

Standard errors of marginal effects in the heteroskedastic probit model

Thomas Cornelißen*

DISCUSSION PAPER NO. 320

AUGUST 2005

ISSN: 0949–9962

Abstract

In non-linear regression models, such as the heteroskedastic probit model, coefficients cannot be interpreted as marginal effects. Marginal effects can be computed as a non-linear combination of the regression coefficients. Standard errors of the marginal effects needed for inference and hypothesis testing have to be derived by approximation using methods such as the delta method. This paper applies the delta method to derive analytically the standard errors of marginal effects in a heteroskedastic probit model. The computation is implemented as a Stata ado-file called *mehetprob* which can be downloaded from the internet. This allows to compute marginal effects at means and their standard errors in a heteroskedastic probit model faster than by numerical calculation which is implemented in the *mfx* routine currently available in Stata for that purposes.

KEYWORDS: heteroskedastic probit model, marginal effects, Stata

JEL-CLASSIFICATION: C25, C87

*Institute of Quantitative Economic Research, University of Hannover, Germany,
cornelissen@mbox.iqw.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 obtained by calculating the derivative of the outcome probability with respect to the regressor. The derivative can be calculated analytically or numerically.

Standard errors of marginal effects also need to be derived to allow inference and hypothesis testing. If the marginal effect is a non-linear transformation of the regression coefficients, the standard error of the marginal effect can only be calculated approximately by methods such as the delta method (see section 3). This involves calculating the derivatives of the marginal effect with respect to all coefficients. Again, the derivatives can be computed analytically or numerically.

For the heteroskedastic probit model, the analytical form of the marginal effect is commonly stated (see for example Greene 2003, p. 680), but the derivatives of the marginal effect needed for calculating its standard errors are not. In the statistical software package Stata, the *hetprobit* command computes coefficients and standard errors of coefficients of the heteroskedastic probit model. When marginal effects are needed, Stata offers the *mf* command which numerically computes marginal effects and their standard errors after regression commands. In the case of the heteroskedastic probit model and when many regressors are involved, this procedure is somewhat time consuming.¹

The aim of this paper is to derive analytically the standard errors for the marginal effects of a heteroskedastic probit regression. The analytical calculation of the marginal effect and the standard error is implemented as a Stata ado-file that is downloadable for public use.

The paper proceeds as follows. Section 2 describes the marginal effect of a regressor in the heteroskedastic probit model. Section 3 covers the delta method and section 4 applies the method in order to derive the standard errors of the heteroskedastic probit model. Section 5 provides information on how to access the Stata ado-file and presents an application which compares the analytical calculation with the numerical one. Section 6 concludes.

¹This depends very much on the size of the model and the data set and on which release of Stata is used. In Stata 8, the computation of marginal effects with *mf* after *hetprobit* can take hours, while in Stata 9 it has been speeded up to the range of minutes. See section 5 for running times of a practical application.

2 The marginal effect of the heteroskedastic probit model

The heteroskedastic probit model extends the simple probit model by introducing heteroskedasticity of the error term of the latent variable. Let the latent variable be

$$y_i^* = x_i' \beta + \epsilon_i, \quad (1)$$

where i indexes observations, x is an $s \times 1$ vector of covariates including the regressor 1, β a corresponding coefficient vector and ϵ a normally distributed error term satisfying

$$E[\epsilon_i] = 0, \quad (2)$$

$$V[\epsilon_i] = \sigma_i^2 = [\exp(z_i' \gamma)]^2, \quad (3)$$

and

$$\text{Cov}[\epsilon_i, \epsilon_j] = 0, \quad i \neq j, \quad (4)$$

where z is a $t \times 1$ vector of regressors determining the variance of the error term, and γ is the corresponding coefficient vector.

Let the observed binary variable y depend on the latent variable y^* such that

$$y_i = \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & \text{if } y_i^* \leq 0, \end{cases} \quad (5)$$

then the probability of a success is given by

$$P(Y_i = 1) = P(\epsilon_i > -x_i' \beta) = 1 - \Phi \left(\frac{-x_i' \beta}{\exp(z_i' \gamma)} \right) = \Phi \left(\frac{x_i' \beta}{\exp(z_i' \gamma)} \right) \quad (6)$$

where Φ is the c.d.f of the standard normal distribution.

