solution typically does not exist and numerical estimates can be obtained by applying the Newton-Raphson algorithm. For example, by expanding $\mathbf{w}_n(\boldsymbol{\beta})$ around the estimating equation estimate using a stochastic Taylor series expansion and ignoring the higher order terms, we have

$$\mathbf{w}_n\left(\boldsymbol{\beta}^{(k+1)}\right) = \mathbf{w}_n\left(\boldsymbol{\beta}^{(k)}\right) + \left(\frac{\partial}{\partial \boldsymbol{\beta}}\mathbf{w}_n\left(\boldsymbol{\beta}^{(k)}\right)\right)^{\top}\left(\boldsymbol{\beta}^{(k+1)} - \boldsymbol{\beta}^{(k)}\right)$$

By setting $\mathbf{w}_n\left(\boldsymbol{\beta}^{(k+1)}\right) = 0$ and solving for $\boldsymbol{\theta}^{(k+1)}$, we obtain

$$\boldsymbol{\beta}^{(k+1)} = \boldsymbol{\beta}^{(k)} + \left[-\frac{\partial}{\partial \boldsymbol{\beta}}\mathbf{w}_n\left(\boldsymbol{\beta}^{(k)}\right)\right]^{-\top}\mathbf{w}_n\left(\boldsymbol{\beta}^{(k)}\right) \qquad (2.87)$$

where $A^{-\top} = \left(A^{-1}\right)^{\top}$ for a matrix A. Starting with some initial $\boldsymbol{\beta}^{(0)}$, we use (2.87) to iterate until some convergence criterion is reached and the limit $\widehat{\boldsymbol{\beta}}$ is the estimating equation estimate of $\boldsymbol{\beta}$. The calculation of $\frac{\partial}{\partial \boldsymbol{\beta}}\mathbf{w}_n(\boldsymbol{\beta})$ may be complicated in some cases. An alternative is to compute the estimate $\widehat{\boldsymbol{\beta}}$ using the following algorithm:

$$\boldsymbol{\beta}^{(k+1)} = \boldsymbol{\beta}^{(k)} + \left(\sum_{i=1}^{n} D_i V_i^{-1} D_i^{\top}\right)^{-1}\mathbf{w}_n\left(\boldsymbol{\beta}^{(k)}\right) \qquad (2.88)$$

While this algorithm may converge at a slower rate, it avoids the calculation of $\frac{\partial}{\partial \boldsymbol{\beta}}\mathbf{w}_n$.

Example 8. Consider the distribution-free log-linear model specified by only the systematic component in Example 3. Let $V_i = \mu_i$. The estimating equation is given by

$$\mathbf{w}_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} D_i V_i^{-1} S_i = \sum_{i=1}^{n} (y_i - \mu_i)\,\mathbf{x}_i = \mathbf{0}$$

It is readily checked (see exercise) that

$$\frac{\partial}{\partial \boldsymbol{\beta}}\mathbf{w}_n(\boldsymbol{\beta}) = -\sum_{i=1}^{n}\mathbf{x}_i\left(\frac{\partial \mu_i}{\partial \boldsymbol{\beta}}\right)^{\top} = \sum_{i=1}^{n}\mu_i \mathbf{x}_i \mathbf{x}_i^{\top} \qquad (2.89)$$

$$D_i V_i^{-1} D_i^{\top} = \mu_i \frac{\partial \mu_i}{\partial \boldsymbol{\beta}}\left(\frac{\partial \mu_i}{\partial \boldsymbol{\beta}}\right)^{\top} = \mu_i \mathbf{x}_i \mathbf{x}_i^{\top}$$

Thus, for this particular example, the two updating schemes (2.87) and (2.88) are identical.