On regression modelling with dummy variables versus separate regressions per group: comment on Holgersson et al

On regression modelling with dummy variables versus separate regressions per group: Comment on Holgersson *et al.*

Jan Schepers
Faculty of Psychology and Neuroscience
Maastricht University

Email: jan.schepers@maastrichtuniversity.nl
Address: Faculty of Psychology and Neuroscience, Maastricht University, P.O.box 616, 6200 MD Maastricht, The Netherlands

Abstract

In a recent issue of this journal, Holgersson *et al.* [11] compared the use of dummy coding in regression analysis to the use of category-wise models (i.e., estimating separate regression models for each group) with respect to estimating and testing group differences in intercept and in slope. They presented three objections against the use of dummy variables in a single regression equation, which could be overcome by the category-wise approach. In this note, I first comment on each of these three objections and next draw attention to some other issues in comparing these two approaches. This commentary further clarifies the differences and similarities between dummy variable and category-wise approaches.

Keywords: regression analysis; dummy variables; equivalency of OLS estimates; variance heterogeneity

1 Comments on objections raised by Holgersson *et al.*

1.1 Non-invariance to coding scheme

One of the objections raised by Holgersson *et al.* [11] states that the dummy variable approach is not invariant with respect to the coding of zeros and ones and therefore inferences are not invariant with respect to the choice of baseline. Consider the following regression model:

$$E(Y_i|X_i, D_i) = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 X_i D_i, \tag{1.1}$$

where Y_i $(i = 1, ..., n)$ is a quantitative response variable, X_i is a quantitative covariate, and $D_i \in \{0, 1\}$ is a dummy variable indicating group membership for observation i. Furthermore, assume that $Y_i = E(Y_i|X_i, D_i) + \epsilon_i$, where the error terms ϵ_i are $iid\, \mathcal{N}(0, \sigma^2)$ variables. Model (1.1) allows for interaction (i.e., the slope of X may differ across the two groups) and exemplifies the so-called dummy variable approach [11]. Labeling of the groups is essentially arbitrary, and therefore the regression model

$$E(Y_i|X_i, \dot{D}_i) = \beta_0^* + \beta_1^* X_i + \beta_2^* \dot{D}_i + \beta_3^* X_i \dot{D}_i, \tag{1.2}$$

where $\dot{D}_i = 1 - D_i$, describes the data equally well as model (1.1), that is, $E(Y_i|X_i, D_i) = E(Y_i|X_i, \dot{D}_i)$. In [11], models (1.1) and (1.2) were both estimated for a data set by Gujarati [6, 7] and it was found that the parameter estimates and their standard errors differ between the two models. The authors therefore concluded that inferences about the parameters are not invariant with respect to the coding scheme (i.e., the choice of baseline). Indeed, some of the individual estimates of β parameters and their standard errors depend on the coding scheme of the dummy variables. However, I wish to point out that the parameters themselves change as a function of this coding scheme. In my opinion, Holgersson *et al.* [11] did not discuss this issue thoroughly enough, although they do mention that "... the model itself is invariant to the coding of zeros and ones ...". In particular, β_0 in model (1.1) is the mean of Y in group 0 for X fixed at 0, whereas β_0^* in (1.2) is the mean of Y in group 1 for X fixed at 0. Likewise, β_1 in (1.1) is the slope of X for group 0, whereas β_1^* in (1.2) is the slope of X for group 1. The relation between the parameters of models (1.1) and (1.2) is clarified by writing model (1.2) in terms of the variable D:

$$\begin{aligned} E(Y_i|X_i, D_i) &= \beta_0^* + \beta_1^* X_i + \beta_2^*(1 - D_i) + \beta_3^* X_i(1 - D_i) \\ &= (\beta_0^* + \beta_2^*) + (\beta_1^* + \beta_3^*)X_i + (-\beta_2^*)D_i + (-\beta_3^*)X_i D_i. \end{aligned}$$

