Marginal effects in the probit model with a triple dummy variable interaction term

Thomas Cornelißen and Katja Sonderhof*

Leibniz Universität Hannover, Discussion Paper No. 386

January 2008

ISSN: 0949–9962

Abstract

In non-linear regression models, such as the probit model, coefficients cannot be interpreted as marginal effects. The marginal effects are usually non-linear combinations of all regressors and regression coefficients of the model. This paper derives the marginal effects in a probit model with a triple dummy variable interaction term. A frequent application of this model is the regression-based difference-in-difference-in-differences estimator with a binary outcome variable. The formulae derived here are implemented in a Stata program called `inteff3` which applies the delta method in order to compute also the standard errors of the marginal effects.

Keywords: difference-in-difference-in-differences, probit model, interaction terms, marginal effects, Stata

JEL-Classification: C25, C87

*Institute of Empirical Economics and Institute of Labour Economics, Leibniz Universität Hannover, Germany, cornelissen@ewifo.uni-hannover.de, sonderhof@aoek.uni-hannover.de

1 Introduction

Regression analysis usually aims at estimating the marginal effect of a regressor on the outcome variable controlling for the influence of other regressors. In the linear regression model the regression coefficients can be interpreted as marginal effects. In non-linear regression models, such as the probit model, coefficients cannot be interpreted as marginal effects. The marginal effect of a regressor is given by the partial derivative of the expected value of the outcome variable with respect to the regressor.

When an interaction term of two variables is included in the model, the interaction effect of the two variables is given by the cross-partial derivative of the expectation of the dependent variable with respect to the two interacted variables. In a linear model this is simply the coefficient on the interaction term. In a non-linear model, the cross-derivative is usually a non-linear combination of all regressors and all coefficients of the model. Ai and Norton (2003) and Norton et al. (2004) derive the formulae of interaction effects of two interacted variables in a logit and probit model.

In this paper we look at the case of a triple dummy variable interaction in a probit model. A common application of a model with three interacted dummy variables is the difference-in-difference-in-differences (DDD) estimator (Gruber 1994). When the dependent variable is binary, the regression based DDD model can be estimated as a probit model with a triple dummy variable interaction term (Gruber and Poterba 1994). We derive the marginal effects in this model in an analogous way as Ai and Norton (2003) and Norton et al. (2004). The standard errors of the marginal effects can be computed using the delta method (see e.g. Davidson/MacKinnon 2004, p.202). We implemented the computation of the marginal effect and their standard errors in a downloadable Stata program `inteff3` (Cornelißen and Sonderhof 2008)[1].

The paper proceeds as follows. Section 2 derives the marginal effects of the three dummy variables and their interactions in a probit model. Section 3 describes the Stata ado-file `inteff3` and presents a short empirical application. Section 4 concludes.

[1]Type *net search inteff3* in Stata or visit http://ideas.repec.org/c/boc/bocode/s456903.html. The program requires at least Stata version 9.

2 The marginal interaction effects in a probit model with a triple dummy variable interaction term

The probit model with a triple dummy variable interaction term is

$$\begin{aligned} P(y=1|x_1,x_2,x_3,\tilde{\mathbf{x}}) &= \Phi(\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 \\ &\quad + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_1 x_2 x_3 + \tilde{\mathbf{x}}\tilde{\beta}) \\ &= \Phi(\mathbf{x}\beta) \end{aligned} \tag{1}$$

where subscripts for observations are dropped for simplicity, y is the binary dependent variable, Φ is the standard normal cumulative density function, x_1, x_2 and x_3 are dummy variables to be interacted, β_j are the associated coefficients, and $\tilde{\mathbf{x}}\tilde{\beta}$ denotes the linear combination of all remaining explanatory variables and coefficients.

Taking into account that the dummy variables x_1, x_2 and x_3 and their interactions are discrete variables, the marginal effects of interest are the following:

$$\begin{aligned} g_1 = \frac{\Delta\Phi(\mathbf{x}\beta)}{\Delta x_1} &= \Phi(\beta_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_2 + \beta_{13} x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_2 x_3 + \tilde{\mathbf{x}}\tilde{\beta}) \\ &\quad - \Phi(\beta_2 x_2 + \beta_3 x_3 + \beta_{23} x_2 x_3 + \tilde{\mathbf{x}}\tilde{\beta}) \end{aligned} \tag{2}$$

$$\begin{aligned} g_2 = \frac{\Delta\Phi(\mathbf{x}\beta)}{\Delta x_2} &= \Phi(\beta_1 x_1 + \beta_2 + \beta_3 x_3 + \beta_{12} x_1 + \beta_{13} x_1 x_3 + \beta_{23} x_3 + \beta_{123} x_1 x_3 + \tilde{\mathbf{x}}\tilde{\beta}) \\ &\quad - \Phi(\beta_1 x_1 + \beta_3 x_3 + \beta_{13} x_1 x_3 + \tilde{\mathbf{x}}\tilde{\beta}) \end{aligned} \tag{3}$$

$$\begin{aligned} g_3 = \frac{\Delta\Phi(\mathbf{x}\beta)}{\Delta x_3} &= \Phi(\beta_1 x_1 + \beta_2 x_2 + \beta_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 + \beta_{23} x_2 + \beta_{123} x_1 x_2 + \tilde{\mathbf{x}}\tilde{\beta}) \\ &\quad - \Phi(\beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \tilde{\mathbf{x}}\tilde{\beta}) \end{aligned} \tag{4}$$