Unite the x and z regressors in vector $w' = (x', z')$ of length $l = s + t$. For each regressor w_k , $k = 1, \dots, l$, a marginal effect has to be derived.

Corresponding to the $l \times 1$ vector w define coefficient vectors b and g of the same length l such that

$$b_k = \begin{cases} \beta_j & \text{if } w_k \text{ is the } j\text{th element of } x \\ 0 & \text{otherwise,} \end{cases} \quad (7)$$

and

$$g_k = \begin{cases} \gamma_j & \text{if } w_k \text{ is the } j\text{th element of } z \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

The k th element of b contains the β coefficient associated with w_k if w_k is part of the x vector, otherwise it is 0. Similarly, the k th element of g contains the γ coefficient associated with w_k if w_k is part of the z vector, otherwise it is 0.

Suppose a regressor that is part of x and z , then it will appear two times in w , say $w_i = w_j$. Consequently, $b_i = b_j$ and $g_i = g_j$. The marginal effects and standard errors of the marginal effects derived for w_i and w_j will then be the same.

The marginal effect of one regressor w_k is function of all β and γ parameters that may be assembled in a vector $\theta' = (\beta', \gamma')$ of same length ($l = s + t$) as w .

If w_k is to be treated as a continuous variable, the marginal effect of the outcome probability with respect to this regressor, $g(w_k)$, is found by deriving (6) with respect to w_k , which yields

$$g_C(w_k) = \frac{\partial P(Y = 1)}{\partial w_k} = \phi \left(\frac{x' \beta}{\exp(z' \gamma)} \right) \frac{b_k - x' \beta \cdot g_k}{\exp(z' \gamma)}, \quad (9)$$

(see Greene 2003, p.680).

If w_k is element of both, the x and the z vector, then b_k and g_k are the coefficients associated with the regressor w_k in the mean equation (1) and the variance equation (3) respectively.

If w_k is only element of x but not of z , then g_k is zero by definition, i.e. only the first part of (9) applies, and if w_k is only element of z but not of x , then b_k is zero by definition, i.e. only the last part of (9) applies.

If w_k is a dummy variable, the marginal effect of the outcome probability with respect to this regressor, $g(w_k)$, can alternatively, i.e. if the dummy variable is not to be treated as a continuous variable, be calculated as

$$g_D(w_k) = \Phi \left(\frac{x'_1 \beta}{\exp(z'_1 \gamma)} \right) - \Phi \left(\frac{x'_0 \beta}{\exp(z'_0 \gamma)} \right), \quad (10)$$

where w_k is set to 1 in x_1 and z_1 and set to 0 in x_0 and z_0 (provided that the regressor is part of the respective vectors).

The marginal effect can be evaluated at different values of the regressors, i.e. a marginal effect for each individual can be calculated. Usually a summary statistic is warranted and thus the marginal effect is evaluated at a specific point, e.g. at means or at the median. Alternatively, a (weighted) average of the individual specific marginal effects can be calculated.

There is an important distinction between the normal probit model and the heteroskedastic probit model. In the normal probit model the marginal effect of a regressor x_k for individual i is given by $\phi(x_i'\beta) \cdot \beta_k$. As $\phi(\cdot)$ is always positive, the marginal effect has the same sign as the regression coefficient for all individuals in the sample. In the heteroskedastic probit model, the sign of the marginal effect depends on $b_k - x'\beta \cdot g_k$ (see equation (9)), which can switch sign in the sample for regressors that are part of both x and z and for regressors that are only part of z . Therefore it can be of interest to calculate the individual marginal effects and check for which fraction of the sample the effect is positive or negative.

For the following application of the delta method suppose that one marginal effect for each regressor w_k , $k = 1, \dots, l$ is calculated at specific values of the regressors (e.g. at means). This leads to a $l \times 1$ vector of estimated marginal effects $\hat{\lambda}$.