It can be seen that β_0 in model (1.1) is equal to $\beta_0^* + \beta_2^*$ in model (1.2), and so on. The same identities hold for the ordinary least squares (OLS) parameter estimates (i.e., $\hat{\beta}_0 = \hat{\beta}_0^* + \hat{\beta}_2^*$, etc.). Furthermore, one can derive the standard errors for any of the parameter estimates of model (1.1) from the covariance matrix of the parameter estimates of model (1.2), and vice versa. For instance, it holds that $\hat{\sigma}_{\hat{\beta}_0} = \hat{\sigma}_{\hat{\beta}_0^* + \hat{\beta}_2^*} = \sqrt{\hat{\sigma}^2_{\hat{\beta}_0^*} + \hat{\sigma}^2_{\hat{\beta}_2^*} + 2\hat{\sigma}_{\hat{\beta}_0^*\hat{\beta}_2^*}}$.

Hence, although the interpretation of the individual parameter estimates and corresponding hypothesis tests are not invariant with respect to the coding scheme, *inferences* about population characteristics of interest are.

1.2 Precision of point estimates

A second objection raised in [11] against the dummy variable approach states that because multicollinearity is introduced into the regression model (i.e., the dummy by covariate interaction term is correlated with the other predictor variables) the precision of the point estimates is decreased, that is, the parameter mean square error (MSE) is increased, as compared to the so-called category-wise approach. This approach implies that a separate regression model is formulated for each of the two groups:

$$E(Y_i|X_i, D_i = 0) = \gamma_0 + \gamma_1 X_i, \tag{1.3}$$

$$E(Y_i|X_i, D_i = 1) = \delta_0 + \delta_1 X_i, \tag{1.4}$$

where it is further assumed that if $D_i = 0$ then $Y_i = \gamma_0 + \gamma_1 X_i + \tau_i$, where the error terms τ_i are *iid* $\mathcal{N}(0, \sigma_0^2)$ variables, whereas if $D_i = 1$ then $Y_i = \delta_0 + \delta_1 X_i + \upsilon_i$, where the error terms υ_i are *iid* $\mathcal{N}(0, \sigma_1^2)$ variables.

Holgersson *et al.* [11] ran a simulation study to compare the two approaches in terms of Type I error rate and power for testing hypotheses on the following two *quantities of interest*: $\delta_0 - \gamma_0$ (i.e., difference between the means of the two groups at $X = 0$) and $\delta_1 - \gamma_1$ (i.e., difference between the two groups in terms of the slope of X). The corresponding parameters in model (1.1) are β_2 and β_3, respectively [2]. Interestingly, it can be shown (see Appendix A.1 and [15] for an earlier and alternative derivation) that OLS point estimates of group differences in intercept and slope are *identitical* between the two approaches, that is:

$$\hat{\beta}_2 = \hat{\delta}_0 - \hat{\gamma}_0, \tag{1.5}$$

$$\hat{\beta}_3 = \hat{\delta}_1 - \hat{\gamma}_1. \tag{1.6}$$

Hence, in terms of these quantities of interest, the second objection cannot be correct. Then why do the simulations reported in [11] show a difference between the dummy variable approach and the category-wise one in terms of parameter MSE? Let us have a look at how MSE was defined for each of the two approaches. For the dummy variable approach, Holgersson *et al.* [11] defined

MSE as the Monte Carlo average (over all simulated data sets) of the squared distances $((\hat{\beta}_0-\beta_0)^2+(\hat{\beta}_1-\beta_1)^2+(\hat{\beta}_2-\beta_2)^2+(\hat{\beta}_3-\beta_3)^2)$. For the category-wise approach, they defined MSE as the Monte Carlo average (over all simulated data sets) of the squared distances $\left([(\hat{\delta}_0-\hat{\gamma}_0)-(\delta_0-\gamma_0)]^2+[(\hat{\delta}_1-\hat{\gamma}_1)-(\delta_1-\gamma_1)]^2\right)$. In the latter definition, MSE is based on squared deviations between point estimates and population values of group differences in intercept and slope (i.e., $\delta_0-\gamma_0=\beta_2$ and $\delta_1-\gamma_1=\beta_3$). For the dummy variable approach, however, MSE is based on squared deviations between point estimates and population values of these same two quantitities of interest as well as squared deviations between point estimates and population values of the quantities β_0 (i.e., the mean of Y in group 0 for X fixed at 0) and β_1 (i.e., the slope of X for group 0). As their study is about comparing groups (either by a dummy variable approach or by a category-wise one), I do not understand why Holgersson *et al.* [11] chose to include quantitities not concerned with group comparisons in their definition of MSE for one of the two approaches. Had they compared MSE values defined only on group differences in intercept and slope, the authors would not have found any difference between the two approaches.