$$\begin{aligned} g_4 = \frac{\Delta^2\Phi(\mathbf{x}\beta)}{\Delta x_1 \Delta x_2} &= \Phi(\beta_1 + \beta_2 + \beta_3 x_3 + \beta_{12} + \beta_{13} x_3 + \beta_{23} x_3 + \beta_{123} x_3 + \tilde{\mathbf{x}}\tilde{\beta}) \\ &\quad - \Phi(\beta_1 + \beta_3 x_3 + \beta_{13} x_3 + \tilde{\mathbf{x}}\tilde{\beta}) - \Phi(\beta_2 + \beta_3 x_3 + \beta_{23} x_3 + \tilde{\mathbf{x}}\tilde{\beta}) \\ &\quad + \Phi(\beta_3 x_3 + \tilde{\mathbf{x}}\tilde{\beta}) \end{aligned} \tag{5}$$

$$\begin{aligned} g_5 = \frac{\Delta^2\Phi(\mathbf{x}\beta)}{\Delta x_1 \Delta x_3} &= \Phi(\beta_1 + \beta_2 x_2 + \beta_3 + \beta_{12} x_2 + \beta_{13} + \beta_{23} x_2 + \beta_{123} x_2 + \tilde{\mathbf{x}}\tilde{\beta}) \\ &\quad - \Phi(\beta_1 + \beta_2 x_2 + \beta_{12} x_2 + \tilde{\mathbf{x}}\tilde{\beta}) - \Phi(\beta_2 x_2 + \beta_3 + \beta_{23} x_2 + \tilde{\mathbf{x}}\tilde{\beta}) \\ &\quad + \Phi(\beta_2 x_2 + \tilde{\mathbf{x}}\tilde{\beta}) \end{aligned} \tag{6}$$

$$g_6 = \frac{\Delta^2\Phi(\mathbf{x}\beta)}{\Delta x_2 \Delta x_3} = \Phi(\beta_1 x_1 + \beta_2 + \beta_3 + \beta_{12}x_1 + \beta_{13}x_1 + \beta_{23} + \beta_{123}x_1 + \tilde{\mathbf{x}}\tilde{\beta})$$
$$-\Phi(\beta_1 x_1 + \beta_2 + \beta_{12}x_1 + \tilde{\mathbf{x}}\tilde{\beta}) - \Phi(\beta_1 x_1 + \beta_3 + \beta_{13}x_1 + \tilde{\mathbf{x}}\tilde{\beta})$$
$$+\Phi(\beta_1 x_1 + \tilde{\mathbf{x}}\tilde{\beta}) \quad (7)$$

$$g_7 = \frac{\Delta^3\Phi(\mathbf{x}\beta)}{\Delta x_1 \Delta x_2 \Delta x_3} = \Phi(\beta_1 + \beta_2 + \beta_3 + \beta_{12} + \beta_{13} + \beta_{23} + \beta_{123} + \tilde{\mathbf{x}}\tilde{\beta})$$
$$-\Phi(\beta_1 + \beta_2 + \beta_{12} + \tilde{\mathbf{x}}\tilde{\beta}) - \Phi(\beta_1 + \beta_3 + \beta_{13} + \tilde{\mathbf{x}}\tilde{\beta})$$
$$-\Phi(\beta_2 + \beta_3 + \beta_{23} + \tilde{\mathbf{x}}\tilde{\beta}) + \Phi(\beta_3 + \tilde{\mathbf{x}}\tilde{\beta})$$
$$+\Phi(\beta_2 + \tilde{\mathbf{x}}\tilde{\beta}) + \Phi(\beta_1 + \tilde{\mathbf{x}}\tilde{\beta}) - \Phi(\tilde{\mathbf{x}}\tilde{\beta}) \quad (8)$$

With given estimates of the coefficients of the probit model, $\hat{\beta}$, (2)-(8) can be used to derive estimates of the marginal effects. As the marginal effects $\hat{g}_1$ through $\hat{g}_7$ are non-linear functions of the underlying parameters estimates $\hat{\beta}$, their standard errors can be computed using the delta method (see e.g. Davidson/MacKinnon 2004, p.202). Let $\hat{\mathbf{g}}(\hat{\beta})$ be a column vector of the marginal effects represented by (2)-(8). Then, for the given estimated covariance matrix of the probit regression coefficients, $\hat{\mathbf{V}}(\hat{\beta})$, the covariance matrix of $\hat{\mathbf{g}}$, can be estimated according to the delta method by

$$\hat{V}(\hat{\mathbf{g}}) = \hat{\mathbf{G}}\hat{\mathbf{V}}(\hat{\beta})\hat{\mathbf{G}}', \quad (9)$$

where $\hat{\mathbf{G}} \equiv \mathbf{G}(\hat{\beta})$ is the matrix $\partial\mathbf{g}(\beta)/\partial\beta'$. The ith row of $\mathbf{G}(\hat{\beta})$ is the vector of partial derivatives of the ith function with respect to $\hat{\beta}'$ or, in other words, the typical element in row i and column j of $\mathbf{G}(\hat{\beta})$ is $\partial g_i(\beta)/\partial\beta_j$ (Davidson/MacKinnon 2004, p. 208).
Hence, the method requires the derivatives of the functions g_1 through g_7 with respect to the underlying probit coefficients β. These derivatives are presented in the appendix.
We have implemented the computation of the marginal effects and their standard errors in the Stata program `inteff3`. The program computes marginal effects at means, at values specified by the user, or the average marginal effects, which are computed by averaging over the marginal effects for each observation in the sample.

