Appendix A Variances of the Modified GMC Method

The variance-covariance matrix of the averaging method is discussed in detail by Yun (2008). The same technique can be applied for the modified GMC method with slight modification.

The variance-covariance matrix of the normalized regression coefficients is computed as $\boldsymbol{\Sigma}_{b^*} = \boldsymbol{W}\boldsymbol{\Sigma}_{B^0}\boldsymbol{W}'$ where $\boldsymbol{W}$ is a weight matrix and $\boldsymbol{\Sigma}_{B^0}$ is a reformatted variance-covariance matrix of the original regression coefficients ($\boldsymbol{\Sigma}_{B}$). The weight matrix $\boldsymbol{W}$ for the averaging method used in this study is defined as:

$$\boldsymbol{W} = \begin{bmatrix} 1 & (1/J)\cdot \mathbf{1}_{1\times J} & \mathbf{0}_{1\times K} & (1/L)\cdot \mathbf{1}_{1\times L} \\ \mathbf{0}_{J\times 1} & \boldsymbol{M}_{J\times J} & \mathbf{0}_{J\times K} & \mathbf{0}_{J\times L} \\ \mathbf{0}_{K\times 1} & \mathbf{0}_{K\times J} & \boldsymbol{I}_{K\times K} & \mathbf{0}_{J\times L} \\ \mathbf{0}_{L\times 1} & \mathbf{0}_{L\times J} & \mathbf{0}_{L\times K} & \boldsymbol{M}_{L\times L} \end{bmatrix}$$

where $\boldsymbol{M}_{J\times J} = \boldsymbol{I}_{J\times J} - (1/J)\cdot \mathbf{1}_{J\times J}$ in which $\boldsymbol{J}$ represents for the number of education variables (i.e., four) and 0 and 1 are a matrix of zeros and a matrix of ones, respectively. $\boldsymbol{I}_{K\times K}$ refers to the $\boldsymbol{K}$ by $\boldsymbol{K}$ identity matrix for two age variables. $\boldsymbol{L}$ represents the number of marriage variables (i.e., two).

The reformatted variance-covariance matrix of the regression coefficients ($\boldsymbol{\Sigma}_{B^0}$) is attained by adding zero vectors to the variance-covariance matrix of the original regression coefficients, as follows:

$$\boldsymbol{\Sigma}_{B^0} = \begin{bmatrix} \sigma_a^2 & 0 & \Sigma_{a,b'_J} & \Sigma_{a,b'_K} & 0 & \Sigma_{a,b'_L} \\ 0 & 0 & \mathbf{0}_{1\times(J-1)} & \mathbf{0}_{1\times(K-1)} & 0 & \mathbf{0}_{1\times(L-1)} \\ \Sigma_{b_J,a} & \mathbf{0}_{(J-1)\times 1} & \Sigma_{b_J,b'_J} & \Sigma_{b_J,b'_K} & \mathbf{0}_{(J-1)\times 1} & \Sigma_{b_J,b'_L} \\ \Sigma_{b_K,a} & \mathbf{0}_{(K-1)\times 1} & \Sigma_{b_K,b'_J} & \Sigma_{b_K,b'_K} & \mathbf{0}_{(K-1)\times 1} & \Sigma_{b_K,b'_L} \\ 0 & 0 & \mathbf{0}_{1\times(J-1)} & \mathbf{0}_{1\times(K-1)} & 0 & \mathbf{0}_{1\times(L-1)} \\ \Sigma_{b_L,a} & \mathbf{0}_{(L-1)\times 1} & \Sigma_{b_L,b'_J} & \Sigma_{b_L,b'_K} & \mathbf{0}_{(L-1)\times 1} & \Sigma_{b_L,b'_L} \end{bmatrix}$$

where $\boldsymbol{\sigma_a^2}$ is the residual variance and $\boldsymbol{\Sigma}$ is a partial covariance matrix. For example, $\boldsymbol{\Sigma_{b_K,b'_J}}$ is a covariance matrix between education coefficients $\boldsymbol{b_J}$ and age coefficients $\boldsymbol{b_K}$; 0 refers to a zero matrix.