1.3 Homogeneity of error variance across groups

A third and final objection raised in [11] against the dummy variable approach states that it implies the assumption of homogeneous error variances across groups and therefore may lead to incorrect Type I error rates for testing hypotheses on group differences if this assumption is violated. Indeed, the simulation results reported in [11] show that for moderate violations of homogeneity, given a nominal significance level $\alpha=0.05$, the empirical Type I error rate for the dummy variable approach can be up to 0.25 or down to 0.01. That is, if homogeneity is violated, the standard errors obtained by means of the dummy variable approach are either too small or too large, depending on the relative sizes of the groups. In contrast, the results in [11] show that the empirical Type I error rate for the category-wise approach is in all situations close to the nominal level. Hence, whether error variances can be assumed homogeneous across groups is definitely an important aspect to consider when faced with choosing between an OLS dummy variable approach or an OLS category-wise one for a data set at hand. Moreover, it can be shown (see Appendix A.2) that, if one were to assume homogeneity of error variances in the category-wise approach, the standard errors of $\hat{\delta}_0 - \hat{\gamma}_0$ and $\hat{\delta}_1 - \hat{\gamma}_1$, respectively, are identical to those of $\hat{\beta}_2$ and $\hat{\beta}_2$ as obtained by means of an OLS dummy variable approach based on model (1.1).

The impact of heterogeneity of error variance on conclusions drawn by means of OLS regression is an important issue that has been studied by several authors [1, 3, 4]. Furthermore, a number of solutions which account for variance heterogeneity in testing for equality of regression slopes, including weighted least squares and Welch [17] procedures, have been studied (see e.g., [8, 10, 14, 15]).

2 Other considerations

In this section I discuss some additional matters that in some cases should be considered when comparing dummy variable approaches to category-wise ones.

2.1 Misspecification of dummy regression model

It is common in applications of regression analysis to exclude product interaction terms, even when dummy variables are involved. If population slopes are unequal across groups, the dummy regression model is in that case misspecified. It is interesting to study how the common slope parameter in the misspecified regression model is related to the population slopes of each group. Here I shall do this for the case of one dummy variable and one covariate X. Specifically, consider models (1.1) and (1.3) and (1.4), and assume $\beta_3 \neq 0$, which implies $\gamma_1 \neq \delta_1$. Now consider the following (dummy) regression model, in which the interaction term is ignored:

$$E(Y_i|X_i, D_i) = \eta_0 + \eta_1 X_i + \eta_2 D_i, \tag{2.1}$$

where η_1 is the common slope parameter of X. From (2.1) it follows that

$$\begin{aligned} \eta_1 &= \frac{E[\sigma(X,Y)|D]}{E[\sigma^2(X)|D]} \\ &= \frac{P(D=0)\sigma(X_0,Y_0) + P(D=1)\sigma(X_1,Y_1)}{P(D=0)\sigma^2(X_0) + P(D=1)\sigma^2(X_1)} \\ &= \frac{P(D=0)\sigma(X_0,Y_0)\frac{\sigma^2(X_0)}{\sigma^2(X_0)} + P(D=1)\sigma(X_1,Y_1)\frac{\sigma^2(X_1)}{\sigma^2(X_1)}}{P(D=0)\sigma^2(X_0) + P(D=1)\sigma^2(X_1)} \\ &= \frac{P(D=0)\sigma^2(X_0)\gamma_1 + P(D=1)\sigma^2(X_1)\delta_1}{P(D=0)\sigma^2(X_0) + P(D=1)\sigma^2(X_1)}. \end{aligned} \tag{2.2}$$

Hence, η_1 is a weighted average of γ_1 and δ_1, where the weights are products of relative group size and group-specific variance of the covariate. It is interesting to see that η_1 may be biased toward the slope of the smaller group if the variance of the covariate in this group is much larger than that in the larger group. Note that the point estimate $\hat{\eta}_1$, as yielded by a dummy regression analysis based on model (2.1), can alternatively be obtained by means of expression (2.2) by substituting the category-wise point estimates $\hat{\gamma}_1$ and $\hat{\delta}_1$ for γ_1 and δ_1, respectively, and n_j/n for $P(D = j)$ $(j = 0, 1)$. Furthermore, since X is considered fixed, the variances $\sigma^2(X_j)$ $(j = 0, 1)$ are to be calculated by normalizing by n_j instead of $n_j - 1$.