3 The Stata ado-file inteff3 and an empirical application

We illustrate the computation of the interaction effect of a triple dummy variable interaction term in a probit model in a regression with simulated data on 100,000 observations. Our data generating process mimics a typical evaluation context of a reform, where the effect of the reform on a binary dependent variable is to be estimated by a DDD estimator. As regressors we generate three dummy variables `after`, `treated` and `group` and a continuous regressor `x`. The covariances between the generated regressors are 0. The continuous regressor has mean 0 and standard deviation $\sqrt{2}$. The dummy variables `after`, `treated` and `group` have mean values of .5, .5 and .7. We think of the dummy variable `after` as being 0 before the reform, and 1 after the reform. The variable `treated` is 1 for the treatment group (say a certain region or group of individuals to which the reform applies) and 0 for the control group. The dummy `group` could be for example a male/female dummy in case we want to estimate separate treatment effects for one of these groups. We abbreviate the interactions in the following way: `a_t=after*treated`, `a_g=after*group`, `t_g=treated*group` and `atg=after*treated*group`. Further we draw a standard normally distributed error term u and implement the following data generating process:

$$y^* = 1 + .5 \cdot \text{after} + .2 \cdot \text{treated} + .5 \cdot \text{group} + .8 \cdot \text{a\_t} + .3 \cdot \text{a\_g} + .1 \cdot \text{t\_g} + .5 \cdot \text{atg} - 0.7 \cdot \text{x} + u$$

$$\text{and } y = \begin{cases} 1 & \text{if } y^* > 0 \\ 0 & \text{if } y^* \leq 0. \end{cases}$$

A different way to write down the model is

$$P(y=1|X) = \Phi(1 + .5 \cdot \text{after} + .2 \cdot \text{treated} + .5 \cdot \text{group} + .8 \cdot \text{a\_t} + .3 \cdot \text{a\_g} + .1 \cdot \text{t\_g} + .5 \cdot \text{atg} - 0.7 \cdot \text{x}).$$

Based on this model we can compute the treatment effect by applying the triple difference given in (8). This gives us a true treatment effect of -.0372. Running a probit model on the data gives:

```
Iteration 7:   log likelihood = -20269.238
Probit regression                                 Number of obs   =     100000
                                                  LR chi2(8)      =   23608.08
                                                  Prob > chi2     =     0.0000
Log likelihood = -20269.238                       Pseudo R2       =     0.3680
```

```
------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       after |   .5069591   .0287085    17.66   0.000     .4506915    .5632267
     treated |   .2041092   .0271926     7.51   0.000     .1508127    .2574057
       group |   .4745062   .0234567    20.23   0.000      .428532    .5204805
         a_t |   .8145441   .0486727    16.74   0.000     .7191473    .9099408
         a_g |   .3207281   .0374531     8.56   0.000     .2473214    .3941348
         t_g |   .0706887   .0342361     2.06   0.039     .0035872    .1377903
         atg |   .5811146   .0885926     6.56   0.000     .4074763    .7547529
           x |   -.700084   .0066222  -105.72   0.000    -.7130632   -.6871047
       _cons |   1.018685   .0192915    52.80   0.000     .9808741    1.056495
------------------------------------------------------------------------------
Note: 0 failures and 334 successes completely determined.
```

A naive way of estimating the marginal effects at means would be using Stata's dprobit:

```
Iteration 7:   log likelihood = -20269.238
Probit regression, reporting marginal effects           Number of obs = 100000
                                                        LR chi2(8)    =23608.08
                                                        Prob > chi2   = 0.0000
Log likelihood = -20269.238                             Pseudo R2     = 0.3680

------------------------------------------------------------------------------
       y |      dF/dx   Std. Err.      z    P>|z|     x-bar  [    95% C.I.   ]
---------+--------------------------------------------------------------------
  after*|   .0209206    .001422    17.66   0.000         .5   .018134  .023708
treated*|    .008161   .0011298     7.51   0.000         .5   .005947  .010375
  group*|   .0237043   .0016065    20.23   0.000         .7   .020556  .026853
    a_t*|    .023044   .0012749    16.74   0.000     .25023   .020545  .025543
    a_g*|   .0116574   .0013007     8.56   0.000     .34877   .009108  .014207
    t_g*|   .0027468    .001304     2.06   0.039      .3485   .000191  .005303
    atg*|   .0160393   .0014806     6.56   0.000     .17368   .013137  .018941
      x |  -.0278173   .0008487  -105.72   0.000   -.000312  -.029481 -.026154
---------+--------------------------------------------------------------------
  obs. P |     .90197
 pred. P |   .9841367  (at x-bar)
------------------------------------------------------------------------------
(*) dF/dx is for discrete change of dummy variable from 0 to 1
    z and P>|z| correspond to the test of the underlying coefficient being 0
```

Here we get a positive and statistically significant marginal effect at means of the triple dummy variable interaction term of about .016. However, this effect is equal to

$$\frac{\Delta^3\Phi(\mathbf{x}\beta)}{\Delta(x_1x_2x_3)} = \Phi(\beta_1+\beta_2+\beta_3+\beta_{12}+\beta_{13}+\beta_{23}+\beta_{123}+\tilde{\mathbf{x}}\tilde{\beta}) - \Phi(\beta_1+\beta_2+\beta_3+\beta_{12}+\beta_{13}+\beta_{23}+\tilde{\mathbf{x}}\tilde{\beta}), \quad (10)$$

which is different from the treatment effect g_7 given in (8) which we are interested in. There is no guarantee that (8) and (10) are of equal sign and we see that the effect estimated by `dprobit` is of opposite sign than the true treatment effect of -.0372.