3 The delta method

The delta method is a popular way to estimate standard errors of non-linear functions of model parameters. While it is straightforward to calculate the variance of a linear function of a random variable, it is not for a nonlinear function. The delta method therefore relies on finding a linear approximation of the function by using a first-order Taylor expansion (e.g. Davidson/MacKinnon 2004, p.202). In the multivariate case, let $\hat{\lambda} = \mathbf{g}(\hat{\theta})$ be a $l \times 1$ vector of monotonic continuously differentiable functions of the $n \times 1$ coefficient estimator $\hat{\theta}$. Then, for a given estimated covariance matrix of the model parameters, $\hat{V}(\hat{\theta})$, the covariance matrix of $\hat{\lambda}$, can be estimated according to the delta method by

$$\hat{V}(\hat{\lambda}) = \hat{\mathbf{G}}\hat{V}(\hat{\theta})\hat{\mathbf{G}}', \quad (11)$$

where $\hat{\mathbf{G}} \equiv \mathbf{G}(\hat{\theta})$ is the $l \times n$ matrix $\partial \mathbf{g}(\theta)/\partial \theta'$. The i th row of $\mathbf{G}(\hat{\theta})$ is the vector of partial derivatives of the i th function with respect to $\hat{\theta}'$ or, in other words, the typical element in row i and column j of $\mathbf{G}(\hat{\theta})$ is $\partial g_i(\theta)/\partial \theta_j$ (Davidson/MacKinnon 2004, p. 208).

4 Applying the delta method to standard errors of marginal effects of the heteroskedastic probit model

In the heteroskedastic probit model as defined above there is a $(l = s + t) \times 1$ vector w of regressors entering the model. For each regressor w_k a marginal effect is derived, which is function of all β and γ parameters that may be assembled in a $1 \times (l = s + t)$ vector $\theta' = (\beta', \gamma')$.²

The $l \times 1$ vector of marginal effects is $\hat{\lambda} = \mathbf{g}(\hat{\theta})$ with its k th element being the marginal effect $g(w_k)$ as defined in (9) and (10).

The element (k, j) of the $l \times l$ matrix $\mathbf{G}(\hat{\theta})$ contains the derivative of $g(w_k)$ with respect to the coefficient θ_j , whereby θ_j may be (i) a β coefficient not associated with w_k , (ii) a β coefficient associated with w_k , (iii) a γ coefficient not associated with w_k , or (iv) a γ coefficient associated with w_k .³ For these four cases, and if the regressors are treated as continuous, the derivatives are:

(i) If θ_j is a β coefficient not associated with w_k :

$$\begin{aligned} \partial g_C(w_k)/\partial \theta_j &= -b_k w_j (x' \beta) \frac{\phi\left(\frac{x' \beta}{\exp(z' \gamma)}\right)}{[\exp(z' \gamma)]^3} \\ &\quad - g_k w_j \frac{\phi\left(\frac{x' \beta}{\exp(z' \gamma)}\right)}{\exp(z' \gamma)} \left(1 - \frac{(x' \beta)^2}{[\exp(z' \gamma)]^2}\right) \end{aligned} \quad (12)$$

(ii) If θ_j is a β coefficient associated with w_k :

$$\begin{aligned} \partial g_C(w_k)/\partial \theta_j &= \frac{\phi\left(\frac{x' \beta}{\exp(z' \gamma)}\right)}{\exp(z' \gamma)} - b_k w_j (x' \beta) \frac{\phi\left(\frac{x' \beta}{\exp(z' \gamma)}\right)}{[\exp(z' \gamma)]^3} \\ &\quad - g_k w_j \frac{\phi\left(\frac{x' \beta}{\exp(z' \gamma)}\right)}{\exp(z' \gamma)} \left(1 - \frac{(x' \beta)^2}{[\exp(z' \gamma)]^2}\right) \end{aligned} \quad (13)$$

(iii) If θ_j is a γ coefficient not associated with w_k :

$$\begin{aligned} \partial g_C(w_k)/\partial \theta_j &= (b_k - x' \beta \cdot g_k) \frac{w_j \cdot \phi\left(\frac{x' \beta}{\exp(z' \gamma)}\right)}{\exp(z' \gamma)} \\ &\quad \cdot \left(\frac{(x' \beta)^2}{[\exp(z' \gamma)]^2} - 1\right) \end{aligned} \quad (14)$$

²Recall that s is the length of x and β , t is the length of z and γ .