As a category-wise approach implies estimation of the parameters in each group separately, one may argue that incorrectly assuming equality of population slopes across groups is an issue that is not applicable to this approach.

2.2 More than two groups

By estimating a regression model in each group k ($k = 1, ..., K$) separately, as implied by a category-wise approach, each of the k regressions yield hypothesis tests that tell whether the intercept/slope of X for group k differs from 0. What is not tested is whether the intercept/slope of X differs between any pair of groups. Researchers sometimes incorrectly conclude that groups differ in the effect of X based on the slope of X being significantly different from 0 in one group but not in another. However, any conclusion about group differences requires additional tests. For example, Holgersson *et al.* [11] discuss a t test for comparing coefficients between groups k and j ($k \neq j$).

For K groups, a dummy variable approach requires a total of $K - 1$ dummy variables. For instance, for three groups, labeled 0, 1, and 2, respectively, and a single covariate X, a full model (i.e., which allows all slopes to differ) reads as follows:

$$E(Y_i|X_i, D_{i1}, D_{i2}) = \beta_0 + \beta_1 X_i + \beta_2 D_{i1} + \beta_3 D_{i2} + \beta_4 X_i D_{i1} + \beta_5 X_i D_{i2}, \quad (2.3)$$

where $D_{i1} \in \{0, 1\}$ is a first dummy variable indicating whether observation i is a member of group 1 and $D_{i2} \in \{0, 1\}$ is a second dummy variable indicating whether observation i is a member of group 2. This then implies that group 0 is the reference group. For instance, the parameter β_3 represents the difference in intercept between group 2 and group 0 (for $X = 0$), and β_5 represents the difference in slope of X between group 2 and group 0. In general, a dummy regression analysis yields hypothesis tests for all coefficients of terms involving dummy variable k ($k = 1, ..., K-1$), which tell whether the intercept/slope of X for group k differs from that of the reference group. What is not tested is whether the intercept/slope of X differs between group k and any of the other groups k' ($k' \neq k$ and $k' \neq 0$), nor whether the intercept/slope of X for group k differs from 0. To draw those conclusions, one should test either the difference in two coefficients (e.g., $H_0 : \beta_4 - \beta_5 = 0$) or the sum of two coefficients (e.g., $H_0 : \beta_1 + \beta_4 = 0$). Any of these inferences can be made by making use of the estimated coefficients of the dummy regression in question and the covariance matrix of these estimates, which are easily obtained by means of statistical software packages. It is also important to note that dummy variables do not have to be binary coded. Alternative ways of coding are contrast and effect coding [9], which alter the interpretation of the coefficients. For instance, effect coding may be used to express each group's coefficients as deviations from an overall "average", which is often of interest from a substantive point of view [16]. As this type of comparison inherently implies that at some stage of the analysis all groups are considered simultaneously, a category-wise approach is not suited to this purpose.

Finally, I wish to note that if there are three or more groups, the t test proposed by Holgersson *et al.* [11] cannot be applied to perform a single omnibus test of the null hypothesis that all slopes of X are equal. However, Welch-type tests for comparing slopes by means of a category-wise approach have been proposed by others [1, 3, 4]. More generally, Wald type, likelihood ratio type or Langrangian multiplier tests may be defined of any linear combination of parameters [12] and can therefore be applied to compare the slopes across groups. Within the dummy regression framework, an omnibus test for equality of all slopes can easily be performed by comparing full model (2.3) to a restricted model (in which all dummy by covariate product terms are omitted)

$$E(Y_i|X_i, D_{i1}, D_{i2}) = \beta_0 + \beta_1 X_i + \beta_2 D_{i1} + \beta_3 D_{i2},$$

by means of an F-statistic (see e.g., [13]). Many statistical software packages include built-in routines for performing this type of model comparison test.