We can use the program `inteff3` in order to estimate the treatment effects at means we are interested in correctly. (Before being able to use `inteff3` we must re-estimate the original probit model, because `inteff3` can only run after `probit` and not after `dprobit`.)

```
. inteff3
The 7 Diff in Diff in Diff Dummies must be the first dependent variables of the
preceding probit model.
They must be in the order x, y, z, x*y, x*z, y*z, x*y*z.

Diff in Diff in Diff Dummies are: after, treated, group, a_t, a_g, t_g, atg.
Control variables are: x , constant term.

Marginal effect at means of probit estimation sample:
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       after |   .0673222   .0013512    49.82   0.000     .0646738    .0699706
     treated |   .0380835   .0010632    35.82   0.000     .0359997    .0401673
       group |   .0489316   .0015462    31.65   0.000     .0459011    .0519622
         a_t |   -.015593   .0023681    -6.58   0.000    -.0202343   -.0109517
         a_g |  -.0595704   .0032588   -18.28   0.000    -.0659576   -.0531832
         t_g |  -.0455653   .0029802   -15.29   0.000    -.0514064   -.0397243
         atg |  -.0336326   .0068336    -4.92   0.000    -.0470262   -.0202389
------------------------------------------------------------------------------
```

In fact we now get a negative and statistically significant treatment effect of -0.0336 much closer to the true treatment effect of -.0372. Instead of computing the effects at means we compute the marginal effects for each individual in the sample and then compute its average. According to Greene (2003, p. 668) this is more advisable than just computing the effect at means. This is possible with `inteff3` by specifying:

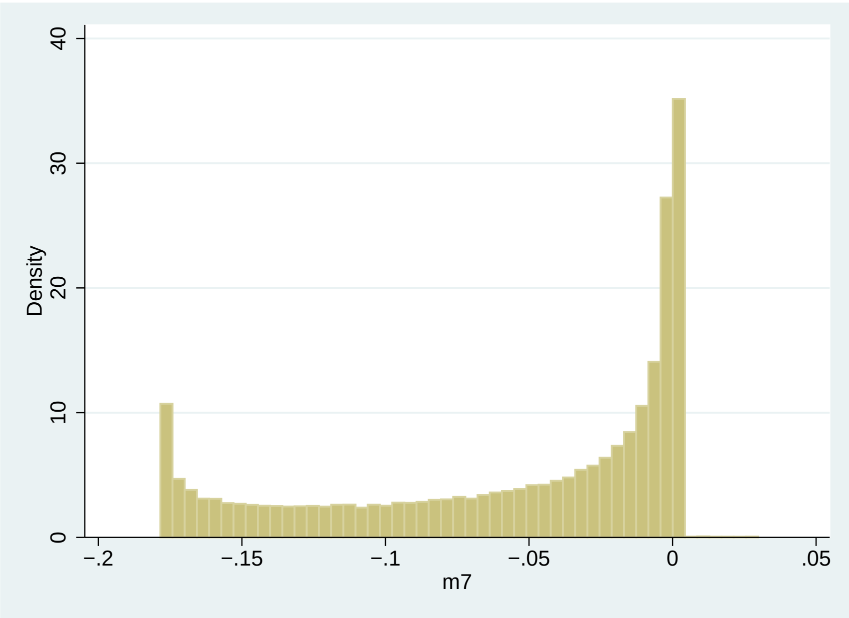
```
. inteff3, average me(m1 m2 m3 m4 m5 m6 m7)
The 7 Diff in Diff in Diff Dummies must be the first dependent variables of the
preceding probit model.
They must be in the order x, y, z, x*y, x*z, y*z, x*y*z.

Diff in Diff in Diff Dummies are: after, treated, group, a_t, a_g, t_g, atg.
Control variables are: x , constant term.
Average marginal effect:
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       after |   .1066692   .0015556    68.57   0.000     .1036202    .1097182
     treated |   .0524024   .0015524    33.76   0.000     .0493597     .055445
       group |   .0750136   .0018947    39.59   0.000        .0713    .0787272
         a_t |   .0238329   .0031032     7.68   0.000     .0177507    .0299151
         a_g |  -.0273591   .0037947    -7.21   0.000    -.0347967   -.0199216
         t_g |   -.029282   .0037815    -7.74   0.000    -.0366936   -.0218703
         atg |  -.0539799   .0075763    -7.12   0.000    -.0688291   -.0391307
------------------------------------------------------------------------------
```

Here we get a quantitatively quite different estimate of -0.054. Again there is no guarantee that the treatment effect computed at means and the average treatment effect have the same sign. Both are computed by applying equation (8) but they are computed using different values for the regressors.

A more complete description of the sample distribution of the estimated treatment effect than just reporting the average would be to report quantiles or to graph the distribution of the treatment effect. The option `me(m1 m2 m3 m4 m5 m6 m7)` of `inteff3` saves the effects for each individual as a variable and allows to describe or graph their distribution. For the effect $\hat{g}_7$ of the given estimation the distribution has a median of -.0305 is graphically represented by the following histogram.

```
histogram m7
```



4 Conclusion

This paper has derived the marginal effects in a probit model with a triple dummy variable interaction term. A frequent application of this model is the regression-based difference-in-difference-in-differences estimator with a binary outcome variable.

The computation of the marginal effects and their standard errors has been implemented in the Stata program `inteff3` which applies the delta method in order to compute also the standard errors of the marginal effects. In an empirical application we have demonstrated the theoretical result that the sign and the significance of the interaction effects can be very different from the sign and significance of the simple marginal effect of the interaction term in the probit model.

