one with the smaller variance may be preferable in terms of treatment generalizability and cost effectiveness, both of which are of great importance in effectiveness research. Also, in applications of the analysis of variance (ANOVA) model to real study data, it is critically important to check variance homogeneity to ensure valid inference, since the F-test in ANOVA for inference about difference in group means are sensitive to this assumption of variance homogeneity. However, testing for differences in variance is beyond the mean based, distribution-free GLM discussed in Chapter 2. In Chapter 3, we discussed a U-statistic based distribution-free model for testing equality of variance between two groups. With the introduction of multivariate U-statistics, we can readily extend this approach to compare both the mean and variance simultaneously between two groups, as we consider in the next example.

Example 12 (*Model for mean and variance between two groups*). Consider two i.i.d. samples y_{ki} with mean μ_k and variance σ_k^2 ($1 \le i \le n_k$, $1 \le k \le 2$). Let

$$\mathbf{h}\left(y_{1i}, y_{1j}; y_{2l}, y_{2m}\right) = \begin{pmatrix} y_{1i} \\ \frac{1}{2}\left(y_{1i} - y_{1j}\right)^2 \end{pmatrix} - \begin{pmatrix} y_{2l} \\ \frac{1}{2}\left(y_{2l} - y_{2m}\right)^2 \end{pmatrix} \tag{5.47}$$

It follows that

$$\boldsymbol{\theta} = E\left[\mathbf{h}\left(y_{1i}, y_{1j}; y_{2l}, y_{2m}\right)\right] = \begin{pmatrix} \mu_1 - \mu_2 \\ \sigma_1^2 - \sigma_2^2 \end{pmatrix}$$

Thus, the two components of $\boldsymbol{\theta}$ represent the difference in the mean and variance between the two groups. If the null $H_0 : \boldsymbol{\theta} = \mathbf{0}$ holds true, then y_{ki} have the same mean and same variance and vice versa.

Since $\mathbf{h}$ is not symmetric (e.g. $\mathbf{h}\left(y_{11}, y_{12}; y_{21}, y_{22}\right) \neq \mathbf{h}\left(y_{12}, y_{11}; y_{21}, y_{22}\right)$), we symmetrize it to obtain

$$\mathbf{g}\left(y_{1i}, y_{1j}; y_{2l}, y_{2m}\right) = \begin{pmatrix} \frac{1}{2}\left(y_{1i} + y_{1j}\right) \\ \frac{1}{2}\left(y_{1i} - y_{1j}\right)^2 \end{pmatrix} - \begin{pmatrix} \frac{1}{2}\left(y_{2l} + y_{2m}\right) \\ \frac{1}{2}\left(y_{2l} - y_{2m}\right)^2 \end{pmatrix}$$

For notational brevity, we denote $\mathbf{g}$ by $\mathbf{h}$. The U-statistic vector based on