3 Final Remarks

Categorical variables can be included in a regression approach by means of dummy variables. How one chooses to code these dummy variables is arbitrary, but by convention we use 0 and 1. Regardless of the coding used, the model fit and our understanding of group differences remains the same. The coding is, however, a necessary consideration in one's interpretation of the parameter estimates. While the dummy variable approach (partly) depends on the coding, the category-wise approach does not involve any binary variables and is therefore free from such concerns.

Contrary to the claim by Holgersson *et al.* [11], OLS point estimates of group comparisons as yielded by estimating model (1.1) are equivalent to differences between OLS point estimates of coefficients of models (1.3) and (1.4). Hence, precision of these point estimates cannot be an argument for choosing one approach over the other.

By using a category-wise approach, an error variance estimate is obtained in each group separately. Variances of the coefficients in group k - which are in a next step used to obtain standard errors for differences in coefficients between two independent groups - are therefore based on an estimate of the error variance in group k only. Hence, variance heterogeneity across groups is naturally dealt with by using a category-wise approach. A standard OLS dummy regression approach implies that homogeneity of error variance across groups is assumed. However, it is possible to apply the dummy regression approach within other frameworks (e.g., [8, 10, 14]) such that heterogeneity of error variance can be taken into account. One may finally note that if homogeneity of variances across groups is a valid assumption, the dummy variable approach yields a more precise estimate of the common variance than any of the separate regressions in a category-wise approach since the former is based on more degrees of freedom.

In conclusion, choosing between the two approaches considered here ultimately appears to depend on the ease with which one's research questions can be dealt with. In this regard, neither a dummy variable approach nor a category-wise one automatically yield inference tests that match every potential research question of interest. For either of the two approaches, it may therefore be necessary to do some additional calculations.

References

[1] R.A. Alexander and R.P. DeShon, *Effect of error variance heterogeneity on the power of tests for regression slope differences*, Psychol. Bull. 115(2) (1994), pp. 308-314.

[2] P.D. Allison, *Testing for interaction in multiple regression*, Am. J. Sociol. 83(1) (1977), pp. 144-153.

[3] R.P. DeShon and R.A. Alexander, *Alternative procedures for testing regression slope homogeneity when group error variances are unequal*, Psychol. Methods 1(3) (1996), pp. 261-277.

[4] B.J. Dretzke, J.R. Levin, and R.C. Serlin, *Testing for regression homogeneity under variance heterogeneity*, Psychol. Bull. 91(2) (1982), pp. 376-383.

[5] G.H. Golub and C.F. Van Loan, *Matrix computations*, The Johns Hopkins University Press, Baltimore: MD, 1996.

[6] D. Gujarati, *Use of dummy variables in testing for equality between sets of coefficients in linear regressions: A Note*, Am. Stat. 24(1) (1970a), pp. 50-52.

[7] D. Gujarati, *Use of dummy variables in testing for equality between sets of coefficients in linear regressions: A Generalization*, Am. Stat. 24(5) (1970b), pp. 18-22.

[8] J.S. Long and L.H. Ervin, *Using heteroscedasticity-consistent standard errors in the linear regression model*, Am. Stat. 54 (3) (2000), pp. 217-224.

[9] M.A. Hardy, *Regression with dummy variables*, Sage University Paper series on Quantitative Applications in the Social Sciences, Newbury Park: CA, 1993.

[10] A.F. Hayes and L. Cai, *Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation*, Behav. Res. Methods, 39(4) (2008), pp. 709-722.

[11] H.E.T. Holgersson, L. Nordström, and Ö. Öner, *Dummy variables vs. category-wise models*, J. Appl. Stat. 41(2) (2014), pp. 233-241. doi: 10.1080/02664763.2013.838665

[12] G.G. Judge, W.E. Griffiths, R. Carter Hill, H. Lütkepohl and T.-C. Lee, *The Theory and Practice of Econometrics (2nd ed.)*, Wiley, New York: NY, 1985.

[13] S.E. Maxwell and H.D. Delaney, *Designing experiments and analyzing data: A model comparison perspective*, Brooks/Cole Publishing Company, Pacific Grove: CA, 1990.