In an analogous way as presented here and as presented in Ai and Norton (2003) and Norton et al. (2004), the effects can be computed for the case of an interaction of three continuous variables or for a mixture of continuous and dummy variables.

References

[1] Ai, C. and E. C. Norton (2003): Interaction terms in logit and probit models, *Economics Letters*, 80, pp. 123-129.

[2] Cornelißen, T. and K. Sonderhof (2008). inteff3: Stata module to compute marginal effects in a probit model with a triple dummy variable interaction term. http://ideas.repec.org/c/boc/bocode/s456903.html.

[3] Davidson, R. and J. MacKinnon (2004): *Econometric Theory and Methods*, Oxford University Press, New York.

[4] Greene, W. (2003): *Econometric Analysis*, Prentice Hall, New Jersey.

[5] Gruber, J. (1994): The Incidence of Mandated Maternity Benefits, *The American Economic Review*, Vol. 84, No. 3, pp. 622-641.

[6] Gruber, J. and J. Poterba (1994): Tax Incentives and the Decision to Purchase Health Insurance: Evidence from the Self-Employed, *The Quarterly Journal of Economics*, Vol. 109, No. 3, pp. 701-733.

[7] Norton, E. C., H. Wang and C. Ai (2004): Computing interaction effects and standard errors in logit and probit models, *The Stata Journal*, 4(2), 154-167.