³The distinction whether w_k is part of x , z or both enters through the components of b and g being zero if w_k is not in x or z , respectively.

(iv) If θ_j is a γ coefficient associated with w_k :

$$\begin{aligned} \partial g_C(w_k)/\partial \theta_j &= (b_k - x'\beta \cdot g_k) \frac{w_j \cdot \phi\left(\frac{x'\beta}{\exp(z'\gamma)}\right)}{\exp(z'\gamma)} \left(\frac{(x'\beta)^2}{[\exp(z'\gamma)]^2} - 1 \right) \\ &\quad - \frac{\phi\left(\frac{x'\beta}{\exp(z'\gamma)}\right) \cdot x'\beta}{\exp(z'\gamma)} \end{aligned} \quad (15)$$

Hereby, all regressors are evaluated at the same values that have been chosen to evaluate the marginal effect, e.g. at means. Recall that b_k (g_k) is zero by definition if w_k is only element of x (z). In that case, the derivatives simplify accordingly.

If regressor w_k is a dummy variable that is not to be treated as a continuous variables, the derivative of $g_D(w_k)$ in (10) is used, which is

(i) If θ_j is a β coefficient:

$$\partial g_D(w_k)/\partial \theta_j = w_{1j} \cdot \frac{\phi\left(\frac{x'_1\beta}{\exp(z'_1\gamma)}\right)}{\exp(z'_1\gamma)} - w_{0j} \cdot \frac{\phi\left(\frac{x'_0\beta}{\exp(z'_0\gamma)}\right)}{\exp(z'_0\gamma)} \quad (16)$$

(ii) If θ_j is a γ coefficient:

$$\begin{aligned} \partial g_D(w_k)/\partial \theta_j &= \frac{-w_{1j} \cdot (x'_1\beta) \phi\left(\frac{x'_1\beta}{\exp(z'_1\gamma)}\right)}{\exp(z'_1\gamma)} \\ &\quad + \frac{w_{0j} \cdot (x'_0\beta) \phi\left(\frac{x'_0\beta}{\exp(z'_0\gamma)}\right)}{\exp(z'_0\gamma)}, \end{aligned} \quad (17)$$

where w_{0j} and w_{1j} are the j th element of the vectors $w'_0 = (x'_0, z'_0)$ and $w'_1 = (x'_1, z'_1)$, respectively. The vectors x_0, z_0, x_1 and z_1 have been defined in the discussion of equation (10).

Once $\hat{\mathbf{G}} \equiv \mathbf{G}(\hat{\theta}) = \partial \mathbf{g}(\hat{\theta})/\partial \hat{\theta}'$ is derived, the standard errors of the marginal effects can be calculated by (11).

5 Stata ado-file and application

The computation of marginal effects and their standard errors that has been derived analytically in the preceding section has been programmed as a Stata

ado-file called *mehetprob*. This file and the associated help file are available for download at the Statistical Software Components (SSC) archive, often also called the Boston College Archive. The recommended way to download it is by typing *ssc install mehetprob* in the Stata command line. Alternatively, to view a description type *ssc describe mehetprob*.⁴

In the following, an application is presented. A heteroskedastic probit regression is run with data from the German Socio-Economic Panel. An overview of the data set is provided by Haisken-DeNew/Frick (2003). West German private sector workers are sampled. The dependent variable *quit1* takes on the value 1 if a worker quits his job in the next period and 0 otherwise. Explanatory variables are the hourly wage (*w_h4*), the actual working hours (*hoursact*), sex (*male*), dummies for increasing firm size categories (*fsize2*, *fsize3*, *fsize4*), a dummy that indicates foreign nationality (*foreign*), age, age squared (*agesq*), tenure and the regional unemployment rate (*regunemp*). All variables enter the main equation (the *x* vector in terms of the notation used earlier), tenure and the hourly wage also enter the equation for modelling the variance of the error term (the *z* vector).