[14] R.C. Overton, *Moderated multiple regression for interactions involving categorical variables: A statistical control for heterogeneous variance across two groups*, Psychol. Methods 6(3) (2001), pp. 218-233.

[15] G. Shieh, *Detection of interactions between a dichotomous moderator and a continuous predictor in moderated multiple regression with heterogeneous error variance*, Behav. Res. Methods. 41(1) (2009), pp. 61-74. doi: 10.3758/BRM.41.1.61

[16] D.B. Suits, *Dummy variables: Mechanics v. interpretation*, Rev. Econ. Stat. 66 (1983), pp. 177-180.

[17] B.L. Welch, *On the comparison of several mean values: An alternative approach*, Biometrika 38 (191951), pp. 330-336.

A Appendices

A.1 Proof of identities (1.5) and (1.6)

Proof. Assume, without loss of generality, that the first n_0 observations belong to group 0, and the next n_1 observations belong to group 1, so that the total number of observations is $n = n_0+n_1$. Let $\mathbf{W}_0 = [\mathbf{1}_{n_0}, \mathbf{X}_0]$ and $\mathbf{W}_1 = [\mathbf{1}_{n_1}, \mathbf{X}_1]$, where $\mathbf{X}_0 = [X_1, ..., X_{n_0}]'$ and $\mathbf{X}_1 = [X_{n_0+1}, ..., X_n]'$, and $\mathbf{1}_k$ and $\mathbf{0}_k$ are k-dimensional column vectors of 1's and 0's, respectively. Similarly, let $\mathbf{Y}_0 = [Y_1, ..., Y_{n_0}]'$ and $\mathbf{Y}_1 = [Y_{n_0+1}, ..., Y_n]'$. Given this notation, it follows that the design matrix for the dummy variable approach in (1.1) is (in block matrix notation)

$$\mathbf{R} = \begin{pmatrix} \mathbf{W}_0 & \mathbf{0}_{n_0} & \mathbf{0}_{n_0} \\ \mathbf{W}_1 & \mathbf{1}_{n_1} & \mathbf{X}_1 \end{pmatrix}.$$

Finally, let $\tilde{\mathbf{R}} = [\mathbf{1}_n, \mathbf{X}] = [\mathbf{W}_0', \mathbf{W}_1']'$, $\mathbf{U}_0 = \mathbf{W}_0'\mathbf{W}_0$, $\mathbf{U}_1 = \mathbf{W}_1'\mathbf{W}_1$ and $\tilde{\mathbf{S}} = \tilde{\mathbf{R}}'\tilde{\mathbf{R}}$. We further assume that $\mathbf{U}_0$, $\mathbf{U}_1$ and $\tilde{\mathbf{S}}$ are invertible, which is the case if $\mathbf{W}_0$, $\mathbf{W}_1$ and $\tilde{\mathbf{R}}$, respectively, are of full column rank. These conditions are necessarily fulfilled if the OLS point estimates of the parameters in models (1.1), (1.3) and (1.4) are to be uniquely defined.

It then holds that

$$
\begin{aligned}
(\mathbf{R}'\mathbf{R})^{-1} &= \begin{pmatrix} \tilde{\mathbf{S}} & \mathbf{U}_1 \\ \mathbf{U}_1 & \mathbf{U}_1 \end{pmatrix}^{-1} \\
&= \begin{pmatrix} (\tilde{\mathbf{S}} - \mathbf{U}_1\mathbf{U}_1^{-1}\mathbf{U}_1)^{-1} & -\tilde{\mathbf{S}}^{-1} - \mathbf{U}_1(\mathbf{U}_1 - \mathbf{U}_1\tilde{\mathbf{S}}^{-1}\mathbf{U}_1)^{-1} \\ -\mathbf{U}_1^{-1}\mathbf{U}_1(\tilde{\mathbf{S}} - \mathbf{U}_1\mathbf{U}_1^{-1}\mathbf{U}_1)^{-1} & (\mathbf{U}_1 - \mathbf{U}_1\tilde{\mathbf{S}}^{-1}\mathbf{U}_1)^{-1} \end{pmatrix} \\
&= \begin{pmatrix} (\tilde{\mathbf{S}} - \mathbf{U}_1)^{-1} & -\tilde{\mathbf{S}}^{-1} - \mathbf{U}_1(\mathbf{U}_1 - \mathbf{U}_1\tilde{\mathbf{S}}^{-1}\mathbf{U}_1)^{-1} \\ -(\tilde{\mathbf{S}} - \mathbf{U}_1)^{-1} & (\mathbf{U}_1 - \mathbf{U}_1\tilde{\mathbf{S}}^{-1}\mathbf{U}_1)^{-1} \end{pmatrix}, \qquad \text{(A.1)}
\end{aligned}
$$

