sum_{i=1}^{N} (Y_i - \sigma(X_i\beta))X_i = 0$

It's just-identified since we have k parameters and k moment conditions

(2) $L_i(\beta|y_i, x_i) = \sigma(x_i\beta)^{y_i}(1-\sigma(x_i\beta))^{1-y_i}$ $y_i = 0$ or 1.

Then $L(\beta|y, x) = \prod_{i=1}^{N} L_i(\beta|y_i, x_i) = \prod_{i=1}^{N} \sigma(x_i\beta)^{y_i}(1-\sigma(x_i\beta))^{1-y_i}$

(3) Take the log: $\log L = \sum_{i=1}^{N} (y_i \log(\sigma(x_i\beta)) + (1-y_i)\log(1-\sigma(x_i\beta)))$

$S_N(\beta) = \frac{\partial}{\partial\beta} \log L = \sum_{i=1}^{N} (y_i \frac{\partial}{\partial\beta} \log(\sigma(x_i\beta)) + (1-y_i)\frac{\partial}{\partial\beta}(\log(1-\sigma(x_i\beta)))$

$= \sum_{i=1}^{N} (y_i \frac{x_i e^{x_i\beta}}{\sigma(x_i\beta)} - (1-y_i) \cdot \frac{x_i e^{x_i\beta}}{1-\sigma(x_i\beta)})$

$= \sum_{i=1}^{N} (y_i x_i (1-\sigma(x_i\beta)) - (1-y_i)x_i\sigma(x_i\beta))$

$= \sum_{i=1}^{N} (y_i x_i - \sigma(x_i\beta)x_i)$

Then $S_N(b) = \sum_{i=1}^{N} (y_i x_i - \sigma(x_i b)x_i)$

$E[S_N(\beta)] = E[\sum_{i=1}^{N} (y_i x_i - \sigma(x_i\beta)x_i)] = \sum_{i=1}^{N} E(y_i x_i - \sigma(x_i\beta)x_i]$

$= \sum_{i=1}^{N} (x_i E(y_i) - x_i E(\sigma(x_i\beta))) = \sum_{i=1}^{N} (x_i E(y_i - \sigma(x_i\beta))) = 0$