Appendix

$$
\begin{aligned}
\frac{\partial g_1}{\partial \beta_1} &= \phi(\beta_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_2 + \beta_{13} x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_2 x_3 + \tilde{\mathbf{x}}\tilde{\beta}) \\
\frac{\partial g_1}{\partial \beta_2} &= (\phi(\beta_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_2 + \beta_{13} x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_2 x_3 + \tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_2 x_2 + \beta_3 x_3 + \beta_{23} x_2 x_3 + \tilde{\mathbf{x}}\tilde{\beta}))x_2 \\
\frac{\partial g_1}{\partial \beta_3} &= (\phi(\beta_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_2 + \beta_{13} x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_2 x_3 + \tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_2 x_2 + \beta_3 x_3 + \beta_{23} x_2 x_3 + \tilde{\mathbf{x}}\tilde{\beta}))x_3 \\
\frac{\partial g_1}{\partial \beta_{12}} &= \phi(\beta_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_2 + \beta_{13} x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_2 x_3 + \tilde{\mathbf{x}}\tilde{\beta})x_2 \\
\frac{\partial g_1}{\partial \beta_{13}} &= \phi(\beta_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_2 + \beta_{13} x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_2 x_3 + \tilde{\mathbf{x}}\tilde{\beta})x_3 \\
\frac{\partial g_1}{\partial \beta_{23}} &= (\phi(\beta_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_2 + \beta_{13} x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_2 x_3 + \tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_2 x_2 + \beta_3 x_3 + \beta_{23} x_2 x_3 + \tilde{\mathbf{x}}\tilde{\beta}))x_2 x_3 \\
\frac{\partial g_1}{\partial \beta_{123}} &= \phi(\beta_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_2 + \beta_{13} x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_2 x_3 + \tilde{\mathbf{x}}\tilde{\beta})x_2 x_3 \\
\frac{\partial g_1}{\partial \beta_j} &= (\phi(\beta_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_2 + \beta_{13} x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_2 x_3 + \tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_2 x_2 + \beta_3 x_3 + \beta_{23} x_2 x_3 + \tilde{\mathbf{x}}\tilde{\beta}))x_j \\
\frac{\partial g_2}{\partial \beta_1} &= (\phi(\beta_1 x_1 + \beta_2 + \beta_3 x_3 + \beta_{12} x_1 + \beta_{13} x_3 x_1 + \beta_{23} x_3 + \beta_{123} x_3 x_1 + \tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_1 x_1 + \beta_3 x_3 + \beta_{13} x_3 x_1 + \tilde{\mathbf{x}}\tilde{\beta}))x_1 \\
\frac{\partial g_2}{\partial \beta_2} &= \phi(\beta_1 x_1 + \beta_2 + \beta_3 x_3 + \beta_{12} x_1 + \beta_{13} x_3 x_1 + \beta_{23} x_3 + \beta_{123} x_3 x_1 + \tilde{\mathbf{x}}\tilde{\beta}) \\
\frac{\partial g_2}{\partial \beta_3} &= (\phi(\beta_1 x_1 + \beta_2 + \beta_3 x_3 + \beta_{12} x_1 + \beta_{13} x_3 x_1 + \beta_{23} x_3 + \beta_{123} x_3 x_1 + \tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_1 x_1 + \beta_3 x_3 + \beta_{13} x_3 x_1 + \tilde{\mathbf{x}}\tilde{\beta}))x_3 \\
\frac{\partial g_2}{\partial \beta_{12}} &= \phi(\beta_1 x_1 + \beta_2 + \beta_3 x_3 + \beta_{12} x_1 + \beta_{13} x_3 x_1 + \beta_{23} x_3 + \beta_{123} x_3 x_1 + \tilde{\mathbf{x}}\tilde{\beta})x_1 \\
\frac{\partial g_2}{\partial \beta_{13}} &= (\phi(\beta_1 x_1 + \beta_2 + \beta_3 x_3 + \beta_{12} x_1 + \beta_{13} x_3 x_1 + \beta_{23} x_3 + \beta_{123} x_3 x_1 + \tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_1 x_1 + \beta_3 x_3 + \beta_{13} x_3 x_1 + \tilde{\mathbf{x}}\tilde{\beta}))x_1 x_3 \\
\frac{\partial g_2}{\partial \beta_{23}} &= \phi(\beta_1 x_1 + \beta_2 + \beta_3 x_3 + \beta_{12} x_1 + \beta_{13} x_3 x_1 + \beta_{23} x_3 + \beta_{123} x_3 x_1 + \tilde{\mathbf{x}}\tilde{\beta})x_3 \\
\frac{\partial g_2}{\partial \beta_{123}} &= \phi(\beta_1 x_1 + \beta_2 + \beta_3 x_3 + \beta_{12} x_1 + \beta_{13} x_3 x_1 + \beta_{23} x_3 + \beta_{123} x_3 x_1 + \tilde{\mathbf{x}}\tilde{\beta})x_1 x_3 \\