where the inverse has been solved by making use of block matrix inversion. Only the lower two 2×2 submatrices in (A.1) are of relevance for this proof. Firstly, one may observe that

$$
\begin{aligned}
-(\tilde{\mathbf{S}} - \mathbf{U}_1)^{-1} &= -\left(\begin{pmatrix} n & \mathbf{1}_n'\mathbf{X} \\ \mathbf{1}_n'\mathbf{X} & \mathbf{X}'\mathbf{X} \end{pmatrix} - \begin{pmatrix} n_1 & \mathbf{1}_{n_1}'\mathbf{X}_1 \\ \mathbf{1}_{n_1}'\mathbf{X}_1 & \mathbf{X}_1'\mathbf{X}_1 \end{pmatrix} \right)^{-1} \\
&= -\begin{pmatrix} n_0 & \mathbf{1}_{n_0}'\mathbf{X}_0 \\ \mathbf{1}_{n_0}'\mathbf{X}_0 & \mathbf{X}_0'\mathbf{X}_0 \end{pmatrix}^{-1} \\
&= -\mathbf{U}_0^{-1} \\
&= -(\mathbf{W}_0'\mathbf{W}_0)^{-1}. \qquad \text{(A.2)}
\end{aligned}
$$

Secondly, by making use of the matrix inversion lemma [5], we have

$$
\begin{aligned}
(\mathbf{U}_1 - \mathbf{U}_1\tilde{\mathbf{S}}^{-1}\mathbf{U}_1)^{-1} &= \mathbf{U}_1^{-1} + (\mathbf{I} - \mathbf{U}_1^{-1}\mathbf{U}_1\tilde{\mathbf{S}}^{-1}\mathbf{U}_1)^{-1}\mathbf{U}_1^{-1}\mathbf{U}_1\tilde{\mathbf{S}}^{-1}\mathbf{U}_1\mathbf{U}_1^{-1} \\
&= \mathbf{U}_1^{-1} + (\mathbf{I} - \tilde{\mathbf{S}}^{-1}\mathbf{U}_1)^{-1}\tilde{\mathbf{S}}^{-1} \\
&= \mathbf{U}_1^{-1} + (\tilde{\mathbf{S}} - \mathbf{U}_1)^{-1} \\
&= \mathbf{U}_1^{-1} + \mathbf{U}_0^{-1} \\
&= (\mathbf{W}_1'\mathbf{W}_1)^{-1} + (\mathbf{W}_0'\mathbf{W}_0)^{-1}. \qquad \text{(A.3)}
\end{aligned}
$$

The OLS point estimates $\hat{\boldsymbol{\gamma}} = (\hat{\gamma}_0, \hat{\gamma}_1)'$ and $\hat{\boldsymbol{\delta}} = (\hat{\delta}_0, \hat{\delta}_1)'$ for models (1.3) and (1.4), respectively, are $(\mathbf{W}_0'\mathbf{W}_0)^{-1}\mathbf{W}_0'\mathbf{Y}_0$ and $(\mathbf{W}_1'\mathbf{W}_1)^{-1}\mathbf{W}_1'\mathbf{Y}_1$. For the OLS point estimate $\hat{\boldsymbol{\beta}} = (\mathbf{R}'\mathbf{R})^{-1}\mathbf{R}'\mathbf{Y}$ of the parameters in model (1.1), making use of (A.1)-(A.3), we now have (focussing only on the relevant submatrices):