For the modified GMC method, everything except the weighting matrix $\boldsymbol{W^*}$ is the same as the averaging method. The weighting matrix needs to be rebuilt as follows:

$$\boldsymbol{W^*} = \begin{bmatrix} \boldsymbol{1} & \boldsymbol{G_{1\times J}} & \boldsymbol{0_{1\times K}} & \boldsymbol{G_{1\times L}} \\ \boldsymbol{0_{J\times 1}} & \boldsymbol{N_{J\times J}} & \boldsymbol{0_{J\times K}} & \boldsymbol{0_{J\times L}} \\ \boldsymbol{0_{K\times 1}} & \boldsymbol{0_{K\times J}} & \boldsymbol{I_{K\times K}} & \boldsymbol{0_{J\times L}} \\ \boldsymbol{0_{L\times 1}} & \boldsymbol{0_{L\times J}} & \boldsymbol{0_{L\times K}} & \boldsymbol{N_{L\times L}} \end{bmatrix}$$

where $\boldsymbol{G_{1\times J}}$ refers to a $\boldsymbol{1\times J}$ column vector of grand means for education variables and $\boldsymbol{G_{1\times L}}$ refers to a $\boldsymbol{1\times L}$ column vector of grand means for marriage variables. $\boldsymbol{N_{J\times J}}$ denotes $\boldsymbol{I_{J\times J} - 1_{J\times J}\cdot D_{J\times J}}$ where $\boldsymbol{D_{J\times J}}$ refers to a diagonal matrix converted from the column vector of the grand means of education.

The new coefficients for the modified GMC method, ($\boldsymbol{B^*}$) can be obtained by taking diagonal values of $\boldsymbol{W^*B}$. The variance-covariance matrix of the new coefficients is computed as $\boldsymbol{\Sigma_{b^*} = W^*\Sigma_{B^0}W^{*\prime}}$. The variance of any predicted value $\boldsymbol{X\hat{B}}$ is estimated as $\boldsymbol{X\Sigma_{\hat{B}}X'}$. Likewise, the variance of $\boldsymbol{\bar{X}\hat{B}^*}$ can be estimated as $\boldsymbol{\bar{X}\Sigma_{\hat{B}^*}\bar{X}'}$.

Once the variance-covariance matrix of the coefficients of the modified GMC method is computed, the standard errors for the detailed decomposition can be obtained relatively easily. Using matrix notation, D1B in Equation 2 can be expressed as $\boldsymbol{\bar{X}_B\hat{B}^*_{D1B}}$. The variance of $\boldsymbol{\bar{X}_B\hat{B}^*_{D1B}}$ is therefore calculated by $\boldsymbol{\bar{X}\Sigma_{\hat{C}^*}\bar{X}}$, where $\boldsymbol{\Sigma_{\hat{C}^*}}$ is an addition of the two variance-covariance matrices of coefficients, $\boldsymbol{\Sigma_{\hat{b}^*_W} + \Sigma_{\hat{b}^*_B}}$. The variances of D2 are estimated in the same way. As we estimate the variances of the decomposed components, the t-tests of the estimated decomposition components can be conducted simply; for example, by $\boldsymbol{t_{D1B} = \frac{D1B}{\hat{\sigma}_{D1B}}}$. Note that here we assume the $\boldsymbol{\bar{X}}$'s to be fixed. if $\boldsymbol{\bar{X}}$'s are stochastic, the estimation of variances should include the variance of $\boldsymbol{\bar{X}}$ and the interaction effects (Lin 2007; Jann 2008).

References

Jann, Ben. 2008. "The BlinderOaxaca Decomposition for Linear Regression Models." *The Stata Journal* 8:453–479.

Lin, Eric S. 2007. "On the Standard Errors of Oaxaca-type Decompositions for Interindustry Gender Wage Differentials." *Economics Bulletin* 10:1–11.

Yun, Myeong-Su. 2008. "Identification Problem and Detailed Oaxaca Decomposition: A General Solution and Inference." *Journal of Economic and Social Measurement* 33:27–38.