\frac{\partial g_2}{\partial \beta_j} &= (\phi(\beta_1 x_1 + \beta_2 + \beta_3 x_3 + \beta_{12} x_1 + \beta_{13} x_3 x_1 + \beta_{23} x_3 + \beta_{123} x_3 x_1 + \tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_1 x_1 + \beta_3 x_3 + \beta_{13} x_3 x_1 + \tilde{\mathbf{x}}\tilde{\beta}))x_j
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial g_3}{\partial \beta_1} &= (\phi(\beta_1 x_1 + \beta_2 x_2 + \beta_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 + \beta_{23} x_2 + \beta_{123} x_2 x_1 + \tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \tilde{\mathbf{x}}\tilde{\beta}))x_1 \\
\frac{\partial g_3}{\partial \beta_2} &= (\phi(\beta_1 x_1 + \beta_2 x_2 + \beta_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 + \beta_{23} x_2 + \beta_{123} x_2 x_1 + \tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \tilde{\mathbf{x}}\tilde{\beta}))x_2 \\
\frac{\partial g_3}{\partial \beta_3} &= \phi(\beta_1 x_1 + \beta_2 x_2 + \beta_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 + \beta_{23} x_2 + \beta_{123} x_2 x_1 + \tilde{\mathbf{x}}\tilde{\beta}) \\
\frac{\partial g_3}{\partial \beta_{12}} &= (\phi(\beta_1 x_1 + \beta_2 x_2 + \beta_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 + \beta_{23} x_2 + \beta_{123} x_2 x_1 + \tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \tilde{\mathbf{x}}\tilde{\beta}))x_1 x_2 \\
\frac{\partial g_3}{\partial \beta_{13}} &= \phi(\beta_1 x_1 + \beta_2 x_2 + \beta_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 + \beta_{23} x_2 + \beta_{123} x_2 x_1 + \tilde{\mathbf{x}}\tilde{\beta})x_1 \\
\frac{\partial g_3}{\partial \beta_{23}} &= \phi(\beta_1 x_1 + \beta_2 x_2 + \beta_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 + \beta_{23} x_2 + \beta_{123} x_2 x_1 + \tilde{\mathbf{x}}\tilde{\beta})x_2 \\
\frac{\partial g_3}{\partial \beta_{123}} &= \phi(\beta_1 x_1 + \beta_2 x_2 + \beta_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 + \beta_{23} x_2 + \beta_{123} x_2 x_1 + \tilde{\mathbf{x}}\tilde{\beta})x_1 x_2 \\
\frac{\partial g_3}{\partial \beta_j} &= (\phi(\beta_1 x_1 + \beta_2 x_2 + \beta_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 + \beta_{23} x_2 + \beta_{123} x_2 x_1 + \tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \tilde{\mathbf{x}}\tilde{\beta}))x_j \\
\frac{\partial g_4}{\partial \beta_1} &= (\phi(\beta_1 + \beta_2 + \beta_3 x_3 + \beta_{12} + \beta_{13} x_3 + \beta_{23} x_3 + \beta_{123} x_3 + \tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_1 + \beta_3 x_3 + \beta_{13} x_3 + \tilde{\mathbf{x}}\tilde{\beta})) \\
\frac{\partial g_4}{\partial \beta_2} &= (\phi(\beta_1 + \beta_2 + \beta_3 x_3 + \beta_{12} + \beta_{13} x_3 + \beta_{23} x_3 + \beta_{123} x_3 + \tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_2 + \beta_3 x_3 + \beta_{23} x_3 + \tilde{\mathbf{x}}\tilde{\beta})) \\
\frac{\partial g_4}{\partial \beta_3} &= (\phi(\beta_1 + \beta_2 + \beta_3 x_3 + \beta_{12} + \beta_{13} x_3 + \beta_{23} x_3 + \beta_{123} x_3 + \tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_2 + \beta_3 x_3 + \beta_{23} x_3 + \tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_1 + \beta_3 x_3 + \beta_{13} x_3 + \tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad +\phi(\beta_3 x_3 + \tilde{\mathbf{x}}\tilde{\beta}))x_3 \\
\frac{\partial g_4}{\partial \beta_{12}} &= \phi(\beta_1 + \beta_2 + \beta_3 x_3 + \beta_{12} + \beta_{13} x_3 + \beta_{23} x_3 + \beta_{123} x_3 + \tilde{\mathbf{x}}\tilde{\beta}) \\
\frac{\partial g_4}{\partial \beta_{13}} &= (\phi(\beta_1 + \beta_2 + \beta_3 x_3 + \beta_{12} + \beta_{13} x_3 + \beta_{23} x_3 + \beta_{123} x_3 + \tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_1 + \beta_3 x_3 + \beta_{13} x_3 + \tilde{\mathbf{x}}\tilde{\beta}))x_3 \\
\frac{\partial g_4}{\partial \beta_{23}} &= (\phi(\beta_1 + \beta_2 + \beta_3 x_3 + \beta_{12} + \beta_{13} x_3 + \beta_{23} x_3 + \beta_{123} x_3 + \tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_2 + \beta_3 x_3 + \beta_{23} x_3 + \tilde{\mathbf{x}}\tilde{\beta}))x_3 \\
\frac{\partial g_4}{\partial \beta_{123}} &= \phi(\beta_1 + \beta_2 + \beta_3 x_3 + \beta_{12} + \beta_{13} x_3 + \beta_{23} x_3 + \beta_{123} x_3 + \tilde{\mathbf{x}}\tilde{\beta})x_3
\end{aligned}
$$