$$
\begin{aligned}
\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} &= \begin{pmatrix} \cdots & \cdots \\ -(\mathbf{W}_0'\mathbf{W}_0)^{-1} & (\mathbf{W}_0'\mathbf{W}_0)^{-1} + (\mathbf{W}_1'\mathbf{W}_1)^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{W}_0'\mathbf{Y}_0 + \mathbf{W}_1'\mathbf{Y}_1 \\ \mathbf{W}_1'\mathbf{Y}_1 \end{pmatrix} \\
&= \begin{pmatrix} \cdots \\ -(\mathbf{W}_0'\mathbf{W}_0)^{-1}\mathbf{W}_0'\mathbf{Y}_0 + (\mathbf{W}_1'\mathbf{W}_1)^{-1}\mathbf{W}_1'\mathbf{Y}_1 \end{pmatrix} \\
&= \begin{pmatrix} \cdots \\ \cdots \\ \hat{\delta}_0 - \hat{\gamma}_0 \\ \hat{\delta}_1 - \hat{\gamma}_1 \end{pmatrix}.
\end{aligned}
$$

□

A.2 Equivalence of standard errors of point estimates of group differences if homogeneity of error variance is assumed in both approaches

Consider the variance of the category-wise point estimate $\hat{\boldsymbol{\delta}} - \hat{\boldsymbol{\gamma}} = (\hat{\delta}_0 - \hat{\gamma}_0, \hat{\delta}_1 - \hat{\gamma}_1)'$, as proposed by Holgersson *et al.* [11]. It is the diagonal of

$$\mathbf{V} = \hat{\sigma}_0^2(\mathbf{W}_0'\mathbf{W}_0)^{-1} + \hat{\sigma}_1^2(\mathbf{W}_1'\mathbf{W}_1)^{-1}, \tag{A.4}$$

where $\hat{\sigma}_0^2 = (n_0 - 2)^{-1}(\mathbf{Y}_0 - \mathbf{W}_0\hat{\boldsymbol{\gamma}})'(\mathbf{Y}_0 - \mathbf{W}_0\hat{\boldsymbol{\gamma}})$ and $\hat{\sigma}_1^2 = (n_1 - 2)^{-1}(\mathbf{Y}_1 - \mathbf{W}_1\hat{\boldsymbol{\delta}})'(\mathbf{Y}_1 - \mathbf{W}_1\hat{\boldsymbol{\delta}})$. Thus, the variance of $\hat{\boldsymbol{\delta}} - \hat{\boldsymbol{\gamma}}$ equals the sum of the variances of the point estimates $\hat{\boldsymbol{\delta}}$ and $\hat{\boldsymbol{\gamma}}$, which are obtained in each group separately (note that the groups are independent). They are therefore based on group-specific estimates of the error variance (i.e., $\hat{\sigma}_0^2$ and $\hat{\sigma}_1^2$, respectively). On the other hand, the variance of $\hat{\boldsymbol{\beta}}$ as obtained by means of the dummy variable approach is the diagonal of $\tilde{\mathbf{V}} = \hat{\sigma}^2(\mathbf{R}'\mathbf{R})^{-1}$, where $\hat{\sigma}^2 = (n-4)^{-1}(\mathbf{Y} - \mathbf{R}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{R}\hat{\boldsymbol{\beta}}) = [(n_0 - 2)\hat{\sigma}_0^2 + (n_1 - 2)\hat{\sigma}_1^2]/(n-4)$ is a pooled estimate of the error variance. Let $\tilde{\mathbf{V}}^*$ be the lower right 2×2 submatrix of $\tilde{\mathbf{V}}$. Making use of (A.3) it follows that $\tilde{\mathbf{V}}^*$, which contains the variances of $\hat{\beta}_2$ and $\hat{\beta}_3$ on its diagonal, can be written as follows:

$$\tilde{\mathbf{V}}^* = \hat{\sigma}^2(\mathbf{W}_0'\mathbf{W}_0)^{-1} + \hat{\sigma}^2(\mathbf{W}_1'\mathbf{W}_1)^{-1}. \tag{A.5}$$

Hence, if in the OLS category-wise approach based on models (1.3) and (1.4) one were to assume homogeneity of error variances across groups, (A.4) would simplify to (A.5).