The output from the regression is:

```
Heteroskedastic probit model                Number of obs   =       57294
                                             Zero outcomes   =       54277
                                             Nonzero outcomes =       3017

                                             Wald chi2(11)   =       829.77
Log likelihood = -10478.78                 Prob > chi2     =       0.0000

-----+-----
quit1 |          Coef.   Std. Err.      z    P>|z|    [95% Conf. Interval]
-----+-----
quit1 |
   w_h4 |   .0117582   .0022307     5.27  0.000   .0073862   .0161303
 hoursact | .0023257   .0011014     2.11  0.035   .000167   .0044845
   male |  -.0577722   .0223677    -2.58  0.010  -.101612  -.0139323
  fsize2 |  -.100491   .0245272    -4.10  0.000  -.1485634  -.0524186
  fsize3 |  -.2425828   .029201    -8.31  0.000  -.2998157  -.1853498
  fsize4 |  -.3415201   .032397   -10.54  0.000  -.4050171  -.2780231
  foreign |  -.210218   .0249594    -8.42  0.000  -.2591375  -.1612984
   age |   .0608655   .0074646     8.15  0.000   .0462352   .0754959
  agesq |  -.0010025   .0001049    -9.56  0.000  -.0012081  -.0007969
  tenure |  -.1485081   .0070428   -21.09  0.000  -.1623117  -.1347045
 regunemp | -.0199462   .0036634    -5.44  0.000  -.0271263  -.0127662
   _cons | -1.646515   .1288987   -12.77  0.000  -1.899152  -1.393879
-----+-----
lnsigma2 |
  tenure |   .033426   .0015828    21.12  0.000   .0303239   .0365282
   w_h4 |  -.0083866   .0018145    -4.62  0.000  -.011943  -.0048302
-----+-----
Likelihood-ratio test of lnsigma2=0: chi2(2) =   269.38   Prob > chi2 = 0.0000
```

⁴Type *help ssc* to get information on the *ssc* command that allows to manage the user written software components from the SSC archive.

All coefficients are highly significant and have the expected sign, at least for those variables that are only in the main equation, as the sign of their marginal effect on the propensity to quit must equal the sign of the regression coefficients. For tenure and the hourly wage the sign of the marginal effect is not clear, because they enter the mean and the variance equation and thus the sign can switch depending on the regressor values at which the marginal effect is evaluated. The likelihood-ratio test reported at the bottom of the regression output rejects a model without heteroskedasticity.

Below the output from *mfx* and *mehetprob* for the computation of marginal effects and their standard errors is presented.

```
. mfx compute;

Marginal effects after hetprob
y = Pr(quit1) (predict)
= .02091509
```

variable	dy/dx	Std. Err.	z	P> z	[95% C.I.]	X
w_h4	-.0003846	.00013	-2.91	0.004	-.000644 -.000126	11.8481
hoursact	.0000937	.00004	2.10	0.036	6.2e-06 .000181	39.5301
male*	-.0023583	.00092	-2.55	0.011	-.004168 -.000549	.63504
fsize2*	-.0039146	.00094	-4.18	0.000	-.005752 -.002077	.289297
fsize3*	-.008871	.00101	-8.75	0.000	-.010859 -.006883	.244615
fsize4*	-.011988	.00107	-11.25	0.000	-.014076 -.0099	.237774
foreign*	-.0078578	.00092	-8.51	0.000	-.009667 -.006049	.271756
age	.0024525	.00028	8.63	0.000	.001895 .00301	37.9823
agesq	-.0000404	.00000	-10.21	0.000	-.000048 -.000033	1572.87
tenure	-.0025629	.00011	-22.95	0.000	-.002782 -.002344	9.60174
regunemp	-.0008037	.00015	-5.38	0.000	-.001096 -.000511	8.54533