$$\begin{aligned}
\frac{\partial g_4}{\partial \beta_j} &= (\phi(\beta_1+\beta_2+\beta_3 x_3+\beta_{12}+\beta_{13}x_3+\beta_{23}x_3+\beta_{123}x_3+\tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_2+\beta_3 x_3+\beta_{23}x_3+\tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_1+\beta_3 x_3+\beta_{13}x_3+\tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad +\phi(\beta_3 x_3+\tilde{\mathbf{x}}\tilde{\beta}))x_j \\
\frac{\partial g_5}{\partial \beta_1} &= (\phi(\beta_1+\beta_2 x_2+\beta_3+\beta_{12}x_2+\beta_{13}+\beta_{23}x_2+\beta_{123}x_2+\tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_1+\beta_2 x_2+\beta_{12}x_2+\tilde{\mathbf{x}}\tilde{\beta})) \\
\frac{\partial g_5}{\partial \beta_2} &= (\phi(\beta_1+\beta_2 x_2+\beta_3+\beta_{12}x_2+\beta_{13}+\beta_{23}x_2+\beta_{123}x_2+\tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_2 x_2+\beta_3+\beta_{23}x_2+\tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_1+\beta_2 x_2+\beta_{12}x_2+\tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad +\phi(\beta_2 x_2+\tilde{\mathbf{x}}\tilde{\beta}))x_2 \\
\frac{\partial g_5}{\partial \beta_3} &= (\phi(\beta_1+\beta_2 x_2+\beta_3+\beta_{12}x_2+\beta_{13}+\beta_{23}x_2+\beta_{123}x_2+\tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_2 x_2+\beta_3+\beta_{23}x_2+\tilde{\mathbf{x}}\tilde{\beta})) \\
\frac{\partial g_5}{\partial \beta_{12}} &= (\phi(\beta_1+\beta_2 x_2+\beta_3+\beta_{12}x_2+\beta_{13}+\beta_{23}x_2+\beta_{123}x_2+\tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_1+\beta_2 x_2+\beta_{12}x_2+\tilde{\mathbf{x}}\tilde{\beta}))x_2 \\
\frac{\partial g_5}{\partial \beta_{13}} &= \phi(\beta_1+\beta_2 x_2+\beta_3+\beta_{12}x_2+\beta_{13}+\beta_{23}x_2+\beta_{123}x_2+\tilde{\mathbf{x}}\tilde{\beta}) \\
\frac{\partial g_5}{\partial \beta_{23}} &= (\phi(\beta_1+\beta_2 x_2+\beta_3+\beta_{12}x_2+\beta_{13}+\beta_{23}x_2+\beta_{123}x_2+\tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_2 x_2+\beta_3+\beta_{23}x_2+\tilde{\mathbf{x}}\tilde{\beta}))x_2 \\
\frac{\partial g_5}{\partial \beta_{123}} &= \phi(\beta_1+\beta_2 x_2+\beta_3+\beta_{12}x_2+\beta_{13}+\beta_{23}x_2+\beta_{123}x_2+\tilde{\mathbf{x}}\tilde{\beta})x_2 \\
\frac{\partial g_5}{\partial \beta_j} &= (\phi(\beta_1+\beta_2 x_2+\beta_3+\beta_{12}x_2+\beta_{13}+\beta_{23}x_2+\beta_{123}x_2+\tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_2 x_2+\beta_3+\beta_{23}x_2+\tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_1+\beta_2 x_2+\beta_{12}x_2+\tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad +\phi(\beta_2 x_2+\tilde{\mathbf{x}}\tilde{\beta}))x_j
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial g_6}{\partial \beta_1} &= (\phi(\beta_1 x_1 + \beta_2 + \beta_3 + \beta_{12} x_1 + \beta_{13} x_1 + \beta_{23} + \beta_{123} x_1 + \tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_1 x_1 + \beta_2 + \beta_{12} x_1 + \tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_1 x_1 + \beta_3 + \beta_{13} x_1 + \tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad +\phi(\beta_1 x_1 + \tilde{\mathbf{x}}\tilde{\beta}))x_1 \\
\frac{\partial g_6}{\partial \beta_2} &= (\phi(\beta_1 x_1 + \beta_2 + \beta_3 + \beta_{12} x_1 + \beta_{13} x_1 + \beta_{23} + \beta_{123} x_1 + \tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_1 x_1 + \beta_2 + \beta_{12} x_1 + \tilde{\mathbf{x}}\tilde{\beta})) \\
\frac{\partial g_6}{\partial \beta_3} &= (\phi(\beta_1 x_1 + \beta_2 + \beta_3 + \beta_{12} x_1 + \beta_{13} x_1 + \beta_{23} + \beta_{123} x_1 + \tilde{\mathbf{x}}\tilde{\beta}) - \phi(\beta_1 x_1 + \beta_3 + \beta_{13} x_1 + \tilde{\mathbf{x}}\tilde{\beta})) \\
\frac{\partial g_6}{\partial \beta_{12}} &= (\phi(\beta_1 x_1 + \beta_2 + \beta_3 + \beta_{12} x_1 + \beta_{13} x_1 + \beta_{23} + \beta_{123} x_1 + \tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_1 x_1 + \beta_2 + \beta_{12} x_1 + \tilde{\mathbf{x}}\tilde{\beta}))x_1 \\
\frac{\partial g_6}{\partial \beta_{13}} &= (\phi(\beta_1 x_1 + \beta_2 + \beta_3 + \beta_{12} x_1 + \beta_{13} x_1 + \beta_{23} + \beta_{123} x_1 + \tilde{\mathbf{x}}\tilde{\beta}) - \phi(\beta_1 x_1 + \beta_3 + \beta_{13} x_1 + \tilde{\mathbf{x}}\tilde{\beta}))x_1 \\
\frac{\partial g_6}{\partial \beta_{23}} &= \phi(\beta_1 x_1 + \beta_2 + \beta_3 + \beta_{12} x_1 + \beta_{13} x_1 + \beta_{23} + \beta_{123} x_1 + \tilde{\mathbf{x}}\tilde{\beta}) \\
\frac{\partial g_6}{\partial \beta_{123}} &= \phi(\beta_1 x_1 + \beta_2 + \beta_3 + \beta_{12} x_1 + \beta_{13} x_1 + \beta_{23} + \beta_{123} x_1 + \tilde{\mathbf{x}}\tilde{\beta})x_1 \\
\frac{\partial g_6}{\partial \beta_j} &= (\phi(\beta_1 x_1 + \beta_2 + \beta_3 + \beta_{12} x_1 + \beta_{13} x_1 + \beta_{23} + \beta_{123} x_1 + \tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_1 x_1 + \beta_2 + \beta_{12} x_1 + \tilde{\mathbf{x}}\tilde{\beta}) - \phi(\beta_1 x_1 + \beta_3 + \beta_{13} x_1 + \tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad +\phi(\beta_1 x_1 + \tilde{\mathbf{x}}\tilde{\beta}))x_j \\
\frac{\partial g_7}{\partial \beta_1} &= (\phi(\beta_1 + \beta_2 + \beta_3 + \beta_{12} + \beta_{13} + \beta_{23} + \beta_{123} + \tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_1 + \beta_2 + \beta_{12} + \tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_1 + \beta_3 + \beta_{13} + \tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad +\phi(\beta_1 + \tilde{\mathbf{x}}\tilde{\beta})) \\
\frac{\partial g_7}{\partial \beta_2} &= (\phi(\beta_1 + \beta_2 + \beta_3 + \beta_{12} + \beta_{13} + \beta_{23} + \beta_{123} + \tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_1 + \beta_2 + \beta_{12} + \tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_2 + \beta_3 + \beta_{23} + \tilde{\mathbf{x}}\tilde{\beta}) + \phi(\beta_2 + \tilde{\mathbf{x}}\tilde{\beta}))
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial g_7}{\partial \beta_3} &= (\phi(\beta_1+\beta_2+\beta_3+\beta_{12}+\beta_{13}+\beta_{23}+\beta_{123}+\tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_1+\beta_3+\beta_{13}+\tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_2+\beta_3+\beta_{23}+\tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad +\phi(\beta_3+\tilde{\mathbf{x}}\tilde{\beta})) \\
\frac{\partial g_7}{\partial \beta_{12}} &= (\phi(\beta_1+\beta_2+\beta_3+\beta_{12}+\beta_{13}+\beta_{23}+\beta_{123}+\tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_1+\beta_2+\beta_{12}+\tilde{\mathbf{x}}\tilde{\beta})) \\
\frac{\partial g_7}{\partial \beta_{13}} &= (\phi(\beta_1+\beta_2+\beta_3+\beta_{12}+\beta_{13}+\beta_{23}+\beta_{123}+\tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_1+\beta_3+\beta_{13}+\tilde{\mathbf{x}}\tilde{\beta})) \\
\frac{\partial g_7}{\partial \beta_{23}} &= (\phi(\beta_1+\beta_2+\beta_3+\beta_{12}+\beta_{13}+\beta_{23}+\beta_{123}+\tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_2+\beta_3+\beta_{23}+\tilde{\mathbf{x}}\tilde{\beta})) \\
\frac{\partial g_7}{\partial \beta_{123}} &= \phi(\beta_1+\beta_2+\beta_3+\beta_{12}+\beta_{13}+\beta_{23}+\beta_{123}+\tilde{\mathbf{x}}\tilde{\beta}) \\
\frac{\partial g_7}{\partial \beta_j} &= (\phi(\beta_1+\beta_2+\beta_3+\beta_{12}+\beta_{13}+\beta_{23}+\beta_{123}+\tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_1+\beta_2+\beta_{12}+\tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_1+\beta_3+\beta_{13}+\tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad -\phi(\beta_2+\beta_3+\beta_{23}+\tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad +\phi(\beta_3+\tilde{\mathbf{x}}\tilde{\beta})+\phi(\beta_2+\tilde{\mathbf{x}}\tilde{\beta}) \\
&\quad +\phi(\beta_1+\tilde{\mathbf{x}}\tilde{\beta})-\phi(+\tilde{\mathbf{x}}\tilde{\beta}))x_j
\end{aligned}
$$