(*) dy/dx is for discrete change of dummy variable from 0 to 1

```
. mehprob;
(obs=57294)
(obs=57294)
P(Y=1) in sample: .05265822
P(Y=1) mean of model prediction: .05267644
P(Y=1) predicted at means: .02091509
```

dP/dX - Marginal effect at means after heteroskedastic probit estimation:

Variable	dP/dX	s.e.	z	P
w_h4	-.0003846	.0001321	-2.91	0.004
hoursact	.0000937	.0000446	2.10	0.036
male*	-.0023583	.0009232	-2.55	0.011
fsize2*	-.0039146	.0009374	-4.18	0.000
fsize3*	-.008871	.0010143	-8.75	0.000
fsize4*	-.011988	.0010654	-11.25	0.000
foreign*	-.0078578	.000923	-8.51	0.000
age	.0024525	.0002843	8.63	0.000
agesq	-.0000404	3.96e-06	-10.21	0.000
tenure	-.0025629	.0001117	-22.95	0.000
regunemp	-.0008037	.0001493	-5.38	0.000

(*) dP/dx is for discrete change of dummy variable from 0 to 1

The results are exactly the same. All marginal effects at means are highly significant. Wage and tenure have negative marginal effect at means. As would be expected, the propensity to quit decreases with rising hourly wages and with longer tenure.

Calculating the marginal effects at means with the *mf*x command takes a running time of 3 hours and 25 minutes in Stata 8. In Stata 9, where the *mf*x command has been speeded up, it takes 1 minute and 24 seconds. The calculation with the *mehetprob* routine takes 3 seconds.

Table 1 compares the running time of the calculation of the marginal effects at means between the numerical computation by *mf*x in Stata 8 and Stata 9 and the analytical computation by the *mehetprob* routine for two different models.

Table1: Comparison of computing time in different models

	Small model	Large model
Size		
No. Regressors	11	44
No. Observations	57294	87487
Running Time (hh:mm:ss)^{a)}		
mf	03:25:00	> 39:00:00
mf	00:01:24	00:10:45
mehetprob	00:00:03	00:00:08
^{a)} Hardware specification: CPU 1,92 Ghz, 200 MB RAM allocated to Stata		

Model 1 is the one presented above with 11 regressors and 57,294 observations. In model 2 more regressors, e.g. time and sector dummies, are added and the sample is extended to East German and public sector workers. This leads to 44 regressors and 87,487 observations (regression results not reported here). While *mehetprob* took 8 seconds, *mf*x in Stata 9 took about 10 minutes and *mf*x in Stata 8 more than 39 hours.

Although the *mf*x command is considerably faster in Stata 9 as compared to Stata 8, the *mehetprob* routine still saves running time. It seems that the larger the model and the number of observations, the greater is the factor by which *mehetprob* is faster. For example, comparing *mf*x of Stata 9 with *mehetprob* shows that *mehetprob* is faster by a factor of 28 (3 seconds versus 1 minute 24 seconds) in model 1 and by a factor of 81 in model 2 (8 seconds versus 10 minutes 45 seconds).

6 Conclusion

This paper has derived an analytical form of the standard errors of marginal effects in a heteroskedastic probit model. The computation has been implemented as a Stata ado-file which can be downloaded from the internet.

This allows to compute marginal effects at means and their standard errors in a heteroskedastic probit model faster than by numerical calculation which is implemented in the *mfx* routine currently available in Stata for that purposes. For users of Stata 8 the routine can save hours of computation time, for users of Stata 9 it saves minutes, as the *mfx* command has been speeded up considerably in the Stata 9 release.

References

- [1] Greene, W. (2003): *Econometric Analysis*, Prentice Hall, New Jersey.
- [2] Davidson, R. and J. MacKinnon (2004): *Econometric Theory and Methods*, Oxford University Press, New York.
- [3] Haisken-DeNew, J.P. and J.R. Frick (Eds.) (2003): *Desktop Companion to the German Socio-Economic Panel Study*, Version 7.0, DIW Berlin.